Objectives Lecture 01

1 Introduction
Objectives Lecture 01

1. Introduction
2. Describing life distributions
Objectives Lecture 01

1. Introduction
2. Describing life distributions
3. Survival function

Readings: KM § 1, 2.1, 2.2, 2.3, 2.4, 2.5, 3.1-2, 3.3, 3.4
Objectives Lecture 01

1. Introduction
2. Describing life distributions
3. Survival function
4. Hazard rate
Objectives Lecture 01

1. Introduction
2. Describing life distributions
3. Survival function
4. Hazard rate
5. Mean residual life

Readings: KM § 1, 2.1, 2.2, 2.3, 2.4, 2.5, 3.1-2, 3.3, 3.4
Objectives Lecture 01

1. Introduction
2. Describing life distributions
3. Survival function
4. Hazard rate
5. Mean residual life
6. Common survival distributions

Readings: KM § 1, 2.1, 2.2, 2.3, 2.4, 2.5, 3.1-2, 3.3, 3.4
Objectives Lecture 01

1. Introduction
2. Describing life distributions
3. Survival function
4. Hazard rate
5. Mean residual life
6. Common survival distributions
7. Censoring
Objectives Lecture 01

1. Introduction
2. Describing life distributions
3. Survival function
4. Hazard rate
5. Mean residual life
6. Common survival distributions
7. Censoring
8. Interval censoring
Objectives Lecture 01

1. Introduction
2. Describing life distributions
3. Survival function
4. Hazard rate
5. Mean residual life
6. Common survival distributions
7. Censoring
8. Interval censoring
9. Truncation

Readings: KM § 1, 2.1, 2.2, 2.3, 2.4, 2.5, 3.1-2, 3.3, 3.4
Objectives Lecture 01

1. Introduction
2. Describing life distributions
3. Survival function
4. Hazard rate
5. Mean residual life
6. Common survival distributions
7. Censoring
8. Interval censoring
9. Truncation
10. Readings: KM §1, 2.1, 2.2, 2.3, 2.4, 2.5, 3.1-2, 3.3, 3.4
Section: Introduction

Subsection: Analysis of data that are event times
Examples

1 Medicine

Deaths from certain kinds of disease
Recovery from disease
Time until disease progresses to a certain point

sometimes a surrogate endpoint

Equipment failures
Mortgage prepayments

etc., etc., ...

SAS Code R Code
Examples

1 Medicine
   1 Deaths from certain kinds of disease

Introduction: Analysis of data that are event times
Examples

1. Medicine
   1. Deaths from certain kinds of disease
   2. Recovery from disease
Examples

1. Medicine
   1. Deaths from certain kinds of disease
   2. Recovery from disease
   3. Time until disease progresses to a certain point
Examples

1. Medicine
   1. Deaths from certain kinds of disease
   2. Recovery from disease
   3. Time until disease progresses to a certain point
      1. sometimes a *surrogate endpoint*
Examples

1 Medicine
   1 Deaths from certain kinds of disease
   2 Recovery from disease
   3 Time until disease progresses to a certain point
      1 sometimes a surrogate endpoint

2 Equipment failures
Examples

1 Medicine
   1 Deaths from certain kinds of disease
   2 Recovery from disease
   3 Time until disease progresses to a certain point
      1 sometimes a surrogate endpoint

2 Equipment failures

3 Mortgage prepayments
Examples

1 Medicine
   1 Deaths from certain kinds of disease
   2 Recovery from disease
   3 Time until disease progresses to a certain point
      1 sometimes a *surrogate endpoint*

2 Equipment failures

3 Mortgage prepayments

4 etc., etc., ... SAS Code  R Code

Introduction: Analysis of data that are event times
Life data analysis differs from other forms of data analysis

1. Additional complications arising from missing data
Life data analysis differs from other forms of data analysis

1. Additional complications arising from missing data
   1. Missingness is in the form of censoring
Life data analysis differs from other forms of data analysis

1. Additional complications arising from missing data
   1. Missingness is in the form of censoring
   2. To be defined later this lecture.

Different probabilistic models are appropriate. For ex., wouldn’t be interested in modeling times as normal asymmetric (skewed right). Different effects of interventions are usually expected. For ex., horizontal shifts are usually not of interest. We are probably interested in estimating not just a mean but all quantiles of a distribution.
Life data analysis differs from other forms of data analysis

1. Additional complications arising from missing data
   1. Missingness is in the form of censoring
   2. To be defined later this lecture.

2. Different probabilistic models are appropriate
Life data analysis differs from other forms of data analysis

1. Additional complications arising from missing data
   1. Missingness is in the form of censoring
   2. To be defined later this lecture.

2. Different probabilistic models are appropriate
   1. For ex., wouldn’t be interested in modeling times as normal Asymmetric (skewed right)

Different effects of interventions are usually expected
For ex., horizontal shifts are usually not of interest
We are probably interested in estimating not just a mean but all quantiles of a distribution
Life data analysis differs from other forms of data analysis

1. Additional complications arising from missing data
   1. Missingness is in the form of censoring
   2. To be defined later this lecture.

2. Different probabilistic models are appropriate
   1. For ex., wouldn’t be interested in modeling times as normal
   2. Asymmetric (skewed right)
Life data analysis differs from other forms of data analysis

1. Additional complications arising from missing data
   1. Missingness is in the form of censoring
   2. To be defined later this lecture.

2. Different probabilistic models are appropriate
   1. For ex., wouldn’t be interested in modeling times as normal
   2. Asymmetric (skewed right)

3. Different effects of interventions are usually expected
Life data analysis differs from other forms of data analysis

1. Additional complications arising from missing data
   - Missingness is in the form of censoring
   - To be defined later this lecture.

2. Different probabilistic models are appropriate
   - For ex., wouldn’t be interested in modeling times as normal
   - Asymmetric (skewed right)

3. Different effects of interventions are usually expected
   - For ex., horizontal shifts are usually not of interest
Life data analysis differs from other forms of data analysis

1. Additional complications arising from missing data
   1. Missingness is in the form of censoring
   2. To be defined later this lecture.

2. Different probabilistic models are appropriate
   1. For ex., wouldn’t be interested in modeling times as normal
   2. Asymmetric (skewed right)

3. Different effects of interventions are usually expected
   1. For ex., horizontal shifts are usually not of interest

4. We are probably interested in estimating not just a mean but all quantiles of a distribution
Section: Introduction

Subsection: Describing life distributions
Most elemental description of distributions is CDF

Let $F(x)$ be the CDF.
Most elemental description of distributions is CDF

1. Let $F(x)$ be the CDF.
   1. Defined to be $P[X \leq x]$. 

All distributions have these properties:

1. Non-increasing
2. $S(x) = 1 \forall x < 0$
3. $S(0) = 1$ if $P[X = 0] = 0$

For some applications, like reliability, $S(0) < 1$. We won't see many such applications.

$S(\infty) = 0$ typically, unless individual can last for ever.
Most elemental description of distributions is CDF

1. Let $F(x)$ be the CDF.
   - Defined to be $P[X \leq x]$.

2. Most survival applications use instead the survival function $S(x) = 1 - F(x) = P[X > x]$. 

Most elemental description of distributions is CDF

1. Let $F(x)$ be the CDF.
   
   Defined to be $P[X \leq x]$.

2. Most survival applications use instead the survival function $S(x) = 1 - F(x) = P[X > x]$.

3. All distributions have these
Most elemental description of distributions is CDF

1. Let $F(x)$ be the CDF.
   - Defined to be $P[X \leq x]$.

2. Most survival applications use instead the survival function $S(x) = 1 - F(x) = P[X > x]$.

3. All distributions have these properties:

   - Non-increasing
   - $S(x) = 1$ for all $x < 0$
   - $S(0) = 1$ if $P[X = 0] = 0$

   For some applications, like reliability, $S(0) < 1$. We won't see many such applications.

   - $S(\infty) = 0$ typically, unless individual can last for ever.
Most elemental description of distributions is CDF

1. Let $F(x)$ be the CDF.
   - Defined to be $P[X \leq x]$.

2. Most survival applications use instead the survival function $S(x) = 1 - F(x) = P[X > x]$.

3. All distributions have these

4. Properties:
   - Non-increasing
Most elemental description of distributions is CDF

1. Let $F(x)$ be the CDF.
   - Defined to be $P[X \leq x]$.

2. Most survival applications use instead the survival function $S(x) = 1 - F(x) = P[X > x]$.

3. All distributions have these properties:
   - Non-increasing
   - $S(x) = 1 \forall x < 0$
Most elemental description of distributions is CDF

1. Let $F(x)$ be the CDF.
   - Defined to be $P[X \leq x]$.

2. Most survival applications use instead the survival function $S(x) = 1 - F(x) = P[X > x]$.

3. All distributions have these properties:
   - Non-increasing
   - $S(x) = 1 \forall x < 0$
   - $S(0) = 1$ if $P[X = 0] = 0$
Most elemental description of distributions is CDF

1. Let $F(x)$ be the CDF.
   - Defined to be $P[X \leq x]$.

2. Most survival applications use instead the survival function $S(x) = 1 - F(x) = P[X > x]$.

3. All distributions have these Properties:
   - Non-increasing
   - $S(x) = 1$ for all $x < 0$.
   - $S(0) = 1$ if $P[X = 0] = 0$.
   - For some applications, like reliability, $S(0) < 1$. 

Introduction: Describing life distributions
Most elemental description of distributions is CDF

1. Let $F(x)$ be the CDF.
   1. Defined to be $P[X \leq x]$.

2. Most survival applications use instead the survival function $S(x) = 1 - F(x) = P[X > x]$

3. All distributions have these properties:
   1. Non-increasing
   2. $S(x) = 1 \forall x < 0$
   3. $S(0) = 1$ if $P[X = 0] = 0$

   - For some applications, like reliability, $S(0) < 1$
   - We won’t see many such applications
Most elemental description of distributions is CDF

1. Let $F(x)$ be the CDF.
   1. Defined to be $P[X \leq x]$.

2. Most survival applications use instead the *survival function* $S(x) = 1 - F(x) = P[X > x]$

3. All distributions have these Properties:

   1. Non-increasing
   2. $S(x) = 1 \forall x < 0$
   3. $S(0) = 1$ if $P[X = 0] = 0$
      1. For some applications, like reliability, $S(0) < 1$
      2. We won’t see many such applications
   4. $S(\infty) = 0$ typically, unless individual can last for ever.
Density of life times, if distribution is continuous

Here $S(x) = \int_x^\infty f(s) \, ds$
Density of life times, if distribution is continuous

1. Here \( S(x) = \int_x^\infty f(s) \, ds \)
2. \( f(x) = -S'(x) \)
Density of life times, if distribution is continuous

1. Here $S(x) = \int_{x}^{\infty} f(s) \, ds$
2. $f(x) = -S'(x)$
3. Some distributions are made up of smooth parts pasted together
Density of life times, if distribution is continuous

1. Here $S(x) = \int_x^\infty f(s) \, ds$
2. $f(x) = -S'(x)$
3. Some distributions are made up of smooth parts pasted together
   1. Relation $f(x) = -S'(x)$ doesn’t necessarily hold at junction
The *hazard rate*:

1. **Definition**: Chance of failing in the next small interval conditional on lasting until now, divided by length of interval.
The *hazard rate*:

1. Definition: Chance of failing in the next small interval conditional on lasting until now, divided by length of interval.

2. Assume distribution continuous

\[
h(x) = \lim_{\Delta \to 0} \frac{P[X < x + \Delta | X \geq x]}{\Delta}
\]

Quantity inside limit is \(h(x) \approx \frac{f(x)}{S(x)} \Delta \to \frac{f(x)}{S(x)} = -\frac{S'(x)}{S(x)} = -\frac{d}{dx} \log(S(x))\) if \(S'(x)\) exists

Recover \(S(x) = \exp\left(-\int_0^x h(s) ds\right)\), by noting that \(\int_0^x h(s) ds = -\log(S(x))\).

\(H(x) = \int_0^x h(s) ds\) is called *integrated hazard*.

Discrete case: suppose possible event times are \(x_j\) counts of days, months, etc.

\(h(x_j) = \frac{p(x_j)}{S(x_j - 1)}\)

\(p(x_j) = P[X = x_j] = S(x_j - 1) - S(x_j)\).
The *hazard rate*:

1. **Definition:** Chance of failing in the next small interval conditional on lasting until now, divided by length of interval.

2. **Assume distribution continuous**

3. \[ h(x) = \lim_{\Delta \to 0} \frac{P[X < x + \Delta | X \geq x]}{\Delta} \]
The hazard rate:

1. Definition: Chance of failing in the next small interval conditional on lasting until now, divided by length of interval.

2. Assume distribution continuous

3. \( h(x) = \lim_{\Delta \to 0} P[X < x + \Delta | X \geq x] / \Delta \)

4. Quantity inside limit is \( h(x) \approx f(x) \Delta / [S(x) \Delta] \to f(x) / S(x) = -S'(x) / S(x) = -\frac{d}{dx} \log(S(x)) \) if \( S'(x) \) exist
The **hazard rate**:

1. Definition: Chance of failing in the next small interval conditional on lasting until now, divided by length of interval.

2. Assume distribution continuous

3. \( h(x) = \lim_{\Delta \to 0} \frac{P[X < x + \Delta | X \geq x]}{\Delta} \)

4. Quantity inside limit is \( h(x) \approx \frac{f(x)\Delta}{S(x)\Delta} \to \frac{f(x)}{S(x)} = -\frac{S'(x)}{S(x)} = -\frac{d}{dx} \log(S(x)) \) if \( S'(x) \) exist

5. \( h(x) \geq 0 \forall x \)
The hazard rate:

1. Definition: Chance of failing in the next small interval conditional on lasting until now, divided by length of interval.

2. Assume distribution continuous

3. \( h(x) = \lim_{\Delta \to 0} P[X < x + \Delta | X \geq x] / \Delta \)

4. Quantity inside limit is \( h(x) \approx f(x) \Delta / [S(x) \Delta] \to f(x)/S(x) = -S'(x)/S(x) = -\frac{d}{dx} \log(S(x)) \) if \( S'(x) \) exist

5. \( h(x) \geq 0 \forall x \)

6. Recover \( S(x) = \exp(- \int_0^x h(s) \, ds) \),
The hazard rate:

1. Definition: Chance of failing in the next small interval conditional on lasting until now, divided by length of interval.

2. Assume distribution continuous

3. \( h(x) = \lim_{\Delta \to 0} \frac{P[X < x + \Delta \mid X \geq x]}{\Delta} \)

4. Quantity inside limit is \( h(x) \approx \frac{f(x)\Delta}{S(x)\Delta} \to \frac{f(x)}{S(x)} = -\frac{S'(x)}{S(x)} = -\frac{d}{dx} \log(S(x)) \) if \( S'(x) \) exist

5. \( h(x) \geq 0 \forall x \)

6. Recover \( S(x) = \exp(-\int_0^x h(s) \, ds) \),
   - by noting that \( \int_0^x h(s) \, ds = -\log(S(x)) \).
The **hazard rate**: 

1. **Definition**: Chance of failing in the next small interval conditional on lasting until now, divided by length of interval.

2. Assume distribution continuous

3. \( h(x) = \lim_{\Delta \to 0} P[X < x + \Delta | X \geq x] / \Delta \)

4. Quantity inside limit is \( h(x) \approx f(x) \Delta /[S(x) \Delta] \to f(x)/S(x) = -S'(x)/S(x) = -\frac{d}{dx} \log(S(x)) \) if \( S'(x) \) exist

5. \( h(x) \geq 0 \forall x \)

6. Recover \( S(x) = \exp(- \int_{0}^{x} h(s) \, ds) \),
   
   1. by noting that \( \int_{0}^{x} h(s) \, ds = - \log(S(x)) \).
   
   2. \( H(x) = \int_{0}^{x} h(s) \, ds \) is called **integrated hazard**.
The hazard rate:

1. Definition: Chance of failing in the next small interval conditional on lasting until now, divided by length of interval.

2. Assume distribution continuous

3. \[ h(x) = \lim_{\Delta \to 0} P [X < x + \Delta | X \geq x] / \Delta \]

4. Quantity inside limit is \[ h(x) \approx f(x) \Delta / [S(x) \Delta] \to f(x)/S(x) = -S'(x)/S(x) = -\frac{d}{dx} \log(S(x)) \] if \( S'(x) \) exist

5. \( h(x) \geq 0 \forall x \)

6. Recover \( S(x) = \exp(-\int_0^x h(s) \, ds) \)

    1. by noting that \( \int_0^x h(s) \, ds = -\log(S(x)) \).

    2. \( H(x) = \int_0^x h(s) \, ds \) is called integrated hazard.

7. Discrete case: suppose possible event times are \( x_j \)

Introduction: Describing life distributions Lecture 01
The hazard rate:

1. Definition: Chance of failing in the next small interval conditional on lasting until now, divided by length of interval.

2. Assume distribution continuous

3. \( h(x) = \lim_{\Delta \to 0} P[X < x + \Delta | X \geq x] / \Delta \)

4. Quantity inside limit is \( h(x) \approx f(x)\Delta/[S(x)\Delta] \to f(x)/S(x) = -S'(x)/S(x) = -\frac{d}{dx} \log(S(x)) \) if \( S'(x) \) exist

5. \( h(x) \geq 0 \forall x \)

6. Recover \( S(x) = \exp(-\int_0^x h(s) \, ds) \),
   - by noting that \( \int_0^x h(s) \, ds = -\log(S(x)) \).
   - \( H(x) = \int_0^x h(s) \, ds \) is called integrated hazard.

7. Discrete case: suppose possible event times are \( x_j \)
   - counts of days, months, etc.
The hazard rate:

1. Definition: Chance of failing in the next small interval conditional on lasting until now, divided by length of interval.

2. Assume distribution continuous

3. \( h(x) = \lim_{\Delta \to 0} P[X < x + \Delta \mid X \geq x] / \Delta \)

4. Quantity inside limit is \( h(x) \approx f(x)\Delta/[S(x)\Delta] \to f(x)/S(x) = -S'(x)/S(x) = -\frac{d}{dx} \log(S(x)) \) if \( S'(x) \) exist

5. \( h(x) \geq 0 \forall x \)

6. Recover \( S(x) = \exp(-\int_0^x h(s) \, ds) \),
   - by noting that \( \int_0^x h(s) \, ds = -\log(S(x)) \).
   - \( H(x) = \int_0^x h(s) \, ds \) is called integrated hazard.

7. Discrete case: suppose possible event times are \( x_j \)
   - counts of days, months, etc.
   - Hazard is \( h(x_j) = p(x_j)/S(x_{j-1}) \)
The hazard rate:

1. **Definition**: Chance of failing in the next small interval conditional on lasting until now, divided by length of interval.

2. **Assume distribution continuous**

3. \( h(x) = \lim_{\Delta \to 0} \frac{P[X < x + \Delta | X \geq x]}{\Delta} \)

4. Quantity inside limit is \( h(x) \approx \frac{f(x) \Delta}{S(x)\Delta} \rightarrow \frac{f(x)}{S(x)} = -\frac{S'(x)}{S(x)} = -\frac{d}{dx} \log(S(x)) \) if \( S'(x) \) exist

5. \( h(x) \geq 0 \forall x \)

6. Recover \( S(x) = \exp\left(-\int_0^x h(s) \, ds\right) \),

   - by noting that \( \int_0^x h(s) \, ds = -\log(S(x)) \).

   - \( H(x) = \int_0^x h(s) \, ds \) is called **integrated hazard**.

7. **Discrete case**: suppose possible event times are \( x_j \)

   - counts of days, months, etc.

   - **Hazard is** \( h(x_j) = \frac{p(x_j)}{S(x_{j-1})} \)

   - \( p(x_j) = P[X = x_j] = S(x_{j-1}) - S(x_j) \).
Some possible hazard shapes:
Some possible hazard shapes:

1. Constant: Same “risk” of event at every time point:

- Many electronics components
- Event of one radioactive atom decaying
- Increasing
- Decreasing
- Bathtub: Highest at beginning and late
- Human life spans
- Products under warranty

- Bump shaped:
  - Mortgages
  - Not many others

Will be used later to give model diagnostics
Some possible hazard shapes:

- Constant: Same “risk” of event at every time point:
  - Many electronics components

- Event of one radioactive atom decaying
- Increasing
- Decreasing
- Bathtub: Highest at beginning and late
  - Human life spans
  - products under warranty

- Bump shaped:
  - Mortgages
  - Not many others

Will be used later to give model diagnostics
Some possible hazard shapes:

1. Constant: Same “risk” of event at every time point:
   1. Many electronics components
   2. Event of one radioactive atom decaying
Some possible hazard shapes:

1. **Constant:** Same “risk” of event at every time point:
   - Many electronics components
   - Event of one radioactive atom decaying

2. **Increasing**
Some possible hazard shapes:

1. Constant: Same “risk” of event at every time point:
   - Many electronics components
   - Event of one radioactive atom decaying

2. Increasing

3. Decreasing

Will be used later to give model diagnostics
Some possible hazard shapes:

1. **Constant**: Same “risk” of event at every time point:
   - Many electronics components
   - Event of one radioactive atom decaying
2. **Increasing**
3. **Decreasing**
4. **Bathtub**: Highest at beginning and late

Introduction: Describing life distributions
Some possible hazard shapes:

1. Constant: Same “risk” of event at every time point:
   - Many electronics components
   - Event of one radioactive atom decaying

2. Increasing

3. Decreasing

4. Bathtub: Highest at beginning and late
   - Human life spans
Some possible hazard shapes:

1. **Constant**: Same "risk" of event at every time point:
   - Many electronics components
   - Event of one radioactive atom decaying

2. **Increasing**

3. **Decreasing**

4. **Bathtub**: Highest at beginning and late
   - Human life spans
   - Products under warranty

Introduction: Describing life distributions
Some possible hazard shapes:

1. **Constant**: Same “risk” of event at every time point:
   - Many electronics components
   - Event of one radioactive atom decaying

2. **Increasing**

3. **Decreasing**

4. **Bathtub**: Highest at beginning and late
   - Human life spans
   - Products under warranty

5. **Bump shaped**: 

Introduction: Describing life distributions
Some possible hazard shapes:

1. **Constant:** Same “risk” of event at every time point:
   1. Many electronics components
   2. Event of one radioactive atom decaying

2. **Increasing**

3. **Decreasing**

4. **Bathtub:** Highest at beginning and late
   1. Human life spans
   2. Products under warranty

5. **Bump shaped:**
   1. Mortgages
Some possible hazard shapes:

1. **Constant**: Same “risk” of event at every time point:
   - Many electronics components
   - Event of one radioactive atom decaying

2. **Increasing**

3. **Decreasing**

4. **Bathtub**: Highest at beginning and late
   - Human life spans
   - Products under warranty

5. **Bump shaped**:
   - Mortgages
   - Not many others
Some possible hazard shapes:

1. **Constant**: Same “risk” of event at every time point:
   - Many electronics components
   - Event of one radioactive atom decaying
2. **Increasing**
3. **Decreasing**
4. **Bathtub**: Highest at beginning and late
   - Human life spans
   - Products under warranty
5. **Bump shaped**:
   - Mortgages
   - Not many others
6. Will be used later to give model diagnostics
The *mean residual life*:

1. Consider distribution with finite expectation.
The *mean residual life*:

1. Consider distribution with finite expectation.
2. Expectation of how much life remains, conditional on no event so far.
The *mean residual life*:

1. Consider distribution with finite expectation.
2. Expectation of how much life remains, conditional on no event so far.
3. For continuous distributions, is $\text{m.r.l.}(x) = \int_x^\infty (s - x)f(s)\, ds / S(x)$
The mean residual life:

1. Consider distribution with finite expectation.
2. Expectation of how much life remains, conditional on no event so far.
3. For continuous distributions, is \( \text{m.r.l.}(x) = \int_x^\infty (s - x) f(s) \, ds / S(x) \)
4. Integrate by parts: \( \int_x^\infty (s - x) f(s) \, ds = (s - x) S(s)|_x^\infty + \int_x^\infty S(s) \, ds = \int_x^\infty S(s) \, ds \).
The mean residual life:

1. Consider distribution with finite expectation.
2. Expectation of how much life remains, conditional on no event so far.
3. For continuous distributions, is m.r.l.\((x) = \int_x^\infty (s - x)f(s) \, ds / S(x)\)
4. Integrate by parts: 
   \[
   \int_x^\infty (s - x)f(s) \, ds = (s - x)S(s)|_x^\infty + \int_x^\infty S(s) \, ds = \int_x^\infty S(s) \, ds
   \]
5. Hence m.r.l.\((x) = \int_x^\infty S(s) \, ds / S(x)\)

Since \(\lim_{s \to \infty} sS(s) = 0\); otherwise the expectation is infinite.

\[\int_a^b u \, dv = uv|_a^b - \int_a^b v \, du.\]
The *mean residual life*: 

1. Consider distribution with finite expectation.
2. Expectation of how much life remains, conditional on no event so far.
3. For continuous distributions, is m.r.l.(x) = \( \int_x^\infty (s - x) f(s) \, ds / S(x) \)
4. Integrate by parts: 
   \[ \int_x^\infty (s - x) f(s) \, ds = (s - x) S(s)|_x^\infty + \int_x^\infty S(s) \, ds = \int_x^\infty S(s) \, ds \]

   1. \( \int_a^b u \, dv = uv|_a^b - \int_a^b v \, du \).
   2. \( u(s) = (s - x), \ dv = f(s) \, ds \) and so \( du = ds, \ v(s) = -S(s) \).
The mean residual life:

1. Consider distribution with finite expectation.
2. Expectation of how much life remains, conditional on no event so far.
3. For continuous distributions, is \( \text{m.r.l.}(x) = \int_x^\infty (s - x) f(s) \, ds / S(x) \)
4. Integrate by parts: \( \int_x^\infty (s - x) f(s) \, ds = (s - x) S(s)|_x^\infty + \int_x^\infty S(s) \, ds = \int_x^\infty S(s) \, ds \)

\[ \int_a^b u \, dv = uv|_a^b - \int_a^b v \, du. \]
\[ u(s) = (s - x), \, dv = f(s) \, ds \text{ and so } du = ds, \, v(s) = -S(s). \]
\[ \text{since } \lim_{s \to \infty} sS(s) = 0; \text{ otherwise the expectation is infinite.} \]
The mean residual life:

1. Consider distribution with finite expectation.
2. Expectation of how much life remains, conditional on no event so far.
3. For continuous distributions, is m.r.l.($x$) = $\int_x^\infty (s - x)f(s) \, ds / S(x)$
4. Integrate by parts: $\int_x^\infty (s - x)f(s) \, ds = (s - x)S(s)|_x^\infty + \int_x^\infty S(s) \, ds = \int_x^\infty S(s) \, ds$
   1. $\int_a^b u \, dv = uv|_a^b - \int_a^b v \, du$.
   2. $u(s) = (s - x)$, $dv = f(s) \, ds$ and so $du = ds$, $v(s) = -S(s)$.
   3. since $\lim_{s \to \infty} sS(s) = 0$; otherwise the expectation is infinite.
5. Hence m.r.l.($x$) = $\int_x^\infty S(s) \, ds / S(x)$
Continued

1. When calculating m.r.l.'(x), note \( \frac{d}{dx} \int_x^\infty S(s)ds = -S'(x) \)

Because

\[
\frac{d}{dx} \int_x^\infty S(s)ds = \lim_{\delta \to 0} \left[ \frac{\int_x^{x+\delta} S(s)ds - \int_x^\infty S(s)ds}{\delta} \right] = \lim_{\delta \to 0} \left[ - \int_x^{x+\delta} S(s)ds \right]/\delta = \lim_{\delta \to 0} \left[ -S'(x^*)\delta \right]/\delta
\]

for \( x^* \in [x, x+\delta] \)

\[
= -S'(x)
\]

since \( S \) continuous
When calculating m.r.l.’(x), note \( \frac{d}{dx} \int_x^\infty S(s)ds = -S'(x) \)

Because

\[
\frac{d}{dx} \int_x^\infty S(s)ds = \lim_{\delta \to 0} \left[ \int_x^{x+\delta} S(s)ds - \int_x^\infty S(s)ds \right] / \delta \\
= \lim_{\delta \to 0} \left[ - \int_x^{x+\delta} S(s)ds \right] / \delta \\
= \lim_{\delta \to 0} \left[ -S(x^*) \delta \right] / \delta \text{ for } x^* \in [x, x + \delta] \\
= \lim_{\delta \to 0} \left[ -S(x^*) \right] \text{ for } x^* \in [x, x + \delta] \\
= -S(x) \text{ since } S \text{ continuous}
When calculating m.r.l.'(x), note \( \frac{d}{dx} \int_x^\infty S(s)ds = -S'(x) \)

Because

\[
\frac{d}{dx} \int_x^\infty S(s)ds = \lim_{\delta \to 0} \left[ \int_{x+\delta}^\infty S(s)ds - \int_x^\infty S(s)ds \right]/\delta
\]

\[
= \lim_{\delta \to 0} \left[ -\int_x^{x+\delta} S(s)ds \right]/\delta
\]

\[
= \lim_{\delta \to 0} \left[ -S(x^*)\delta \right]/\delta \text{ for } x^* \in [x, x + \delta]
\]

\[
= \lim_{\delta \to 0} \left[ -S(x^*) \right] \text{ for } x^* \in [x, x + \delta]
\]

\[
= -S(x) \text{ since } S \text{ continuous}
\]

m.r.l.(0) is original mean
When calculating m.r.l.'(x), note \( \frac{d}{dx} \int_x^\infty S(s)ds = -S'(x) \)

Because

\[
\frac{d}{dx} \int_x^\infty S(s)ds = \lim_{\delta \to 0} \left[ \int_x^{x+\delta} S(s)ds - \int_x^\infty S(s)ds \right]/\delta
\]

\[
= \lim_{\delta \to 0} \left[ - \int_x^{x+\delta} S(s)ds \right]/\delta
\]

\[
= \lim_{\delta \to 0} [ -S(x^*)\delta]/\delta \text{ for } x^* \in [x, x + \delta]
\]

\[
= \lim_{\delta \to 0} [ -S(x^*)] \text{ for } x^* \in [x, x + \delta]
\]

\[
= -S(x) \text{ since } S \text{ continuous}
\]

m.r.l.(0) is original mean

You can give formulae for survival function, density, and hazard in terms of mean residual life.
Section: Introduction

Subsection: Parametric Models for Life Distributions
lifetime is sum of large number of independent contributions
Normal

1. lifetime is sum of large number of independent contributions
2. Bad one: negative values illegal
What distribution has constant hazard?

- Exponential
What distribution has constant hazard?

1. Exponential

   \[ -\frac{d}{dx} \log(S(x)) = \lambda \]
What distribution has constant hazard?

1. **Exponential**
   1. \(-\frac{d}{dx} \log(S(x)) = \lambda\)
   2. \(\log(S(x)) - \log(S(0)) = -\lambda x\)
What distribution has constant hazard?

1. Exponential

1. \(- \frac{d}{dx} \log(S(x)) = \lambda\)
2. \(\log(S(x)) - \log(S(0)) = -\lambda x\)
3. \(\log(S(x)) = -\lambda x\)
What distribution has constant hazard?

1. **Exponential**
   
   1. $-\frac{d}{dx} \log(S(x)) = \lambda$
   2. $\log(S(x)) - \log(S(0)) = -\lambda x$
   3. $\log(S(x)) = -\lambda x$
   4. $S(x) = \exp(-\lambda x)$
What distribution has constant hazard?

1. Exponential
   
   1. \(-\frac{d}{dx} \log(S(x)) = \lambda\)
   2. \(\log(S(x)) - \log(S(0)) = -\lambda x\)
   3. \(\log(S(x)) = -\lambda x\)
   4. \(S(x) = \exp(-\lambda x)\)
   5. \(f(x) = \lambda \exp(-\lambda x)\)
What distribution has constant hazard?

1. **Exponential**
   
   1. $- \frac{d}{dx} \log(S(x)) = \lambda$
   2. $\log(S(x)) - \log(S(0)) = -\lambda x$
   3. $\log(S(x)) = -\lambda x$
   4. $S(x) = \exp(-\lambda x)$
   5. $f(x) = \lambda \exp(-\lambda x)$
   6. Distribution of remaining life is always the same: Memoryless
What distribution has constant hazard?

**Exponential**

1. $-\frac{d}{dx} \log(S(x)) = \lambda$
2. $\log(S(x)) - \log(S(0)) = -\lambda x$
3. $\log(S(x)) = -\lambda x$
4. $S(x) = \exp(-\lambda x)$
5. $f(x) = \lambda \exp(-\lambda x)$

6. Distribution of remaining life is always the same: Memoryless
   - Take $s > x$. 

Note joint probability

$P[X > s \text{ and } X > t] = P[X > s]$

$P[X > s | X > x] = \exp(-\lambda s) \exp(-\lambda x) = \exp(-\lambda (s-x)) = P[X > s-x]$
What distribution has constant hazard?

**Exponential**

1. \(-\frac{d}{dx} \log(S(x)) = \lambda\)
2. \(\log(S(x)) - \log(S(0)) = -\lambda x\)
3. \(\log(S(x)) = -\lambda x\)
4. \(S(x) = \exp(-\lambda x)\)
5. \(f(x) = \lambda \exp(-\lambda x)\)
6. Distribution of remaining life is always the same: Memoryless
   1. Take \(s > x\).
   2. Note joint probability \(P[X > s \text{ and } X > t] = P[X > s]\)
What distribution has constant hazard?

1. **Exponential**
   1. \[-\frac{d}{dx} \log(S(x)) = \lambda\]
   2. \[\log(S(x)) - \log(S(0)) = -\lambda x\]
   3. \[\log(S(x)) = -\lambda x\]
   4. \[S(x) = \exp(-\lambda x)\]
   5. \[f(x) = \lambda \exp(-\lambda x)\]
   6. Distribution of remaining life is always the same: Memoryless

   1. Take \(s > x\).
   2. Note joint probability \(P[X > s \text{ and } X > t] = P[X > s]\)
   3. \[P[X > s \mid X > x] = \frac{\exp(-\lambda s)}{\exp(-\lambda x)} = \exp(-(s - x)\lambda) = P[X > s - x]\]
Generalization of Exponential

1 Sum of $k$ independent exponentials with same (hazard) rate $\lambda$ is $\Gamma(\lambda, k)$. 

Density is $\lambda^k x^{k-1} \exp(-\lambda x) / \Gamma(k)$.

Survival function given by incomplete gamma function $h(x) = \lambda + 1 - k x$.

See Fig. 1.
Generalization of Exponential

1. Sum of $k$ independent exponentials with same (hazard) rate $\lambda$ is $\Gamma(\lambda, k)$.
2. Density is $\lambda^k x^{k-1} \exp(-\lambda x) / \Gamma(k)$
Generalization of Exponential

1. Sum of $k$ independent exponentials with same (hazard) rate $\lambda$ is $\Gamma(\lambda, k)$.

2. Density is $\lambda^k x^{k-1} \exp(-\lambda x)/\Gamma(k)$

3. Survival function given by incomplete gamma function
Generalization of Exponential

1. Sum of \( k \) independent exponentials with same (hazard) rate \( \lambda \) is \( \Gamma(\lambda, k) \).
2. Density is \( \lambda^k x^{k-1} \exp(-\lambda x)/\Gamma(k) \)
3. Survival function given by incomplete gamma function
4. \( h(x) = \lambda + \frac{1-k}{x} \).
Generalization of Exponential

1. Sum of $k$ independent exponentials with same (hazard) rate $\lambda$ is $\Gamma(\lambda, k)$.

2. Density is $\lambda^k x^{k-1} \exp(-\lambda x)/\Gamma(k)$.

3. Survival function given by incomplete gamma function

4. $h(x) = \lambda + \frac{1-k}{x}$.

See Fig. 1.
Generalization: *Weibull distribution.*

Suppose that $Y$ has an exponential distribution with rate $\lambda$. Let $X = Y^{1/\alpha}$ for some $\alpha > 0$. Survival function for $X$ is $S(x) = \exp(-\lambda x^\alpha)$. Density is $\exp(-\lambda x^\alpha) \alpha \lambda x^{\alpha-1}$. Hazard Function $\alpha \lambda x^{\alpha-1}$ is increasing if $\alpha > 1$ and decreasing if $\alpha < 1$. See Fig. 2.

Fig. 2: Weibull Hazards

<table>
<thead>
<tr>
<th>Random Variable Value</th>
<th>Hazard</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>
Generalization: \textit{Weibull distribution}.

1. Suppose that $Y$ has an exponential distribution with rate $\lambda$.
2. Let $X = Y^{1/\alpha}$ for some $\alpha > 0$. 

---

Introduction: Parametric Models for Life Distributions
Generalization: *Weibull distribution.*

1. Suppose that $Y$ has an exponential distribution with rate $\lambda$.
2. Let $X = Y^{1/\alpha}$ for some $\alpha > 0$.
3. Survival function for $X$ is $S(x) = \exp(-\lambda x^\alpha)$
Generalization: \textit{Weibull distribution.}

1. Suppose that $Y$ has an exponential distribution with rate $\lambda$.
2. Let $X = Y^{1/\alpha}$ for some $\alpha > 0$.
3. Survival function for $X$ is $S(x) = \exp(-\lambda x^\alpha)$
4. Density $\exp(-\lambda x^\alpha)\alpha \lambda x^{\alpha-1}$
Generalization: **Weibull distribution.**

1. Suppose that $Y$ has an exponential distribution with rate $\lambda$.
2. Let $X = Y^{1/\alpha}$ for some $\alpha > 0$.
3. Survival function for $X$ is $S(x) = \exp(-\lambda x^{\alpha})$.
4. Density $\exp(-\lambda x^{\alpha}) \alpha \lambda x^{\alpha-1}$.
5. Hazard Function $\alpha \lambda x^{\alpha-1}$.
Generalization: **Weibull distribution.**

1. Suppose that $Y$ has an exponential distribution with rate $\lambda$.
2. Let $X = Y^{1/\alpha}$ for some $\alpha > 0$.
3. Survival function for $X$ is $S(x) = \exp(-\lambda x^\alpha)$.
4. Density $\exp(-\lambda x^\alpha)\alpha \lambda x^{\alpha-1}$.
5. Hazard Function $\alpha \lambda x^{\alpha-1}$

1. Increasing if $\alpha > 1$ and decreasing if $\alpha < 1$. See Fig. 2.

*Fig. 2: Weibull Hazards*
Log Normal

1. Maybe $T$ is product of large number of independent contributions

$S(x) = 1 - \Phi((\log(x) - \mu)/\sigma) = 1 - \Phi(\log(x/\exp(\mu))/\sigma)$ some $\mu, \sigma$

Hence $\mu$ is equivalent to a scale parameter

$f(x) = \phi((\log(x) - \mu)/\sigma)/\sigma x$

$\phi(x) = \exp(-x^2/2)/\sqrt{2\pi}$ the standard normal density, $\Phi(x) = \int_{-\infty}^{x} \exp(-y^2/2)/\sqrt{2\pi} dy/\sqrt{2\pi}$ the standard normal CDF.

MRL is pretty complicated.
Log Normal

1. Maybe $T$ is product of large number of independent contributions
2. Log is normal
Log Normal

1. Maybe $T$ is product of large number of independent contributions
2. Log is normal
3. Gives the log normal distribution

\[
S(x) = 1 - \Phi\left(\frac{\log(x) - \mu}{\sigma}\right) = 1 - \Phi\left(\log\left(\frac{x}{\exp(\mu)}\right)\right)
\]

Hence $\mu$ is equivalent to a scale parameter

\[
f(x) = \frac{\phi\left(\frac{\log(x) - \mu}{\sigma}\right)}{\sigma x}
\]

\[
\phi(x) = \frac{\exp\left(-\frac{x^2}{2}\right)}{\sqrt{2\pi}}
\]

the standard normal density, $\Phi(x) = \frac{\int_{-\infty}^{x} \exp\left(-\frac{y^2}{2}\right) dy}{\sqrt{2\pi}}$ the standard normal CDF.

MRL is pretty complicated.
Log Normal

1. Maybe $T$ is product of large number of independent contributions
2. Log is normal
3. Gives the log normal distribution
4. $S(x) = 1 - \Phi((\log(x) - \mu)/\sigma) = 1 - \Phi(\log(x/\exp(\mu))/\sigma)$ some $\mu, \sigma$
Log Normal

1. Maybe $T$ is product of large number of independent contributions
2. Log is normal
3. Gives the log normal distribution
4. $S(x) = 1 - \Phi((\log(x) - \mu)/\sigma) = 1 - \Phi(\log(x/\exp(\mu))/\sigma)$ some $\mu, \sigma$

Hence $\mu$ is equivalent to a scale parameter
Maybe $T$ is product of large number of independent contributions

Log is normal

Gives the log normal distribution

$S(x) = 1 - \Phi((\log(x) - \mu)/\sigma) = 1 - \Phi(\log(x/\exp(\mu))/\sigma)$ some $\mu, \sigma$

Hence $\mu$ is equivalent to a scale parameter

$f(x) = \phi((\log(x) - \mu)/\sigma)/(\sigma x)$
Log Normal

1. Maybe $T$ is product of large number of independent contributions

2. Log is normal

3. Gives the log normal distribution

4. $S(x) = 1 - \Phi((\log(x) - \mu)/\sigma) = 1 - \Phi(\log(x/\exp(\mu))/\sigma)$ some $\mu, \sigma$

   Hence $\mu$ is equivalent to a scale parameter

5. $f(x) = \phi((\log(x) - \mu)/\sigma)/(\sigma x)$

6. $\phi(x) = \exp(-x^2/2)/\sqrt{2\pi}$ the standard normal density, $\Phi(x) = \int_{-\infty}^{x} \exp(-y^2/2) \; dy/\sqrt{2\pi}$ the standard normal CDF.
Log Normal

1. Maybe $T$ is product of large number of independent contributions
2. Log is normal
3. Gives the log normal distribution
4. $S(x) = 1 - \Phi((\log(x) - \mu)/\sigma) = 1 - \Phi(\log(x/\exp(\mu))/\sigma)$ some $\mu, \sigma$
   - Hence $\mu$ is equivalent to a scale parameter
5. $f(x) = \phi((\log(x) - \mu)/\sigma)/(\sigma x)$
6. $\phi(x) = \exp(-x^2/2)/\sqrt{2\pi}$ the standard normal density, $\Phi(x) = \int_{-\infty}^{x} \exp(-y^2/2) \, dy/\sqrt{2\pi}$ the standard normal CDF.
7. MRL is pretty complicated.
Continued

Hazard is no simpler. See Fig. 3.

Fig. 3: Lognormal Hazard

- Mean 0, Variance 0.25
- Mean 0, Variance 1
- Mean 0, Variance 4
Generalization of Gamma:

\[
f(x) = \alpha \lambda^k x^{\alpha k-1} \exp(-\lambda x^\alpha) / \Gamma(k)
\]
Generalization of Gamma:

1. \[ f(x) = \alpha \lambda^k x^{\alpha k - 1} \exp(-\lambda x^\alpha) / \Gamma(k) \]
2. Integral \( S(x) \) involves incomplete gamma function
Generalization of Gamma:

1. \( f(x) = \alpha \lambda^k x^{\alpha k - 1} \exp(-\lambda x^\alpha)/\Gamma(k) \)

2. Integral \( S(x) \) involves incomplete gamma function

3. Hazard is complicated.
Continued

Contains Weibull \((k = 1)\) and log normal as special cases.
Continued

Contains Weibull ($k = 1$) and log normal as special cases.

See Fig. 4.

Fig. 4: Relationships between distributions

- **Gamma**: Parameters: shape $\gamma$, scale $\lambda$
  - $\gamma = 1$  
  - $\alpha = 1$

- **Exponential**: Parameters: scale $\lambda$
  - $\alpha = 1$

- **Generalized Gamma**: Parameters: shape $\gamma$, power $\alpha$, scale $\lambda$
  - $\gamma = 1$  
  - $\lambda \rightarrow \infty$; $\alpha$, $\lambda$ move accordingly

- **Weibull**: Parameters: scale $\lambda$, power $\alpha$

- **Log Normal**: Scale, shape parameter
Pareto distribution:

1. \( h(x) = \frac{\theta}{x} \) for \( x > \lambda \): decreasing.

\[
S(x) = S(\lambda) = \exp \left( -\int_{\lambda}^{x} \frac{\theta}{s} \, ds \right) = \exp \left( \theta \log(\lambda) - \theta \log(x) \right) = \frac{\lambda^\theta}{x^\theta}
\]

\[
f(x) = \frac{d}{dx} S(x) = \frac{\theta}{x} - \frac{\theta}{x} - 1 = \frac{\theta}{x} - 1
\]
Pareto distribution:

1. \( h(x) = \frac{\theta}{x} \) for \( x > \lambda \): decreasing.

2. \( \frac{S(x)}{S(\lambda)} = \exp \left( - \int_{\lambda}^{x} \frac{\theta}{s} \, ds \right) = \exp(\theta \log(\lambda) - \theta \log(x)) = \frac{\lambda^\theta}{x^\theta} \)
Pareto distribution:

1. $h(x) = \frac{\theta}{x}$ for $x > \lambda$: decreasing.

2. $\frac{S(x)}{S(\lambda)} = \exp \left( - \int_{\lambda}^{x} \frac{\theta}{s} \, ds \right) = \exp(\theta \log(\lambda) - \theta \log(x)) = \frac{\lambda^\theta}{x^\theta}$

3. $f(x) = -\frac{d}{dx} S(x) = \theta x^{-\theta-1} \lambda^\theta$
Section: Censoring and Truncation

Subsection: Types of censoring
Right censoring

1 know that a realization of $X$ exceeds some value, rather than knowing it exactly.

Observe $\min(X,C)$ and indicator for $X \geq C$.

Probabilistic structure relevant

If censoring mechanism has nothing to do with event you are trying to study, censored events give no additional information.

For ex., life of a car before theft censored because of a serious accident doesn't tell you anything.

If censoring mechanism is related to point in life, you know more than just that life exceeds some value.

For ex., scrapping a car because of poor condition might tell you that no one would bother to steal it.

Taxonomy

Type I censoring: $C$ fixed and known

Ex., medical study designed to follow people for a year censors them after a year.
Right censoring

1. Know that a realization of $X$ exceeds some value, rather than knowing it exactly.

2. Observe $\min(X, C)$ and indicator for $X \geq C$. 

3. Probabilistic structure relevant

   If censoring mechanism has nothing to do with event you are trying to study, censored events give no additional information. For ex., life of a car before theft censored because of a serious accident doesn't tell you anything.

   If censoring mechanism is related to point in life, you know more than just that life exceeds some value. For ex., scrapping a car because of poor condition might tell you that no one would bother to steal it.

4. Taxonomy

   Type I censoring: $C$ fixed and known. Ex., medical study designed to follow people for a year censors them after a year.
Right censoring

1. Know that a realization of $X$ exceeds some value, rather than knowing it exactly.
2. Observe $\min(X, C)$ and indicator for $X \geq C$.
3. Probabilistic structure relevant
**Right censoring**

1. know that a realization of $X$ exceeds some value, rather than knowing it exactly
2. Observe $\min(X, C)$ and indicator for $X \geq C$.
3. Probabilistic structure relevant
   1. If censoring mechanism has nothing to do with event you are trying to study, censored events give no additional information
Right censoring

1. know that a realization of \( X \) exceeds some value, rather than knowing it exactly

2. Observe \( \min(X, C) \) and indicator for \( X \geq C \).

3. Probabilistic structure relevant
   - If censoring mechanism has nothing to do with event you are trying to study, censored events give no additional information
     - For ex., life of a car before theft censored because of a serious accident doesn’t tell you anything
   - If censoring mechanism is related to point in life, you know more than just that life exceeds some value.
     - For ex., scrapping a car because of poor condition might tell you that no one would bother to steal it.
Right censoring

1. know that a realization of $X$ exceeds some value, rather than knowing it exactly

2. Observe $\min(X, C)$ and indicator for $X \geq C$.

3. Probabilistic structure relevant
   - If censoring mechanism has nothing to do with event you are trying to study, censored events give no additional information
     - For ex., life of a car before theft censored because of a serious accident doesn’t tell you anything
   - If censoring mechanism is related to point in life, you know more than just that life exceeds some value.
Right censoring

1. know that a realization of $X$ exceeds some value, rather than knowing it exactly

2. Observe $\min(X, C)$ and indicator for $X \geq C$.

3. Probabilistic structure relevant
   - If censoring mechanism has nothing to do with event you are trying to study, censored events give no additional information
     - For ex., life of a car before theft censored because of a serious accident doesn’t tell you anything
   - If censoring mechanism is related to point in life, you know more than just that life exceeds some value.
     - For ex., scrapping a car because of poor condition might tell you that no one would bother to steal it.
**Right censoring**

1. Know that a realization of $X$ exceeds some value, rather than knowing it exactly.
2. Observe $\min(X, C)$ and indicator for $X \geq C$.
3. Probabilistic structure relevant
   - If censoring mechanism has nothing to do with event you are trying to study, censored events give no additional information.
     - For ex., life of a car before theft censored because of a serious accident doesn’t tell you anything.
   - If censoring mechanism is related to point in life, you know more than just that life exceeds some value.
     - For ex., scrapping a car because of poor condition might tell you that no one would bother to steal it.
4. Taxonomy
Right censoring

1. know that a realization of $X$ exceeds some value, rather than knowing it exactly

2. Observe $\min(X, C)$ and indicator for $X \geq C$.

3. Probabilistic structure relevant

   1. If censoring mechanism has nothing to do with event you are trying to study, censored events give no additional information
      
      For ex., life of a car before theft censored because of a serious accident doesn’t tell you anything

   2. If censoring mechanism is related to point in life, you know more than just that life exceeds some value.
      
      For ex., scrapping a car because of poor condition might tell you that no one would bother to steal it.

4. Taxonomy

   1. *Type I censoring*: $C$ fixed and known
Right censoring

1. Know that a realization of $X$ exceeds some value, rather than knowing it exactly.

2. Observe $\min(X, C)$ and indicator for $X \geq C$.

3. Probabilistic structure relevant
   - If censoring mechanism has nothing to do with event you are trying to study, censored events give no additional information.
     - For ex., life of a car before theft censored because of a serious accident doesn’t tell you anything.
   - If censoring mechanism is related to point in life, you know more than just that life exceeds some value.
     - For ex., scrapping a car because of poor condition might tell you that no one would bother to steal it.

4. Taxonomy
   - **Type I censoring**: $C$ fixed and known.
     - Ex., medical study designed to follow people for a year censors them after a year.
Censoring times might not all be the same. Might be the time between enrollment and the fixed end of a study. See Fig. 5.
Censoring times might not all be the same. Might be the time between enrollment and the fixed end of a study. See Fig. 5.

Fig. 5: Calendar Time Diagram

Calendar Time

End of Study
Censoring times might not all be the same.
Censoring times might not all be the same.

Might be the time between an enrollment and the fixed end of a study. See Fig. 5.

Fig. 5: Calendar Time Diagram
Continued

Continued

Type II censoring: Study proceeds until $r < n$ events.

Random censoring: For each $X_i$ associate a censoring time $C_i$. easiest is when censoring is ⊥ mechanism under study.
Continued

Continued

Continued

Continued

Figure 6: Time on Study Diagram

Time on Study

Type II censoring: Study proceeds until \( r < n \) events.

\[
\begin{align*}
C_j &= X(r) \\
C_i &= X_i
\end{align*}
\]

random censoring: For each \( X_i \) associate a censoring time \( C_i \).

Easiest is when censoring is \( \perp \) mechanism under study.
Continued

Continued

Continued

Makes \textit{time on study} more relevant than \textit{calendar time}. See Fig. 6.

\textit{Fig. 6: Time on Study Diagram}

\begin{center}
\begin{tikzpicture}
\node at (0,0) {$X_1$};
\node at (0,-1) {$C_2$};
\node at (0,-2) {$C_3$};
\node at (0,-3) {$X_4$};
\end{tikzpicture}
\end{center}

Time on Study
Makes *time on study* more relevant than *calendar time*. See Fig. 6.

**Fig. 6: Time on Study Diagram**

\[
\begin{align*}
X_1 & \quad C_2 \\
& \quad C_3 \\
& \quad X_4
\end{align*}
\]

**Time on Study**

2. *Type II censoring*: Study proceeds until \( r < n \) events.

Censoring and Truncation: Types of censoring
Continued

1. Continued
   1. Continued
      1. Makes *time on study* more relevant than *calendar time*. See Fig. 6.

*Fig. 6: Time on Study Diagram*

```
X_1
    
    C_2
    
    C_3

X_4
```

Time on Study

2. *Type II censoring*: Study proceeds until $r < n$ events.
   1. $C_j = X_{(r)}$

Censoring and Truncation: Types of censoring
Continued

1. Makes *time on study* more relevant than *calendar time*. See Fig. 6.

*Fig. 6: Time on Study Diagram*

| \( X_1 \) | \( C_2 \) | \( C_3 \) | \( X_4 \) |

Time on Study

2. **Type II censoring**: Study proceeds until \( r < n \) events.
3. \( C_j = X_{(r)} \)
4. *random censoring*:
Continued

Continued

- Makes time on study more relevant than calendar time. See Fig. 6.

*Fig. 6: Time on Study Diagram*

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$C_2$</th>
<th>$C_3$</th>
<th>$X_4$</th>
</tr>
</thead>
</table>

Time on Study

2. **Type II censoring**: Study proceeds until $r < n$ events.

1. $C_j = X_{(r)}$

3. **Random censoring**:

   - For each $X_i$ associate a censoring time $C_i$
Continued

Continued

Continued

Makes *time on study* more relevant than *calendar time*. See Fig. 6.

*Fig. 6: Time on Study Diagram*

![Diagram](attachment:image.png)

Time on Study

*Type II censoring:* Study proceeds until $r < n$ events.

- $C_j = X_{(r)}$

*Random censoring:*

1. For each $X_i$ associate a censoring time $C_i$
2. easiest is when censoring is $\perp$ mechanism under study
Continued

1. Continued
May have a mixture of these mechanisms.
Continued

1. May have a mixture of these mechanisms.
2. Display on a *Lexis diagram*: Time on study by calendar time. See Fig. 7.

![Lexis Diagram](image-url)

**Fig. 7: Lexis Diagram**
**Left censoring:**

1. Knowledge that failure time is less than some value replaces knowledge of exact time.
**Left censoring:**

1. Knowledge that failure time is less than some value replaces knowledge of exact time.
2. Ex., disease onset age for those who have disease at first examination.
**Left censoring:**

1. Knowledge that failure time is less than some value replaces knowledge of exact time.
2. Ex., disease onset age for those who have disease at first examination.
3. Ex., disease onset age for those who forgot when it came on.
More exotic censoring mechanisms

1. *double censoring* if either may happen
More exotic censoring mechanisms

1. *double censoring* if either may happen
2. *interval censoring* if your information is an interval.
More exotic censoring mechanisms

1. *double censoring* if either may happen
2. *interval censoring* if your information is an interval.
   - Ex., if a person is periodically screened for a disease.
Truncation:

1. Certain subjects omitted from data set.

left truncation: result of delayed entry
Those who have event before start are not recorded

right truncation: Those who haven't had event are not recorded
Ex.: data from death records.
Truncation:

1. Certain subjects omitted from data set.
2. left truncation:

   - result of delayed entry
   - Those who have event before start are not recorded
   - Those who haven't had event are not recorded

Ex.: data from death records.
**Truncation:**

1. Certain subjects omitted from data set.
2. *left truncation:*
   1. result of delayed entry
**Truncation:**

1. Certain subjects omitted from data set.
2. *left truncation:*
   1. result of delayed entry
   2. Those who have event before start are not recorded

Ex.: data from death records.
Truncation:

1. Certain subjects omitted from data set.
2. *left truncation*:
   1. result of delayed entry
   2. Those who have event before start are not recorded
3. *right truncation*: Those who haven’t had event are not recorded
Truncation:

1. Certain subjects omitted from data set.
2. *left truncation:*
   1. result of delayed entry
   2. Those who have event before start are not recorded
3. *right truncation:* Those who haven’t had event are not recorded
   1. Ex.: data from death records.
Objectives Lecture 02

1. Likelihood for censored data

Readings: KM §3.5, 4.1-4.2a, 4.2b, 4.2c, 4.3, 4.4

Censoring and Truncation: Types of censoring
Objectives Lecture 02

1. Likelihood for censored data
2. Product – Limit Estimation of the Survival Function

Readings: KM §3.5, 4.1-4.2a, 4.2b, 4.2c, 4.3, 4.4
Objectives Lecture 02

1. Likelihood for censored data
2. Product – Limit Estimation of the Survival Function
3. Variance of the survival function

Readings: KM §3.5, 4.1-4.2a, 4.2b, 4.2c, 4.3, 4.4
Objectives Lecture 02

1. Likelihood for censored data
2. Product – Limit Estimation of the Survival Function
3. Variance of the survival function
4. Estimating the Cumulative Hazard
Objectives Lecture 02

1. Likelihood for censored data
2. Product – Limit Estimation of the Survival Function
3. Variance of the survival function
4. Estimating the Cumulative Hazard
5. Survival function confidence intervals

Readings: KM § 3.5, 4.1-4.2a, 4.2b, 4.2c, 4.3, 4.4
Objectives Lecture 02

1. Likelihood for censored data
2. Product – Limit Estimation of the Survival Function
3. Variance of the survival function
4. Estimating the Cumulative Hazard
5. Survival function confidence intervals
6. Simultaneous bounds
Objectives Lecture 02

1. Likelihood for censored data
2. Product – Limit Estimation of the Survival Function
3. Variance of the survival function
4. Estimating the Cumulative Hazard
5. Survival function confidence intervals
6. Simultaneous bounds
7. Readings: KM §3.5, 4.1-4.2a, 4.2b, 4.2c, 4.3, 4.4
Section: Likelihood for censored data

Subsection: Likelihood Theory
Likelihood is density or mass function for data viewed as function of unknown parameters. Parameter values that make this large are considered more likely than those making it small. Pick as estimate the values making likelihood as large as possible: maximum likelihood. Log likelihood components for independent observations add.
Likelihood Definition

1. Likelihood is density or mass function for data viewed as function of unknown parameters.
Likelihood Definition

1. Likelihood is density or mass function for data viewed as function of unknown parameters.
2. Parameter values that make this large are considered more likely than those making it small.

Likelihood for censored data: Likelihood Theory Lecture 02
Likelihood Definition

1. Likelihood is density or mass function for data viewed as function of unknown parameters

2. Parameter values that make this large are considered more likely than those making it small

3. Pick as estimate the values making likelihood as large as possible: *maximum likelihood*

Likelihood for censored data:  Likelihood Theory Lecture 02
Likelihood Definition

1. Likelihood is density or mass function for data viewed as function of unknown parameters
2. Parameter values that make this large are considered more likely than those making it small
3. Pick as estimate the values making likelihood as large as possible: maximum likelihood
4. Log likelihood components for independent observations add.
Section: Likelihood for censored data

Subsection: Application to censored data
Contributions

1 Exact lifetimes:
1. **Exact lifetimes:**
   1. Continuous: \( f(T) = h(T)S(T) \)
Contributions

1. Exact lifetimes:
   1. Continuous: \( f(T) = h(T)S(T) \)
   2. Discrete: \( p(T) = h(T)S(T) \)
Contributions

1. Exact lifetimes:
   - Continuous: \( f(T) = h(T)S(T) \)
   - Discrete: \( p(T) = h(T)S(T) \)

2. Right censored observations: \( S(T) \)
Contributions

1. **Exact lifetimes:**
   - Continuous: \( f(T) = h(T)S(T) \)
   - Discrete: \( p(T) = h(T)S(T) \)

2. **Right censored observations:** \( S(T) \)

3. **Left censored observations:** \( 1 - S(T) \)
Contributions

1. **Exact lifetimes:**
   - Continuous: \( f(T) = h(T)S(T) \)
   - Discrete: \( p(T) = h(T)S(T) \)

2. **Right censored observations:** \( S(T) \)

3. **Left censored observations:** \( 1 - S(T) \)

4. **Interval Censored:** \( S(R) - S(L) \)
Contributions

1. Exact lifetimes:
   1. Continuous: \( f(T) = h(T)S(T) \)
   2. Discrete: \( p(T) = h(T)S(T) \)

2. Right censored observations: \( S(T) \)

3. Left censored observations: \( 1 - S(T) \)

4. Interval Censored: \( S(R) - S(L) \)

5. (Potentially) truncated observations: Divide by \( P[\text{No Truncation}] \)
Section: Estimation

Subsection: Survival function with Right Censoring
Notation

1. Pick $0 = t_0 < t_1 < t_2 < \cdots < t_D$ to contain observed event times.
Notation

1. Pick $0 = t_0 < t_1 < t_2 < \cdots < t_D$ to contain observed event times.
2. Let $Y_i$ be number *at risk* (alive and uncensored) at time $t_i$.
Notation

1. Pick $0 = t_0 < t_1 < t_2 < \cdots < t_D$ to contain observed event times
2. Let $Y_i$ be number \textit{at risk} (alive and uncensored) at time $t_i$
   - Includes those still alive and uncensored
Notation

1. Pick $0 = t_0 < t_1 < t_2 < \cdots < t_D$ to contain observed event times
2. Let $Y_i$ be number at risk (alive and uncensored) at time $t_i$
   1. Includes those still alive and uncensored
   2. Number censored at $t_i$ are included in risk set
Notation

1. Pick $0 = t_0 < t_1 < t_2 < \cdots < t_D$ to contain observed event times
2. Let $Y_i$ be number at risk (alive and uncensored) at time $t_i$
   - 1. Includes those still alive and uncensored
   - 2. Number censored at $t_i$ are included in risk set
   - 3. $\Rightarrow$ if event time $=$ censor time, censoring considered after event.

Estimation: Survival function with Right Censoring
Notation

1. Pick $0 = t_0 < t_1 < t_2 < \cdots < t_D$ to contain observed event times
2. Let $Y_i$ be number at risk (alive and uncensored) at time $t_i$
   1. Includes those still alive and uncensored
   2. Number censored at $t_i$ are included in risk set
   3. $\Rightarrow$ if event time $=$ censor time, censoring considered after event.
3. Let $d_j$ be number having event at $t_j$
Notation

1. Pick \(0 = t_0 < t_1 < t_2 < \cdots < t_D\) to contain observed event times

2. Let \(Y_i\) be number \textit{at risk} (alive and uncensored) at time \(t_i\)
   - Includes those still alive and uncensored
   - Number censored at \(t_i\) are included in risk set
   - \(\Rightarrow\) if event time = censor time, censoring considered after event.

3. Let \(d_j\) be number having event at \(t_j\)

4. Let \(h_j\) be hazard at \(t_j\)
Notation

1. Pick $0 = t_0 < t_1 < t_2 < \cdots < t_D$ to contain observed event times
2. Let $Y_i$ be number at risk (alive and uncensored) at time $t_i$
   - Includes those still alive and uncensored
   - Number censored at $t_i$ are included in risk set
   - $\Rightarrow$ if event time = censor time, censoring considered after event.
3. Let $d_j$ be number having event at $t_j$
4. Let $h_j$ be hazard at $t_j$
5. Then $Y_j - Y_{j+1} - d_j$ is number censored at $t_j$
Likelihood

\[ L = \prod_{j=1}^{D} f(t_j)^{d_j} S(t_j)^{Y_j - Y_{j+1} - d_j} \]
Likelihood

1. \[ L = \prod_{j=1}^{D} f(t_j)^{d_j} S(t_j)^{Y_j-Y_{j+1}-d_j} \]

2. For fixed \( t_j \), and with \( S(t_j) \) all fixed, \( L \) maximized when all the weight in \( (t_{j-1}, t_j] \) is on \( t_j \)
Likelihood

1. \[ L = \prod_{j=1}^{D} f(t_j)^{d_j} S(t_j)^{Y_j-Y_{j+1}-d_j} \]

2. For fixed \( t_j \), and with \( S(t_j) \) all fixed, \( L \) maximized when all the weight in \((t_{j-1}, t_j]\) is on \( t_j \)

3. Hence we can restrict search for maximizer to discrete distributions
Represent limit of continuous likelihoods as discrete likelihood.

1. For $\epsilon$ small, $f(t_j) \approx P[T \in (t_{j-1}, t_j)] / \epsilon$.  

$L \approx \prod_{Dj=1} P[T \in (t_{j-1}, t_j)] d_j S(t_j) Y_j - Y_j + 1 - d_j$.  

Let $h_j = P[T \in (t_{j-1}, t_j)] / S(t_{j-1})$.  

Ignore factor of $\epsilon$ because it does not include parameter.  

$L \approx \prod_{Dj=1} P[T \in (t_{j-1}, t_j)] d_j S(t_j) Y_j - Y_j + 1 - d_j$.  

For distribution with all of probability on the $t_j$, 

$L = D \prod_{j=1} P[T = t_j] d_j S(t_j) Y_j - Y_j + 1 - d_j$.  

Estimation : Survival function with Right Censoring Lecture 02
Represent limit of continuous likelihoods as discrete likelihood.

1. For $\epsilon$ small, $f(t_j) \approx P[T \in (t_{j-1}, t_j)]/\epsilon$.
2. $L \approx \prod_{j=1}^{D} P[T \in (t_j - \epsilon, t_j)]^{d_j} S(t_j)^{Y_j-Y_{j+1}-d_j} \epsilon^{-d_j}$.
Represent limit of continuous likelihoods as discrete likelihood.

1. For $\epsilon$ small, $f(t_j) \approx P[T \in (t_{j-1}, t_j)] / \epsilon$.
2. $L \approx \prod_{j=1}^{D} P[T \in (t_j - \epsilon, t_j)]^{d_j} S(t_j)^{Y_j-1-d_j} \epsilon^{-d_j}$.
3. Let $h_j = P[T \in (t_j - \epsilon, t_j)] / S(t_{j-1})$. 
Represent limit of continuous likelihoods as discrete likelihood.

1. For $\epsilon$ small, $f(t_j) \approx P[T \in (t_{j-1}, t_j)] / \epsilon$.
2. $L \approx \prod_{j=1}^{D} P[T \in (t_j - \epsilon, t_j)]^{d_j} S(t_j)^{Y_j-Y_{j+1}-d_j}\epsilon^{-d_j}$.
3. Let $h_j = P[T \in (t_j - \epsilon, t_j)] / S(t_{j-1})$.
4. Ignore factor of $\epsilon$ because it does not include parameter.
Represent limit of continuous likelihoods as discrete likelihood.

1. For $\epsilon$ small, $f(t_j) \approx P[T \in (t_{j-1}, t_j)] / \epsilon$.
2. $L \approx \prod_{j=1}^{D} P[T \in (t_j - \epsilon, t_j)]^{d_j} S(t_j)^{Y_j - Y_{j+1} - d_j} \epsilon^{-d_j}$.
3. Let $h_j = P[T \in (t_j - \epsilon, t_j)] / S(t_{j-1})$.
4. Ignore factor of $\epsilon$ because it does not include parameter.
5. $L \approx \prod_{j=1}^{D} P[T \in (t_j - \epsilon, t_j)]^{d_j} S(t_j)^{Y_j - Y_{j+1} - d_j}$.
Represent limit of continuous likelihoods as discrete likelihood.

1. For $\epsilon$ small, $f(t_j) \approx P[T \in (t_{j-1}, t_j)] / \epsilon$.
2. $L \approx \prod_{j=1}^{D} P[T \in (t_j - \epsilon, t_j)]^{d_j} S(t_j)^{Y_j-Y_{j+1}-d_j}$. 
3. Let $h_j = P[T \in (t_j - \epsilon, t_j)] / S(t_{j-1})$.
4. Ignore factor of $\epsilon$ because it does not include parameter.
5. $L \approx \prod_{j=1}^{D} P[T \in (t_j - \epsilon, t_j)]^{d_j} S(t_j)^{Y_j-Y_{j+1}-d_j}$.
6. For distribution with all of probability on the $t_j$,

$$L = \prod_{j=1}^{D} P[T = t_j]^{d_j} S(t_j)^{Y_j-Y_{j+1}-d_j}.$$
For this discrete distribution:

1. \[ S(t_j) = S(t_{j-1})(1 - h_j). \]

2. By induction, \[ S(t_j) = \prod_{i \leq j} (1 - h_i). \]

3. \[ p_j = P[T = t_j] = h_j S(t_j - 1) = h_j \prod_{i < j} (1 - h_i). \]

4. Multiply:
   \[ L = \prod_{D_j=1} h_d_j \left[ \prod_{i < j} (1 - h_i) \right] d_j \left[ \prod_{i \leq j} (1 - h_i) \right] Y_j - Y_j + 1 - d_j. \]

5. Interchange ordering:
   \[ L = \prod_{D_j=1} h_d_j \left[ \prod_{i \leq j} (1 - h_i) \right] Y_j - Y_j + 1 - d_j. \]

6. Log likelihood:
   \[ \ell = \sum_{D_j=1} d_j \log \left( \frac{h_j}{1 - h_j} \right) + \sum_{D_j=1} \sum_{D_i=1} (Y_j - Y_j + 1) \log (1 - h_i). \]

7. Distribute sum:
   \[ \ell = \sum_{D_j=1} d_j \log \left( \frac{h_j}{1 - h_j} \right) + \sum_{D_j=1} \sum_{D_i=1} (Y_j - Y_j + 1) \log (1 - h_i). \]

8. Interchange order of two sums:
   \[ \ell = \sum_{D_j=1} d_j \log \left( \frac{h_j}{1 - h_j} \right) + \sum_{D_i=1} \sum_{D_j=1} (Y_j - Y_j + 1) \log (1 - h_i). \]

9. Adjacent terms cancel:
   \[ \ell = \sum_{D_j=1} d_j \log \left( \frac{h_j}{1 - h_j} \right) + \sum_{D_j=1} \sum_{D_i=1} Y_i \log (1 - h_i). \]
For this discrete distribution:

1. \( S(t_j) = S(t_{j-1})(1 - h_j). \)
2. By induction, \( S(t_j) = \prod_{i \leq j}(1 - h_i). \)
For this discrete distribution:

1. \( S(t_j) = S(t_{j-1})(1 - h_j) \).
2. By induction, \( S(t_j) = \prod_{i \leq j}(1 - h_i) \).
3. \( p_j = P[T = t_j] = h_j S(t_{j-1}) = h_j \prod_{i < j}(1 - h_i) \)
For this discrete distribution:

1. \( S(t_j) = S(t_{j-1})(1 - h_j) \).
2. By induction, \( S(t_j) = \prod_{i \leq j}(1 - h_i) \).
3. \( p_j = P[T = t_j] = h_j S(t_{j-1}) = h_j \prod_{i < j}(1 - h_i) \)
4. Multiply: \( L = \prod_{j=1}^{D} h_j^{d_j} [\prod_{i < j}(1 - h_i)]^{d_j} [\prod_{i \leq j}(1 - h_i)]^{Y_j - Y_{j+1} - d_j} \).
For this discrete distribution:

1. \( S(t_j) = S(t_{j-1})(1 - h_j) \).
2. By induction, \( S(t_j) = \prod_{i \leq j} (1 - h_i) \).
3. \( p_j = \mathbb{P}[T = t_j] = h_j S(t_{j-1}) = h_j \prod_{i < j} (1 - h_i) \).
4. Multiply: \( L = \prod_{j=1}^{D} h_j^{d_j} \left[ \prod_{i < j} (1 - h_i) \right]^{d_j} \left[ \prod_{i \leq j} (1 - h_i) \right] Y_j - Y_{j+1} - d_j \).
5. Interchange ordering: \( L = \prod_{j=1}^{D} h_j^{d_j} \left[ (1 - h_j) \right]^{-d_j} \left[ \prod_{i \leq j} (1 - h_i) \right] Y_j - Y_{j+1} \)
For this discrete distribution:

1. \( S(t_j) = S(t_{j-1})(1 - h_j) \).
2. By induction, \( S(t_j) = \prod_{i \leq j} (1 - h_i) \).
3. \( p_j = P[T = t_j] = h_j S(t_{j-1}) = h_j \prod_{i < j} (1 - h_i) \).
4. Multiply: \( L = \prod_{j=1}^{D} h_j^{d_j} \left[ \prod_{i < j} (1 - h_i) \right]^{d_j} \left[ \prod_{i \leq j} (1 - h_i) \right]^{Y_j - Y_{j+1} - d_j} \).

   - Interchange ordering: \( L = \prod_{j=1}^{D} h_j^{d_j} \left[ (1 - h_j) \right]^{-d_j} \left[ \prod_{i \leq j} (1 - h_i) \right]^{Y_j - Y_{j+1}} \).
   - Log likelihood
     \[
     \ell = \sum_{j=1}^{D} [d_j \log(h_j/[1 - h_j])] + \sum_{i=1}^{j} (Y_j - Y_{j+1}) \log(1 - h_i)].
     \]
For this discrete distribution:

1. \( S(t_j) = S(t_{j-1})(1 - h_j). \)
2. By induction, \( S(t_{j}) = \prod_{i \leq j}(1 - h_i). \)
3. \( p_j = P[T = t_j] = h_j S(t_{j-1}) = h_j \prod_{i < j}(1 - h_i) \)
4. Multiply: \( L = \prod_{j=1}^{D} h_j^{d_j} [\prod_{i < j}(1 - h_i)]^{d_j} [\prod_{i \leq j}(1 - h_i)]^{Y_j - Y_{j+1} - d_j}. \)
   - Interchange ordering: \( L = \prod_{j=1}^{D} h_j^{d_j} [(1 - h_j)]^{-d_j} [\prod_{i \leq j}(1 - h_i)]^{Y_j - Y_{j+1}} \)
   - Log likelihood
     \[ \ell = \sum_{j=1}^{D} [d_j \log(h_j/[1 - h_j]) + \sum_{i=1}^{j} (Y_j - Y_{j+1}) \log(1 - h_i)]. \]
5. Distribute sum:
     \[ \ell = \sum_{j=1}^{D} d_j \log(h_j/[1 - h_j]) + \sum_{j=1}^{D} \sum_{i=1}^{j} (Y_j - Y_{j+1}) \log(1 - h_i) \]
For this discrete distribution:

1. \( S(t_j) = S(t_{j-1})(1 - h_j). \)
2. By induction, \( S(t_j) = \prod_{i \leq j}(1 - h_i). \)
3. \( p_j = P[T = t_j] = h_j S(t_{j-1}) = h_j \prod_{i < j}(1 - h_i) \)
4. Multiply: \( L = \prod_{j=1}^{D} h_j^{d_j} \left[ \prod_{i < j}(1 - h_i) \right]^{d_j} \left[ \prod_{i \leq j}(1 - h_i) \right]^{Y_j - Y_{j+1} - d_j} \)

1. Interchange ordering: \( L = \prod_{j=1}^{D} h_j^{d_j} \left[ (1 - h_j) \right]^{-d_j} \left[ \prod_{i \leq j}(1 - h_i) \right]^{Y_j - Y_{j+1}} \)
2. Log likelihood
   \[
   \ell = \sum_{j=1}^{D} d_j \log \left( \frac{h_j}{1 - h_j} \right) + \sum_{i=1}^{j} (Y_j - Y_{j+1}) \log(1 - h_i) \]
3. Distribute sum:
   \[
   \ell = \sum_{j=1}^{D} d_j \log \left( \frac{h_j}{1 - h_j} \right) + \sum_{j=1}^{D} \sum_{i=1}^{j} (Y_j - Y_{j+1}) \log(1 - h_i) \]
4. Interchange order of two sums
   \[
   \ell = \sum_{j=1}^{D} d_j \log \left( \frac{h_j}{1 - h_j} \right) + \sum_{i=1}^{D} \sum_{j=i}^{D} (Y_j - Y_{j+1}) \log(1 - h_i) \]

5. Adjacent terms cancel:
6. Hence:
   \[
   \ell = \sum_{j=1}^{D} d_j \log \left( \frac{h_j}{1 - h_j} \right) + \sum_{j=1}^{D} \sum_{i=1}^{j} (Y_j - Y_{j+1}) \log(1 - h_i) \]
For this discrete distribution:

1. \( S(t_j) = S(t_{j-1})(1 - h_j) \).
2. By induction, \( S(t_j) = \prod_{i \leq j}(1 - h_i) \).
3. \( p_j = P[T = t_j] = h_j S(t_{j-1}) = h_j \prod_{i < j}(1 - h_i) \)
4. Multiply: \( L = \prod_{j=1}^{D} h_j^{d_j} \left[ \prod_{i < j}(1 - h_i) \right]^{d_j} \left[ \prod_{i \leq j}(1 - h_i) \right]^{Y_j - Y_{j+1} - d_j} \).
   - Interchange ordering: \( L = \prod_{j=1}^{D} h_j^{d_j} \left[ (1 - h_j) \right]^{-d_j} \left[ \prod_{i \leq j}(1 - h_i) \right]^{Y_j - Y_{j+1}} \)
   - Log likelihood
     \[ \ell = \sum_{j=1}^{D} d_j \log(h_j/[1 - h_j]) + \sum_{j=1}^{D} (Y_j - Y_{j+1}) \log(1 - h_j). \]
5. Distribute sum:
   \[ \ell = \sum_{j=1}^{D} d_j \log(h_j/[1 - h_j]) + \sum_{j=1}^{D} \sum_{i=1}^{j} (Y_j - Y_{j+1}) \log(1 - h_i). \]
6. Interchange order of two sums
   \[ \ell = \sum_{j=1}^{D} d_j \log(h_j/[1 - h_j]) + \sum_{i=1}^{D} \sum_{j=i}^{D} (Y_j - Y_{j+1}) \log(1 - h_i). \]
7. Adjacent terms cancel:
For this discrete distribution:

1. \( S(t_j) = S(t_{j-1})(1 - h_j). \)
2. By induction, \( S(t_j) = \prod_{i \leq j} (1 - h_i). \)
3. \( p_j = P[T = t_j] = h_j S(t_{j-1}) = h_j \prod_{i < j} (1 - h_i) \)
4. Multiply: \( L = \prod_{j=1}^{D} h_j^{d_j} \left[ \prod_{i < j} (1 - h_i) \right]^{d_j} \left[ \prod_{i \leq j} (1 - h_i) \right]^{Y_j - Y_{j+1} - d_j}. \)

   1. Interchange ordering: \( L = \prod_{j=1}^{D} h_j^{d_j} \left[ (1 - h_j) \right]^{-d_j} \left[ \prod_{i \leq j} (1 - h_i) \right]^{Y_j - Y_{j+1}} \)
   2. Log likelihood
   \[ \ell = \sum_{j=1}^{D} d_j \log(h_j/[1 - h_j]) + \sum_{i=1}^{j} (Y_j - Y_{j+1}) \log(1 - h_i). \]
   3. Distribute sum:
   \[ \ell = \sum_{j=1}^{D} d_j \log(h_j/[1 - h_j]) + \sum_{j=1}^{D} \sum_{i=1}^{j} (Y_j - Y_{j+1}) \log(1 - h_i) \]
   4. Interchange order of two sums
   \[ \ell = \sum_{j=1}^{D} d_j \log(h_j/[1 - h_j]) + \sum_{i=1}^{D} \sum_{j=i}^{D} (Y_j - Y_{j+1}) \log(1 - h_i). \]
   5. Adjacent terms cancel:
   
   1. Use \( Y_{D+1} = 0 \): No one is at risk after least death or censoring time.
For this discrete distribution:

1. \( S(t_j) = S(t_{j-1})(1 - h_j) \).
2. By induction, \( S(t_j) = \prod_{i \leq j} (1 - h_i) \).
3. \( p_j = P[T = t_j] = h_j S(t_{j-1}) = h_j \prod_{i < j} (1 - h_i) \).
4. Multiply: \( L = \prod_{j=1}^D h_j^{d_j} \left[ \prod_{i < j} (1 - h_i) \right]^{d_j} \left[ \prod_{i \leq j} (1 - h_i) \right]^{Y_j - Y_{j+1} - d_j} \).

Interchange ordering: \( L = \prod_{j=1}^D h_j^{d_j} \left[ (1 - h_j) \right]^{-d_j} \left[ \prod_{i \leq j} (1 - h_i) \right]^{Y_j - Y_{j+1}} \).

Log likelihood

\[ \ell = \sum_{j=1}^D [d_j \log(h_j/[1 - h_j]) + \sum_{i=1}^j (Y_j - Y_{j+1}) \log(1 - h_i)] \].

Distribute sum:

\[ \ell = \sum_{j=1}^D d_j \log(h_j/[1 - h_j]) + \sum_{j=1}^D \sum_{i=1}^j (Y_j - Y_{j+1}) \log(1 - h_i) \].

Interchange order of two sums

\[ \ell = \sum_{j=1}^D d_j \log(h_j/[1 - h_j]) + \sum_{i=1} \sum_{j=i}^D (Y_j - Y_{j+1}) \log(1 - h_i) \].

Adjacent terms cancel:

1. Use \( Y_{D+1} = 0 \): No one is at risk after least death or censoring time.
2. Hence

\[ \ell = \sum_{j=1}^D d_j \log(h_j/[1 - h_j]) + \sum_{i=1} Y_i \log(1 - h_i) \].
For this discrete distribution:

1. \( S(t_j) = S(t_{j-1})(1 - h_j). \)

2. By induction, \( S(t_j) = \prod_{i \leq j}(1 - h_i). \)

3. \( p_j = P[T = t_j] = h_j S(t_{j-1}) = h_j \prod_{i < j}(1 - h_i) \)

4. Multiply: \( L = \prod_{j=1}^{D} h_j^{d_j} \left[ \prod_{i < j}(1 - h_i) \right]^{d_j} \left[ \prod_{i \leq j}(1 - h_i) \right]^{Y_j - Y_{j+1} - d_j}. \)

   - Interchange ordering: \( L = \prod_{j=1}^{D} h_j^{d_j} \left[ (1 - h_j) \right]^{-d_j} \left[ \prod_{i \leq j}(1 - h_i) \right]^{Y_j - Y_{j+1}} \)
   - Log likelihood
     \[ \ell = \sum_{j=1}^{D} d_j \log(\frac{h_j}{1 - h_j}) + \sum_{j=1}^{j} (Y_j - Y_{j+1}) \log(1 - h_i). \]

   - Distribute sum:
     \[ \ell = \sum_{j=1}^{D} d_j \log(\frac{h_j}{1 - h_j}) + \sum_{j=1}^{D} \sum_{i=1}^{j} (Y_j - Y_{j+1}) \log(1 - h_i) \]

   - Interchange order of two sums
     \[ \ell = \sum_{j=1}^{D} d_j \log(\frac{h_j}{1 - h_j}) + \sum_{i=1}^{D} \sum_{j=i}^{D} (Y_j - Y_{j+1}) \log(1 - h_i). \]

   - Adjacent terms cancel:
     - Use \( Y_{D+1} = 0 \): Noone is at risk after least death or censoring time.
     - Hence
     \[ \ell = \sum_{j=1}^{D} d_j \log(\frac{h_j}{1 - h_j}) + \sum_{i=1}^{D} Y_i \log(1 - h_i) \]

   - Hence: \( \ell = \sum_{j=1}^{D} d_j \log(h_j) + \sum_{j=1}^{D} (Y_j - d_j) \log(1 - h_j) \)

Estimation: Survival function with Right Censoring Lecture 02
Survival Function MLE

1. From before: \( \ell = \sum_{j=1}^{D} d_j \log(h_j) + \sum_{j=1}^{D} (Y_j - d_j) \log(1 - h_j) \)
Survival Function MLE

1. From before: \( \ell = \sum_{j=1}^{D} d_j \log(h_j) + \sum_{j=1}^{D} (Y_j - d_j) \log(1 - h_j) \)

2. \( \frac{d}{dh_j} \ell = \frac{d_j}{h_j} - \frac{(Y_j - d_j)}{(1 - h_j)} \)
Survival Function MLE

1. From before: \( \ell = \sum_{j=1}^{D} d_j \log(h_j) + \sum_{j=1}^{D} (Y_j - d_j) \log(1 - h_j) \)

2. \( \frac{d}{dh_j} \ell = \frac{d_j}{h_j} - \frac{(Y_j - d_j)}{(1 - h_j)} \)

3. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

4. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

5. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

6. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

7. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

8. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

9. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

10. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

11. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

12. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

13. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

14. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

15. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

16. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

17. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

18. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

19. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

20. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

21. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

22. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

23. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

24. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

25. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

26. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

27. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

28. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

29. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

30. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)
Survival Function MLE

1. From before: $\ell = \sum_{j=1}^{D} d_j \log(h_j) + \sum_{j=1}^{D} (Y_j - d_j) \log(1 - h_j)$

2. $\frac{d}{dh_j} \ell = \frac{d_j}{h_j} - \frac{(Y_j - d_j)}{(1 - h_j)}$

3. Hazard estimates satisfy $d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)}$

4. $\hat{h}_j = \frac{d_j}{Y_j}$ while $Y_j > 0$
Survival Function MLE

1. From before: \( \ell = \sum_{j=1}^{D} d_j \log(h_j) + \sum_{j=1}^{D} (Y_j - d_j) \log(1 - h_j) \)

2. \( \frac{d}{dh_j} \ell = d_j/h_j - (Y_j - d_j)/(1 - h_j) \)

3. Hazard estimates satisfy \( d_j/\hat{h}_j = (Y_j - d_j)/(1 - \hat{h}_j) \)

4. \( \hat{h}_j = d_j/Y_j \) while \( Y_j > 0 \)

5. \( \hat{S}(t_j) = \prod_{i \leq j}(1 - \hat{h}_i) = \prod_{i \leq j}(1 - d_j/Y_j) \) for \( t_j \) such that \( Y_j > 0 \)
Survival Function MLE

1. From before: \( \ell = \sum_{j=1}^{D} d_j \log(h_j) + \sum_{j=1}^{D} (Y_j - d_j) \log(1 - h_j) \)

2. \( \frac{d}{dh_j} \ell = \frac{d_j}{h_j} - \frac{(Y_j - d_j)}{(1 - h_j)} \)

3. Hazard estimates satisfy \( d_j/\hat{h}_j = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

4. \( \hat{h}_j = \frac{d_j}{Y_j} \) while \( Y_j > 0 \)

5. \( \hat{S}(t_j) = \prod_{i \leq j} (1 - \hat{h}_i) = \prod_{i \leq j} (1 - d_j/Y_j) \) for \( t_j \) such that \( Y_j > 0 \)

6. The product-limit estimator or Kaplan–Meier estimator.
Survival Function MLE

1. From before: \( \ell = \sum_{j=1}^{D} d_j \log(h_j) + \sum_{j=1}^{D} (Y_j - d_j) \log(1 - h_j) \)

2. \( \frac{d}{dh_j} \ell = \frac{d_j}{h_j} - \frac{(Y_j - d_j)}{(1 - h_j)} \)

3. Hazard estimates satisfy \( \frac{d_j}{\hat{h}_j} = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

4. \( \hat{h}_j = \frac{d_j}{Y_j} \) while \( Y_j > 0 \)

5. \( \hat{S}(t_j) = \prod_{i \leq j}(1 - \hat{h}_i) = \prod_{i \leq j}(1 - \frac{d_j}{Y_j}) \) for \( t_j \) such that \( Y_j > 0 \)

6. The product - limit estimator or Kaplan – Meier estimator.

7. Undefined if \( Y_j = 0 \)
Survival Function MLE

1. From before: \( \ell = \sum_{j=1}^{D} d_j \log(h_j) + \sum_{j=1}^{D} (Y_j - d_j) \log(1 - h_j) \)

2. \( \frac{d}{dh_j} \ell = \frac{d_j}{h_j} - \frac{(Y_j - d_j)}{(1 - h_j)} \)

3. Hazard estimates satisfy \( \frac{d_j}{\hat{h}_j} = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

4. \( \hat{h}_j = \frac{d_j}{Y_j} \) while \( Y_j > 0 \)

5. \( \hat{S}(t_j) = \prod_{i \leq j}(1 - \hat{h}_i) = \prod_{i \leq j}(1 - d_j/Y_j) \) for \( t_j \) such that \( Y_j > 0 \)

6. The product-limit estimator or Kaplan–Meier estimator.

7. Undefined if \( Y_j = 0 \)

8. Relevant if last event is censoring.
Survival Function MLE

1. From before: \( \ell = \sum_{j=1}^{D} d_j \log(h_j) + \sum_{j=1}^{D} (Y_j - d_j) \log(1 - h_j) \)

2. \( \frac{d}{dh_j} \ell = \frac{d_j}{h_j} - \frac{(Y_j - d_j)}{(1 - h_j)} \)

3. Hazard estimates satisfy \( \frac{d_j}{\hat{h}_j} = \frac{(Y_j - d_j)}{(1 - \hat{h}_j)} \)

4. \( \hat{h}_j = \frac{d_j}{Y_j} \) while \( Y_j > 0 \)

5. \( \hat{S}(t_j) = \prod_{i \leq j}(1 - \hat{h}_i) = \prod_{i \leq j}(1 - d_j/Y_j) \) for \( t_j \) such that \( Y_j > 0 \)

6. The product-limit estimator or Kaplan–Meier estimator.

7. Undefined if \( Y_j = 0 \)

8. Relevant if last event is censoring.

Heuristic interpretation: redistribution to the right
Survival Function MLE

1. From before: $\ell = \sum_{j=1}^{D} d_j \log(h_j) + \sum_{j=1}^{D} (Y_j - d_j) \log(1 - h_j)$
2. $\frac{d}{dh_j} \ell = d_j/h_j - (Y_j - d_j)/(1 - h_j)$
3. Hazard estimates satisfy $d_j/\hat{h}_j = (Y_j - d_j)/(1 - \hat{h}_j)$
4. $\hat{h}_j = d_j/Y_j$ while $Y_j > 0$
5. $\hat{S}(t_j) = \prod_{i \leq j} (1 - \hat{h}_i) = \prod_{i \leq j} (1 - d_j/Y_j)$ for $t_j$ such that $Y_j > 0$
6. The product-limit estimator or Kaplan–Meier estimator.
7. Undefined if $Y_j = 0$
   1. Relevant if last event is censoring.
8. Heuristic interpretation: redistribution to the right
   1. All censored observations have their probability redistributed to events to right
Alternative Method of Moments derivation:

From definition

\[ S(t_j) = P[X > t_j] = P[X > t_{j-1}] P[X > t_j | X > t_{j-1}] \]
Alternative Method of Moments derivation:

1. From definition

\[ S(t_j) = P[X > t_j] = P[X > t_{j-1}] P[X > t_j | X > t_{j-1}] \]

2. Continuing,

\[ S(t_j) = P[X > t_1] P[X > t_2 | X > t_1] \cdots P[X > t_j | X > t_{j-1}] \]
Alternative Method of Moments derivation:

1. From definition

\[ S(t_j) = P[X > t_j] = P[X > t_{j-1}] P[X > t_j | X > t_{j-1}] \]

2. Continuing,

\[ S(t_j) = P[X > t_1] P[X > t_2 | X > t_1] \cdots P[X > t_j | X > t_{j-1}] \]

3. Estimate \( P[X > t_i|X > t_{i-1}] \) as \( (1 - d_j/Y_j) \)
Continued

\[ \hat{S}(t_j) = \prod_{i \leq j} (1 - d_i / Y_i). \]
\[ \hat{S}(t_j) = \prod_{i \leq j} \left(1 - \frac{d_j}{Y_j}\right). \]

See Fig. 8.

**Fig. 8: Kaplan-Meier Estimator for Stage 2 Larynx Cancer Data**

Segment from \( t = 0 \) to \( t = .1 \). \( \hat{S}(.1) = 1. \)
\( \hat{S}(t_j) = \prod_{i \leq j} (1 - d_j / Y_j). \)

See Fig. 8.

**Fig. 8: Kaplan-Meier Estimator for Stage 2 Larynx Cancer Data**

\[ \hat{S}(0.2) = \hat{S}(0.1) \times \frac{16}{17} = 1 \times \frac{16}{17} = \frac{16}{17} = 0.94. \]

- Curve approaches, does not meet this point.
- Curve meets this point.

Estimation: Survival function with Right Censoring
\[
\hat{S}(t_j) = \prod_{i \leq j} \left(1 - \frac{d_j}{Y_j}\right).
\]

See Fig. 8.

**Fig. 8: Kaplan-Meier Estimator for Stage 2 Larynx Cancer Data**

\[t = .2 \text{ to } t = 1.8, \text{ 2nd event.}\]
\[\hat{S}(1.8) = \hat{S}(1.7) \times \frac{15}{16} = \frac{16}{17} \times \frac{15}{16} = \frac{15}{17} = .88.\]

- Curve approaches, does not meet this point.
- Curve meets this point.

Survival function with Right Censoring Lecture 02 42 / 260
Continued

\[ \hat{S}(t_j) = \prod_{i \leq j} \left(1 - \frac{d_j}{Y_j}\right). \]

See Fig. 8.

**Fig. 8: Kaplan-Meier Estimator for Stage 2 Larynx Cancer Data**

<table>
<thead>
<tr>
<th>Time (Months to nearest tenth)</th>
<th>Survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2, 1.8, 2.0, 2.2+, 2.6+, 3.3+, 3.6, 3.6+, 4.0, 4.3+, 4.3+, 5.0+, 6.2, 7.0, 7.5+, 7.6+, 9.3+</td>
<td></td>
</tr>
</tbody>
</table>

\[ \hat{S}(2.0) = \frac{15}{17} \times \frac{14}{15} = \frac{14}{17} = .82. \]
\[ \hat{S}(t_j) = \prod_{i \leq j} (1 - d_j/Y_j). \]

See Fig. 8.

**Fig. 8: Kaplan-Meier Estimator for Stage 2 Larynx Cancer Data**

- **Surveillance**
  - 0.2, 1.8, 2.0, 2.2+, 2.6+, 3.3+, 3.6, 3.6+, 4.0, 4.3+, 4.3+, 5.0+, 6.2, 7.0, 7.5+, 7.6+, 9.3+.
  - Curve approaches, does not meet this point.
  - Curve meets this point.

- **Time (Months to nearest tenth)**
  - 0.2, 1.8, 2.0, 2.2+, 2.6+, 3.3+, 3.6, 3.6+, 4.0, 4.3+, 4.3+, 5.0+, 6.2, 7.0, 7.5+, 7.6+, 9.3+.

- **Survival**
  - 0.2, 0.4, 0.6, 0.8, 1.0

- **Time intervals and events**
  - \( t = 2.0 \) to \( t = 2.2 \), 1st censoring.
  - \( \hat{S} \) unchanged.
  - \( t = 2.0 \) to \( t = 2.2 \), 1st censoring.
  - \( \hat{S} \) unchanged.
  - \( t = 2.2 \) to \( t = 2.6 \), 2nd event.
  - \( \hat{S} \) unchanged.
  - \( t = 2.6 \) to \( t = 3.3 \), 3rd event.
  - \( \hat{S} \) unchanged.
  - \( t = 3.3 \) to \( t = 3.6 \), 4th event.
  - \( \hat{S} \) unchanged.
  - \( t = 3.6 \) to \( t = 4.0 \), 5th event.
  - \( \hat{S} \) unchanged.
  - \( t = 4.0 \) to \( t = 4.3 \), 6th event.
  - \( \hat{S} \) unchanged.
  - \( t = 4.3 \) to \( t = 5.0 \), 7th event.
  - \( \hat{S} \) unchanged.
  - \( t = 5.0 \) to \( t = 6.2 \), 8th event.
  - \( \hat{S} \) unchanged.
  - \( t = 6.2 \) to \( t = 7.0 \), 9th event.
  - \( \hat{S} \) unchanged.
  - \( t = 7.0 \) to \( t = 7.5 \), 10th event.
  - \( \hat{S} \) unchanged.
  - \( t = 7.5 \) to \( t = 7.6 \), 11th event.
  - \( \hat{S} \) unchanged.
  - \( t = 7.6 \) to \( t = 9.3 \), 12th event.
  - \( \hat{S} \) unchanged.

- **Legend**
  - ● Curve meets this point.
  - ○ Curve approaches, does not meet this point.
  - ≤ 17 at risk
  - ≤ 16 at risk
  - ≤ 15 at risk
\( \hat{S}(t_j) = \prod_{i \leq j} (1 - d_j / Y_j). \)

See Fig. 8.

**Fig. 8: Kaplan-Meier Estimator for Stage 2 Larynx Cancer Data**

- \( t = 2.2 \) to \( t = 3.3 \), through 2nd, 3rd censoring. \( \hat{S} \) unchanged

- Curve approaches, does not meet this point.
- Curve meets this point.

0.2, 1.8, 2.0, 2.2+, 2.6+, 3.3+, 3.6, 3.6+, 4.0, 4.3+, 4.3+, 5.0+, 6.2, 7.0, 7.5+, 7.6+, 9.3+. 
Continued

\[ \hat{S}(t_j) = \prod_{i \leq j} (1 - d_j / Y_j). \]

See Fig. 8.

**Fig. 8: Kaplan-Meier Estimator for Stage 2 Larynx Cancer Data**

\[ t = 3.3 \text{ to } t = 3.6, \text{ to 4th event.} \]

\[ \hat{S}(3.6) = \hat{S}(3.5) \times \frac{10}{11} = \frac{14}{17} \times \frac{10}{11} = \frac{140}{187} = .75. \]

Note lack of cancellation in fractions.

Another subject is censored just after 3.6, but this does not impact \( \hat{S} \) until next event time.

- Curve approaches, does not meet this point.
- Curve meets this point.

Estimation: Survival function with Right Censoring
\( \hat{S}(t_j) = \prod_{i \leq j} (1 - d_j/Y_j) \).

See Fig. 8.

**Fig. 8: Kaplan-Meier Estimator for Stage 2 Larynx Cancer Data**

\( t = 3.6 \) to \( t = 4.0 \), to 5th event.

\( \hat{S}(4.0) = \frac{140}{187} \times \frac{8}{9} = .67 \)

- Curve approaches, does not meet this point.
- Curve meets this point.

0.2, 1.8, 2.0, 2.2+, 2.6+, 3.3+, 3.6, 3.6+, 4.0, 4.3+, 4.3+, 5.0+, 6.2, 7.0, 7.5+, 7.6+, 9.3+. 
\( \hat{S}(t_j) = \prod_{i \leq j} (1 - d_j/Y_j). \)

See Fig. 8.

Fig. 8: Kaplan-Meier Estimator for Stage 2 Larynx Cancer Data

3 more censored and then an event at 6.2.

- Curve approaches, does not meet this point.
- Curve meets this point.
\[ \hat{S}(t_j) = \prod_{i \leq j} (1 - d_j/Y_j). \]

See Fig. 8.

**Fig. 8: Kaplan-Meier Estimator for Stage 2 Larynx Cancer Data**

- Next event at 7.0, no intermediate censoring

- Curve approaches, does not meet this point.
- Curve meets this point.

0.2, 1.8, 2.0, 2.2+, 2.6+, 3.3+, 3.6, 3.6+, 4.0, 4.3+, 4.3+, 5.0+, 6.2, 7.0, 7.5+, 7.6+, 9.3+.
\[ \hat{S}(t_j) = \prod_{i \leq j} (1 - d_j / Y_j). \]

See Fig. 8.

**Fig. 8: Kaplan-Meier Estimator for Stage 2 Larynx Cancer Data**

3 censorings, 0 events, so \( \hat{S}(t) \) always \( > 0 \).
- Curve approaches, does not meet this point.
- Curve meets this point.
Time points with no event contribute nothing
Time points with no event contribute nothing.

Hence don’t worry about missing potential time(s) above.
Time points with no event contribute nothing

Hence don’t worry about missing potential time(s) above.

Argument implies \( \hat{S}(t) \) unbiased without censoring.
Time points with no event contribute nothing

1. Hence don't worry about missing potential time(s) above.

Argument implies \( \hat{S}(t) \) unbiased without censoring.

2. But it is biased in the case of censoring.
Time points with no event contribute nothing
  Hence don't worry about missing potential time(s) above.

Argument implies $\hat{S}(t)$ unbiased without censoring.
  But it is biased in the case of censoring.

$\hat{S}(t)$ is empirical estimator without censoring.
Suppose $g$ a differentiable function, with domain containing the support of $X$. 

The Taylor series approximation of $g(x)$ is:

$$g(x) \approx g(\mu) + (x - \mu) g'(\mu)$$

Hence

$$\text{Var}[g(X)] \approx \text{Var}[g(E[X]) + (X - E[X]) g'(E[X])]$$

Using rule for affine transformations,

$$\text{Var}[g(X)] \approx g'(E[X])^2 \text{Var}[X] \approx g'(X)^2 \text{Var}[X].$$

This is called the Delta method or Propagation of Errors.
Variance of Transformation of a Random Variable

1. Suppose $g$ a differentiable function, with domain containing the support of $X$.
2. Taylor series approximation

\[
T(x) \approx g(\mu) + (x - \mu) g'(\mu)
\]

Hence
\[
\text{Var}[g(X)] \approx \text{Var}[g(E[X]) + (X - E[X]) g'(E[X])]
\]

Using rule for affine transformations,
\[
\text{Var}[g(X)] \approx g'(E[X])^2 \text{Var}[X] \approx g'(X)^2 \text{Var}[X]
\]

Called Delta method or Propagation of Errors
Variance of Transformation of a Random Variable

1. Suppose $g$ a differentiable function, with domain containing the support of $X$.
2. Taylor series approximation
   
   $g(x) \approx g(\mu) + (x - \mu)g'(\mu)$
Variance of Transformation of a Random Variable

1. Suppose $g$ a differentiable function, with domain containing the support of $X$.

2. Taylor series approximation
   - $g(x) \approx g(\mu) + (x - \mu)g'(\mu)$
   - Hence $\text{Var}[g(X)] \approx \text{Var}[g(\text{E}[X]) + (X - \text{E}[X])g'(\text{E}[X])].$
Suppose $g$ a differentiable function, with domain containing the support of $X$.

Taylor series approximation

1. $g(x) \approx g(\mu) + (x - \mu)g'(\mu)$
2. Hence $\text{Var}[g(X)] \approx \text{Var}[g(\text{E}[X]) + (X - \text{E}[X])g'(\text{E}[X])]$.

Using rule for affine transformations,

$$\text{Var}[g(X)] \approx g'(\text{E}[X])^2\text{Var}[X] \approx g'(X)^2\text{Var}[X].$$
1. Suppose $g$ a differentiable function, with domain containing the support of $X$.

2. Taylor series approximation

   1. $g(x) \approx g(\mu) + (x - \mu)g'(\mu)$
   2. Hence $\text{Var}[g(X)] \approx \text{Var}[g(\mathbb{E}[X]) + (X - \mathbb{E}[X])g'(\mathbb{E}[X])]$.

3. Using rule for affine transformations,

   $\text{Var}[g(X)] \approx g'(\mathbb{E}[X])^2 \text{Var}[X] \approx g'(X)^2 \text{Var}[X]$.

4. Called *Delta method* or *Propagation of Errors*


Continued

1. Appropriate if \( \text{Var} [X] \) small relative to \( g''(x) \).

Fig. 9: Illustration of Propagation of Errors for Log Function
Appropriate if $\text{Var}[X]$ small relative to $g''(x)$.

See Fig. 9.

*Fig. 9: Illustration of Propagation of Errors for Log Function*
Variance of Survival Estimate

1. Easier to get \( \text{Var} \left[ \log(\hat{S}(t)) \right] = \text{Var} \left[ \sum_{j=1}^{i} \log(1 - d_j/Y_j) \right] \)
Variance of Survival Estimate

1. Easier to get $\text{Var} \left[ \log(\hat{S}(t)) \right] = \text{Var} \left[ \sum_{j=1}^{i} \log(1 - d_j/Y_j) \right]$

2. Addends close to independent
Easier to get $\text{Var} \left[ \log(\hat{S}(t)) \right] = \text{Var} \left[ \sum_{j=1}^{i} \log(1 - d_j/Y_j) \right]$

2. Addends close to independent

3. Variances add
Variance of Survival Estimate

1. Easier to get \( \text{Var} \left[ \log \left( \hat{S}(t) \right) \right] = \text{Var} \left[ \sum_{j=1}^{i} \log (1 - d_j / Y_j) \right] \)
2. Addends close to independent
3. Variances add
4. \( \text{Var} \left[ d_j / Y_j \right] = h_j (1 - h_j) / Y_j \)

Approximated by \( d_j / (Y_j - d_j) Y_j \), using \( \hat{h}_j = d_j / Y_j \).
Variance of Survival Estimate

1. Easier to get $\text{Var} \left[ \log(\hat{S}(t)) \right] = \text{Var} \left[ \sum_{j=1}^{i} \log(1 - d_j/Y_j) \right]$

2. Addends close to independent

3. Variances add

4. $\text{Var} \left[ d_j/Y_j \right] = h_j(1 - h_j)/Y_j$
   - $g(p) = \log(1 - p)$. $g'(p) = -1/(1 - p)$.

5. Approximated by $d_j/Y_j - \left( Y_j - d_j \right)/Y_j$, using $\hat{h}_j = d_j/Y_j$.

6. $\text{Var} \left[ \log(\hat{S}(t)) \right] \approx \sum_{i,j=1}^{i} d_j/Y_j - \left( Y_j - d_j \right)/Y_j$

7. $\text{Var} \left[ \hat{S}(t) \right] \approx \hat{S}(t)^2 \sum_{i,j=1}^{i} d_j/Y_j - \left( Y_j - d_j \right)/Y_j$
Variance of Survival Estimate

1. Easier to get $\text{Var} \left[ \log(\hat{S}(t)) \right] = \text{Var} \left[ \sum_{j=1}^{i} \log(1 - d_j/Y_j) \right]$

2. Addends close to independent

3. Variances add

4. $\text{Var} \left[ d_j/Y_j \right] = h_j (1 - h_j)/Y_j$
   - $g(p) = \log(1 - p)$, $g'(p) = -1/(1 - p)$.
   - $\text{Var} \left[ \log(1 - d_j/Y_j) \right] \approx h_j (1 - h_j)/[(1 - h_j)^2 Y_j] = h_j/[(1 - h_j) Y_j]$

5. Approximated by $d_j/[(Y_j - d_j) Y_j]$, using $\hat{h}_j = d_j/Y_j$.

6. $\text{Var} \left[ \log\hat{S}(t_i) \right] \approx \sum_{j=1}^{i} d_j/[(Y_j - d_j) Y_j]$

7. Formula called Greenwood's Formula for $t_i > \text{maximal time}$ if last event was censoring
Variance of Survival Estimate

1. Easier to get \( \text{Var} \left[ \log(\hat{S}(t)) \right] = \text{Var} \left[ \sum_{j=1}^{i} \log(1 - d_j/Y_j) \right] \)
2. Addends close to independent
3. Variances add
4. \( \text{Var} \left[ d_j/Y_j \right] = h_j(1 - h_j)/Y_j \)
   1. \( g(p) = \log(1 - p) \). \( g'(p) = -1/(1 - p) \).
   2. \( \text{Var} \left[ \log(1 - d_j/Y_j) \right] \approx h_j(1 - h_j)/[(1 - h_j)^2 Y_j] = h_j/[(1 - h_j) Y_j] \)
   3. Approximated by \( d_j/[(Y_j - d_j) Y_j] \), using \( \hat{h}_j = d_j/Y_j \).
Variance of Survival Estimate

1. Easier to get \( \text{Var} \left[ \log(\hat{S}(t)) \right] = \text{Var} \left[ \sum_{j=1}^{i} \log(1 - d_j/Y_j) \right] \)

2. Addends close to independent

3. Variances add

4. \( \text{Var} \left[ d_j/Y_j \right] = h_j(1 - h_j)/Y_j \)
   - \( g(p) = \log(1 - p), \quad g'(p) = -1/(1 - p). \)
   - \( \text{Var} \left[ \log(1 - d_j/Y_j) \right] \approx h_j(1 - h_j)/[(1 - h_j)^2 Y_j] = h_j/[(1 - h_j) Y_j] \)
   - Approximated by \( d_j/[(Y_j - d_j) Y_j] \), using \( \hat{h}_j = d_j/Y_j. \)

5. \( \text{Var} \left[ \log(\hat{S}(t_i)) \right] \approx \sum_{j=1}^{i} d_j/[(Y_j - d_j) Y_j] \)
Variance of Survival Estimate

1. Easier to get $\text{Var} \left[ \log(\hat{S}(t)) \right] = \text{Var} \left[ \sum_{j=1}^{i} \log(1 - d_j/Y_j) \right]$

2. Addends close to independent

3. Variances add

4. $\text{Var} \left[ d_j/Y_j \right] = h_j(1 - h_j)/Y_j$
   
   1. $g(p) = \log(1 - p)$. $g'(p) = -1/(1 - p)$.
   2. $\text{Var} \left[ \log(1 - d_j/Y_j) \right] \approx h_j(1 - h_j)/[(1 - h_j)^2 Y_j] = h_j/[(1 - h_j) Y_j]$
   3. Approximated by $d_j/[(Y_j - d_j) Y_j]$, using $\hat{h}_j = d_j/Y_j$.

5. $\text{Var} \left[ \log(\hat{S}(t_i)) \right] \approx \sum_{j=1}^{i} d_j/[(Y_j - d_j) Y_j]$

6. $\text{Var} \left[ \hat{S}(t_i) \right] \approx \hat{S}(t_i)^2 \sum_{j=1}^{i} d_j/[(Y_j - d_j) Y_j]$
Variance of Survival Estimate

1. Easier to get $\text{Var} \left[ \log(\hat{S}(t)) \right] = \text{Var} \left[ \sum_{j=1}^{i} \log(1 - d_j/Y_j) \right]$

2. Addends close to independent

3. Variances add

4. $\text{Var} \left[ d_j/Y_j \right] = h_j(1 - h_j)/Y_j$
   - $g(p) = \log(1 - p)$. $g'(p) = -1/(1 - p)$.
   - $\text{Var} \left[ \log(1 - d_j/Y_j) \right] \approx h_j(1 - h_j)/[(1 - h_j)^2 Y_j] = h_j/[(1 - h_j)Y_j]$
   - Approximated by $d_j/[(Y_j - d_j)Y_j]$, using $\hat{h}_j = d_j/Y_j$.

5. $\text{Var} \left[ \log(\hat{S}(t_i)) \right] \approx \sum_{j=1}^{i} d_j/[(Y_j - d_j)Y_j]$

6. $\text{Var} \left[ \hat{S}(t_i) \right] \approx \hat{S}(t_i)^2 \sum_{j=1}^{i} d_j/[(Y_j - d_j)Y_j]$

7. Formula called Greenwood’s Formula
Variance of Survival Estimate

1. Easier to get \( \text{Var} \left[ \log(\hat{S}(t)) \right] = \text{Var} \left[ \sum_{j=1}^{i} \log(1 - d_{j}/Y_{j}) \right] \)

2. Addends close to independent

3. Variances add

4. \( \text{Var} \left[ d_{j}/Y_{j} \right] = h_{j}(1 - h_{j})/Y_{j} \)
   1. \( g(p) = \log(1 - p) \). \( g'(p) = -1/(1 - p) \).
   2. \( \text{Var}[\log(1 - d_{j}/Y_{j})] \approx h_{j}(1 - h_{j})/[(1 - h_{j})^{2} Y_{j}] = \hat{h}_{j}/[(1 - \hat{h}_{j}) Y_{j}] \)
   3. Approximated by \( d_{j}/[(Y_{j} - d_{j}) Y_{j}], \) using \( \hat{h}_{j} = d_{j}/Y_{j} \).

5. \( \text{Var} \left[ \log(\hat{S}(t_{i})) \right] \approx \sum_{j=1}^{i} d_{j}/[(Y_{j} - d_{j}) Y_{j}] \)

6. \( \text{Var} \left[ \hat{S}(t_{i}) \right] \approx \hat{S}(t_{i})^{2} \sum_{j=1}^{i} d_{j}/[(Y_{j} - d_{j}) Y_{j}] \)

7. Formula called Greenwood’s Formula
   1. 0 for \( t > \) maximal time if last event was censoring
Case with no censoring

1. Kaplan-Meier curve degenerates to 1—empirical CDF
Case with no censoring

1. Kaplan-Meier curve degenerates to 1—empirical CDF
2. Jump of $1/n$ at each failure
Case with no censoring

1. Kaplan-Meier curve degenerates to $1$—empirical CDF
2. Jump of $1/n$ at each failure
3. Since previous numerator cancels with each denominator.
Case with no censoring

1. Kaplan-Meier curve degenerates to 1—empirical CDF
2. Jump of $1/n$ at each failure
3. Since previous numerator cancels with each denominator.
4. Greenwood’s formula is $n[1 - \hat{S}(t)]\hat{S}(t)$, since
Case with no censoring

1. Kaplan-Meier curve degenerates to $1$—empirical CDF
2. Jump of $1/n$ at each failure
3. Since previous numerator cancels with each denominator.
4. Greenwood’s formula is $n[1 - \hat{S}(t)]\hat{S}(t)$, since
   - $Y_i = n + 1 - i$ and $d_i = 1$
Case with no censoring

1. Kaplan-Meier curve degenerates to $1$—empirical CDF
2. Jump of $1/n$ at each failure
3. Since previous numerator cancels with each denominator.
4. Greenwood’s formula is $n[1 - \hat{S}(t)]\hat{S}(t)$, since
   1. $Y_i = n + 1 - i$ and $d_i = 1$
   2. $\sum_{i=1}^{m} 1/[(n + 1 - i)(n - i)] = 1/(n - m) - 1/n = m/[n(n - m)] = n[1 - \hat{S}(t)]/\hat{S}(t)$
Case with no censoring

1. Kaplan-Meier curve degenerates to $1$—empirical CDF
2. jump of $1/n$ at each failure
3. Since previous numerator cancels with each denominator.
4. Greenwood’s formula is $n[1 - \hat{S}(t)]\hat{S}(t)$, since
   
   $\begin{align*}
   Y_i &= n + 1 - i \text{ and } d_i = 1 \\
   \sum_{i=1}^{m} \frac{1}{(n + 1 - i)(n - i)} &= 1/(n - m) - 1/n = m/[n(n - m)] = n[1 - \hat{S}(t)]/\hat{S}(t)
   \end{align*}$

   exactly what it should be for binomial
Confidence intervals for $S(t)$

1. point-wise
Confidence intervals for $S(t)$

1. **point-wise**

2. Use $\hat{S}(t) \pm 1.96 \times \sqrt{\text{Var}[\hat{S}(t)]}$
Confidence intervals for $S(t)$

1. **point-wise**

2. Use $\hat{S}(t) \pm 1.96 \times \sqrt{\text{Var}[\hat{S}(t)]}$

3. Problems:
Confidence intervals for $S(t)$

1. **point-wise**

2. Use $\hat{S}(t) \pm 1.96 \times \sqrt{\text{Var}\left[\hat{S}(t)\right]}$

3. Problems:
   - Heuristically, lower limit ought to fall when we have censoring,
Confidence intervals for $S(t)$

1. **point-wise**

2. Use $\hat{S}(t) \pm 1.96 \times \sqrt{\text{Var}[\hat{S}(t)]}$

3. Problems:
   1. Heuristically, lower limit ought to fall when we have censoring,
   2. since we might have more events there that we can't see than at the last event

SAS Code  R Code
Continuous CI can fall outside [0, 1]. See Fig. 10.

Fig. 10: Product Limit Estimator for Stage 2, with Raw Confidence Intervals

Estimation: Survival function with Right Censoring
CI can fall outside $[0, 1]$. 

See Fig. 10.

Fig. 10: Product Limit Estimator for Stage 2, with Raw Confidence Intervals
CI can fall outside $[0, 1]$.

See Fig. 10.

Fig. 10: Product Limit Estimator for Stage 2, with Raw Confidence Intervals
Variance estimated as 0 when $\hat{S}(t) = 0$: unrealistic

Solution:

1. Do CI for $\log(\hat{S}(t))$

   \[\exp(\log(\hat{S}(t)) \pm 1.96 \times \sqrt{\text{Var}[\log \hat{S}(t)]})\]

   CI won't fall below zero

   Variance won't go to zero

2. Use another transformation like arcsin or logistic to map the real line to $[0, 1]$ to solve problem at 1 as well.
Continued

1. Variance estimated as 0 when $\hat{S}(t) = 0$: unrealistic
Continued

1. Continued
   1. Variance estimated as 0 when $\hat{S}(t) = 0$: unrealistic
   2. Solution:

\[
\text{Do CI for } \log(\hat{S}(t)) = \exp(\log(\hat{S}(t)) \pm 1.96 \times \sqrt{\text{Var} \left[ \log \hat{S}(t) \right]})
\]

CI won't fall below zero

Variance won't go to zero

Use another transformation like arcsin or logistic to map the real line to $[0, 1]$ to solve problem at 1 as well.
Continued

1. Continued
   1. Variance estimated as 0 when $\hat{S}(t) = 0$: unrealistic

2. Solution:
   1. Do CI for $\log(S(t))$

Variance estimated as 0 when $\hat{S}(t) = 0$: unrealistic

Solution:

Do CI for $\log(S(t))$

$$\exp(\log(\hat{S}(t))) \pm 1.96 \times \sqrt{\text{Var} \left[ \log \hat{S}(t) \right]}$$
Continued

1. **Continued**
   1. Variance estimated as 0 when $\hat{S}(t) = 0$: unrealistic
2. **Solution:**
   1. Do CI for $\log(S(t))$
      1. $\exp(\log(\hat{S}(t)) \pm 1.96 \times \sqrt{\text{Var} \left[ \log \hat{S}(t) \right]})$
   2. CI won’t fall below zero
Continued

1. Variance estimated as 0 when $\hat{S}(t) = 0$: unrealistic

2. Solution:
   1. Do CI for $\log(S(t))$
      - $\exp(\log(\hat{S}(t))) \pm 1.96 \times \sqrt{\text{Var} \left[ \log \hat{S}(t) \right]}$
   2. CI won’t fall below zero
   3. Variance won’t go to zero

Use another transformation like arcsin or logistic to map the real line to $[0, 1]$ to solve problem at 1 as well.
Continued

1. Variance estimated as 0 when $\hat{S}(t) = 0$: unrealistic

2. Solution:
   1. Do CI for $\log(S(t))$
      
      \[ \exp(\log(\hat{S}(t))) \pm 1.96 \times \sqrt{\text{Var}[\log \hat{S}(t)]} \]

   2. CI won’t fall below zero
   3. Variance won’t go to zero

2. Use another transformation like arcsin or logistic to map the real line to $[0, 1]$ to solve problem at 1 as well. R Code  SAS Code
Simultaneous intervals:

1. Bonferroni inadequate, since bands must hold at an infinite number of points.
Simultaneous intervals:

1. Bonferroni inadequate, since bands must hold at an infinite number of points.
2. Use correlation of estimates of survival function at nearby points.
Simultaneous intervals:

1. Bonferroni inadequate, since bands must hold at an infinite number of points
2. Use correlation of estimates of survival function at nearby points.
3. Bound using sophisticated results from Stochastic Processes
Simultaneous intervals:

1. Bonferroni inadequate, since bands must hold at an infinite number of points
2. Use correlation of estimates of survival function at nearby points.
3. Bound using sophisticated results from Stochastic Processes
4. Does it depend on life distribution?

No, because we could re-scale time to make lives exponential.

Does depend on relationship between life and censoring distribution.

For instance, if most of censoring happens before median of life times estimate will be more variable than if it happens after.
Simultaneous intervals:

1. Bonferroni inadequate, since bands must hold at an infinite number of points
2. Use correlation of estimates of survival function at nearby points.
3. Bound using sophisticated results from Stochastic Processes
4. Does it depend on life distribution?
   1. No, because we could re-scale time to make lives exponential
Simultaneous intervals:

1. Bonferroni inadequate, since bands must hold at an infinite number of points.
2. Use correlation of estimates of survival function at nearby points.
4. Does it depend on life distribution?
   1. No, because we could re-scale time to make lives exponential.
   2. Does depend on relationship between life and censoring distribution.
Simultaneous intervals:

1. Bonferroni inadequate, since bands must hold at an infinite number of points
2. Use correlation of estimates of survival function at nearby points.
3. Bound using sophisticated results from Stochastic Processes
4. Does it depend on life distribution?
   1. No, because we could re-scale time to make lives exponential
   2. Does depend on relationship between life and censoring distribution.
   3. For instance, if most of censoring happens before median of life times estimate will be more variable than if it happens after
Monte Carlo Experiment to check simultaneous Coverage

1. Setup

- Setup
  - Life distribution exponential
  - Censoring distribution exponential with twice the mean
  - 20 obs
  - A large number (here 1000) of times, draw random data set and evaluate interval
    - Draw 20 exponentials
    - Draw 20 exponentials with twice the mean
    - Run routine to calculate confidence intervals
  - Check how often interval covers truth
    - Point-wise: at some pre-specified time
    - Simultaneous: all times
    - Need only check at jumps

Simultaneous coverage about 75%
Monte Carlo Experiment to check simultaneous Coverage

1 Setup
   1 Life distribution exponential
Monte Carlo Experiment to check simultaneous Coverage

1 Setup
   1 Life distribution exponential
   2 Censoring distribution exponential with twice the mean
Monte Carlo Experiment to check simultaneous Coverage

Setup

1. Life distribution exponential
2. Censoring distribution exponential with twice the mean
3. 20 obs
Monte Carlo Experiment to check simultaneous Coverage

1. Setup
   1. Life distribution exponential
   2. Censoring distribution exponential with twice the mean
   3. 20 obs

2. A large number (here 1000) of times, draw random data set and evaluate interval
Monte Carlo Experiment to check simultaneous Coverage

1. **Setup**
   1. Life distribution exponential
   2. Censoring distribution exponential with twice the mean
   3. 20 obs

2. **A large number (here 1000) of times, draw random data set and evaluate interval**
   1. Draw 20 exponentials

R Code

Estimation : Survival function with Right Censoring
Monte Carlo Experiment to check simultaneous Coverage

1. Setup
   1. Life distribution exponential
   2. Censoring distribution exponential with twice the mean
   3. 20 obs

2. A large number (here 1000) of times, draw random data set and evaluate interval
   1. Draw 20 exponentials
   2. Draw 20 exponentials with twice the mean
Monte Carlo Experiment to check simultaneous Coverage

1. Setup
   1. Life distribution exponential
   2. Censoring distribution exponential with twice the mean
   3. 20 obs

2. A large number (here 1000) of times, draw random data set and evaluate interval
   1. Draw 20 exponentials
   2. Draw 20 exponentials with twice the mean
   3. Run routine to calculate confidence intervals
Monte Carlo Experiment to check simultaneous Coverage

1. Setup
   1. Life distribution exponential
   2. Censoring distribution exponential with twice the mean
   3. 20 obs

2. A large number (here 1000) of times, draw random data set and evaluate interval
   1. Draw 20 exponentials
   2. Draw 20 exponentials with twice the mean
   3. Run routine to calculate confidence intervals

3. Check how often interval covers truth
Monte Carlo Experiment to check simultaneous Coverage

1. Setup
   1. Life distribution exponential
   2. Censoring distribution exponential with twice the mean
   3. 20 obs

2. A large number (here 1000) of times, draw random data set and evaluate interval
   1. Draw 20 exponentials
   2. Draw 20 exponentials with twice the mean
   3. Run routine to calculate confidence intervals

3. Check how often interval covers truth
   1. point-wise: at some pre-specified time
Monte Carlo Experiment to check simultaneous Coverage

1 Setup
   1 Life distribution exponential
   2 Censoring distribution exponential with twice the mean
   3 20 obs

2 A large number (here 1000) of times, draw random data set and evaluate interval
   1 Draw 20 exponentials
   2 Draw 20 exponentials with twice the mean
   3 Run routine to calculate confidence intervals

3 Check how often interval covers truth
   1 point-wise: at some pre-specified time
   2 simultaneous: all times
Monte Carlo Experiment to check simultaneous Coverage

1. Setup
   1. Life distribution exponential
   2. Censoring distribution exponential with twice the mean
   3. 20 obs

2. A large number (here 1000) of times, draw random data set and evaluate interval
   1. Draw 20 exponentials
   2. Draw 20 exponentials with twice the mean
   3. Run routine to calculate confidence intervals

3. Check how often interval covers truth
   1. point-wise: at some pre-specified time
   2. simultaneous: all times
   3. Need only check at jumps
Monte Carlo Experiment to check simultaneous Coverage

1. Setup
   1. Life distribution exponential
   2. Censoring distribution exponential with twice the mean
   3. 20 obs

2. A large number (here 1000) of times, draw random data set and evaluate interval
   1. Draw 20 exponentials
   2. Draw 20 exponentials with twice the mean
   3. Run routine to calculate confidence intervals

3. Check how often interval covers truth
   1. point-wise: at some pre-specified time
   2. simultaneous: all times
   3. Need only check at jumps

4. Simultaneous coverage about 75%  

R Code
Section: Estimation

Subsection: Estimate Cumulative Hazard
Definition: Integral of hazard function for continuous distributions

1. Use Approximate Log of Kaplan-Meier Estimate

\[ H(t_i) = -\log(S(t_i)) \]

Could use \( \hat{H}(t_i) = -\sum_{j=1}^{i} \log(1 - d_j/Y_j) \) from above

Note

\[-\log(1 - x) = x + x^2/2 + x^3/3 + \cdots\]

Use \( \hat{H}(t_i) = \sum_{j=1}^{i} d_j/Y_j \): Nelson–Aalen estimate

As before, \( \text{Var}[\hat{H}(t_i)] = \sum_{j=1}^{i} d_j(Y_j - d_j)/Y_j^3 \)

Typically too small; people usually substitute \( Y_j \) in place of \( Y_j - d_j \) to get \( \text{Var}[\hat{H}(t_i)] = d_j/Y_j^2 \)

Estimating \( S \) by \( \exp(-H) \) is called Altshuler’s estimate
Definition: Integral of hazard function for continuous distributions

1. Use Approximate Log of Kaplan-Meier Estimate

\[ H(t_i) = -\log(S(t_i)) \]

Note

\[ -\log(1 - x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots \]

Typically too small; people usually substitute \( Y_j \) in place of \( Y_j - d_j \) to get

\[ \text{Var}[\hat{H}(t_i)] = \frac{d_j}{Y_j^2} \]

Estimating \( S \) by \( \exp(-H) \) is called Altshuler’s estimate.
Definition: Integral of hazard function for continuous distributions

1. Use Approximate Log of Kaplan-Meier Estimate

\[ H(t_i) = -\log(S(t_i)) \]

2. Could use \( \hat{H}(t_i) = -\sum_{j=1}^{i} \log(1 - d_j/Y_j) \) from above

Note

\[-\log(1 - x) = x + x^2/2 + x^3/3 + \cdots\]

Use \( \hat{H}(t_i) = \sum_{j=1}^{i} d_j/Y_j \): Nelson–Aalen estimate

As before, \( \text{Var}[\hat{H}(t_i)] = \sum_{j=1}^{i} d_j/Y_j^3 \)

Typically too small; people usually substitute \( Y_j \) in place of \( Y_j - d_j \) to get \( \text{Var}[\hat{H}(t_i)] = d_j/Y_j^2 \)

Estimating \( S \) by \( \exp(-H) \) is called Altshuler's estimate

R Code

SAS

Code

Estimation : Estimate Cumulative Hazard Lecture 02
Definition: Integral of hazard function for continuous distributions

1. Use Approximate Log of Kaplan-Meier Estimate
2. \( H(t_i) = -\log(S(t_i)) \)
3. Could use \( \hat{H}(t_i) = -\sum_{j=1}^{i} \log(1 - d_j/Y_j) \) from above
4. Note \(-\log(1 - x) = x + x^2/2 + x^3/3 + \cdots\)
Definition: Integral of hazard function for continuous distributions

1. Use Approximate Log of Kaplan-Meier Estimate
   \[ H(t_i) = - \log(S(t_i)) \]

2. Could use \( \hat{H}(t_i) = - \sum_{j=1}^{i} \log(1 - d_j/Y_j) \) from above

3. Note \(-\log(1 - x) = x + x^2/2 + x^3/3 + \cdots\)

4. Use \( \hat{H}(t_i) = \sum_{j=1}^{i} d_j/Y_j; \) Nelson–Aalen estimate
Definition: Integral of hazard function for continuous distributions

1. Use Approximate Log of Kaplan-Meier Estimate

\[ H(t_i) = -\log(S(t_i)) \]

2. \[ H(t_i) = -\log(S(t_i)) \]

3. Could use \( \hat{H}(t_i) = -\sum_{j=1}^{i} \log(1 - d_j / Y_j) \) from above

4. Note \( -\log(1 - x) = x + x^2/2 + x^3/3 + \cdots \)

5. Use \( \hat{H}(t_i) = \sum_{j=1}^{i} d_j / Y_j \): Nelson–Aalen estimate

6. As before, \( \text{Var} \left[ \hat{H}(t_i) \right] = \sum_{j=1}^{i} d_j (Y_j - d_j)/Y_j^3 \)
Definition: Integral of hazard function for continuous distributions

1. Use Approximate Log of Kaplan-Meier Estimate
   \[ H(t_i) = -\log(S(t_i)) \]

2. Could use \( \hat{H}(t_i) = -\sum_{j=1}^{i} \log(1 - d_j/Y_j) \) from above

3. Note \(-\log(1 - x) = x + x^2/2 + x^3/3 + \cdots\)

4. Use \( \hat{H}(t_i) = \sum_{j=1}^{i} d_j/Y_j \): Nelson–Aalen estimate

5. As before, \( \text{Var} \left[ \hat{H}(t_i) \right] = \sum_{j=1}^{i} d_j(Y_j - d_j)/Y_j^3 \)

6. Typically too small; people usually substitute \( Y_j \) in place of \( Y_j - d_j \) to get \( \text{Var} \left[ \hat{H}(t_i) \right] = d_j/Y_j^2 \)
Definition: Integral of hazard function for continuous distributions

1. Use Approximate Log of Kaplan-Meier Estimate

\[ H(t_i) = -\log(S(t_i)) \]

2. Could use \( \hat{H}(t_i) = -\sum_{j=1}^{i} \log(1 - d_j/Y_j) \) from above

3. Note \(-\log(1-x) = x + x^2/2 + x^3/3 + \cdots\)

4. Use \( \hat{H}(t_i) = \sum_{j=1}^{i} d_j/Y_j \): Nelson–Aalen estimate

5. As before, \( \text{Var} \left[ \hat{H}(t_i) \right] = \sum_{j=1}^{i} d_j(Y_j - d_j)/Y_j^3 \)

6. Typically too small; people usually substitute \( Y_j \) in place of \( Y_j - d_j \) to get \( \text{Var} \left[ \hat{H}(t_i) \right] = d_j/Y_j^2 \)

7. Estimating \( S \) by \( \exp(-H) \) is called Altshuler’s estimate

R Code

SAS Code
Objectives Lecture 03

1. Quantile confidence intervals
Objectives Lecture 03

1. Quantile confidence intervals
2. Estimating means
Objectives Lecture 03

1. Quantile confidence intervals
2. Estimating means
3. Estimation with truncation
Objectives Lecture 03

1. Quantile confidence intervals
2. Estimating means
3. Estimation with truncation
4. Life table estimates
Objectives Lecture 03

1. Quantile confidence intervals
2. Estimating means
3. Estimation with truncation
4. Life table estimates
5. Introduction to survival curve hypothesis testing
Objectives Lecture 03

1. Quantile confidence intervals
2. Estimating means
3. Estimation with truncation
4. Life table estimates
5. Introduction to survival curve hypothesis testing
6. One sample testing
Objectives Lecture 03

1. Quantile confidence intervals
2. Estimating means
3. Estimation with truncation
4. Life table estimates
5. Introduction to survival curve hypothesis testing
6. One sample testing
7. Log rank statistic variance
Objectives Lecture 03

1. Quantile confidence intervals
2. Estimating means
3. Estimation with truncation
4. Life table estimates
5. Introduction to survival curve hypothesis testing
6. One sample testing
7. Log rank statistic variance
8. Tests based on ranks
Objectives Lecture 03

1. Quantile confidence intervals
2. Estimating means
3. Estimation with truncation
4. Life table estimates
5. Introduction to survival curve hypothesis testing
6. One sample testing
7. Log rank statistic variance
8. Tests based on ranks
9. Readings: KM §4.5a, 4.5b, 4.6, 5.4, 7.1, 7.2, 7.3b, L §8.2
Section: Estimation

Subsection: Estimating location measures:
Define $p$ quantile $\nu$ to satisfy

$F(\nu) \geq p, \quad F(\nu-) \leq p$
Define $p$ quantile $\nu$ to satisfy

1. $F(\nu) \geq p, \ F(\nu-) \leq p$
2. $S(\nu) \leq 1 - p, \ S(\nu-) \geq 1 - p$
Define $p$ quantile $\nu$ to satisfy

1. $F(\nu) \geq p$, $F(\nu-) \leq p$
2. $S(\nu) \leq 1 - p$, $S(\nu-) \geq 1 - p$
3. $\nu = S^{-1}(1 - p)$ as long as distribution is continuous; Assume this.
Estimating quantiles

1 Behaves better than mean because it does not depend on extremes: robust
Estimating quantiles

1. Behaves better than mean because it does not depend on extremes: robust
2. Estimate as $U$ such that $\hat{S}(U) \leq 1 - p$, $\hat{S}(U-) \geq 1 - p$
Estimating quantiles

1. Behaves better than mean because it does not depend on extremes: robust

2. Estimate as $U$ such that $\hat{S}(U) \leq 1 - p$, $\hat{S}(U-) \geq 1 - p$
   - Uniquely defined unless curve has a flat spot with value $p$
Estimating quantiles

1. Behaves better than mean because it does not depend on extremes: robust
2. Estimate as $U$ such that $\hat{S}(U) \leq 1 - p$, $\hat{S}(U-) \geq 1 - p$
   1. Uniquely defined unless curve has a flat spot with value $p$
   2. Value is place where you jump through $p$
Estimating quantiles

1. Behaves better than mean because it does not depend on extremes: robust
2. Estimate as $U$ such that $\hat{S}(U) \leq 1 - p$, $\hat{S}(U-) \geq 1 - p$
   1. Uniquely defined unless curve has a flat spot with value $p$
   2. Value is place where you jump through $p$
   3. Otherwise can be any value on that flat spot.
Estimating quantiles

1. Behaves better than mean because it does not depend on extremes: robust

2. Estimate as $U$ such that $\hat{S}(U) \leq 1 - p$, $\hat{S}(U-) \geq 1 - p$
   - 1. Uniquely defined unless curve has a flat spot with value $p$
   - 2. Value is place where you jump through $p$
   - 3. Otherwise can be any value on that flat spot.
   - 4. We’ll estimate as midpoint
Behaves better than mean because it does not depend on extremes: robust

Estimate as $U$ such that $\hat{S}(U) \leq 1 - p$, $\hat{S}(U -) \geq 1 - p$

1. Uniquely defined unless curve has a flat spot with value $p$
2. Value is place where you jump through $p$
3. Otherwise can be any value on that flat spot.
4. We’ll estimate as midpoint

Quantile Confidence Interval is inversion of Survival Interval
Estimating quantiles

1. Behaves better than mean because it does not depend on extremes: robust
2. Estimate as $U$ such that $\hat{S}(U) \leq 1 - p$, $\hat{S}(U-) \geq 1 - p$
   1. Uniquely defined unless curve has a flat spot with value $p$
   2. Value is place where you jump through $p$
   3. Otherwise can be any value on that flat spot.
   4. We’ll estimate as midpoint
3. Quantile Confidence Interval is inversion of Survival Interval
   1. Draw (point-wise) CI for $S$
Estimating quantiles

1. Behaves better than mean because it does not depend on extremes: robust
2. Estimate as $U$ such that $\hat{S}(U) \leq 1 - p$, $\hat{S}(U-) \geq 1 - p$
   1. Uniquely defined unless curve has a flat spot with value $p$
   2. Value is place where you jump through $p$
   3. Otherwise can be any value on that flat spot.
   4. We’ll estimate as midpoint
3. Quantile Confidence Interval is inversion of Survival Interval
   1. Draw (point-wise) CI for $S$
      1. May be on log or arcsine scale
Estimating quantiles

1. Behaves better than mean because it does not depend on extremes: robust

2. Estimate as $U$ such that $\hat{S}(U) \leq 1 - p$, $\hat{S}(U-) \geq 1 - p$
   - Uniquely defined unless curve has a flat spot with value $p$
   - Value is place where you jump through $p$
   - Otherwise can be any value on that flat spot.
   - We’ll estimate as midpoint

3. Quantile Confidence Interval is inversion of Survival Interval
   - Draw (point-wise) CI for $S$
     - May be on log or arcsine scale
   - Draw horizontal line at $p$
Estimating quantiles

1. Behaves better than mean because it does not depend on extremes: robust

2. Estimate as $U$ such that $\hat{S}(U) \leq 1 - p$, $\hat{S}(U-) \geq 1 - p$
   - Uniquely defined unless curve has a flat spot with value $p$
   - Value is place where you jump through $p$
   - Otherwise can be any value on that flat spot.
   - We’ll estimate as midpoint

3. Quantile Confidence Interval is inversion of Survival Interval
   - Draw (point-wise) CI for $S$
     - May be on log or arcsine scale
   - Draw horizontal line at $p$
   - CI is parts of horizontal line inside confidence band
May approximate interval using normal distribution
May approximate interval using normal distribution

For true quantile $\nu$, $P \left[ \frac{|\hat{S}(\nu) - S(\nu)|}{\sqrt{\text{Var}[\hat{S}(\nu)]}} \geq 1.96 \right] \approx 0.05$
May approximate interval using normal distribution

For true quantile $v$, \[ P\left(\left|\frac{\hat{S}(v) - S(v)}{\sqrt{\text{Var}[\hat{S}(v)]}}\right| \geq 1.96 \right) \approx .05 \]

We say this statistic is *pivotal*.
May approximate interval using normal distribution

For true quantile $\nu$, $P \left[ \left| \frac{\hat{S}(\nu) - S(\nu)}{\sqrt{\text{Var}[\hat{S}(\nu)]}} \right| \geq 1.96 \right] \approx .05$

We say this statistic is *pivotal*.

Intervals for $S(t)$ give vertical lines
May approximate interval using normal distribution

For true quantile \( \nu \),

\[
P \left[ \left| \frac{\hat{S}(\nu) - S(\nu)}{\sqrt{\text{Var}[\hat{S}(\nu)]}} \right| \geq 1.96 \right] \approx 0.05
\]

We say this statistic is *pivotal*.

Intervals for \( S(t) \) give vertical lines

Intervals for \( \nu \) give horizontal lines
May approximate interval using normal distribution

For true quantile \( \nu \),

\[
P \left( \left| \frac{\hat{S}(\nu) - S(\nu)}{\sqrt{\text{Var}[\hat{S}(\nu)]}} \right| \geq 1.96 \right) \approx .05
\]

We say this statistic is \textit{pivotal}.

Intervals for \( S(t) \) give vertical lines

Intervals for \( \nu \) give horizontal lines

Shouldn’t we be doing this with simultaneous intervals?
May approximate interval using normal distribution

For true quantile $\nu$, $P\left[ \left| \frac{\hat{S}(\nu) - S(\nu)}{\sqrt{\text{Var}[\hat{S}(\nu)]}} \right| \geq 1.96 \right] \approx 0.05$

We say this statistic is *pivotal*.

Intervals for $S(t)$ give vertical lines

Intervals for $\nu$ give horizontal lines

Shouldn’t we be doing this with simultaneous intervals?

No, because we only use half an interval at two places
May approximate interval using normal distribution

For true quantile $\nu$, $P \left[ \left| \frac{\hat{S}(\nu) - S(\nu)}{\sqrt{\text{Var}[\hat{S}(\nu)]}} \right| \geq 1.96 \right] \approx .05$

We say this statistic is *pivotal*.

Intervals for $S(t)$ give vertical lines

Intervals for $\nu$ give horizontal lines

Shouldn’t we be doing this with simultaneous intervals?

No, because we only use half an interval at two places

No guarantee that interval is connected
May approximate interval using normal distribution

For true quantile $\nu$, $P \left( \left| \frac{\hat{S}(\nu) - S(\nu)}{\sqrt{\text{Var}[\hat{S}(\nu)]}} \right| \geq 1.96 \right) \approx 0.05$

We say this statistic is *pivotal*.

Intervals for $S(t)$ give vertical lines

Intervals for $\nu$ give horizontal lines

Shouldn’t we be doing this with simultaneous intervals?

No, because we only use half an interval at two places

No guarantee that interval is connected

Since upper confidence bound might go back up
Alternative approach via standard error
Alternative approach via standard error

\[ \hat{F}(\hat{\nu}) - \hat{F}(\nu) \approx f(\nu)(\hat{\nu} - \nu) \]
Alternative approach via standard error

1. \( \hat{F}(\hat{\upsilon}) - \hat{F}(\upsilon) \approx f(\upsilon)(\hat{\upsilon} - \upsilon) \)
2. \( p - \hat{F}(\upsilon) \approx f(\upsilon)(\hat{\upsilon} - \upsilon) \)
Alternative approach via standard error

1. \( \hat{F}(\hat{v}) - \hat{F}(v) \approx f(v) (\hat{v} - v) \)
2. \( p - \hat{F}(v) \approx f(v)(\hat{v} - v) \)
3. \( \text{Var} \left[ \hat{F}(v) \right] \approx f(v)^2 \text{Var} [\hat{v}] \)
Alternative approach via standard error

1. \( \hat{F}(\hat{\nu}) - \hat{F}(\nu) \approx f(\nu)(\hat{\nu} - \nu) \)
2. \( p - \hat{F}(\nu) \approx f(\nu)(\hat{\nu} - \nu) \)
3. \( \text{Var} \left[ \hat{F}(\nu) \right] \approx f(\nu)^2 \text{Var} \left[ \hat{\nu} \right] \)
4. \( \text{Var} \left[ \hat{\nu} \right] \approx \hat{f}(\hat{\nu})^{-2} \text{Var} \left[ \hat{F}(\hat{\nu}) \right] \)
1. Alternative approach via standard error

1. \( \hat{F}(\hat{v}) - \hat{F}(v) \approx f(v)(\hat{v} - v) \)
2. \( p - \hat{F}(v) \approx f(v)(\hat{v} - v) \)
3. \( \text{Var} \left[ \hat{F}(v) \right] \approx f(v)^2 \text{Var} \left[ \hat{v} \right] \)
4. \( \text{Var} \left[ \hat{v} \right] \approx \hat{f}(\hat{v})^{-2} \text{Var} \left[ \hat{F}(\hat{v}) \right] \)

1. Need estimator of density.  

R Code  SAS Code
Mean

1. Defined as $E[X] = \int_0^\infty tf(t) \, dt$
Mean

Defined as $E[X] = \int_0^\infty tf(t) \, dt$

Integration by parts

Integration by parts gives $E[X] = (\text{constant}) - \int_0^\infty S(t) \, dt$.

Argument can be extended to discrete distributions.

Infinity if $\hat{S}(t)$ hits zero at last event.

Can estimate standard error.

Estimators typically wrong if last event not a events.

Common fix: Estimate restricted mean life $\int_0^K S(t) \, dt$ by $\int_0^K \hat{S}(t) \, dt$.

R Code

SAS Code

Estimation:

Estimating location measures:

Lecture 03
Mean

1. Defined as $E[X] = \int_0^\infty tf(t) \, dt$
2. Integration by parts
   1. $\int_a^b u \, dv = uv|_a^b - \int_a^b v \, du$

Estimation: Estimating location measures: Lecture 03 61/260
Mean

1. Defined as $E[X] = \int_0^\infty tf(t) \, dt$
2. Integration by parts
   1. $\int_a^b u \, dv = uv\big|_a^b - \int_a^b v \, du$
   2. In this case, let $v = -S(t)$, $u = t$. 

Estimation: Estimating location measures: Lecture 03
Mean

1. Defined as \( E[X] = \int_0^\infty tf(t) \, dt \)

2. Integration by parts

   1. \( \int_a^b u \, dv = uv|_a^b - \int_a^b v \, du \)
   2. In this case, let \( v = -S(t) \), \( u = t \).
   3. gives

\[
E[X] = (-S(t))t|_0^\infty - \int_0^\infty (-S(t)) \, dt = \int_0^\infty S(t) \, dt
\]
Mean

1. Defined as $E[X] = \int_0^\infty tf(t) \, dt$

2. Integration by parts

   1. $\int_a^b u \, dv = uv|_a^b - \int_a^b v \, du$
   2. In this case, let $v = -S(t)$, $u = t$.
   3. gives

   $$E[X] = (-S(t))t|_0^\infty - \int_0^\infty (-S(t)) \, dt = \int_0^\infty S(t) \, dt$$

4. Argument can be extended to discrete distributions
## Mean

1. Defined as $E[X] = \int_0^\infty tf(t) \, dt$

2. Integration by parts
   - $\int_a^b u \, dv = uv\big|_a^b - \int_a^b v \, du$
   - In this case, let $v = -S(t)$, $u = t$.
   - Gives
     
     $E[X] = (-S(t))t\big|_0^\infty - \int_0^\infty (-S(t)) \, dt = \int_0^\infty S(t) \, dt$

3. Argument can be extended to discrete distributions

4. Estimate as $\int_0^\infty \hat{S}(t) \, dt$
Mean

1. Defined as $E[X] = \int_{0}^{\infty} tf(t) \, dt$

2. Integration by parts
   
   $\int_{a}^{b} u \, dv = uv\big|_{a}^{b} - \int_{a}^{b} v \, du$

   1. In this case, let $v = -S(t)$, $u = t$.

   3. gives

   $$E[X] = (-S(t))t\big|_{0}^{\infty} - \int_{0}^{\infty} (-S(t)) \, dt = \int_{0}^{\infty} S(t) \, dt$$

4. Argument can be extended to discrete distributions

3. Estimate as $\int_{0}^{\infty} \hat{S}(t) \, dt$

4. $\infty$ if $\hat{S} > 0$ at last event
**Mean**

1. Defined as $E[X] = \int_0^\infty t f(t) \, dt$

2. Integration by parts
   - $\int_a^b u \, dv = uv|_a^b - \int_a^b v \, du$
   - In this case, let $v = -S(t)$, $u = t$.
   - gives
     \[
     E[X] = (-S(t))t|_0^\infty - \int_0^\infty (-S(t)) \, dt = \int_0^\infty S(t) \, dt
     \]

3. Argument can be extended to discrete distributions

4. Estimate as $\int_0^\infty \hat{S}(t) \, dt$

5. $\infty$ if $\hat{S} > 0$ at last event

   - Finite if $\hat{S}$ hit zero at last event
Mean

1. Defined as $E[X] = \int_{0}^{\infty} tf(t) \, dt$

2. Integration by parts
   
   $\int_{a}^{b} u \, dv = uv \bigg|_{a}^{b} - \int_{a}^{b} v \, du$

   In this case, let $v = -S(t)$, $u = t$.

   This gives

   $$E[X] = (-S(t))t \bigg|_{0}^{\infty} - \int_{0}^{\infty} (-S(t)) \, dt = \int_{0}^{\infty} S(t) \, dt$$

3. Argument can be extended to discrete distributions

4. Estimate as $\int_{0}^{\infty} \hat{S}(t) \, dt$

   \(\infty\) if $\hat{S} > 0$ at last event

   1. Finite if $\hat{S}$ hit zero at last event

   1. Could fix using parametric estimate of rest of curve
Mean

1. Defined as $\mathbb{E}[X] = \int_0^\infty tf(t) \, dt$

2. Integration by parts

   $\int_a^b u \, dv = uv\big|_a^b - \int_a^b v \, du$

   In this case, let $v = -S(t)$, $u = t$.

3. gives

   $\mathbb{E}[X] = (-S(t))t\big|_0^\infty - \int_0^\infty (-S(t)) \, dt = \int_0^\infty S(t) \, dt$

4. Argument can be extended to discrete distributions

3. Estimate as $\int_0^\infty \hat{S}(t) \, dt$

4. $\infty$ if $\hat{S} > 0$ at last event

   1. Finite if $\hat{S}$ hit zero at last event

   2. Could fix using parametric estimate of rest of curve

   2. Can estimate standard error
Mean

1. Defined as $E[X] = \int_0^\infty tf(t) \, dt$

2. Integration by parts
   
   $\int_a^b uv \, dv = uv\big|_a^b - \int_a^b v \, du$

   In this case, let $v = -S(t)$, $u = t$.

   3. gives

   $$E[X] = (-S(t))t|_0^\infty - \int_0^\infty (-S(t)) \, dt = \int_0^\infty S(t) \, dt$$

   4. Argument can be extended to discrete distributions

3. Estimate as $\int_0^\infty \hat{S}(t) \, dt$

4. $\infty$ if $\hat{S} > 0$ at last event

   1. Finite if $\hat{S}$ hit zero at last event

      1. Could fix using parametric estimate of rest of curve

   2. Can estimate standard error

   1. Estimators typically wrong if last event not a events.
Mean

1. Defined as $E[X] = \int_0^\infty tf(t) \, dt$

2. Integration by parts
   1. $\int_a^b u \, dv = uv|_a^b - \int_a^b v \, du$
   2. In this case, let $v = -S(t)$, $u = t$.
   3. gives
      $$E[X] = (-S(t))t|_0^\infty - \int_0^\infty (-S(t)) \, dt = \int_0^\infty S(t) \, dt$$

4. Argument can be extended to discrete distributions

3. Estimate as $\int_0^\infty \hat{S}(t) \, dt$

4. $\infty$ if $\hat{S} > 0$ at last event
   1. Finite if $\hat{S}$ hit zero at last event
      1. Could fix using parametric estimate of rest of curve
      2. Can estimate standard error
   2. Estimators typically wrong if last event not a events.

5. Common fix: Estimate restricted mean life $\int_0^K S(t) \, dt$ for some $K$ by $\int_0^K \hat{S}(t) \, dt$
Section: Estimation

Subsection: Other Sampling Schemes
For data grouped in intervals:

1. Summarize data by number $m_j$ censored and $D_j$ with event in intervals $(a_{j-1}, a_j]$
For data grouped in intervals:

1. Summarize data by number $m_j$ censored and $D_j$ with event in intervals $(a_{j-1}, a_j]$

2. Number at risk should be measured somewhere in $(a_{j-1}, a_j]$
For data grouped in intervals:

1. Summarize data by number $m_j$ censored and $D_j$ with event in intervals $(a_{j-1}, a_j]$
2. Number at risk should be measured somewhere in $(a_{j-1}, a_j]$
3. Raise $Y_j$ to $Y'_j = Y_j + m_j/2 + D_j$
For data grouped in intervals:

1. Summarize data by number $m_j$ censored and $D_j$ with event in intervals $(a_{j-1}, a_j]$

2. Number at risk should be measured somewhere in $(a_{j-1}, a_j]$

3. Raise $Y_j$ to $Y'_j = Y_j + m_j/2 + D_j$
   - Approx. equivalent to estimating $h(t)$ on as $D_j/(Y_j + (m_j + D_j)/2)$.

Estimation: Other Sampling Schemes

Lecture 03
For data grouped in intervals:

1. Summarize data by number $m_j$ censored and $D_j$ with event in intervals $(a_{j-1}, a_j]$

2. Number at risk should be measured somewhere in $(a_{j-1}, a_j]$

3. Raise $Y_j$ to $Y'_j = Y_j + m_j/2 + D_j$
   - Approx. equivalent to estimating $h(t)$ on as $D_j/(Y_j + (m_j + D_j)/2)$.
   - $Y_j$ still number at risk at time $a_j$
For data grouped in intervals:

1. Summarize data by number $m_j$ censored and $D_j$ with event in intervals $(a_{j-1}, a_j]$
2. Number at risk should be measured somewhere in $(a_{j-1}, a_j]$
3. Raise $Y_j$ to $Y'_j = Y_j + m_j/2 + D_j$
   - Approx. equivalent to estimating $h(t)$ on as $D_j/(Y_j + (m_j + D_j)/2)$.
   - $Y_j$ still number at risk at time $a_j$
4. Called the actuarial estimate or life table estimate. R Code SAS Code
With truncation:

1. Estimates are conditional on inclusion
With truncation:

1. Estimates are conditional on inclusion
2. If truncation \( \perp \) event time, hazard rate still the same
With truncation:

1. Estimates are conditional on inclusion
2. If truncation \( \perp \) event time, hazard rate still the same
3. Hence can still get \( h \) from slope of Nelson–Aalen estimator \( \hat{H} \).
Section: Testing with survival curves

Subsection: Introduction to survival curve hypothesis testing
Suppose that $K$ populations have survival functions $S_k$, $k = 1, \ldots, K$. 

1. Suppose that $K$ populations have survival functions $S_k$, $k = 1, \ldots, K$. 

Notation
Suppose that $K$ populations have survival functions $S_k$, $k = 1, \ldots, K$

Suppose we have samples of size $n_k$ respectively
Notation

1. Suppose that $K$ populations have survival functions $S_k, k = 1, \ldots, K$
2. Suppose we have samples of size $n_k$ respectively
3. Generally assume that they are independent
Two questions

1. Do (population) survival curves have some relation?

2. If $K = 1$, does it take some simple form like exponential?

3. If $K > 1$, are they all the same?

Two questions

1. Do (population) survival curves have some relation?
   - If $K = 1$, does it take some simple form like exponential?
Two questions

1. Do (population) survival curves have some relation?
   1. If $K = 1$, does it take some simple form like exponential?
   2. If $K > 1$, are they all the same?

Attributes might be:
1. Mean
2. Median
3. Value at some time

Relation might be:
1. Equals a null value if $K = 1$
2. Equal to each other if $K > 1$
Two questions

1. Do (population) survival curves have some relation?
   1. If $K = 1$, does it take some simple form like exponential?
   2. If $K > 1$, are they all the same?
   3. Generally $K = 2$ simpler than $K > 2$
Two questions

1. Do (population) survival curves have some relation?
   1. If $K = 1$, does it take some simple form like exponential?
   2. If $K > 1$, are they all the same?
   3. Generally $K = 2$ simpler than $K > 2$

2. Do (population) survival curves have relation with some attribute?

Attributes might be:
1. Mean
2. Median
3. Value at some time

Relation might be:
1. Equals a null value if $K = 1$
2. Equal to each other if $K > 1$
Two questions

1. Do (population) survival curves have some relation?
   1. If $K = 1$, does it take some simple form like exponential?
   2. If $K > 1$, are they all the same?
   3. Generally $K = 2$ simpler than $K > 2$

2. Do (population) survival curves have relation with some attribute?
   1. Attributes might be
      - Mean
      - Median
      - Value at some time
   2. Relation might be
      - Equals a null value if $K = 1$
      - Equal to each other if $K > 1$
Two questions

1. Do (population) survival curves have some relation?
   1. If $K = 1$, does it take some simple form like exponential?
   2. If $K > 1$, are they all the same?
   3. Generally $K = 2$ simpler than $K > 2$

2. Do (population) survival curves have relation with some attribute?
   1. Attributes might be
      1. Mean
Two questions

1. Do (population) survival curves have some relation?
   1. If $K = 1$, does it take some simple form like exponential?
   2. If $K > 1$, are they all the same?
   3. Generally $K = 2$ simpler than $K > 2$

2. Do (population) survival curves have relation with some attribute?
   1. Attributes might be
      1. Mean
      2. Median
Two questions

1. Do (population) survival curves have some relation?
   1. If $K = 1$, does it take some simple form like exponential?
   2. If $K > 1$, are they all the same?
   3. Generally $K = 2$ simpler than $K > 2$

2. Do (population) survival curves have relation with some attribute?
   1. Attributes might be
      1. Mean
      2. Median
      3. Value at some time
Two questions

1. Do (population) survival curves have some relation?
   - If $K = 1$, does it take some simple form like exponential?
   - If $K > 1$, are they all the same?
   - Generally $K = 2$ simpler than $K > 2$

2. Do (population) survival curves have relation with some attribute?
   - Attributes might be
     - Mean
     - Median
     - Value at some time
   - Relation might be
Two questions

1. Do (population) survival curves have some relation?
   1. If $K = 1$, does it take some simple form like exponential?
   2. If $K > 1$, are they all the same?
   3. Generally $K = 2$ simpler than $K > 2$

2. Do (population) survival curves have relation with some attribute?
   1. Attributes might be
      1. Mean
      2. Median
      3. Value at some time
   2. Relation might be
      1. Equals a null value if $K = 1$
Two questions

1. Do (population) survival curves have some relation?
   1. If $K = 1$, does it take some simple form like exponential?
   2. If $K > 1$, are they all the same?
   3. Generally $K = 2$ simpler than $K > 2$

2. Do (population) survival curves have relation with some attribute?
   1. Attributes might be
      1. Mean
      2. Median
      3. Value at some time
   2. Relation might be
      1. Equals a null value if $K = 1$
      2. Equal to each other if $K > 1$
One-sample Hypothesis Test

Comparing observed number of events to expected number.

Expected value for $o_i = D_i / Y_i$ is $e_i = 1 - S(t_{i+1}) / S(t_i) = 1 - \exp(-\int_{t_i}^{t_{i+1}} h(s) ds) \approx \int_{t_i}^{t_{i+1}} h(s) ds$.

Test statistic is then $T = \sum_{D_i=1} W(t_i)(o_i - e_i) = O - E$ for some weight function $W$ depending on time.

$O = \sum_{D_i=1} W(t_i) o_i$, $E = \sum_{D_i=1} W(t_i) e_i$.

Often use $W(t_i) = Y_i$ to give test statistic $T = (\sum_{D_i=1} D_i - Y_i e_i) / E$ approximately $\chi^2_1$.

Variance is approximately what gets subtracted off (call it $E$) from first part (call it $O$), $2(0 - E)^2 / E$ approximately $\chi^2_1$.

Variance for $K > 1$ also $\approx E$ but we don't need this.

Often $h$ is determined empirically from a very large sample and is effectively non-random.
One-sample Hypothesis Test

1. Comparing observed number of events to expected number.

2. Expected value for \( o_i = D_i / Y_i \) is \( e_i = 1 - S(t_{i+1})/S(t_i) = 1 - \exp(- \int_{t_i}^{t_{i+1}} h(s) \, ds) \approx \int_{t_i}^{t_{i+1}} h(s) \, ds. \)
One-sample Hypothesis Test

1. Comparing observed number of events to expected number.

2. Expected value for \( o_i = D_i / Y_i \) is \( e_i = 1 - S(t_{i+1}) / S(t_i) = 1 - \exp(- \int_{t_i}^{t_{i+1}} h(s) \, ds) \approx \int_{t_i}^{t_{i+1}} h(s) \, ds \).

3. Test statistic is then \( T = \sum_{i=1}^{D} W(t_i)(o_i - e_i) = O - E \) for
One-sample Hypothesis Test

1. Comparing observed number of events to expected number.
2. Expected value for $o_i = D_i / Y_i$ is $e_i = 1 - S(t_{i+1}) / S(t_i) = 1 - \exp(-\int_{t_i}^{t_{i+1}} h(s) \, ds) \approx \int_{t_i}^{t_{i+1}} h(s) \, ds$.
3. Test statistic is then $T = \sum_{i=1}^{D} W(t_i)(o_i - e_i) = O - E$ for some weight function $W$ depending on time.
One-sample Hypothesis Test

1. Comparing observed number of events to expected number.

2. Expected value for \( o_i = D_i/Y_i \) is \( e_i = 1 - S(t_{i+1})/S(t_i) = 1 - \exp(-\int_{t_i}^{t_{i+1}} h(s) \, ds) \approx \int_{t_i}^{t_{i+1}} h(s) \, ds \).

3. Test statistic is then \( T = \sum_{i=1}^{D} W(t_i)(o_i - e_i) = O - E \) for some weight function \( W \) depending on time.

4. \( O = \sum_{i=1}^{D} W(t_i)o_i, \quad E = \sum_{i=1}^{D} W(t_i)e_i. \)
One-sample Hypothesis Test

1. Comparing observed number of events to expected number.

2. Expected value for $o_i = D_i/Y_i$ is $e_i = 1 - S(t_{i+1})/S(t_i) = 1 - \exp(-\int_{t_i}^{t_{i+1}} h(s) \, ds) \approx \int_{t_i}^{t_{i+1}} h(s) \, ds$.

3. Test statistic is then $T = \sum_{i=1}^{D} W(t_i)(o_i - e_i) = O - E$ for some weight function $W$ depending on time.

4. $O = \sum_{i=1}^{D} W(t_i)o_i$, $E = \sum_{i=1}^{D} W(t_i)e_i$.

5. Often use $W(t_i) = Y_i$ to give test statistic $T = (\sum_{i=1}^{D} D_i - Y_i e_i)$.
One-sample Hypothesis Test

1. Comparing observed number of events to expected number.
2. Expected value for \( o_i = D_i / Y_i \) is \( e_i = 1 - S(t_{i+1})/S(t_i) = 1 - \exp(- \int_{t_i}^{t_{i+1}} h(s) \ ds) \approx \int_{t_i}^{t_{i+1}} h(s) \ ds \).
3. Test statistic is then \( T = \sum_{i=1}^{D} W(t_i)(o_i - e_i) = O - E \) for some weight function \( W \) depending on time.
4. \( O = \sum_{i=1}^{D} W(t_i)o_i, \ E = \sum_{i=1}^{D} W(t_i)e_i \).
5. Often use \( W(t_i) = Y_i \) to give test statistic \( T = (\sum_{i=1}^{D} D_i - Y_ie_i) \).
6. Estimate of variance like chi-square example:
One-sample Hypothesis Test

1. Comparing observed number of events to expected number.
2. Expected value for \( o_i = D_i/Y_i \) is \( e_i = 1 - S(t_{i+1})/S(t_i) = 1 - \exp(-\int_{t_i}^{t_{i+1}} h(s) \, ds) \approx \int_{t_i}^{t_{i+1}} h(s) \, ds \).
3. Test statistic is then \( T = \sum_{i=1}^{D} W(t_i)(o_i - e_i) = O - E \) for some weight function \( W \) depending on time.
4. \( O = \sum_{i=1}^{D} W(t_i)o_i, \ E = \sum_{i=1}^{D} W(t_i)e_i. \)
5. Often use \( W(t_i) = Y_i \) to give test statistic \( T = (\sum_{i=1}^{D} D_i - Y_i e_i) \)

6. Estimate of variance like chi-square example:

   1. Variance is approximately what gets subtracted off (call it \( E \)) from first part (call it \( O \))
One-sample Hypothesis Test

1. Comparing observed number of events to expected number.

2. Expected value for $o_i = D_i / Y_i$ is $e_i = 1 - S(t_{i+1}) / S(t_i) = 1 - \exp(- \int_{t_i}^{t_{i+1}} h(s) \ ds) \approx \int_{t_i}^{t_{i+1}} h(s) \ ds$.

3. Test statistic is then $T = \sum_{i=1}^{D} W(t_i)(o_i - e_i) = O - E$ for some weight function $W$ depending on time.

4. $O = \sum_{i=1}^{D} W(t_i)o_i$, $E = \sum_{i=1}^{D} W(t_i)e_i$.

5. Often use $W(t_i) = Y_i$ to give test statistic $T = (\sum_{i=1}^{D} D_i - Y_i e_i)$.

6. Estimate of variance like chi-square example:
   1. Variance is approximately what gets subtracted off (call it $E$) from first part (call it $O$)
   2. $(0 - E)^2 / E$ approximately $\chi^2_1$
One-sample Hypothesis Test

1. Comparing observed number of events to expected number.
2. Expected value for $o_i = D_i / Y_i$ is $e_i = 1 - S(t_{i+1})/S(t_i) = 1 - \exp(-\int_{t_i}^{t_{i+1}} h(s) \, ds) \approx \int_{t_i}^{t_{i+1}} h(s) \, ds$.
3. Test statistic is then $T = \sum_{i=1}^{D} W(t_i)(o_i - e_i) = O - E$ for some weight function $W$ depending on time.
4. $O = \sum_{i=1}^{D} W(t_i) o_i$, $E = \sum_{i=1}^{D} W(t_i) e_i$.
5. Often use $W(t_i) = Y_i$ to give test statistic $T = (\sum_{i=1}^{D} D_i - Y_i e_i)$.
6. Estimate of variance like chi-square example:
   1. Variance is approximately what gets subtracted off (call it $E$) from first part (call it $O$)
   2. $(0 - E)^2 / E$ approximately $\chi_1^2$
   3. Variance for $K > 1$ also $\approx E$ but we don’t need this
One-sample Hypothesis Test

1. Comparing observed number of events to expected number.

2. Expected value for \( o_i = D_i/Y_i \) is \( e_i = 1 - S(t_{i+1})/S(t_i) = 1 - \exp(-\int_{t_i}^{t_{i+1}} h(s) \, ds) \approx \int_{t_i}^{t_{i+1}} h(s) \, ds. \)

3. Test statistic is then \( T = \sum_{i=1}^{D} W(t_i)(o_i - e_i) = O - E \) for some weight function \( W \) depending on time.

4. \( O = \sum_{i=1}^{D} W(t_i)o_i, \ E = \sum_{i=1}^{D} W(t_i)e_i. \)

5. Often use \( W(t_i) = Y_i \) to give test statistic \( T = (\sum_{i=1}^{D} D_i - Y_i e_i) \)

6. Estimate of variance like chi-square example:
   - Variance is approximately what gets subtracted off (call it \( E \)) from first part (call it \( O \))
   - \( (0 - E)^2/E \) approximately \( \chi^2_1 \)
   - Variance for \( K > 1 \) also \( \approx E \) but we don’t need this

7. Often \( h \) is determined empirically from a very large sample and is effectively non-random
The $K > 1$ testing task

1. Pick one group labeled $k$ from $\{1, \ldots, K\}$. 
The $K > 1$ testing task

1. Pick one group labeled $k$ from \{1, \ldots, K\}.
2. Ask whether $S_k$ is different from the rest.
The $K > 1$ testing task

1. Pick one group labeled $k$ from \{1, \ldots, K\}.
2. Ask whether $S_k$ is different from the rest.
   1. Maybe $K = 2$ and $k = 1$ or 2
The $K > 1$ testing task

1. Pick one group labeled $k$ from $\{1, \ldots, K\}$.
2. Ask whether $S_k$ is different from the rest.
   1. Maybe $K = 2$ and $k = 1$ or $2$
   2. Intuition doesn’t require this
The $K > 1$ testing task

1. Pick one group labeled $k$ from $\{1, \ldots, K\}$.
2. Ask whether $S_k$ is different from the rest.
   1. Maybe $K = 2$ and $k = 1$ or $2$
   2. Intuition doesn’t require this
   3. More generality at this stage will be useful.
The $K > 1$ testing task

1. Pick one group labeled $k$ from $\{1, \ldots, K\}$.
2. Ask whether $S_k$ is different from the rest.
   1. Maybe $K = 2$ and $k = 1$ or 2
   2. Intuition doesn’t require this
   3. More generality at this stage will be useful.
3. under the Proportional Hazards Alternative
Null and alternative hypotheses for $K$ samples, $K > 1$.

1. Postulate a common $S$
Null and alternative hypotheses for $K$ samples, $K > 1$.

1. Postulate a common $S$
2. Postulate how distns will differ under $H_A$
Null and alternative hypotheses for $K$ samples, $K > 1$.

1. Postulate a common $S$

2. Postulate how distns will differ under $H_A$
   - Shift for standard $t$ tests, etc.
Null and alternative hypotheses for $K$ samples, $K > 1$.

1. Postulate a common $S$
2. Postulate how distns will differ under $H_A$
   1. Shift for standard $t$ tests, etc.
   2. Assume $-\frac{d}{dt} \log(S_k(t)) = \alpha_k \times (-\frac{d}{dt} \log(S(t)))$:
Null and alternative hypotheses for $K$ samples, $K > 1$.

1. Postulate a common $S$
2. Postulate how distns will differ under $H_A$
   1. Shift for standard $t$ tests, etc.
   2. Assume $-\frac{d}{dt} \log(S_k(t)) = \alpha_k \times (-\frac{d}{dt} \log(S(t)))$:
   3. Called *proportional hazards* in survival analysis lingo, and *Lehmann alternative* in nonparametrics literature
Null and alternative hypotheses for $K$ samples, $K > 1$.

1. Postulate a common $S$
2. Postulate how distns will differ under $H_A$
   1. Shift for standard $t$ tests, etc.
   2. Assume $-\frac{d}{dt} \log(S_k(t)) = \alpha_k \times (-\frac{d}{dt} \log(S(t)))$:
   3. Called proportional hazards in survival analysis lingo, and Lehmann alternative in nonparametrics literature
   4. Implies $\log(S_k(t)) = \alpha_k \log(S(t))$
Null and alternative hypotheses for $K$ samples, $K > 1$.

1. Postulate a common $S$

2. Postulate how distns will differ under $H_A$
   
   1. Shift for standard $t$ tests, etc.
   2. Assume $-\frac{d}{dt} \log(S_k(t)) = \alpha_k \times (\frac{d}{dt} \log(S(t)))$:
   3. Called *proportional hazards* in survival analysis lingo, and *Lehmann alternative* in nonparametrics literature
      
      1. Implies $\log(S_k(t)) = \alpha_k \log(S(t))$
      4. Implies $S_k(t) = S(t)^{\alpha_k}$
Useful Consequences of the Lehmann Alternative

1. If null and alternative distributions for $T$ are related by the Lehmann relation, then so are those for $U = g(T)$ any increasing invertible transformation of $T$.
Useful Consequences of the Lehmann Alternative

1. If null and alternative distributions for $T$ are related by the Lehmann relation, then so are those for $U = g(T)$ any increasing invertible transformation of $T$

   Because $P_A[U > u] = P_A[T > g^{-1}(u)] = P_0[T > g^{-1}(u)]^{\alpha_k} = P_0[U > u]^{\alpha_k}$
Useful Consequences of the Lehmann Alternative

1. If null and alternative distributions for $T$ are related by the Lehmann relation, then so are those for $U = g(T)$ any increasing invertible transformation of $T$

   Because $P_A[U > u] = P_A[T > g^{-1}(u)] = P_0[T > g^{-1}(u)]^{\alpha_k} = P_0[U > u]^{\alpha_k}$

2. Lehmann alternative for exponential is also a shift alternative on the log scale:

   Suppose $U = \log(T)$, then
   
   Alternative distribution of $U$ has survival curve $P_A[U > u] = P[T > \exp(u)] = S_k(\exp(u)) = \exp(-\lambda \exp(u + \log(\alpha_k)))$. 

Useful Consequences of the Lehmann Alternative

1. If null and alternative distributions for $T$ are related by the Lehmann relation, then so are those for $U = g(T)$ any increasing invertible transformation of $T$

   Because $P_A[U > u] = P_A[T > g^{-1}(u)] = P_0[T > g^{-1}(u)]^\alpha_k = P_0[U > u]^\alpha_k$

2. Lehmann alternative for exponential is also a shift alternative on the log scale:

   Suppose $U = \log(T)$,
Useful Consequences of the Lehmann Alternative

1. If null and alternative distributions for $T$ are related by the Lehmann relation, then so are those for $U = g(T)$ any increasing invertible transformation of $T$

   Because $P_A [U > u] = P_A [T > g^{-1}(u)] = P_0 [T > g^{-1}(u)]^{\alpha_k} = P_0 [U > u]^{\alpha_k}$

2. Lehmann alternative for exponential is also a shift alternative on the log scale:

   1. Suppose $U = \log(T)$,
   2. Alternative distribution of $U$ has survival curve $P_A [U \geq u] = P [T \geq \exp(u)] = S_k(\exp(u)) = \exp(-\lambda \exp(u))^{\alpha_k} = \exp(-\lambda \exp(u + \log(\alpha_k)))$. 

Testing with survival curves: Introduction to survival curve hypothesis testing

Lecture 03
Suppose that $T_j$ are times associated with groups $g_j, \ g_j \in \{1, \ldots, K\}, \ j \in \{1, \ldots, K\}$. 
No-censoring general rank test

1. Suppose that $T_j$ are times associated with groups $g_j$, $g_j \in \{1, \ldots, K\}$, $j \in \{1, \ldots, K\}$.
2. Make $n$ scores $a_j$
No-censoring general rank test

1. Suppose that $T_j$ are times associated with groups $g_j$, $g_j \in \{1, \ldots, K\}$, $j \in \{1, \ldots, K\}$.

2. Make $n$ scores $a_j$
   1. Nondecreasing in $j$

3. Let $R_j$ be the rank of $T_j$ from the entire sample.

4. General rank statistic for testing group $k$ different from rest is $W_k = \sum_{j: g_j = k} a_j R_j$. 
No-censoring general rank test

1. Suppose that $T_j$ are times associated with groups $g_j$, $g_j \in \{1, \ldots, K\}$, $j \in \{1, \ldots, K\}$.

2. Make $n$ scores $a_j$
   1. Nondecreasing in $j$
   2. Sum to zero.

3. Let $R_j$ be the rank of $T_j$ from the entire sample.

4. General rank statistic for testing group $k$ different from rest is $W_k = \sum_{j: g_j = k} a_j R_j$. 
No-censoring general rank test

1. Suppose that $T_j$ are times associated with groups $g_j$, $g_j \in \{1, \ldots, K\}$, $j \in \{1, \ldots, K\}$.

2. Make $n$ scores $a_j$
   - Nondecreasing in $j$
   - Sum to zero.

3. Let $R_j$ be the rank of $T_j$ from the entire sample.
Suppose that \( T_j \) are times associated with groups \( g_j, g_j \in \{1, \ldots, K\}, j \in \{1, \ldots, K\} \).

Make \( n \) scores \( a_j \)

1. Nondecreasing in \( j \)
2. Sum to zero.

Let \( R_j \) be the rank of \( T_j \) from the entire sample.

General rank statistic for testing group \( k \) different from rest is

\[
W_k = \sum_{j: g(j) = k} aR_j.
\]
Optimal Test for Equality of Distributions

1. Best means giving best power for very large $n$ and $\alpha$ near 1
Optimal Test for Equality of Distributions

1. Best means giving best power for very large $n$ and $\alpha$ near 1
2. Under a shift alternative, then best scores are expected values of order statistics evaluated at $g'(\cdot)/g(\cdot)$ for $g$ the density.
Best means giving best power for very large $n$ and $\alpha$ near 1

Under a shift alternative, then best scores are expected values of order statistics evaluated at $g'(\cdot)/g(\cdot)$ for $g$ the density.

This works out to expected order statistic of exponential if null distribution is that of $\log(T)$ for $T$ exponential
Optimal Test for Equality of Distributions

1. Best means giving best power for very large $n$ and $\alpha$ near 1
   - Under a shift alternative, then best scores are expected values of order statistics evaluated at $g'(\cdot)/g(\cdot)$ for $g$ the density.
   - This works out to expected order statistic of exponential if null distribution is that of $\log(T)$ for $T$ exponential

2. Best scores are $a_j = -1 + \log(1 - j/(n+1))$
Optimal Test for Equality of Distributions

1. Best means giving best power for very large $n$ and $\alpha$ near 1
   - Under a shift alternative, then best scores are expected values of order statistics evaluated at $g'(\cdot)/g(\cdot)$ for $g$ the density.
   - This works out to expected order statistic of exponential if null distribution is that of $\log(T)$ for $T$ exponential

2. Best scores are $a_j = -1 + \log(1 - j/(n + 1))$

3. $a_j \approx -1 + \sum_{i=n+1-j}^{n}(1/i)$
Optimal Test for Equality of Distributions

1. Best means giving best power for very large $n$ and $\alpha$ near 1
   1. Under a shift alternative, then best scores are expected values of order statistics evaluated at $g'(\cdot)/g(\cdot)$ for $g$ the density.
   2. This works out to expected order statistic of exponential if null distribution is that of log($T$) for $T$ exponential

2. Best scores are $a_j = -1 + \log(1 - j/(n + 1))$

3. $a_j \approx -1 + \sum_{i=n+1}^{n} \frac{1}{i}$

4. Approximately $Log$ rank statistic $\sum_{i=1}^{n} [D_{ki} - (D_i/Y_i)Y_{ki}]$
Best means giving best power for very large \( n \) and \( \alpha \) near 1

Under a shift alternative, then best scores are expected values of order statistics evaluated at \( g'(\cdot)/g(\cdot) \) for \( g \) the density.

This works out to expected order statistic of exponential if null distribution is that of \( \log(T) \) for \( T \) exponential.

Best scores are \( a_j = -1 + \log(1 - j/(n+1)) \)

\[ a_j \approx -1 + \sum_{i=n+1}^{n} (1/i) \]

Approximately Log rank statistic \( \sum_{i=1}^{n} [D_{ki} - (D_i/Y_i)Y_{ki}] \)

Interpretation: Expected \# of events in stratum \( l \) if they were distributed \( \propto \) \# at risk \(-\) \# who actually had event.
Optimal Test for Equality of Distributions

1. Best means giving best power for very large $n$ and $\alpha$ near 1
   - Under a shift alternative, then best scores are expected values of order statistics evaluated at $g'(\cdot)/g(\cdot)$ for $g$ the density.
   - This works out to expected order statistic of exponential if null distribution is that of $\log(T)$ for $T$ exponential

2. Best scores are $a_j = -1 + \log(1 - j/(n+1))$

3. $a_j \approx -1 + \sum_{i=n+1-j}^{n} (1/i)$

4. Approximately Log rank statistic $\sum_{i=1}^{n} [D_{ki} - (D_i/Y_i)Y_{ki}]$
   - Interpretation: Expected # of events in stratum $l$ if they were distributed $\propto$ # at risk $-$ # who actually had event.
   - $D_{ki}|D_i, Y_{ki}, Y_{i}$ is hypergeometric:

\[
\begin{array}{ccc}
D_{ki} & D_i - D_{ki} & D_i \\
Y_{ki} - D_{ki} & Y_i - D_i - (Y_{ki} - D_{ki}) & Y_i - D_i \\
Y_{ki} & Y_i - Y_{ki} & Y_i
\end{array}
\]
Let $Y_j$ be \# at risk when item $j$ of joint sample has event.

1. Let $Y_j$ be \# at risk when item $j$ of joint sample has event.
Phrase in terms of survival quantities.

1. Let $Y_j$ be # at risk when item $j$ of joint sample has event.
2. $Y_{kj}$ be # at risk from group $k$.
Phrase in terms of survival quantities.

1. Let $Y_j$ be # at risk when item $j$ of joint sample has event.
2. $Y_{kj}$ be # at risk from group $l$.
3. $D_{ki} =$ # from strata $l$ dying at time $i$.
Phrase in terms of survival quantities.

1. Let $Y_j$ be # at risk when item $j$ of joint sample has event.
2. $Y_{kj}$ be # at risk from group $l$
3. $D_{ki} = # \text{ from strata } l \text{ dying at time } i$
4. Then $R_j = n + 1 - Y_j$
Phrase in terms of survival quantities.

1. Let $Y_j$ be # at risk when item $j$ of joint sample has event.
2. $Y_{kj}$ be # at risk from group $l$
3. $D_{ki} = $ # from strata $l$ dying at time $i$
4. Then $R_j = n + 1 - Y_j$
5. Use as test statistics $W_k = \sum_{j: g_j = k} aR_j$
Phrase in terms of survival quantities.

1. Let $Y_j$ be $\#$ at risk when item $j$ of joint sample has event.
2. $Y_{kj}$ be $\#$ at risk from group $l$
3. $D_{ki} = \#$ from strata $l$ dying at time $i$
4. Then $R_j = n + 1 - Y_j$
5. Use as test statistics $W_k = \sum_{j:g_j=k} aR_j$
6. Big when the observations for sample $l$ tend to be bigger than those in other groups.
Phrase in terms of survival quantities.

1. Let $Y_j$ be \# at risk when item $j$ of joint sample has event.
2. $Y_{kj}$ be \# at risk from group $l$
3. $D_{ki} = \#$ from strata $l$ dying at time $i$
4. Then $R_j = n + 1 - Y_j$
5. Use as test statistics $W_k = \sum_{j:g_j=k} aR_j$
6. Big when the observations for sample $l$ tend to be bigger than those in other groups.
7. In multi-group case, you can calculate statistic for either group.
Phrase in terms of survival quantities.

1. Let $Y_j$ be $\#$ at risk when item $j$ of joint sample has event.
2. $Y_{kj}$ be $\#$ at risk from group $l$
3. $D_{ki} = \#$ from strata $l$ dying at time $i$
4. Then $R_j = n + 1 - Y_j$
5. Use as test statistics $W_k = \sum_{j: g_j = k} aR_j$
6. Big when the observations for sample $l$ tend to be bigger than those in other groups.
7. In multi-group case, you can calculate statistic for either group.
   - $\sum_k W_k = \sum_k \sum_i (D_{ik} - D_i Y_{ik} / Y_i) = \sum_i \sum_{k=1}^2 (D_{ik} - D_i Y_{ik} / Y_i) = \sum_i (D_i - D_i Y_i / Y_i) \equiv 0$,
Phrase in terms of survival quantities.

1. Let $Y_j$ be $\#$ at risk when item $j$ of joint sample has event.
2. $Y_{kj}$ be $\#$ at risk from group $l$
3. $D_{ki} = \#$ from strata $l$ dying at time $i$
4. Then $R_j = n + 1 - Y_j$
5. Use as test statistics $W_k = \sum_{j: g_j = k} aR_j$
6. Big when the observations for sample $l$ tend to be bigger than those in other groups.
7. In multi-group case, you can calculate statistic for either group.
   1. $\sum_k W_k = \sum_k \sum_i (D_{ik} - D_i Y_{ik}/Y_i) = \sum_i \sum_{k=1}^2 (D_{ik} - D_i Y_{ik}/Y_i) = \sum_i (D_i - D_i Y_i/Y_i) \equiv 0,$
   2. so when $K = 2$ it doesn’t matter which group you use
Phrase in terms of survival quantities.

1. Let $Y_j$ be \# at risk when item $j$ of joint sample has event.
2. $Y_{kj}$ be \# at risk from group $l$
3. $D_{ki}$ = \# from strata $l$ dying at time $i$
4. Then $R_j = n + 1 - Y_j$
5. Use as test statistics $W_k = \sum_{j:g_j=k} a R_j$
6. Big when the observations for sample $l$ tend to be bigger than those in other groups.
7. In multi-group case, you can calculate statistic for either group.
   1. $\sum_k W_k = \sum_k \sum_i (D_{ik} - D_iY_{ik}/Y_i) = \sum_i \sum_{k=1}^2 (D_{ik} - D_iY_{ik}/Y_i) = \sum_i (D_i - D_iY_i/Y_i) \equiv 0$
   2. so when $K = 2$ it doesn’t matter which group you use
   3. Constraints also holds for larger $K$. R Code
Null Distribution:

1. Mean zero

At each time, under $H_0$, contribution to test $D_{ki} \sim \text{Hypergeometric}$.

Marginals are $D_i$, $Y_i - D_i$ and $Y_{ki}$, $Y_i$.

Variance contribution is Hypergeometric variance $Y_{ki}(Y_i - Y_{ki}) Y_i^2 Y_i - D_i Y_i - 1 D_i$.

Var $H_0[W_k] \approx \sum_{D_i=1} Y_{ki}(Y_i - Y_{ki}) Y_i^2 Y_i - D_i Y_i - 1 D_i$.

Test statistic is equivalent to Mantel–Haenzel test.

Variances add as with Greenwood's formula.

Variance formula is exact in no–censoring case, but depends on censoring mechanism.
Null Distribution:

1. Mean zero
2. Shape is approximately normal (Háyek 1960).
Null Distribution:

1. Mean zero
2. Shape is approximately normal (Háyek 1960).
3. At each time, under $H_0$, contribution to test $D_{ki} \sim \text{Hypergeometric}$. 
Null Distribution:

1. Mean zero
2. Shape is approximately normal (Háyek 1960).
3. At each time, under $H_0$, contribution to test $D_{ki} \sim \text{Hypergeometric}$.
   1. Marginals are $D_i$, $Y_i - D_i$ and $Y_{ki}$, $Y_i$. 

Testing with survival curves: Introduction to survival curve hypothesis testing

Lecture 03
Null Distribution:

1. Mean zero
2. Shape is approximately normal (Háyek 1960).
3. At each time, under $H_0$, contribution to test $D_{ki} \sim \text{Hypergeometric}$.
   1. Marginals are $D_i$, $Y_i - D_i$ and $Y_{ki}$, $Y_i$.
   2. Table:

$$
\begin{array}{c|c|c|c}
D_{ki} & D_i - D_{ki} & D_i \\
Y_{ki} & Y_i - D_i - (Y_{ki} - D_{ki}) & Y_i - D_i \\
Y_{ki} & Y_i - Y_{ki} & Y_i \\
\end{array}
$$
Null Distribution:

1. Mean zero
2. Shape is approximately normal (Háyek 1960).
3. At each time, under $H_0$, contribution to test $D_{ki} \sim \text{Hypergeometric}$.
   
   - Marginals are $D_i, Y_i - D_i$ and $Y_{ki}, Y_i$.
   - Table:
     
     $\begin{array}{cccc}
     D_{ki} & D_i - D_{ki} & D_i & Y_i \\
     Y_{ki} - D_{ki} & Y_i - D_i - (Y_{ki} - D_{ki}) & Y_i - D_i & Y_i \\
     \hline
     Y_{ki} & Y_{ki} & Y_{ki} - Y_{ki} & \end{array}$

3. Variance contribution is Hypergeometric variance $\frac{Y_{ki}(Y_i - Y_{ki})}{Y_i^2} \frac{Y_i - D_i}{Y_i - 1} D_i$. 

Testing with survival curves: Introduction to survival curve hypothesis testing
Null Distribution:

1. Mean zero
2. Shape is approximately normal (Háyek 1960).
3. At each time, under \( H_0 \), contribution to test \( D_{ki} \sim \text{Hypergeometric} \).
   1. Marginals are \( D_i, Y_i - D_i \) and \( Y_{ki}, Y_i \).
   2. Table:

\[
\begin{array}{c|c|c}
D_{ki} & D_i - D_{ki} & D_i \\
Y_{ki} - D_{ki} & Y_i - D_i - (Y_{ki} - D_{ki}) & Y_i - Y_{ki} \\
Y_{ki} & Y_i - Y_{ki} & Y_i \\
\end{array}
\]

3. Variance contribution is Hypergeometric variance \( \frac{Y_{ki}(Y_i - Y_{ki})}{Y_i^2} \frac{Y_i - D_i}{Y_i - 1} D_i \).
4. \( \text{Var}_{H_0} [W_k] \approx \sum_{i=1}^{D} \frac{Y_{ki}(Y_i - Y_{ki})}{Y_i^2} \frac{Y_i - D_i}{Y_i - 1} D_i \)
Null Distribution:

1. Mean zero
2. Shape is approximately normal (Háyek 1960).
3. At each time, under $H_0$, contribution to test $D_{ki} \sim \text{Hypergeometric}$.
   a. Marginals are $D_i$, $Y_i - D_i$ and $Y_{ki}$, $Y_i$.
   b. Table:

<table>
<thead>
<tr>
<th>$D_{ki}$</th>
<th>$D_i - D_{ki}$</th>
<th>$D_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{ki} - D_{ki}$</td>
<td>$Y_i - D_i - (Y_{ki} - D_{ki})$</td>
<td>$Y_i - D_i$</td>
</tr>
<tr>
<td>$Y_{ki}$</td>
<td>$Y_i - Y_{ki}$</td>
<td>$Y_i$</td>
</tr>
</tbody>
</table>

4. Variance contribution is Hypergeometric variance \( \frac{Y_{ki}(Y_i - Y_{ki})}{Y_i^2} \frac{Y_i - D_i}{Y_i - 1} D_i \).
5. \( \text{Var}_{H_0} [W_k] \approx \sum_{i=1}^{D} \frac{Y_{ki}(Y_i - Y_{ki})}{Y_i^2} \frac{Y_i - D_i}{Y_i - 1} D_i \)
6. Test statistic is equivalent to Mantel–Haenzel test.
Null Distribution:

1. Mean zero
2. Shape is approximately normal (Háyek 1960).
3. At each time, under $H_0$, contribution to test $D_{ki} \sim \text{Hypergeometric}$.
   1. Marginals are $D_i$, $Y_i - D_i$ and $Y_{ki}$, $Y_i$.
   2. Table:

<table>
<thead>
<tr>
<th></th>
<th>$D_{ki}$</th>
<th>$D_i - D_{ki}$</th>
<th>$D_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{ki}$</td>
<td>$Y_{ki} - D_{ki}$</td>
<td>$Y_i - D_i - (Y_{ki} - D_{ki})$</td>
<td>$Y_i - D_i$</td>
</tr>
<tr>
<td>$Y_{ki}$</td>
<td>$Y_{ki}$</td>
<td>$Y_i - Y_{ki}$</td>
<td>$Y_i$</td>
</tr>
</tbody>
</table>

3. Variance contribution is Hypergeometric variance $\frac{Y_{ki}(Y_i - Y_{ki})}{Y_i^2} \frac{Y_i - D_i}{Y_i - 1} D_i$.

4. $\text{Var}_{H_0} [W_k] \approx \sum_{i=1}^{D} \frac{Y_{ki}(Y_i - Y_{ki})}{Y_i^2} \frac{Y_i - D_i}{Y_i - 1} D_i$

1. Test statistic is equivalent to Mantel–Haenzel test.
2. Variances add
Null Distribution:

1. Mean zero
2. Shape is approximately normal (Háyek 1960).
3. At each time, under $H_0$, contribution to test $D_{ki} \sim \text{Hypergeometric}$.
   - Marginals are $D_i$, $Y_i - D_i$ and $Y_{ki}$, $Y_i$.
   - Table:
     
     \[
     \begin{array}{cccc}
     D_{ki} & D_i - D_{ki} & D_i \\
     Y_{ki} - D_{ki} & Y_i - D_i - (Y_{ki} - D_{ki}) & Y_i - Y_{ki} \\
     Y_{ki} & Y_i - Y_{ki} & Y_i \\
     \end{array}
     \]
4. Variance contribution is Hypergeometric variance $\frac{Y_{ki}(Y_i - Y_{ki})}{Y_i^2} \frac{Y_i - D_i}{Y_i - 1} D_i$.

\[\text{Var}_{H_0} [W_k] \approx \sum_{i=1}^{D} \frac{Y_{ki}(Y_i - Y_{ki})}{Y_i^2} \frac{Y_i - D_i}{Y_i - 1} D_i\]

1. Test statistic is equivalent to Mantel–Haenzel test.
2. Variances add
   - as with Greenwood’s formula
Null Distribution:

1. Mean zero
2. Shape is approximately normal (Háyek 1960).
3. At each time, under $H_0$, contribution to test $D_{ki} \sim \text{Hypergeometric.}$
   1. Marginals are $D_i$, $Y_i - D_i$ and $Y_{ki}$, $Y_i$.
   2. Table:
      
      \[
      \begin{array}{ccc}
      D_{ki} & D_i - D_{ki} & D_i \\
      Y_{ki} - D_{ki} & Y_i - D_i - (Y_{ki} - D_{ki}) & Y_i - D_i \\
      Y_{ki} & Y_i - Y_{ki} & Y_i \\
      \end{array}
      \]

3. Variance contribution is Hypergeometric variance $\frac{Y_{ki}(Y_i - Y_{ki})}{Y_i^2} \frac{Y_i - D_i}{Y_i - 1} D_i$.

4. $\text{Var}_{H_0} [W_k] \approx \sum_{i=1}^{D} \frac{Y_{ki}(Y_i - Y_{ki})}{Y_i^2} \frac{Y_i - D_i}{Y_i - 1} D_i$
   1. Test statistic is equivalent to Mantel–Haenzel test.
   2. Variances add
      1. as with Greenwood’s formula
   3. Variance formula is exact in no–censoring case, but depends on censoring mechanism with censoring

R Code
SAS Code
Objectives Lecture 04

1. Ordered Groups

Tests for $K > 2$ groups
Stratified tests
Matched Pairs
Multi-sample tests based on curve differences
Procedures based on tests of means and medians

Readings: C §2.8, KM §7.3a, 7.3b, 7.5a, 7.5b, 7.6
Objectives Lecture 04

1. Ordered Groups
2. Two–sample hypothesis testing with survival curves using ranks
Objectives Lecture 04

1. Ordered Groups
2. Two-sample hypothesis testing with survival curves using ranks
3. Tests for $K > 2$ groups
Objectives Lecture 04

1. Ordered Groups
2. Two-sample hypothesis testing with survival curves using ranks
3. Tests for $K > 2$ groups
4. Stratified tests
Objectives Lecture 04

1. Ordered Groups
2. Two–sample hypothesis testing with survival curves using ranks
3. Tests for $K > 2$ groups
4. Stratified tests
5. Matched Pairs
Objectives Lecture 04

1. Ordered Groups
2. Two–sample hypothesis testing with survival curves using ranks
3. Tests for $K > 2$ groups
4. Stratified tests
5. Matched Pairs
6. Multi-sample tests based on curve differences
Objectives Lecture 04

1. Ordered Groups
2. Two–sample hypothesis testing with survival curves using ranks
3. Tests for \( K > 2 \) groups
4. Stratified tests
5. Matched Pairs
6. Multi-sample tests based on curve differences
7. Procedures based on tests of means and medians

Readings: C §2.8, KM §7.3a, 7.3b, 7.5a, 7.5b, 7.6
Objectives Lecture 04

1. Ordered Groups
2. Two-sample hypothesis testing with survival curves using ranks
3. Tests for $K > 2$ groups
4. Stratified tests
5. Matched Pairs
6. Multi-sample tests based on curve differences
7. Procedures based on tests of means and medians
8. Readings: C §2.8, KM §7.3a, 7.3b, 7.5a, 7.5b, 7.6
Use with censoring

1. Random censoring \perp lifetime OK

\[ W_k = \sum_{i=1}^{n} w_i \left[ D_{ki} - \left( \frac{D_i}{Y_i} \right) Y_{ki} \right] \]

- \( w_i = 1 \) for log rank
- \( w_i = Y_i \) gives Gehan's Wilcoxon test

Keep in mind that you shouldn't choose test after having looked at the data.
Use with censoring

1. Random censoring \perp lifetime OK
2. Weights can depend on censoring

Recall that the log rank statistic is (approximately) optimal for proportional hazards aka Lehmann alternative.

\[ W_k = \sum_{i=1}^{n} w_i \left[ D_{ki} - \left( D_i / Y_i \right) Y_{ki} \right] \]

\( w_i = 1 \) for log rank

\( w_i = Y_i \) gives Gehan’s Wilcoxon test.

Keep in mind that you shouldn’t choose test after having looked at the data.

SAS Code

R Code
Use with censoring

1. Random censoring \(\perp\) lifetime OK
2. Weights can depend on censoring
3. Variance approx. is poor when

\[
W_k = \sum_{i=1}^{n_i} w_i \left[ D_{ki} - \left( D_i / Y_i \right) Y_{ki} \right]
\]

4. \(w_i = 1\) for log rank
5. \(w_i = Y_i\) gives Gehan's Wilcoxon test.

Keep in mind that you shouldn't choose test after having looked at the data. SAS Code R Code
Use with censoring

1. Random censoring $\perp$ lifetime OK
2. Weights can depend on censoring
3. Variance approx. is poor when
   1. censoring distns are different in different strata and

Recall that the log rank statistic is (approximately) optimal for proportional hazards aka Lehmann alternative.

$W_k = \sum_{n_i=1} w_i [D_{ki} - (D_i/Y_i)Y_{ki}]$

$w_i = 1$ for log rank
$w_i = Y_i$ gives Gehan’s Wilcoxon test.

Keep in mind that you shouldn’t choose test after having looked at the data.

SAS Code

R Code

Testing with survival curves: Introduction to survival curve hypothesis testing
Use with censoring

1. Random censoring $\perp$ lifetime OK
2. Weights can depend on censoring
3. Variance approx. is poor when
   1. censoring distns are different in different strata and
4. Add weights to allow arbitrary weighting
Use with censoring

1. Random censoring $\perp$ lifetime OK
2. Weights can depend on censoring
3. Variance approx. is poor when
   - censoring distns are different in different strata and
4. Add weights to allow arbitrary weighting
   - Recall that the log rank statistic is (approximately) optimal for proportional hazards aka Lehmann alternative.
Use with censoring

1. Random censoring \( \perp \) lifetime OK
2. Weights can depend on censoring
3. Variance approx. is poor when
   1. censoring distns are different in different strata and
4. Add weights to allow arbitrary weighting
   1. Recall that the log rank statistic is (approximately) optimal for proportional hazards aka Lehmann alternative.
   2. \( W_k = \sum_{i=1}^{n} w_i [D_{ki} - (D_i/Y_i)Y_{ki}] \)
Use with censoring

1. Random censoring $\perp$ lifetime OK
2. Weights can depend on censoring
3. Variance approx. is poor when
testing
  1. censoring distns are different in different strata and
4. Add weights to allow arbitrary weighting
   1. Recall that the log rank statistic is (approximately) optimal for proportional hazards aka Lehmann alternative.
   2. $W_k = \sum_{i=1}^{n} w_i [D_{ki} - (D_i/Y_i) Y_{ki}]$
   3. $w_i = 1$ for log rank
Use with censoring

1. Random censoring $\perp$ lifetime OK
2. Weights can depend on censoring
3. Variance approx. is poor when
e. clockwise
4. Add weights to allow arbitrary weighting
   1. Recall that the log rank statistic is (approximately) optimal for
      proportional hazards aka Lehmann alternative.
   2. $W_k = \sum_{i=1}^{n} w_i [D_{ki} - (D_i / Y_i) Y_{ki}]$
   3. $w_i = 1$ for log rank
   4. $w_i = Y_i$ gives Gehan’s Wilcoxon test.
Use with censoring

1. Random censoring \perp lifetime OK
2. Weights can depend on censoring
3. Variance approx. is poor when
   - censoring distns are different in different strata and
4. Add weights to allow arbitrary weighting
   - Recall that the log rank statistic is (approximately) optimal for proportional hazards aka Lehmann alternative.
   \[ W_k = \sum_{i=1}^{n} w_i [D_{ki} - (D_i/Y_i)Y_{ki}] \]
   - \( w_i = 1 \) for log rank
   - \( w_i = Y_i \) gives Gehan’s Wilcoxon test.
   - Keep in mind that you shouldn’t choose test after having looked at the data. SAS Code R Code
Section: Testing with survival curves

Subsection: $K > 2$ sample case:
Notation:

\[ \mathbf{W} = (W_1, \ldots, W_K) \text{ (column vector)} \]
Notation:

1. \( \mathbf{W} = (W_1, \ldots, W_K) \) (column vector)
2. \( \text{Cov}_{H_0} [W_k, W_r] \approx - \sum_{i=1}^{D} w_i^2 \frac{Y_{ki} Y_{ri}}{Y_i^2} \frac{Y_i - D_i}{Y_i - 1} D_i \) if \( k \neq r \)
Notation:

1. $\mathbf{W} = (W_1, \ldots, W_K)$ (column vector)
2. $\text{Cov}_{H_0} [W_k, W_r] \approx - \sum_{i=1}^D w_i^2 \frac{Y_{ki}Y_{ri}}{Y_i^2} \frac{Y_i-D_i}{Y_{i-1}} D_i$ if $k \neq r$
   
   Calculate this by

Testing with survival curves: $K > 2$ sample case:
Notation:

1. \( W = (W_1, \ldots, W_K) \) (column vector)

2. \( \text{Cov}_{H_0} [W_k, W_r] \approx - \sum_{i=1}^{D} w_i^2 \frac{Y_{ki}}{Y_i^2} \frac{Y_{ri}-D_i}{Y_{i-1}} D_i \) if \( k \neq r \)

   Calculate this by

   Noting that \( W_k + W_r \) is the standard log rank statistic with groups \( k \) and \( r \) collapsed.
Notation:

1. \( W = (W_1, \ldots, W_K) \) (column vector)

2. \( \text{Cov}_{H_0}[W_k, W_r] \approx - \sum_{i=1}^{D} w_i^2 \frac{Y_{ki}}{Y_i^2} \frac{Y_{ri}}{Y_{i-1}} D_i \) if \( k \neq r \)

   1. Calculate this by
      1. Noting that \( W_k + W_r \) is the standard log rank statistic with groups \( k \) and \( r \) collapsed.
      2. So the hypergeometric formula applies to \( \text{Var}[W_k + W_r] \)

Testing with survival curves: \( K > 2 \) sample case:
Notation:

1. \( \mathbf{W} = (W_1, \ldots, W_K) \) (column vector)

2. \( \text{Cov}_{H_0} [W_k, W_r] \approx -\sum_{i=1}^{D} w_i^2 \frac{Y_{ki} Y_{ri}}{Y_i^2} \frac{Y_{i-D_i}}{Y_{i-1}} D_i \) if \( k \neq r \)

   1. Calculate this by
      1. Noting that \( W_k + W_r \) is the standard log rank statistic with groups \( k \) and \( r \) collapsed.
      2. So the hypergeometric formula applies to \( \text{Var} [W_k + W_r] \)
      3. Know \( \text{Var} [W_k] \), \( \text{Var} [W_r] \)

Testing with survival curves: \( K > 2 \) sample case:
Notation:

1. $\mathbf{W} = (W_1, \ldots, W_K)$ (column vector)

2. $\text{Cov}_{H_0} [W_k, W_r] \approx - \sum_{i=1}^{D} w_i^2 \frac{Y_{ki} Y_{ri}}{Y_i^2} \frac{Y_i - D_i}{Y_i - 1} D_i$ if $k \neq r$

Calculate this by

1. Noting that $W_k + W_r$ is the standard log rank statistic with groups $k$ and $r$ collapsed.
2. So the hypergeometric formula applies to $\text{Var}[W_k + W_r]$
3. Know $\text{Var}[W_k], \text{Var}[W_r]$
4. Solve $\text{Var}[W_k + W_r] = \text{Var}[W_k] + \text{Var}[W_r] + 2\text{Cov}[W_k, W_r]$ for $\text{Cov}[W_k, W_r]$

Testing with survival curves: $K > 2$ sample case:
Unordered case:

1. Like ANOVA, square $W_k$ and sum up
Unordered case:

1. Like ANOVA, square $W_k$ and sum up
   after adjusting for correlation
Unordered case:

1. Like ANOVA, square $W_k$ and sum up after adjusting for correlation

2. Since $\sum_{k=1}^{K} W_k = 0$, must drop one of these to recover invertible matrix.

Testing with survival curves: $K > 2$ sample case:
Unordered case:

1. Like ANOVA, square $W_k$ and sum up after adjusting for correlation.

2. Since $\sum_{k=1}^{K} W_k = 0$, must drop one of these to recover invertible matrix.
   - Can choose any one you want.
Unordered case:

1. Like ANOVA, square $W_k$ and sum up after adjusting for correlation

2. Since $\sum_{k=1}^{K} W_k = 0$, must drop one of these to recover invertible matrix.
   - Can choose any one you want.
   - Let’s choose first.
Unordered case:

1. Like ANOVA, square $W_k$ and sum up after adjusting for correlation.
2. Since $\sum_{k=1}^{K} W_k = 0$, must drop one of these to recover invertible matrix.
   - Can choose any one you want.
   - Let’s choose first.
   - Subscript $-1$ represents matrix or vector with first entry removed.
Unordered case:

1. Like ANOVA, square $W_k$ and sum up after adjusting for correlation.
2. Since $\sum_{k=1}^{K} W_k = 0$, must drop one of these to recover invertible matrix.
   - Can choose any one you want.
   - Let’s choose first.
   - Subscript $-1$ represents matrix or vector with first entry removed.
3. Make quadratic form for test statistic.
Unordered case:

1. Like ANOVA, square $W_k$ and sum up after adjusting for correlation.
2. Since $\sum_{k=1}^{K} W_k = 0$, must drop one of these to recover invertible matrix.
   - Can choose any one you want.
   - Let’s choose first.
   - Subscript $-1$ represents matrix or vector with first entry removed.
3. Make quadratic form for test statistic.
4. $T = W_{-1}^\top \text{Var}_{H_0} [W_{-1}]^{-1} W_{-1}$
Unordered case:

1. Like ANOVA, square $W_k$ and sum up after adjusting for correlation.

2. Since $\sum_{k=1}^{K} W_k = 0$, must drop one of these to recover invertible matrix.
   - Can choose any one you want.
   - Let’s choose first.
   - Subscript $-1$ represents matrix or vector with first entry removed.

3. Make quadratic form for test statistic.

4. $T = W_{-1}^{\top} \text{Var}_{H_0} [W_{-1}]^{-1} W_{-1}$

5. Under $H_0$, $T \sim \chi^2_{K-1}$
Unordered case:

1. Like ANOVA, square $W_k$ and sum up after adjusting for correlation
2. Since $\sum_{k=1}^{K} W_k = 0$, must drop one of these to recover invertible matrix.
   1. Can choose any one you want.
   2. Let’s choose first.
   3. Subscript $-1$ represents matrix or vector with first entry removed.
3. Make quadratic form for test statistic.
4. $T = W_{-1}^\top \text{Var}_{H_0} [W_{-1}]^{-1} W_{-1}$
5. Under $H_0$, $T \sim \chi^2_{K-1}$

Testing with survival curves: $K > 2$ sample case:
Ordered case

1 Here we test $H_A : \alpha_i \geq \alpha_{i-1} \forall i$ with inequality somewhere vs $H_0 : \alpha_i = \alpha_{i-1} \forall i$
Ordered case

1. Here we test $H_A : \alpha_i \geq \alpha_{i-1} \forall i$ with inequality somewhere vs $H_0 : \alpha_i = \alpha_{i-1} \forall i$

2. Make scores $a_1 < a_2 < \cdots < a_K$ (often $0, \ldots, K - 1$)
Ordered case

1. Here we test $H_A : \alpha_i \geq \alpha_{i-1} \forall i$ with inequality somewhere vs $H_0 : \alpha_i = \alpha_{i-1} \forall i$

2. Make scores $a_1 < a_2 < \cdots < a_K$ (often $0, \ldots, K - 1$)

3. Test with $Z = \sum_k a_k W_k$
Ordered case

1. Here we test $H_A : \alpha_i \geq \alpha_{i-1} \forall i$ with inequality somewhere vs $H_0 : \alpha_i = \alpha_{i-1} \forall i$

2. Make scores $a_1 < a_2 < \cdots < a_K$ (often $0, \ldots, K-1$)

3. Test with $Z = \sum_{k} a_k W_k$

4. $E_0 [Z] = 0$
Ordered case

1. Here we test $H_A : \alpha_i \geq \alpha_{i-1} \forall i$ with inequality somewhere vs $H_0 : \alpha_i = \alpha_{i-1} \forall i$

2. Make scores $a_1 < a_2 < \cdots < a_K$ (often $0, \ldots, K - 1$)

3. Test with $Z = \sum_k a_k W_k$

4. $E_0 [Z] = 0$

5. $\text{Var}_0 [Z] = a^\top \text{Var}_0 [W] a$

In contrast with the unordered case, there is no need to invert $\text{Var}_0 [W]$. Because $\sum_k W_k = 0$, the test is unchanged if a constant is added to each score. Because one divides by the standard deviation before comparing to the standard normal distribution, conclusion is unchanged if a constant is multiplied to each score.
Ordered case

1. Here we test $H_A : \alpha_i \geq \alpha_{i-1} \forall i$ with inequality somewhere vs $H_0 : \alpha_i = \alpha_{i-1} \forall i$

2. Make scores $a_1 < a_2 < \cdots < a_K$ (often $0, \ldots, K - 1$)

3. Test with $Z = \sum_k a_k W_k$

4. $\mathbb{E}_0 [Z] = 0$

5. $\text{Var}_0 [Z] = a^\top \text{Var}_0 [W] a$

In contrast with the unordered case, there is no need to invert $\text{Var}_0 [W]$. 

Testing with survival curves: $K > 2$ sample case:
**Ordered case**

1. Here we test $H_A: \alpha_i \geq \alpha_{i-1} \forall i$ with inequality somewhere vs $H_0: \alpha_i = \alpha_{i-1} \forall i$

2. Make scores $a_1 < a_2 < \cdots < a_K$ (often $0, \ldots, K - 1$)

3. Test with $Z = \sum_k a_k W_k$

4. $E_0 [Z] = 0$

5. $\text{Var}_0 [Z] = a^\top \text{Var}_0 [W] a$

   - In contrast with the unordered case, there is no need to invert $\text{Var}_0 [W]$.

6. Because $\sum_{k=1}^m W_k = 0$, the test is unchanged if a constant is added to each score.
Ordered case

1. Here we test $H_A : \alpha_i \geq \alpha_{i-1} \forall i$ with inequality somewhere vs $H_0 : \alpha_i = \alpha_{i-1} \forall i$
2. Make scores $a_1 < a_2 < \cdots < a_K$ (often $0, \ldots, K - 1$)
3. Test with $Z = \sum_k a_k W_k$
4. $E_0[Z] = 0$
5. $\text{Var}_0[Z] = a^\top \text{Var}_0[W] a$
   - In contrast with the unordered case, there is no need to invert $\text{Var}_0[W]$.
6. Because $\sum_{k=1}^m W_k = 0$, the test is unchanged if a constant is added to each score.
7. Because one divides by the standard deviation before comparing to the standard normal distribution, conclusion is unchanged if a constant is multiplied to each score.  

Testing with survival curves: $K > 2$ sample case:

R Code  SAS Code
Section: Testing with survival curves

Subsection: Two-sample testing
Generally will not base tests on single-number summaries

1. means

Problem: Sometimes mean (and variance) aren't well defined.

Median

Procedure as above

Requires estimate of density

not too much censoring

Value at a fixed time

Procedure as above

Generally it isn't the question we want to answer
Generally will not base tests on single-number summaries

1. means
   1. Procedure

---

Problem: Sometimes mean (and variance) aren’t well defined.

Medians

Procedure as above

Requires

1. estimate of density

2. not too much censoring

Value at a fixed time

Procedure as above

Generally it isn’t the question we want to answer
Generally will not base tests on single-number summaries

1. **means**
   1. **Procedure**
      1. estimate means and SE for populations separately

Problem: Sometimes mean (and variance) aren't well defined.

Medians

Procedure as above

Requires

estimate of density

not too much censoring

Value at a fixed time

Procedure as above

Generally it isn't the question we want to answer
Generally will not base tests on single-number summaries

1. means
   1. Procedure
      1. estimate means and SE for populations separately
      2. Use fact that variance of difference of independent quantities is sum of variances
Generally will not base tests on single-number summaries

1. **means**
   
   **Procedure**
   
   1. estimate means and SE for populations separately
   2. Use fact that variance of difference of independent quantities is sum of variances
   3. Compare to normal table

---

Testing with survival curves: Two-sample testing

Lecture 04
Generally will not base tests on single-number summaries

1 means

1 Procedure
   1 estimate means and SE for populations separately
   2 Use fact that variance of difference of independent quantities is sum of variances
   3 Compare to normal table

2 Problem: Sometimes mean (and variance) aren’t well defined.
Generally will not base tests on single-number summaries

1. **means**
   - Procedure
     1. estimate means and SE for populations separately
     2. Use fact that variance of difference of independent quantities is sum of variances
     3. Compare to normal table
   - Problem: Sometimes mean (and variance) aren’t well defined.

2. **Medians**
Generally will not base tests on single-number summaries

1. **means**
   - **Procedure**
     1. Estimate means and SE for populations separately
     2. Use fact that variance of difference of independent quantities is sum of variances
     3. Compare to normal table
   - **Problem:** Sometimes mean (and variance) aren’t well defined.

2. **Medians**
   - **Procedure** as above
Generally will not base tests on single-number summaries

1. **Means**
   1. Procedure
      1. Estimate means and SE for populations separately
      2. Use fact that variance of difference of independent quantities is sum of variances
      3. Compare to normal table
   2. Problem: Sometimes mean (and variance) aren’t well defined.

2. **Medians**
   1. Procedure as above
   2. Requires
Generally will not base tests on single-number summaries

1. **Means**
   
   **Procedure**
   
   1. Estimate means and SE for populations separately
   2. Use fact that variance of difference of independent quantities is sum of variances
   3. Compare to normal table

   **Problem**: Sometimes mean (and variance) aren’t well defined.

2. **Medians**

   **Procedure as above**

   **Requires**
   
   1. Estimate of density
Generally will not base tests on single-number summaries

1. **Means**
   - **Procedure**
     1. Estimate means and SE for populations separately
     2. Use fact that variance of difference of independent quantities is sum of variances
     3. Compare to normal table
   - **Problem:** Sometimes mean (and variance) aren’t well defined.

2. **Medians**
   - **Procedure as above**
   - **Requires**
     1. Estimate of density
     2. Not too much censoring
Generally will not base tests on single-number summaries

1. means
   1. Procedure
      1. estimate means and SE for populations separately
      2. Use fact that variance of difference of independent quantities is sum of variances
      3. Compare to normal table
   2. Problem: Sometimes mean (and variance) aren’t well defined.

2. Medians
   1. Procedure as above
   2. Requires
      1. estimate of density
      2. not too much censoring

3. Value at a fixed time
Generally will not base tests on single-number summaries

1. Means
   1. Procedure
      1. Estimate means and SE for populations separately
      2. Use fact that variance of difference of independent quantities is sum of variances
      3. Compare to normal table
   2. Problem: Sometimes mean (and variance) aren’t well defined.

2. Medians
   1. Procedure as above
   2. Requires
      1. Estimate of density
      2. Not too much censoring

3. Value at a fixed time
   1. Procedure as above
Generally will not base tests on single-number summaries

1. Means
   - Procedure
     1. Estimate means and SE for populations separately
     2. Use fact that variance of difference of independent quantities is sum of variances
     3. Compare to normal table

2. Problem: Sometimes mean (and variance) aren’t well defined.

3. Medians
   - Procedure as above
   - Requires
     1. Estimate of density
     2. Not too much censoring

4. Value at a fixed time
   - Procedure as above
   - Generally it isn’t the question we want to answer
Section: Testing with survival curves

Subsection: Stratified Tests
Assumptions

1. Each population is composed of strata
Assumptions

1. Each population is composed of strata
2. Effect of population is same in each strata
Assumptions

1. Each population is composed of strata
2. Effect of population is same in each strata
3. Null survival curve might be different
$h_{ij}(t)$ is hazard for population $i$ in strata $j$
Mathematical Formulation

1. $h_{ij}(t)$ is hazard for population $i$ in strata $j$

2. $h_{ij}(t) = \alpha_i h_j(t)$
Mathematical Formulation

1. \( h_{ij}(t) \) is hazard for population \( i \) in strata \( j \)
2. \( h_{ij}(t) = \alpha_i h_j(t) \)
3. \( H_0 : \alpha_i \) all equal
Mathematical Formulation

1. \( h_{ij}(t) \) is hazard for population \( i \) in strata \( j \)
2. \( h_{ij}(t) = \alpha_i h_j(t) \)
3. \( H_0 : \alpha_i \) all equal
4. \( H_A : \alpha_i \) not all equal
Solution:

1. Do test separately on each strata
Solution:

1. Do test separately on each strata
   4. Using for ex log rank
Solution:

1. Do test separately on each strata
   1. Using for ex log rank

2. Let $W_{kj} = \text{statistic for population } k \text{ in stratum } j$
Solution:

1. Do test separately on each strata
   1. Using for ex log rank

2. Let $W_{kj} =$ statistic for population $k$ in stratum $j$

3. $W_k = \sum_j W_{kj}$
Solution:

1. Do test separately on each strata
   1. Using for ex log rank

2. Let $W_{kj} =$ statistic for population $k$ in stratum $j$

3. $W_k = \sum_j W_{kj}$

4. $\text{Var}[W_k] = \sum_j \text{Var}[W_{kj}]$
Solution:

1. Do test separately on each strata
   - Using for ex log rank
2. Let $W_{kj} =$ statistic for population $k$ in stratum $j$
3. $W_k = \sum_j W_{kj}$
4. $\text{Var} [W_k] = \sum_j \text{Var} [W_{kj}]$
5. Proceed as before  R Code  SAS Code
Extreme case of Stratification: matched pairs

1. Each pair is a stratum

Each test is now calculated on two observations. Each test statistic is the sum of two terms. The second term is always 1. The first term is 

\[ 1 - \frac{D_2}{D_1} \]  

if the event is in group 1 or 

\[ -1 + \frac{D_1}{D_2} \]  

if the event is in other group. The statistic is half of (event in group 1 - event in group 2).

Variance approximation: Each contribution is \( \frac{1}{4} \). Variance is \( \frac{1}{4} \) times the number of strata.

\[ W_1 = \frac{(D_1 - D_2)}{\sqrt{D_1 + D_2}} \]

Weights cancel.

\( D_i \) is the number of pairs in which group 1 patient dies first.

SAS does this with TEST.
Extreme case of Stratification: matched pairs

1. Each pair is a stratum
2. Each test is now calculated on two observations

Each contribution is \( \frac{1}{n_1} \) for variance, where \( n_1 \) is the number of strata.

Weights cancel.

\[ W_i = \frac{D_1 - D_2}{\sqrt{D_1 + D_2}} \]

\( D_i \) is number of pairs in which group 1 patient dies first.
Extreme case of Stratification: matched pairs

1. Each pair is a stratum
2. Each test is now calculated on two observations
3. Each test statistic is sum of two terms

- Second term is always 1
- First term is \((1 - 1)^2\) if first event is in group or 
- 
  \(-1^2\) if event is in other group
- Statistic is half \((\text{event in group } 1 - \text{event in group } 2)\)

Variance approximation

- Each contribution is \(\frac{1}{4}\)
- Variance is \(\frac{1}{4}\) times number of strata

\[ W_1 = \frac{(D_1 - D_2)}{\sqrt{D_1 + D_2}} \]

Weights cancel.

\(D_i\) is number of pairs in which group 1 patient dies first

SAS does this with TEST
Extreme case of Stratification: matched pairs

1. Each pair is a stratum
2. Each test is now calculated on two observations
3. Each test statistic is sum of two terms
   - Second term is always $1 - 1 = 0$ or $0 - 0 = 0$
4. Variance approximation
   - Each contribution is $\frac{1}{4}$
   - Variance is $\frac{1}{4}$ times number of strata
5. $W = \frac{D_1 - D_2}{\sqrt{D_1 + D_2}}$
6. $D_i$ is number of pairs in which group 1 patient dies first
Extreme case of Stratification: matched pairs

1. Each pair is a stratum
2. Each test is now calculated on two observations
3. Each test statistic is sum of two terms
   - Second term is always $1 - 1 = 0$ or $0 - 0 = 0$
   - First term is $(1 - \frac{1}{2})$ if first event is in group or $-\frac{1}{2}$ if event is in other group

Statistic is half (event in group 1 - event in group 2)

Variance approximation

Each contribution is $\frac{1}{4}$

Variance is $\frac{1}{4}$ times number of strata

$W_1 = \frac{D_1 - D_2}{\sqrt{D_1 + D_2}}$

Weights cancel.

$D_i$ is number of pairs in which group 1 patient dies first

SAS does this with TEST
Extreme case of Stratification: matched pairs

1. Each pair is a stratum
2. Each test is now calculated on two observations
3. Each test statistic is sum of two terms
   1. Second term is always $1 - 1 = 0$ or $0 - 0 = 0$
   2. First term is $(1 - \frac{1}{2})$ if first event is in group or $-\frac{1}{2}$ if event is in other group
   3. Statistic is half (event in group 1 - event in group 2)
Extreme case of Stratification: matched pairs

1. Each pair is a stratum
2. Each test is now calculated on two observations
3. Each test statistic is sum of two terms
   - Second term is always $1 - 1 = 0$ or $0 - 0 = 0$
   - First term is $(1 - \frac{1}{2})$ if first event is in group or $-\frac{1}{2}$ if event is in other group
   - Statistic is half (event in group 1 - event in group 2)
4. Variance approximation

$W_i = \frac{(D_1 - D_2)}{\sqrt{D_1 + D_2}}$
Extreme case of Stratification: matched pairs

1. Each pair is a stratum
2. Each test is now calculated on two observations
3. Each test statistic is sum of two terms
   1. Second term is always $1 - 1 = 0$ or $0 - 0 = 0$
   2. First term is $(1 - \frac{1}{2})$ if first event is in group or $-\frac{1}{2}$ if event is in other group
   3. Statistic is half (event in group 1 - event in group 2)
4. Variance approximation
   1. Each contribution is $\frac{1}{4}$
Extreme case of Stratification: matched pairs

1. Each pair is a stratum
2. Each test is now calculated on two observations
3. Each test statistic is sum of two terms
   1. Second term is always $1 - 1 = 0$ or $0 - 0 = 0$
   2. First term is $(1 - \frac{1}{2})$ if first event is in group or $-\frac{1}{2}$ if event is in other group
   3. Statistic is half (event in group 1 - event in group 2)
4. Variance approximation
   1. Each contribution is $\frac{1}{4}$
   2. Variance is $\frac{1}{4}$ times number of strata
Extreme case of Stratification: matched pairs

1. Each pair is a stratum
2. Each test is now calculated on two observations
3. Each test statistic is sum of two terms
   1. Second term is always $1 - 1 = 0$ or $0 - 0 = 0$
   2. First term is $(1 - \frac{1}{2})$ if first event is in group or $-\frac{1}{2}$ if event is in other group
   3. Statistic is half (event in group 1 - event in group 2)
4. Variance approximation
   1. Each contribution is $\frac{1}{4}$
   2. Variance is $\frac{1}{4}$ times number of strata
5. $W_1 = (D_1 - D_2)/\sqrt{D_1 + D_2}$
Extreme case of Stratification: matched pairs

1. Each pair is a stratum
2. Each test is now calculated on two observations
3. Each test statistic is sum of two terms
   - Second term is always $1 - 1 = 0$ or $0 - 0 = 0$
   - First term is $(1 - \frac{1}{2})$ if first event is in group or $-\frac{1}{2}$ if event is in other group
   - Statistic is half (event in group 1 - event in group 2)
4. Variance approximation
   - Each contribution is $\frac{1}{4}$
   - Variance is $\frac{1}{4}$ times number of strata
5. $W_1 = (D_1 - D_2)/\sqrt{D_1 + D_2}$
   - Weights cancel.
Extreme case of Stratification: matched pairs

1. Each pair is a stratum
2. Each test is now calculated on two observations
3. Each test statistic is sum of two terms
   - Second term is always $1 - 1 = 0$ or $0 - 0 = 0$
   - First term is $(1 - \frac{1}{2})$ if first event is in group or $-\frac{1}{2}$ if event is in other group
   - Statistic is half (event in group 1 - event in group 2)
4. Variance approximation
   - Each contribution is $\frac{1}{4}$
   - Variance is $\frac{1}{4}$ times number of strata
5. $W_1 = \frac{(D_1 - D_2)}{\sqrt{D_1 + D_2}}$
   - Weights cancel.
6. $D_i$ is number of pairs in which group 1 patient dies first
Extreme case of Stratification: matched pairs

1. Each pair is a stratum
2. Each test is now calculated on two observations
3. Each test statistic is sum of two terms
   1. Second term is always $1 - 1 = 0$ or $0 - 0 = 0$
   2. First term is $(1 - \frac{1}{2})$ if first event is in group or $-\frac{1}{2}$ if event is in other group
   3. Statistic is half (event in group 1 - event in group 2)
4. Variance approximation
   1. Each contribution is $\frac{1}{4}$
   2. Variance is $\frac{1}{4}$ times number of strata
5. $W_1 = (D_1 - D_2) / \sqrt{D_1 + D_2}$
   1. Weights cancel.
6. $D_i$ is number of pairs in which group 1 patient dies first
7. SAS does this with TEST
Section: Testing with survival curves
Subsection: Tests based on Curve Differences
Review of general unequal alternative sans Censoring

1 Approach 1: Maximal difference between CDFs gives *Kolmogorov–Smirnov test*.

\[ \text{Calculation is easy: maximal difference happens at event times.} \]

\[ \text{Distribution approximation is hard: use permutation distribution.} \]

2 Approach 2: integrated squared difference gives *Cramér-von Mises test*.

\[ \int |S_1(x) - S_2(x)|^2 f(x) \, dx = \int |S_1(x) - S_2(x)|^2 dF(x) \]

\[ \text{Sample quantity:} \]

\[ \sum |\hat{S}_1(t_i) - \hat{S}_2(t_i)|^2 \hat{p}(t_i) / \hat{S}(t_i) \]

\[ \text{Variant: Drop} \]

\[ \text{Variant: Weight by } S(x)(1 - S(x)) \]

\[ \text{Anderson-Darling test.} \]
Review of general unequal alternative sans Censoring

1. **Approach 1:** Maximal difference between CDFs gives *Kolmogorov–Smirnov test*.
   1. Same idea could work for Kaplan-Meier curves.
Approach 1: Maximal difference between CDFs gives *Kolmogorov–Smirnov test*.

1. Same idea could work for Kaplan-Meier curves
2. Calculation is easy: maximal difference happens at event times.
Approach 1: Maximal difference between CDFs gives *Kolmogorov–Smirnov test*.

1. Same idea could work for Kaplan-Meier curves
2. Calculation is easy: maximal difference happens at event times.
3. Distribution approximation is hard: use permutation distribution.
Review of general unequal alternative sans Censoring

1. Approach 1: Maximal difference between CDFs gives *Kolmogorov–Smirnov test*.
   - Same idea could work for Kaplan-Meier curves
   - Calculation is easy: maximal difference happens at event times.
   - Distribution approximation is hard: use permutation distribution.

Approach 1: Maximal difference between CDFs gives *Kolmogorov–Smirnov test*.  
1. Same idea could work for Kaplan-Meier curves  
2. Calculation is easy: maximal difference happens at event times.  
3. Distribution approximation is hard: use permutation distribution.

Approach 2: integrated squared difference gives *Cramér -von Mises test*.  
1. Population quantity weighted by common null density:  
\[ \int |S_1(x) - S_2(x)|^2 f(x) \, dx = \int |S_1(x) - S_2(x)|^2 dF(x) \]  
2. Sample quantity:  
\[ \sum |\hat{S}_1(t_i) - \hat{S}_2(t_i)|^2 \hat{p}(t_i) / (\hat{S}(t_i) - \hat{S}(t_i)) \], the fitted pooled probability.
Approach 1: Maximal difference between CDFs gives *Kolmogorov–Smirnov* test.

1. Same idea could work for Kaplan-Meier curves
2. Calculation is easy: maximal difference happens at event times.
3. Distribution approximation is hard: use permutation distribution.

Approach 2: integrated squared difference gives *Cramér -von Mises* test.

1. Population quantity weighted by common null density: \( \int |S_1(x) - S_2(x)|^2 f(x) \, dx = \int |S_1(x) - S_2(x)|^2 dF(x) \)
2. Sample quantity: \( \sum |\hat{S}_1(t_i) - \hat{S}_2(t_i)|^2 \hat{p}(t_i) \) for \( \hat{p}(t_i) = \hat{S}(t_i -) - \hat{S}(t_i) \), the fitted pooled probability.
Approach 1: Maximal difference between CDFs gives Kolmogorov–Smirnov test.

- Same idea could work for Kaplan-Meier curves
- Calculation is easy: maximal difference happens at event times.
- Distribution approximation is hard: use permutation distribution.

Approach 2: integrated squared difference gives Cramér-von Mises test.

- Population quantity weighted by common null density: \[ \int |S_1(x) - S_2(x)|^2 f(x) \, dx = \int |S_1(x) - S_2(x)|^2 dF(x) \]
- Sample quantity: \[ \int |\hat{S}_1(x) - \hat{S}_2(x)|^2 d\hat{F}(x) = \sum |\hat{S}_1(t_i) - \hat{S}_2(t_i)|^2 \hat{p}(t_i) \]
  for \( \hat{p}(t_i) = \hat{S}(t_i-) - \hat{S}(t_i) \), the fitted pooled probability.
- Variant: Drop \( \hat{S}_2(t_i) \)
Approach 1: Maximal difference between CDFs gives *Kolmogorov–Smirnov test*.  
1. Same idea could work for Kaplan-Meier curves  
2. Calculation is easy: maximal difference happens at event times.  
3. Distribution approximation is hard: use permutation distribution.

Approach 2: integrated squared difference gives *Cramér-von Mises test*.  
1. Population quantity weighted by common null density:  
   \[ \int |S_1(x) - S_2(x)|^2 f(x) \, dx = \int |S_1(x) - S_2(x)|^2 dF(x) \]
2. Sample quantity:  
   \[ \sum |\hat{S}_1(t_i) - \hat{S}_2(t_i)|^2 \hat{p}(t_i) \]
   for \( \hat{p}(t_i) = \hat{S}(t_i-) - \hat{S}(t_i) \), the fitted pooled probability.  
3. Variant: Drop  
4. Variant: Weight by \( S(x)(1 - S(x)) \):
Review of general unequal alternative sans Censoring

1 Approach 1: Maximal difference between CDFs gives *Kolmogorov–Smirnov test*.

1 Same idea could work for Kaplan-Meier curves
2 Calculation is easy: maximal difference happens at event times.
3 Distribution approximation is hard: use permutation distribution.

2 Approach 2: integrated squared difference gives *Cramér-von Mises test*.

1 Population quantity weighted by common null density: \( \int |S_1(x) - S_2(x)|^2 f(x) \, dx = \int |S_1(x) - S_2(x)|^2 dF(x) \)
2 Sample quantity: \( \sum |\hat{S}_1(t_i) - \hat{S}_2(t_i)|^2 \hat{p}(t_i) \) for \( \hat{p}(t_i) = \hat{S}(t_i-) - \hat{S}(t_i) \), the fitted pooled probability.
3 Variant: Drop \( S(x)(1 - S(x)) \):
4 Variant: Weight by \( S(x)(1 - S(x)) \):

1 Population quantity \( \int |S_1(x) - S_2(x)|^2(F(x)(1 - F(x)))^{-1} dF(x) \)
Review of general unequal alternative sans Censoring

1 Approach 1: Maximal difference between CDFs gives *Kolmogorov–Smirnov test*.
   1 Same idea could work for Kaplan-Meier curves
   2 Calculation is easy: maximal difference happens at event times.
   3 Distribution approximation is hard: use permutation distribution.

2 Approach 2: integrated squared difference gives *Cramér-von Mises test*.
   1 Population quantity weighted by common null density:
      \[ \int |S_1(x) - S_2(x)|^2 f(x) \, dx = \int |S_1(x) - S_2(x)|^2 dF(x) \]
   2 Sample quantity:
      \[ \int |\hat{S}_1(x) - \hat{S}_2(x)|^2 d\hat{F}(x) = \sum |\hat{S}_1(t_i) - \hat{S}_2(t_i)|^2 \hat{p}(t_i) \]
      for \( \hat{p}(t_i) = \hat{S}(t_i-) - \hat{S}(t_i) \), the fitted pooled probability.
   3 Variant: Drop \( \hat{S}(t_i) \)
   4 Variant: Weight by \( S(x)(1 - S(x)) \):
      1 Population quantity
         \[ \int |S_1(x) - S_2(x)|^2 (F(x)(1 - F(x)))^{-1} dF(x) \]
      2 Sample quantity
         \[ \sum |\hat{S}_1(t_i) - \hat{S}_2(t_i)|^2 \hat{p}(t_i)/(|\hat{S}(t_i)(1 - \hat{S}(t_i))|): \text{Anderson-Darling test.} \]
Objectives Lecture 05

1. Sample size calculations
Objectives Lecture 05

1. Sample size calculations
2. Regression models
Objectives Lecture 05

1. Sample size calculations
2. Regression models
3. Tests based on integrated differences

Testing with survival curves: Tests based on Curve Differences
Objectives Lecture 05

1. Sample size calculations
2. Regression models
3. Tests based on integrated differences
4. Cox model
Objectives Lecture 05

1. Sample size calculations
2. Regression models
3. Tests based on integrated differences
4. Cox model
5. Proportional hazards likelihood

Readings: C § 9.1–9.3, KM § 2.6a, 7.7, 8.1a, 8.3
Objectives Lecture 05

1. Sample size calculations
2. Regression models
3. Tests based on integrated differences
4. Cox model
5. Proportional hazards likelihood
6. Readings: C §9.1–9.3, KM §2.6a, 7.7, 8.1a, 8.3
General unequal alternative with censoring

1 approach 1: maximum of partial log rank statistic gives Rényi test.
General unequal alternative with censoring

1 approach 1: maximum of partial log rank statistic gives Rényi test.
   Look for maximal size of one of the log rank statistics
General unequal alternative with censoring

1. approach 1: maximum of partial log rank statistic gives \textit{Rényi test}.
   1. Look for maximal size of one of the log rank statistics
   2. Let $Q = \sup_{t \leq \tau} |W_k(t)|/\sigma_k$
General unequal alternative with censoring

approach 1: maximum of partial log rank statistic gives Rényi test.

1. Look for maximal size of one of the log rank statistics
2. Let $Q = \sup_{t \leq \tau} \left| W_k(t) \right| / \sigma_k$
   for $\tau$ = last event time
General unequal alternative with censoring

1 approach 1: maximum of partial log rank statistic gives Rényi test.

1. Look for maximal size of one of the log rank statistics
2. Let $Q = \sup_{t \leq \tau} \left| W_k(t) \right| / \sigma_k$
   1. for $\tau = \text{last event time}$
   2. for $W_k(t) = \text{statistic for times up to } t$
approach 1: maximum of partial log rank statistic gives Rényi test.  

1. Look for maximal size of one of the log rank statistics
2. Let $Q = \sup_{t \leq \tau} |W_k(t)|/\sigma_k$
   - for $\tau =$ last event time
   - for $W_k(t) =$ statistic for times up to $t$
3. $W_k(t) = \sum_{i \mid t_i \leq t} w_i [D_{ki} - (D_i/Y_i)Y_{ki}]$
General unequal alternative with censoring

1. **approach 1:** maximum of partial log rank statistic gives *Rényi test.*
   1. Look for maximal size of one of the log rank statistics
   2. Let $Q = \sup_{t \leq \tau} |W_k(t)|/\sigma_k$
      1. for $\tau =$ last event time
      2. for $W_k(t) =$ statistic for times up to $t$
      3. $W_k(t) = \sum_{i|t_i \leq t} w_i \left[D_{ki} - (D_i/Y_i)Y_{ki}\right]$
      4. for $\sigma_k = \sqrt{\text{variance with all contributions}}$
Significance levels involve point-wise asymptotic normality correction for multiple comparisons approximation to correlation between values at different time points very high if points are close together.

One-sided Significance Levels drop $|\cdot|$ in definition of $Q$: Let $Q = \sup_{t \leq \tau} \frac{W_k(t)}{\sigma_k}$

Consider paths for $W_k(t)$

After any time $s$, remainder is approximately symmetric

Every path crossing a point $q$ before $\tau$ and ending over $q$ at $\tau$ has a counterpart below $q$

Hence $P_0[\text{above } q \text{ sometime}] = 2P_0[\text{above } q \text{ at } \tau] = 2[1 - \Phi(q)]$

Two-sided Significance Levels

$P_0[|Q| > q] = P_0[\max(W_k(t)) > q] + P_0[\min(W_k(t)) < -q] - P_0[\max(W_k(t)) > q] and \min(W_k(t)) < -q]$

Formula and table given in table in book
Significance levels involve point-wise asymptotic normality correction for multiple comparisons approximation to correlation between values at different time points very high if points are close together.

One–sided Significance Levels

Drop $|·|$ in definition of $Q$: Let $Q = \sup_{t \leq \tau} W_k(t) / \sigma_k$. Consider paths for $W_k(t)$ after any time $s$, remainder is approximately symmetric every path crossing a point $q$ before $\tau$ and ending over $q$ at $\tau$ has a counterpart below $q$.

Hence $P_0[\text{above } q \text{ sometime}] = 2P_0[\text{above } q \text{ at } \tau] = 2[1 - \Phi(q)]$.

Two–sided Significance Levels

$< 2 \times 1$-sided value, because path can cross $\pm q$ both ways in same run.

$P_0[|Q| > q] = P_0[\max(W_k(t)) > q] + P_0[\min(W_k(t)) < -q] - P_0[\max(W_k(t)) > q]$ and $P_0[\min(W_k(t)) < -q]$.

Formula and table given in table in book.
Significance levels involve point-wise asymptotic normality.
Significance levels involve

1. point-wise asymptotic normality
2. correction for multiple comparisons

One-sided Significance Levels

Consider paths for $W_k(t)$.

After any time $s$, remainder is approximately symmetric.

Every path crossing a point $q$ before $\tau$ and ending over $q$ at $\tau$ has a counterpart below $q$.

Hence $P_0[\text{above } q \text{ sometime}] = 2P_0[\text{above } q \text{ at } \tau] = 2[1 - \Phi(q)]$

Two-sided Significance Levels

$P_0[|Q| > q] = P_0[\max(W_k(t)) > q] + P_0[\min(W_k(t)) < -q] - P_0[\max(W_k(t)) > q \text{ and } \min(W_k(t)) < -q]$
Continued

Significance levels involve

1. point-wise asymptotic normality
2. correction for multiple comparisons
3. approximation to correlation between values at different time points
Continued

Significance levels involve

- point-wise asymptotic normality
- correction for multiple comparisons
- approximation to correlation between values at different time points
- very high if points are close together.
Continued

1 Significance levels involve
   1 point-wise asymptotic normality
   2 correction for multiple comparisons
   3 approximation to correlation between values at different time points
   4 very high if points are close together.

2 One–sided Significance Levels

\[
\text{Consider paths for } W_k(t) \text{ after any time } s, \text{ remainder is approximately symmetric.}
\]

\[
\text{Every path crossing a point } q \text{ before } \tau \text{ and ending over } q \text{ at } \tau \text{ has a counterpart below } q.
\]

\[
\Pr_0[\text{above } q \text{ sometime}] = 2 \Pr_0[\text{above } q \text{ at } \tau] = 2[1 - \Phi(q)]
\]
Continued

Significance levels involve
1. point-wise asymptotic normality
2. correction for multiple comparisons
3. approximation to correlation between values at different time points
4. very high if points are close together.

One–sided Significance Levels
1. drop $|\cdot|$ in definition of $Q$: Let $Q = \sup_{t \leq \tau} W_k(t)/\sigma_k$

Testing with survival curves: Tests based on Curve Differences Lecture 05
Continued

1. Continued

   - Significance levels involve
     1. point-wise asymptotic normality
     2. correction for multiple comparisons
     3. approximation to correlation between values at different time points
     4. very high if points are close together.

2. One–sided Significance Levels
   1. drop $|·|$ in definition of $Q$: Let $Q = \sup_{t \leq \tau} \frac{W_k(t)}{\sigma_k}$
   2. Consider paths for $W_k(t)$
Continued

1. Significance levels involve
   1. point-wise asymptotic normality
   2. correction for multiple comparisons
   3. approximation to correlation between values at different time points
   4. very high if points are close together.

2. One–sided Significance Levels
   1. drop $|\cdot|$ in definition of $Q$: Let $Q = \sup_{t \leq \tau} W_k(t)/\sigma_k$
   2. Consider paths for $W_k(t)$
   3. After any time $s$, remainder is approximately symmetric
Continued

1. Significance levels involve
   1. point-wise asymptotic normality
   2. correction for multiple comparisons
   3. approximation to correlation between values at different time points
   4. very high if points are close together.

2. One-sided Significance Levels
   1. drop $|\cdot|$ in definition of $Q$: Let $Q = \sup_{t \leq \tau} \frac{W_k(t)}{\sigma_k}$
   2. Consider paths for $W_k(t)$
   3. After any time $s$, remainder is approximately symmetric
   4. Every path crossing a point $q$ before $\tau$ and ending over $q$ at $\tau$ has a counterpart below $q$
Continued

1. Significance levels involve
   1. point-wise asymptotic normality
   2. correction for multiple comparisons
   3. approximation to correlation between values at different time points
   4. very high if points are close together.

2. One-sided Significance Levels
   1. drop $|·|$ in definition of $Q$: Let $Q = \sup_{t \leq \tau} W_k(t)/\sigma_k$
   2. Consider paths for $W_k(t)$
   3. After any time $s$, remainder is approximately symmetric
   4. Every path crossing a point $q$ before $\tau$ and ending over $q$ at $\tau$ has a counterpart below $q$
   5. Hence $P_0[\text{above } q \text{ sometime}] = 2P_0[\text{above } q \text{ at } \tau] = 2[1 - \Phi(q)]$


Continued

1 Significance levels involve
   1 point-wise asymptotic normality
   2 correction for multiple comparisons
   3 approximation to correlation between values at different time points
   4 very high if points are close together.

2 One-sided Significance Levels
   1 drop $\mid \cdot \mid$ in definition of $Q$: Let $Q = \sup_{t \leq \tau} W_k(t)/\sigma_k$
   2 Consider paths for $W_k(t)$
   3 After any time $s$, remainder is approximately symmetric
   4 Every path crossing a point $q$ before $\tau$ and ending over $q$ at $\tau$ has a counterpart below $q$
   5 Hence $P_0 [\text{above } q \text{ sometime}] = 2P_0 [\text{above } q \text{ at } \tau] = 2[1 - \Phi(q)]$

3 Two-sided Significance Levels
Continued

1 Significance levels involve
   1 point-wise asymptotic normality
   2 correction for multiple comparisons
   3 approximation to correlation between values at different time points
   4 very high if points are close together.

2 One–sided Significance Levels
   1 drop $|\cdot|$ in definition of $Q$: Let $Q = \sup_{t \leq \tau} \frac{W_k(t)}{\sigma_k}$
   2 Consider paths for $W_k(t)$
   3 After any time $s$, remainder is approximately symmetric
   4 Every path crossing a point $q$ before $\tau$ and ending over $q$ at $\tau$ has a counterpart below $q$
   5 Hence $P_0 [\text{above } q \text{ sometime}] = 2P_0 [\text{above } q \text{ at } \tau] = 2[1 - \Phi(q)]$

3 Two–sided Significance Levels
   1 $< 2 \times 1$-sided value, because path can cross $\pm q$ both ways in same run
Significance levels involve
- point-wise asymptotic normality
- correction for multiple comparisons
- approximation to correlation between values at different time points
  very high if points are close together.

One–sided Significance Levels
- drop $|\cdot|$ in definition of $Q$: Let $Q = \sup_{t \leq \tau} W_k(t)/\sigma_k$
- Consider paths for $W_k(t)$
- After any time $s$, remainder is approximately symmetric
- Every path crossing a point $q$ before $\tau$ and ending over $q$ at $\tau$ has a counterpart below $q$
- Hence $P_0[\text{above } q \text{ sometime}] = 2P_0[\text{above } q \text{ at } \tau] = 2[1 - \Phi(q)]$

Two–sided Significance Levels
- $< 2 \times 1$-sided value, because path can cross $\pm q$ both ways in same run
- $P_0[|Q| > q] = P_0[\max(W_k(t)) > q] + P_0[\min(W_k(t)) < -q] - P_0[\max(W_k(t)) > q \text{ and } \min(W_k(t)) < -q]$
Significance levels involve
- point-wise asymptotic normality
- correction for multiple comparisons
- approximation to correlation between values at different time points
  very high if points are close together.

One-sided Significance Levels
- drop $|\cdot|$ in definition of $Q$: Let $Q = \sup_{t \leq \tau} W_k(t)/\sigma_k$
- Consider paths for $W_k(t)$
- After any time $s$, remainder is approximately symmetric
- Every path crossing a point $q$ before $\tau$ and ending over $q$ at $\tau$ has a counterpart below $q$
- Hence $P_0[\text{above } q \text{ sometime}] = 2P_0[\text{above } q \text{ at } \tau] = 2[1 - \Phi(q)]$

Two-sided Significance Levels
- $< 2 \times$ 1-sided value, because path can cross $\pm q$ both ways in same run
- $P_0[|Q| > q] = P_0[\max(W_k(t)) > q] + P_0[\min(W_k(t)) < -q] - P_0[\max(W_k(t)) > q \text{ and } \min(W_k(t)) < -q]$
- Formula and table given in table in book
Approach 2: Look at integrated difference.

\[ \sum_i \left[ \hat{H}_1(t_i) - \hat{H}_2(t_i) \right]^2 \frac{\sigma^4(t_i)}{\hat{\sigma}^4}\]
Approach 2: Look at integrated difference.

- Cramer–von Mises statistic has issues with censoring
Approach 2: Look at integrated difference.

1. Cramer–von Mises statistic has issues with censoring
2. Usually look at differences between cumulative hazard functions instead of survival function
Continued

1. **Approach 2: Look at integrated difference.**
   1. *Cramer–von Mises statistic* has issues with censoring
   2. Usually look at differences between cumulative hazard functions instead of survival function
   3. Use formula $\sum_i [\hat{H}_1(t_i) - \hat{H}_2(t_i)]^2 d\sigma^2(t)/\sigma^4(\tau)$

Testing with survival curves: Tests based on Curve Differences
Approach 2: Look at integrated difference.

1. *Cramer–von Mises statistic* has issues with censoring
2. Usually look at differences between cumulative hazard functions instead of survival function
3. Use formula $\sum_i [\hat{H}_1(t_i) - \hat{H}_2(t_i)]^2 d\sigma^2(t)/\sigma^4(\tau)$
4. Distribution is tabulated in book.
Two-sample testing review

1. Each individual $i$ has response and group indicator.

- Group indicator $G_i \in \{1, 2\}$.
- Censoring time $C_i$ and event time $X_i$, random and unobservable.
- $T_i = \min(C_i, X_i)$ and $\delta_i = \begin{cases} 1 & \text{if } C_i \geq X_i \\ 0 & \text{if } C_i < X_i \end{cases}$, random and observable.

2. Null hypothesis $H_0$: Distribution of $X_i$ the same regardless of $G_i$.

3. Censoring time conditions:
   - Must be independent of event times.
   - Might not be homogenous across groups.


5. For many statistics (e.g., log rank), the above framework ensures that the null distribution is approximately
   - Normal
   - Expectation zero
   - Variance estimable.

6. Distribution is based on the distribution of $C_i$ and $X_i$. 
Each individual $i$ has response and group indicator.

- group indicator $G_i \in \{1, 2\}$. 

Censoring time $C_i$ and event time $X_i$, random and unobservable.

- $T_i = \min(C_i, X_i)$ and $\delta_i = \begin{cases} 1 & \text{if } C_i \geq X_i \\ 0 & \text{if } C_i < X_i \end{cases}$, random and observable.

Null hypothesis $H_0$: Distribution of $X_i$ the same regardless of $G_i$.

Censoring time conditions:
- must be independent of event times.
- Might not be homogeneous across groups.

Construct Test statistic $W$. For many statistics (ex. log rank) the above framework ensures that the null distribution is approximately normal, expectation zero, variance estimable. Distribution is based on the distribution of $C_i$ and $X_i$. 

Testing with survival curves: Tests based on Curve Differences
Two-sample testing review

1. Each individual $i$ has response and group indicator.
   - group indicator $G_i \in \{1, 2\}$.
   - censoring time $C_i$ and event time $X_i$, random and unobservable.

Null hypothesis $H_0$: Distribution of $X_i$ the same regardless of $G_i$.

Censoring time conditions:
1. Must be independent of event times.
2. Might not be homogenous across groups.

Construct Test statistic $W$.
For many statistics (ex. log rank) the above framework ensures that the null distribution is approximately normal with expectation zero and variance estimable.

Distribution is based on the distribution of $C_i$ and $X_i$. 
Two-sample testing review

1. Each individual $i$ has response and group indicator.
   - group indicator $G_i \in \{1, 2\}$.
   - censoring time $C_i$ and event time $X_i$, random and unobservable.
   - $T_i = \min(C_i, X_i)$ and $\delta_i = \begin{cases} 
1 & \text{if } C_i \geq X_i \\
0 & \text{if } C_i < X_i
\end{cases}$, random and observable.

Null hypothesis $H_0$: Distribution of $X_i$ the same regardless of $G_i$.

Censoring time conditions:
1. must be independent of event times.
2. Might not be homogenous across groups.

Construct Test statistic $W$.
For many statistics (ex. log rank) the above framework ensures that the null distribution is approximately
1. normal
2. expectation zero
3. variance estimable.

Distribution is based on the distribution of $C_i$ and $X_i$. 
Two-sample testing review

1. Each individual $i$ has response and group indicator.
   
   1. group indicator $G_i \in \{1, 2\}$.
   2. censoring time $C_i$ and event time $X_i$, random and unobservable.

2. $T_i = \min(C_i, X_i)$ and $\delta_i = \begin{cases} 
1 & \text{if } C_i \geq X_i \\
0 & \text{if } C_i < X_i 
\end{cases}$, random and observable.

3. Null hypothesis $H_0$: Distribution of $X_i$ the same regardless of $G_i$. 

Testing with survival curves: Tests based on Curve Differences
Two-sample testing review

1. Each individual $i$ has response and group indicator.
   - group indicator $G_i \in \{1, 2\}$.
   - censoring time $C_i$ and event time $X_i$, random and unobservable.

2. $T_i = \min(C_i, X_i)$ and $\delta_i = \begin{cases} 1 & \text{if } C_i \geq X_i \\ 0 & \text{if } C_i < X_i \end{cases}$, random and observable.

3. Null hypothesis $H_0$: Distribution of $X_i$ the same regardless of $G_i$.

Censoring time conditions:

- Must be independent of event times.
- Might not be homogenous across groups.

Construct Test statistic $W$ for many statistics (ex. log rank) the above framework ensures that the null distribution is approximately normal with expectation zero and variance estimable.

Distribution is based on the distribution of $C_i$ and $X_i$.

Testing with survival curves: Tests based on Curve Differences
Two-sample testing review

1. Each individual $i$ has response and group indicator. 
   - group indicator $G_i \in \{1, 2\}$.
   - censoring time $C_i$ and event time $X_i$, random and unobservable.

2. $T_i = \min(C_i, X_i)$ and $\delta_i = \begin{cases} 1 & \text{if } C_i \geq X_i \\ 0 & \text{if } C_i < X_i \end{cases}$, random and observable.

3. Null hypothesis $H_0$: Distribution of $X_i$ the same regardless of $G_i$.

4. Censoring time conditions:
   - must be independent of event times.
Two-sample testing review

1. Each individual \( i \) has response and group indicator.
   1. group indicator \( G_i \in \{1, 2\} \).
   2. censoring time \( C_i \) and event time \( X_i \), random and unobservable.

   \[ T_i = \min(C_i, X_i) \text{ and } \delta_i = \begin{cases} 1 & \text{if } C_i \geq X_i \\ 0 & \text{if } C_i < X_i \end{cases}, \text{ random and observable.} \]

2. Null hypothesis \( H_0 \): Distribution of \( X_i \) the same regardless of \( G_i \).

3. Censoring time conditions:
   1. must be independent of event times.
   2. Might not be homogenous across groups.
Two-sample testing review

1. Each individual $i$ has response and group indicator.
   1. group indicator $G_i \in \{1, 2\}$.
   2. censoring time $C_i$ and event time $X_i$, random and unobservable.

   $$T_i = \min(C_i, X_i) \quad \text{and} \quad \delta_i = \begin{cases} 1 & \text{if } C_i \geq X_i \\ 0 & \text{if } C_i < X_i \end{cases},$$

   random and observable.

2. Null hypothesis $H_0$: Distribution of $X_i$ the same regardless of $G_i$.

3. Censoring time conditions:
   1. must be independent of event times.
   2. Might not be homogenous across groups.

4. Construct Test statistic $W$
Each individual $i$ has response and group indicator. 

1. group indicator $G_i \in \{1, 2\}$.
2. censoring time $C_i$ and event time $X_i$, random and unobservable.
3. $T_i = \min(C_i, X_i)$ and $\delta_i = \begin{cases} 1 & \text{if } C_i \geq X_i \\ 0 & \text{if } C_i < X_i \end{cases}$, random and observable.

Null hypothesis $H_0$: Distribution of $X_i$ the same regardless of $G_i$.

Censoring time conditions:

1. must be independent of event times.
2. Might not be homogenous across groups.

Construct Test statistic $W$

For many statistics (ex. log rank) the above framework ensures that the null distribution is approximately

Testing with survival curves: Tests based on Curve Differences
Two-sample testing review

1. Each individual $i$ has response and group indicator.
   - group indicator $G_i \in \{1, 2\}$.
   - censoring time $C_i$ and event time $X_i$, random and unobservable.
   - $T_i = \min(C_i, X_i)$ and $\delta_i = \begin{cases} 1 & \text{if } C_i \geq X_i \\ 0 & \text{if } C_i < X_i \end{cases}$, random and observable.

2. Null hypothesis $H_0$: Distribution of $X_i$ the same regardless of $G_i$.

3. Censoring time conditions:
   - must be independent of event times.
   - Might not be homogenous across groups.

4. Construct Test statistic $W$

5. For many statistics (ex. log rank) the above framework ensures that the null distribution is approximately
   - normal
Two-sample testing review

1. Each individual $i$ has response and group indicator.
   - group indicator $G_i \in \{1, 2\}$.
   - censoring time $C_i$ and event time $X_i$, random and unobservable.
   - $T_i = \min(C_i, X_i)$ and $\delta_i = \begin{cases} 1 & \text{if } C_i \geq X_i \\ 0 & \text{if } C_i < X_i \end{cases}$, random and observable.

2. Null hypothesis $H_0$: Distribution of $X_i$ the same regardless of $G_i$.

3. Censoring time conditions:
   - must be independent of event times.
   - Might not be homogenous across groups.

4. Construct Test statistic $W$

5. For many statistics (ex. log rank) the above framework ensures that the null distribution is approximately
   - normal
   - expectation zero

Testing with survival curves: Tests based on Curve Differences
Two-sample testing review

1. Each individual $i$ has response and group indicator.
   - group indicator $G_i \in \{1, 2\}$.
   - censoring time $C_i$ and event time $X_i$, random and unobservable.
   - $T_i = \min(C_i, X_i)$ and $\delta_i = \begin{cases} 1 & \text{if } C_i \geq X_i \\ 0 & \text{if } C_i < X_i \end{cases}$, random and observable.

2. Null hypothesis $H_0$: Distribution of $X_i$ the same regardless of $G_i$.

3. Censoring time conditions:
   - must be independent of event times.
   - Might not be homogenous across groups.

4. Construct Test statistic $W$

5. For many statistics (ex. log rank) the above framework ensures that the null distribution is approximately
   - normal
   - expectation zero
   - variance estimable.
Two-sample testing review

1. Each individual $i$ has response and group indicator.
   1. group indicator $G_i \in \{1, 2\}$.
   2. censoring time $C_i$ and event time $X_i$, random and unobservable.

2. $T_i = \min(C_i, X_i)$ and $\delta_i = \begin{cases} 1 & \text{if } C_i \geq X_i \\ 0 & \text{if } C_i < X_i \end{cases}$, random and observable.

3. Null hypothesis $H_0$: Distribution of $X_i$ the same regardless of $G_i$.

4. Censoring time conditions:
   1. must be independent of event times.
   2. Might not be homogenous across groups.

5. Construct Test statistic $W$

6. For many statistics (ex. log rank) the above framework ensures that the null distribution is approximately
   1. normal
   2. expectation zero
   3. variance estimable.

Distribution is based on the distribution of $C_i$ and $X_i$. 

Testing with survival curves: Tests based on Curve Differences
Permutation Distribution

1. Treat $C_i, X_i$ as fixed.

2. Null hypothesis implies that group tells you nothing about $X_i$.

3. Hence Null hypothesis implies that $X_i$ tells you nothing about group.

4. So every rearrangement of group indicators is equally likely under $H_0$.

5. Suppose that there are $M_1, M_2$ items in groups 1, 2 respectively.

6. Under $H_0$, every one of the $(M_1 + M_2) M_1$ is equally likely.

7. If this is too many to do, you can sample these randomly.

8. If censoring is related to group, then this fails, because then $(X_i, C_i)$ give information about group, even though $X_i$ by itself does not.

R Code SAS Code

Testing with survival curves: Tests based on Curve Differences Lecture 05
1. Treat $C_i, X_i$ as fixed.
2. Null hypothesis implies that group tells you nothing about $X_i$.
Permutation Distribution

1. Treat $C_i, X_i$ as fixed.
2. Null hypothesis implies that group tells you nothing about $X_i$.
3. Hence Null hypothesis implies that $X_i$ tells you nothing about group.
1. Treat $C_i, X_i$ as fixed.
2. Null hypothesis implies that group tells you nothing about $X_i$.
3. Hence Null hypothesis implies that $X_i$ tells you nothing about group.
4. So every rearrangement of group indicators is equally likely under $H_0$. 

Suppose that there are $M_1, M_2$ items in groups 1, 2 respectively. Under $H_0$, every one of the $(M_1 + M_2 M_1)$ is equally likely.

If this is too many to do, you can sample these randomly.

If censoring is related to group, then this fails, because then $(X_i, C_i)$ give information about group, even though $X_i$ by itself does not.
Permutation Distribution

1. Treat $C_i, X_i$ as fixed.
2. Null hypothesis implies that group tells you nothing about $X_i$.
3. Hence Null hypothesis implies that $X_i$ tells you nothing about group.
4. So every rearrangement of group indicators is equally likely under $H_0$.
5. Suppose that there are $M_1, M_2$ items in groups 1, 2 resp.
Permutation Distribution

1. Treat $C_i, X_i$ as fixed.
2. Null hypothesis implies that group tells you nothing about $X_i$.
3. Hence Null hypothesis implies that $X_i$ tells you nothing about group.
4. So every rearrangement of group indicators is equally likely under $H_0$.
5. Suppose that there are $M_1, M_2$ items in groups 1, 2 resp.
6. Under $H_0$, every one of the $\binom{M_1+M_2}{M_1}$ is equally likely.
Permutation Distribution

1. Treat $C_i$, $X_i$ as fixed.
2. Null hypothesis implies that group tells you nothing about $X_i$.
3. Hence Null hypothesis implies that $X_i$ tells you nothing about group.
4. So every rearrangement of group indicators is equally likely under $H_0$.
5. Suppose that there are $M_1$, $M_2$ items in groups 1, 2 resp.
6. Under $H_0$, every one of the $\binom{M_1+M_2}{M_1}$ is equally likely.
7. If this is too many to do, you can sample these randomly.
**Permutation Distribution**

1. Treat $C_i, X_i$ as fixed.
2. Null hypothesis implies that group tells you nothing about $X_i$.
3. Hence Null hypothesis implies that $X_i$ tells you nothing about group.
4. So every rearrangement of group indicators is equally likely under $H_0$.
5. Suppose that there are $M_1, M_2$ items in groups 1, 2 resp.
6. Under $H_0$, every one of the $\binom{M_1+M_2}{M_1}$ is equally likely.
7. If this is too many to do, you can sample these randomly.
8. If censoring is related to group, then this fails,
Permutation Distribution

1. Treat $C_i, X_i$ as fixed.
2. Null hypothesis implies that group tells you nothing about $X_i$.
3. Hence Null hypothesis implies that $X_i$ tells you nothing about group.
4. So every rearrangement of group indicators is equally likely under $H_0$.
5. Suppose that there are $M_1, M_2$ items in groups 1, 2 resp.
6. Under $H_0$, every one of the $\binom{M_1 + M_2}{M_1}$ is equally likely.
7. If this is too many to do, you can sample these randomly.
8. If censoring is related to group, then this fails,
   because then $(X_i, C_i)$ give information about group,
Treat $C_i, X_i$ as fixed.

2 Null hypothesis implies that group tells you nothing about $X_i$.

3 Hence Null hypothesis implies that $X_i$ tells you nothing about group.

4 So every rearrangement of group indicators is equally likely under $H_0$.

5 Suppose that there are $M_1, M_2$ items in groups 1, 2 resp.

6 Under $H_0$, every one of the $\binom{M_1+M_2}{M_1}$ is equally likely.

7 If this is too many to do, you can sample these randomly.

8 If censoring is related to group, then this fails,
   1 because then $(X_i, C_i)$ give information about group,
   2 even though $X_i$ by itself does not.  R Code  SAS Code
Sample size setup for log rank statistic

1. Can’t do them exactly
Sample size setup for log rank statistic

1. Can’t do them exactly
2. Will do them for large samples.
Sample size setup for log rank statistic

1. Can’t do them exactly
2. Will do them for large samples.
3. One-sided.
User picks

1. test level $\alpha$
User picks

1. Test level $\alpha$
2. Desired power $1 - \beta$
User picks

1. test level $\alpha$
2. Desired power $1 - \beta$
3. Expected hazard ratio $\theta$
User picks

1. Test level $\alpha$
2. Desired power $1 - \beta$
3. Expected hazard ratio $\theta$
4. Ratio of sample sizes $\phi$
Sample sizes for general Gaussian test statistics.

1. Use fact test statistic is approximately normal under either hypothesis.

\[ c = E_0[W] + z_{\alpha} \sqrt{\text{Var}_0[W]} \]

2. Approximate power is \( \bar{\Phi} \left( \frac{c - E_\theta[W]}{\sqrt{\text{Var}_\theta[W]}} \right) = 1 - \beta \).

3. Approximating alternative variance by null variance, \( \bar{\Phi} \left( \frac{E_0[W] - E_\theta[W]}{\sqrt{\text{Var}_0[W]} + z_{\alpha}} \right) = 1 - \beta \).

4. Applying normal quantile function to both sides, \( \frac{E_0[W] - E_\theta[W]}{\sqrt{\text{Var}_0[W]} + z_{\alpha}} = z_{1 - \beta} \).

5. Isolate quantiles \( \frac{E_\theta[W] - E_0[W]}{\sqrt{\text{Var}_0[W]} + z_{\alpha}} = z_{\beta} + z_{\alpha} \).
Sample sizes for general Gaussian test statistics.

1. Use fact test statistic is approximately normal under either hypothesis.
   1. Critical value $c = E_0[W] + z_\alpha \sqrt{\text{Var}_0[W]}$. 

Testing with survival curves: Sample size calculations for log rank statistic
Sample sizes for general Gaussian test statistics.

1. Use fact test statistic is approximately normal under either hypothesis.
   1. Critical value \( c = E_0 [W] + z_\alpha \sqrt{Var_0 [W]} \).
   2. Approximate power is \( \Phi((c - E_\theta [W]) / \sqrt{Var_\theta [W]}) = 1 - \beta \).
Sample sizes for general Gaussian test statistics.

1. Use fact test statistic is approximately normal under either hypothesis.

   1. Critical value \( c = E_0 [W] + z_\alpha \sqrt{\text{Var}_0 [W]} \).
   2. Approximate power is \( \Phi((c - E_\theta [W])/\sqrt{\text{Var}_\theta [W]}) = 1 - \beta \).
   3. Approximating alternative variance by null variance, \( \Phi((E_0 [W] - E_\theta [W])/\sqrt{\text{Var}_0 [W]} + z_\alpha) = 1 - \beta \).
Sample sizes for general Gaussian test statistics.

1. Use fact test statistic is approximately normal under either hypothesis.
   
   1. Critical value \( c = E_0 [W] + z_\alpha \sqrt{\text{Var}_0 [W]} \).
   2. Approximate power is \( \Phi((c - E_\theta [W]) / \sqrt{\text{Var}_\theta [W]}) = 1 - \beta \).
   3. Approximating alternative variance by null variance, \( \Phi((E_0 [W] - E_\theta [W]) / \sqrt{\text{Var}_0 [W]} + z_\alpha ) = 1 - \beta \).
   4. Applying normal quantile function to both sides, \( (E_0 [W] - E_\theta [W]) / \sqrt{\text{Var}_0 [W]} + z_\alpha ) = z_{1-\beta} \).
Sample sizes for general Gaussian test statistics.

1. Use fact test statistic is approximately normal under either hypothesis.

   1. Critical value \( c = E_0[W] + z_\alpha \sqrt{\text{Var}_0[W]} \).
   2. Approximate power is \( \Phi((c - E_\theta[W]) / \sqrt{\text{Var}_\theta[W]}) = 1 - \beta \).
   3. Approximating alternative variance by null variance, \( \Phi((E_0[W] - E_\theta[W]) / \sqrt{\text{Var}_0[W]} + z_\alpha) = 1 - \beta \).
   4. Applying normal quantile function to both sides, \( (E_0[W] - E_\theta[W]) / \sqrt{\text{Var}_0[W]} + z_\alpha) = z_{1-\beta} \).
   5. Isolate quantiles \( (E_\theta[W] - E_0[W]) / \sqrt{\text{Var}_0[W]} = z_\beta + z_\alpha \).
Specific to log rank test.

1. \( D_{ki} | Y_{ki} \sim B(Y_{ki}, \pi_{ki}), \pi_{ki} \text{ small.} \)
Specific to log rank test.

1. $D_{ki} | Y_{ki} \sim B(Y_{ki}, \pi_{ki})$, $\pi_{ki}$ small.
2. $D_{ki} | Y_{ki} \sim \approx P(\lambda_{ki})$, $\lambda_{ki} = \pi_{ki} / Y_{ki}$
Specific to log rank test.

1. $D_{ki} \mid Y_{ki} \sim B(Y_{ki}, \pi_{ki})$, $\pi_{ki}$ small.

2. $D_{ki} \mid Y_{ki} \sim \approx P(\lambda_{ki})$, $\lambda_{ki} = \pi_{ki} / Y_{ki}$

3. Suppose $Y_{1i} \approx \phi Y_{2i}$ for constant $\phi$

Testing with survival curves: Sample size calculations for log rank statistic
Specific to log rank test.

1. $D_{ki} \mid Y_{ki} \sim \mathcal{B}(Y_{ki}, \pi_{ki}), \pi_{ki}$ small.
2. $D_{ki} \mid Y_{ki} \sim \approx \mathcal{P}(\lambda_{ki}), \lambda_{ki} = \pi_{ki} / Y_{ki}$
   1. Suppose $Y_{1i} \approx \phi Y_{2i}$ for constant $\phi$
   2. Then $Y_{1i} / Y_{i} \approx \phi / (1 + \phi), Y_{2i} / Y_{i} \approx 1 / (1 + \phi)$,
Specific to log rank test.

1. $D_{ki} | Y_{ki} \sim \mathcal{B}(Y_{ki}, \pi_{ki})$, $\pi_{ki}$ small.
2. $D_{ki} | Y_{ki} \sim \approx \mathcal{P}(\lambda_{ki})$, $\lambda_{ki} = \pi_{ki} / Y_{ki}$
   
   1. Suppose $Y_{1i} \approx \phi Y_{2i}$ for constant $\phi$
   2. Then $Y_{1i} / Y_i \approx \phi / (1 + \phi)$, $Y_{2i} / Y_i \approx 1 / (1 + \phi)$,

3. Alternative hypothesis: $\lambda_{1i} = \theta \lambda_{2i}$ for $\theta \neq 1$. 

Testing with survival curves: Sample size calculations for log rank statistic Lecture 05
Specific to log rank test.

1. \( D_{ki} | Y_{ki} \sim B(Y_{ki}, \pi_{ki}), \pi_{ki} \) small.
2. \( D_{ki} | Y_{ki} \sim \approx P(\lambda_{ki}), \lambda_{ki} = \pi_{ki} / Y_{ki} \)
   - Suppose \( Y_{1i} \approx \phi Y_{2i} \) for constant \( \phi \)
   - Then \( Y_{1i} / Y_i \approx \phi / (1 + \phi), \ Y_{2i} / Y_i \approx 1 / (1 + \phi) \),
3. Alternative hypothesis: \( \lambda_{1i} = \theta \lambda_{2i} \) for \( \theta \neq 1 \).
4. Conditioning on \( D_i \), distribution of Poissons on sum is multinomial
Specific to log rank test.

1. \( D_{ki} \mid Y_{ki} \sim B(Y_{ki}, \pi_{ki}), \pi_{ki} \) small.
2. \( D_{ki} \mid Y_{ki} \sim \approx P(\lambda_{ki}), \lambda_{ki} = \pi_{ki} / Y_{ki} \)
   - Suppose \( Y_{1i} \approx \phi Y_{2i} \) for constant \( \phi \)
   - Then \( Y_{1i} / Y_i \approx \phi / (1 + \phi), \ Y_{2i} / Y_i \approx 1 / (1 + \phi) \),
3. Alternative hypothesis: \( \lambda_{1i} = \theta \lambda_{2i} \) for \( \theta \neq 1 \).
4. Conditioning on \( D_i \), distribution of Poissons on sum is multinomial
   - If \( K = 2 \) this is binomial.
Specific to log rank test.

1. \(D_{ki} | Y_{ki} \sim B(Y_{ki}, \pi_{ki}), \pi_{ki} \text{ small.}\)

2. \(D_{ki} | Y_{ki} \sim \approx P(\lambda_{ki}), \lambda_{ki} = \pi_{ki} / Y_{ki}\)
   - Suppose \(Y_{1i} \approx \phi Y_{2i}\) for constant \(\phi\)
   - Then \(Y_{1i} / Y_i \approx \phi / (1 + \phi), Y_{2i} / Y_i \approx 1 / (1 + \phi),\)

3. Alternative hypothesis: \(\lambda_{1i} = \theta \lambda_{2i}\) for \(\theta \neq 1.\)

4. Conditioning on \(D_i\), distribution of Poissons on sum is multinomial
   - If \(K = 2\) this is binomial.
   - Probabilities are \(\phi \theta / (1 + \phi \theta)\) and \(1 / (1 + \phi \theta)\) for groups 1 and 2 resp.
Specific to log rank test.

1. \( D_{ki} \mid Y_{ki} \sim B(Y_{ki}, \pi_{ki}), \pi_{ki} \text{ small.} \)
2. \( D_{ki} \mid Y_{ki} \sim \approx P(\lambda_{ki}), \lambda_{ki} = \pi_{ki} / Y_{ki} \)
   1. Suppose \( Y_{1i} \approx \phi Y_{2i} \) for constant \( \phi \)
   2. Then \( Y_{1i} / Y_{i} \approx \phi / (1 + \phi), Y_{2i} / Y_{i} \approx 1 / (1 + \phi), \)
3. Alternative hypothesis: \( \lambda_{1i} = \theta \lambda_{2i} \) for \( \theta \neq 1. \)
4. Conditioning on \( D_{i} \), distribution of Poissons on sum is multinomial
   1. If \( K = 2 \) this is binomial.
   2. Probabilities are \( \phi \theta / (1 + \phi \theta) \) and \( 1 / (1 + \phi \theta) \) for groups 1 and 2 resp.
   3. \( E_{\theta} [D_{1i} - Y_{1i}D_{i} / Y_{i}] = D_{i}(\phi \theta / (1 + \phi \theta) - \phi / (1 + \phi) = D_{i}(\theta - 1)\phi / ((1 + \phi)(\theta + 1)) \).
Specific to log rank test.

1. \( D_{ki} | Y_{ki} \sim B(Y_{ki}, \pi_{ki}), \pi_{ki} \text{ small.} \)

2. \( D_{ki} | Y_{ki} \sim P(\lambda_{ki}), \lambda_{ki} = \pi_{ki} / Y_{ki} \)
   - Suppose \( Y_{1i} \approx \phi Y_{2i} \) for constant \( \phi \)
   - Then \( Y_{1i} / Y_i \approx \phi / (1 + \phi), \ Y_{2i} / Y_i \approx 1 / (1 + \phi), \)

3. Alternative hypothesis: \( \lambda_{1i} = \theta \lambda_{2i} \) for \( \theta \neq 1. \)

4. Conditioning on \( D_i \), distribution of Poissons on sum is multinomial
   - If \( K = 2 \) this is binomial.
   - Probabilities are \( \phi \theta / (1 + \phi \theta) \) and \( 1 / (1 + \phi \theta) \) for groups 1 and 2 resp.
   - \( E_{\theta} [D_{1i} - Y_{1i}D_i / Y_i] = D_i(\phi \theta / (1 + \phi \theta) - \phi / (1 + \phi) = D_i(\theta - 1)\phi / ((1 + \phi)(\theta + 1)). \)
   - Estimate \( \Var_0 [D_i] \) as \( \sum_{i=1}^{D} Y_{ki}(Y_i - Y_{ki})(Y_i - D_i)D_i / (Y_i^2(Y_i - 1)) \approx \phi D_i / (1 + \phi)^2. \)
Specific to log rank test.

1. $D_{ki} \mid Y_{ki} \sim \mathcal{B}(Y_{ki}, \pi_{ki})$, $\pi_{ki}$ small.
2. $D_{ki} \mid Y_{ki} \sim \approx \mathcal{P}(\lambda_{ki})$, $\lambda_{ki} = \pi_{ki} / Y_{ki}$
   1. Suppose $Y_{1i} \approx \phi Y_{2i}$ for constant $\phi$.
   2. Then $Y_{1i} / Y_i \approx \phi / (1 + \phi)$, $Y_{2i} / Y_i \approx 1 / (1 + \phi)$.
3. Alternative hypothesis: $\lambda_{1i} = \theta \lambda_{2i}$ for $\theta \neq 1$.
4. Conditioning on $D_i$, distribution of Poissons on sum is mutlinomial
   1. If $K = 2$ this is binomial.
   2. Probabilities are $\phi \theta / (1 + \phi \theta)$ and $1 / (1 + \phi \theta)$ for groups 1 and 2 resp.
   3. $E_\theta [D_{1i} - Y_{1i} D_i / Y_i] = D_i (\phi \theta / (1 + \phi \theta) - \phi / (1 + \phi) = D_i (\theta - 1) \phi / ((1 + \phi)(\theta + 1))$.
   4. Estimate $\text{Var}_0 [D_i]$ as $\sum_{i=1}^{D} Y_{ki} (Y_i - Y_{ki})(Y_i - D_i) D_i / (Y_i^2 (Y_i - 1)) \approx \phi D_i / (1 + \phi)^2$.
5. Here $E_\theta [W_k] / \sqrt{\text{Var}_\theta [W_k]} \approx \sqrt{\sum_i D_i \frac{\sqrt{\phi (\theta - 1)}}{(1 + \phi \theta)}}$. 
These results to specify number of events

\[ \sum_{i=1}^{D} D_i = (z_{1-\alpha^*} + z_\beta)^2 \left( \frac{1+\theta}{1-\theta} \right)^2 \]

1. Inflate by \(1/[4p(1 - p)]\) if expect fraction \(p\) of events in one of the groups.
Use knowledge of survival probabilities to work backwards to sample size.

R Code  SAS Code
Section: Regression Models

Subsection: General background
Observe:

1. Times $T$

Censoring indicators $z_i$ for obsn $i$

Categorical variables like treatment or demographic indicators

Quantitative variables like weight, age.

May depend on time
Observe:

1. Times $T$
2. Censoring indicators

Categorical variables like treatment or demographic indicators
Quantitative variables like weight, age.
May depend on time
Observe:

1. Times $T$
2. Censoring indicators
3. Explanatory variables $z_i = z_{i1}, \ldots, z_{ip}$ for obsn $i$
Observe:

1. Times $T$
2. Censoring indicators
3. Explanatory variables $z_i = z_{i1}, \ldots, z_{ip}$ for obsn $i$
   - Categorical variables like treatment or demographic indicators
Observe:

1. Times $T$
2. Censoring indicators
3. Explanatory variables $z_i = z_{i1}, \ldots, z_{ip}$ for obsn $i$
   - Categorical variables like treatment or demographic indicators
   - Quantitative variables like weight, age.
Observe:

1. Times $T$
2. Censoring indicators
3. Explanatory variables $z_i = z_{i1}, \ldots, z_{ip}$ for obsn $i$
   1. Categorical variables like treatment or demographic indicators
   2. Quantitative variables like weight, age.
   3. May depend on time
Tasks

1. describe dependence between time and explanatory variables
Tasks

1. describe dependence between time and explanatory variables
   1. for categorical variables we know how to do this with log rank test
Tasks

1. describe dependence between time and explanatory variables
   - for categorical variables we know how to do this with log rank test
2. Control for variable
Tasks

1. describe dependence between time and explanatory variables
   - for categorical variables we know how to do this with log rank test
2. Control for variable
   - Have seen this with stratification
Section: Regression Models

Subsection: Proportional Hazards approach
Model

\[ h_i(t) = h_0(t)c(z_i\beta) \text{ for } c : \mathbb{R} \rightarrow [0, \infty) \]
Model

1. \( h_i(t) = h_0(t)c(z_i\beta) \) for \( c : \mathbb{R} \to [0, \infty) \)

Then hazards are not only \( \propto h_0 \) but to each other as well.
Model

1. \( h_i(t) = h_0(t)c(z_i \beta) \) for \( c : \mathbb{R} \to [0, \infty) \)

2. Then hazards are not only \( \propto h_0 \) but to each other as well.

3. \( S_i(t) = S_0(t)^{c(z_i \beta)} \)

4. Because \( S_i(t) = \exp\left(-\int_0^t h_i(s) \, ds\right) = \exp\left(-\int_0^t h_0(s) c(z_i \beta) \, ds\right) \)

5. Also, \( \log(-\log(S_i(t))) = \log(-\log(S_0(t))) + \log(c(z_i \beta)) \)

6. So \( \log \log \) survival should be shifts of one another

7. We've seen this before for testing:

8. For testing equality of \( m \) groups, \( z \) of length \( m - 1 \), component \( l \) of \( z \) is 1 if item is from population \( l \) and 0 otherwise

9. Model with \( c = \exp \) is called Cox model.

10. Makes \( \log(h_i(t)/h_0(t)) = z_i \beta \): regression model for log hazards

11. Also, \( h_i(t)/h_j(t) = \exp((z_i - z_j) \beta) \)

12. \( \exp((z_i - z_j) \beta) = \text{risk for person with covariates } z_i \text{ relative to person with } z_j \)
1. $h_i(t) = h_0(t) c(z_i \beta)$ for $c : \mathbb{R} \to [0, \infty)$

   Then hazards are not only $\propto h_0$ but to each other as well.

2. $S_i(t) = S_0(t)^{c(z_i \beta)}$

   Because
   
   $$S_i(t) = \exp(-\int_0^t h_i(s) \, ds) = \exp(-\int_0^t h_0(s) c(z_i \beta) \, ds) = \exp(-\int_0^t h_0(s) \, ds)^{c(z_i \beta)}$$
Model

1. \( h_i(t) = h_0(t)c(z_i\beta) \) for \( c : \mathbb{R} \to [0, \infty) \)

Then hazards are not only \( \propto h_0 \) but to each other as well.

2. \( S_i(t) = S_0(t)^{c(z_i\beta)} \)

1. Because \( S_i(t) = \exp(- \int_0^t h_i(s) \, ds) = \exp(- \int_0^t h_0(s)c(z_i\beta) \, ds) = \exp(- \int_0^t h_0(s) \, ds)^{c(z_i\beta)} \)

2. Also \( \log(- \log(S_i(t))) = \log(- \log(S_0(t))) + \log(c(z_i\beta)) \)
Model

1. \( h_i(t) = h_0(t) c(z_i \beta) \) for \( c : \mathbb{R} \to [0, \infty) \)
2. Then hazards are not only \( \propto h_0 \) but to each other as well.

3. \( S_i(t) = S_0(t)^{c(z_i \beta)} \)

1. Because \( S_i(t) = \exp\left(- \int_0^t h_i(s) \, ds\right) = \exp\left(- \int_0^t h_0(s) c(z_i \beta) \, ds\right) = \exp\left(- \int_0^t h_0(s) \, ds\right)^{c(z_i \beta)} \)
2. Also \( \log(- \log(S_i(t))) = \log(- \log(S_0(t))) + \log(c(z_i \beta)) \)
3. So log log survival should be shifts of one another
Model

1. \( h_i(t) = h_0(t)c(z_i\beta) \) for \( c : \mathbb{R} \rightarrow [0, \infty) \)

2. Then hazards are not only \( \propto h_0 \) but to each other as well.

3. \( S_i(t) = S_0(t)^{c(z_i\beta)} \)

4. Because \( S_i(t) = \exp(- \int_0^t h_i(s) \, ds) = \exp(- \int_0^t h_0(s)c(z_i\beta) \, ds) = \exp(- \int_0^t h_0(s) \, ds)^{c(z_i\beta)} \)

5. Also \( \log(- \log(S_i(t))) = \log(- \log(S_0(t))) + \log(c(z_i\beta)) \)

6. So \( \log \log \) survival should be shifts of one another

7. We’ve seen this before for testing:
Model

1. \( h_i(t) = h_0(t)c(z_i \beta) \) for \( c : \mathbb{R} \to [0, \infty) \)
   
   Then hazards are not only \( \propto h_0 \) but to each other as well.

2. \( S_i(t) = S_0(t)^{c(z_i \beta)} \)
   
   Because \( S_i(t) = \exp(- \int_0^t h_i(s) \, ds) = \exp(- \int_0^t h_0(s)c(z_i \beta) \, ds) = \exp(- \int_0^t h_0(s) \, ds)^{c(z_i \beta)} \)

   Also \( \log(- \log(S_i(t))) = \log(- \log(S_0(t))) + \log(c(z_i \beta)) \)

   So \( \log \log \) survival should be shifts of one another

   We’ve seen this before for testing:

   1. For testing equality of \( m \) groups,
Model

$ h_i(t) = h_0(t)c(z_i\beta) $ for $ c : \mathbb{R} \rightarrow [0, \infty) $

Then hazards are not only $ \propto h_0 $ but to each other as well.

$ S_i(t) = S_0(t)^{c(z_i\beta)} $

Because $ S_i(t) = \exp(-\int_0^t h_i(s) \, ds) = \exp(-\int_0^t h_0(s)c(z_i\beta) \, ds) = \exp(-\int_0^t h_0(s) \, ds)^{c(z_i\beta)} $,

Also $ \log(-\log(S_i(t))) = \log(-\log(S_0(t))) + \log(c(z_i\beta)) $,

So $ \log \log $ survival should be shifts of one another

We’ve seen this before for testing:

For testing equality of $ m $ groups,

$ z $ of length $ m - 1 $,
Model

1. $h_i(t) = h_0(t)c(z_i \beta)$ for $c : \mathbb{R} \rightarrow [0, \infty)$
   1. Then hazards are not only $\propto h_0$ but to each other as well.

2. $S_i(t) = S_0(t)^{c(z_i \beta)}$

   1. Because $S_i(t) = \exp(-\int_0^t h_i(s) \, ds) = \exp(-\int_0^t h_0(s)c(z_i \beta) \, ds) = \exp(-\int_0^t h_0(s) \, ds)^{c(z_i \beta)}$
   2. Also $\log(-\log(S_i(t))) = \log(-\log(S_0(t))) + \log(c(z_i \beta))$
   3. So log log survival should be shifts of one another
   4. We’ve seen this before for testing:
      1. For testing equality of $m$ groups,
      2. $z$ of length $m - 1$,
      3. component $l$ of $z$ is 1 if item is from population $l$ and 0 otherwise
Model

1. \( h_i(t) = h_0(t)c(z_i\beta) \) for \( c : \mathbb{R} \to [0, \infty) \)

Then hazards are not only \( \propto h_0 \) but to each other as well.

2. \( S_i(t) = S_0(t)^{c(z_i\beta)} \)

Because \( S_i(t) = \exp(-\int_0^t h_i(s) \, ds) = \exp(-\int_0^t h_0(s)c(z_i\beta) \, ds) = \exp(-\int_0^t h_0(s) \, ds)^{c(z_i\beta)} \)

Also \( \log(-\log(S_i(t))) = \log(-\log(S_0(t))) + \log(c(z_i\beta)) \)

So log log survival should be shifts of one another

4. We’ve seen this before for testing:
   
   1. For testing equality of \( m \) groups,
   2. \( z \) of length \( m - 1 \),
   3. component \( l \) of \( z \) is 1 if item is from population \( l \) and 0 otherwise

3. Model with \( c = \exp \) is called \textit{Cox model}. 

Regression Models: Proportional Hazards approach
Model

1. \( h_i(t) = h_0(t) c(z_i \beta) \) for \( c : \mathcal{R} \rightarrow [0, \infty) \)

   Then hazards are not only \( \propto h_0 \) but to each other as well.

2. \( S_i(t) = S_0(t)^{c(z_i \beta)} \)

   Because \( S_i(t) = \exp(- \int_0^t h_i(s) \, ds) = \exp(- \int_0^t h_0(s) c(z_i \beta) \, ds) = \exp(- \int_0^t h_0(s) \, ds)^{c(z_i \beta)} \)

   Also \( \log(- \log(S_i(t))) = \log(- \log(S_0(t))) + \log(c(z_i \beta)) \)

   So \( \log \log \) survival should be shifts of one another

   We’ve seen this before for testing:

   For testing equality of \( m \) groups,
   
   \( z \) of length \( m-1 \),
   
   component \( l \) of \( z \) is 1 if item is from population \( l \) and 0 otherwise

Model with \( c = \exp \) is called Cox model.

1. Makes \( \log(h_i(t)/h_0(t)) = z_i \beta \): regression model for log hazards
**Model**

1. \( h_i(t) = h_0(t)c(z_i \beta) \) for \( c : \mathbb{R} \to [0, \infty) \)

   Then hazards are not only \( \propto h_0 \) but to each other as well.

2. \( S_i(t) = S_0(t)^{c(z_i \beta)} \)

   Because \( S_i(t) = \exp(- \int_0^t h_i(s) \, ds) = \exp(- \int_0^t h_0(s)c(z_i \beta) \, ds) = \exp(- \int_0^t h_0(s) \, ds)^{c(z_i \beta)} \)

   Also \( \log(- \log(S_i(t))) = \log(- \log(S_0(t))) + \log(c(z_i \beta)) \)

   So log log survival should be shifts of one another

   We’ve seen this before for testing:

   - For testing equality of \( m \) groups,
   - \( z \) of length \( m - 1 \),
   - component \( l \) of \( z \) is 1 if item is from population \( l \) and 0 otherwise

3. Model with \( c = \exp \) is called **Cox model**.

   - Makes \( \log(h_i(t)/h_0(t)) = z_i \beta \): regression model for log hazards
   - Also, \( h_i(t)/h_j(t) = \exp((z_i - z_j) \beta) \)
Model

1. \( h_i(t) = h_0(t)c(z_i/\beta) \) for \( c : \mathbb{R} \to [0, \infty) \)

Then hazards are not only \( \propto h_0 \) but to each other as well.

2. \( S_i(t) = S_0(t)^c(z_i/\beta) \)

Because \( S_i(t) = \exp(-\int_0^t h_i(s) \, ds) = \exp(-\int_0^t h_0(s)c(z_i/\beta) \, ds) = \exp(-\int_0^t h_0(s) \, ds)^c(z_i/\beta) \)

Also \( \log(-\log(S_i(t))) = \log(-\log(S_0(t))) + \log(c(z_i/\beta)) \)

So log log survival should be shifts of one another

We’ve seen this before for testing:

1. For testing equality of \( m \) groups,
2. \( z \) of length \( m - 1 \),
3. component \( l \) of \( z \) is 1 if item is from population \( l \) and 0 otherwise

3. Model with \( c = \exp \) is called Cox model.

1. Makes \( \log(h_i(t)/h_0(t)) = z_i/\beta \): regression model for log hazards
2. Also, \( h_i(t)/h_j(t) = \exp((z_i - z_j)\beta) \)
3. \( \exp((z_i - z_j)\beta) = \) risk for person with covariates \( z_i \) relative to person with \( z_j \)
That is, if $z_{i1} = 1 \forall i$, I will show you two models with different $\beta_1$ but with the same survival curves for all of the observations.
Intercept term would not be identifiable

1. That is, if \( z_i = 1 \forall i \), I will show you two models with different \( \beta_1 \) but with the same survival curves for all of the observations.

2. Suppose the model holds with \( \beta_1 = \beta_1^* \neq 0 \), \( h^*_0(t) \).
Intercept term would not be identifiable

1. That is, if $z_{i1} = 1 \forall i$, I will show you two models with different $\beta_1$ but with the same survival curves for all of the observations.
2. Suppose the model holds with $\beta_1 = \beta_1^* \neq 0$, $h_0^*(t)$.
3. Then model also holds with $\beta_1 = 0$, $h_0(t) = h_0^*(t) \exp(\beta_1^*)$. 
That is, if $z_{i1} = 1 \forall i$, I will show you two models with different $\beta_1$ but with the same survival curves for all of the observations.

Suppose the model holds with $\beta_1 = \beta_1^* \neq 0, h^*_0(t)$.

Then model also holds with $\beta_1 = 0, h_0(t) = h^*_0(t) \exp(\beta_1^*)$.

When fitting such a model via Newton-Raphson, $\ell''$ does not have an inverse.
Let $\delta_i = \begin{cases} 
1 & \text{if } X_i \leq C_i \\
0 & \text{if } T_i > C_i 
\end{cases}$ and $T_i = \min(X_i, C_i)$.
Likelihood via Profile Methods

1. Let \( \delta_i = \begin{cases} 1 & \text{if } X_i \leq C_i \\
0 & \text{if } T_i > C_i \end{cases} \) and \( T_i = \min(X_i, C_i) \).

2. Order data by time: \( 0 < T_1 < \cdots < T_{i-1} < T_i < \cdots \)
Likelihood via Profile Methods

Let $\delta_i = \begin{cases} 
1 & \text{if } X_i \leq C_i \\
0 & \text{if } T_i > C_i 
\end{cases}$ and $T_i = \min(X_i, C_i)$.

Order data by time: $0 < T_1 < \cdots < T_{i-1} < T_i < \cdots$

Likelihood for data is

$$L = \prod_i h_i(T_i)^{\delta_i} S_i(T_i) = \prod_i h_0(T_i)^{\delta_i} \exp(\delta_i z_i \beta) S_0(t)^{\exp(z_i \beta)}$$
Likelihood via Profile Methods

1. Let \( \delta_i = \begin{cases} 
1 & \text{if } X_i \leq C_i \\
0 & \text{if } T_i > C_i 
\end{cases} \) and \( T_i = \min(X_i, C_i) \).

2. Order data by time: \( 0 < T_1 < \cdots < T_i-1 < T_i < \cdots \)

3. Likelihood for data is

\[
L = \prod_{i} h_i(T_i)^{\delta_i} S_i(T_i) = \prod_{i} h_0(T_i)^{\delta_i} \exp(\delta_i z_i \beta) S_0(t) \exp(z_i \beta)
\]

4. The smaller \( h_0 \) is early on, the larger \( S_0 \) is later
Likelihood via Profile Methods

1. Let \( \delta_i = \begin{cases} 1 & \text{if } X_i \leq C_i \\ 0 & \text{if } T_i > C_i \end{cases} \) and \( T_i = \min(X_i, C_i) \).

2. Order data by time: \( 0 < T_1 < \cdots < T_{i-1} < T_i < \cdots \)

3. Likelihood for data is

\[
L = \prod_i h_i(T_i)^{\delta_i} S_i(T_i) = \prod_i h_0(T_i)^{\delta_i} \exp(\delta_i z_i \beta) S_0(t)^{\exp(z_i \beta)}
\]

1. The smaller \( h_0 \) is early on, the larger \( S_0 \) is later
2. For times not observed, reducing \( h_0 \) to 0 there makes last part bigger and leaves first part unchanged.
Likelihood via Profile Methods

1. Let \( \delta_i = \begin{cases} 
1 & \text{if } X_i \leq C_i \\
0 & \text{if } T_i > C_i 
\end{cases} \) and \( T_i = \min(X_i, C_i) \).

2. Order data by time: \( 0 < T_1 < \cdots < T_{i-1} < T_i < \cdots \)

3. Likelihood for data is

\[
L = \prod_i h_i(T_i)^{\delta_i} S_i(T_i) = \prod_i h_0(T_i)^{\delta_i} \exp(\delta_i z_i \beta) S_0(t)^{\exp(z_i \beta)}
\]

1. The smaller \( h_0 \) is early on, the larger \( S_0 \) is later
2. For times not observed, reducing \( h_0 \) to 0 there makes last part bigger and leaves first part unchanged.

4. Let \( C = \{ i | \delta_i = 1 \} \)
Likelihood via Profile Methods

1. Let \( \delta_i = \begin{cases} 1 & \text{if } X_i \leq C_i \\ 0 & \text{if } T_i > C_i \end{cases} \) and \( T_i = \min(X_i, C_i) \).

2. Order data by time: \( 0 < T_1 < \cdots < T_{i-1} < T_i < \cdots \)

3. Likelihood for data is

\[
L = \prod_i h_i(T_i)^{\delta_i} S_i(T_i) = \prod_i h_0(T_i)^{\delta_i} \exp(\delta_i z_i \beta) S_0(t) \exp(z_i \beta)
\]

1. The smaller \( h_0 \) is early on, the larger \( S_0 \) is later
2. For times not observed, reducing \( h_0 \) to 0 there makes last part bigger and leaves first part unchanged.

4. Let \( C = \{ i | \delta_i = 1 \} \)

5. Hence for fixed \( \beta \), \( L \) is maximized at distn putting all weight on \( \{ t_i | i \in C \} \)
Insert maximizing baseline distribution to give profile likelihood
Insert maximizing baseline distribution to give profile likelihood

From our calculations for the survival curve estimate,

\[ L = \prod_{i \in C} [h_i \exp(z_i \beta)] \exp\left(- \sum_{k=1}^{n} H_0(t_k) \exp(z_k \beta)\right) \]
1. Insert maximizing baseline distribution to give profile likelihood

   From our calculations for the survival curve estimate,

   \[ L = \prod_{i \in C} [h_i \exp(z_i \beta)] \exp(- \sum_{k=1}^{n} H_0(t_k) \exp(z_k \beta)) \]

2. Expressing the cumulative hazards in terms of discrete hazards,

   \[ L = \prod_{i \in C} [h_i \exp(z_i \beta)] \exp(- \sum_{k=1}^{n} \sum_{j \leq k, j \in C} h_j \exp(z_k \beta)). \]
Continued

1. Insert maximizing baseline distribution to give profile likelihood

From our calculations for the survival curve estimate,

\[
L = \prod_{i \in C} [h_i \exp(z_i \beta)] \exp\left( - \sum_{k=1}^{n} H_0(t_k) \exp(z_k \beta) \right)
\]

2. Expressing the cumulative hazards in terms of discrete hazards,

\[
L = \prod_{i \in C} [h_i \exp(z_i \beta)] \exp\left( - \sum_{k=1}^{n} \left[ \sum_{j \leq k, j \in C} h_j \right] \exp(z_k \beta) \right).
\]

3. Interchanging summation,

\[
L = \prod_{i \in C} [h_i \exp(z_i \beta)] \exp\left( - \sum_{j \in C} h_j \left[ \sum_{k \geq j} \exp(z_k \beta) \right] \right)
\]
Continued

\[ L = \prod_{i \in C} \left[ h_i \exp(z_i \beta) \right] \exp\left( -h_i \left[ \sum_{k \geq i} \exp(z_k \beta) \right] \right) \]

Hence estimate \( \hat{h}_i \) maximizes \( h_i \exp(-h_i \left[ \sum_{k \geq i} \exp(z_k \beta) \right]) \) after taking logs and differentiating.

Profile likelihood is

\[ L(\beta) = \prod_{i \in C} \frac{\exp(z_i \beta)}{\sum_{k \geq i} \exp(z_k \beta)} \]
Continued

Pulling all of them inside the big product,

\[
L = \prod_{i \in C} \left\{ \begin{array}{l}
[h_i \exp(z_i \beta)] \exp(-h_i \left[ \sum_{k \geq i} \exp(z_k \beta) \right])
\end{array} \right\}
\]
Continued

Pulling all of them inside the big product,

\[
L = \prod_{i \in C} \left\{ [h_i \exp(z_i \beta)] \exp\left(-h_i \left[ \sum_{k \geq i} \exp(z_k \beta) \right] \right) \right\}
\]

Hence estimate \( \hat{h}_i \) maximizes \( h_i \exp\left(-h_i \left[ \sum_{k \geq i} \exp(z_k \beta) \right] \right) \)
Pulling all of them inside the big product,

\[
L = \prod_{i \in C} \left\{ [h_i \exp(z_i \beta)] \exp(-h_i \left[ \sum_{k \geq i} \exp(z_k \beta) \right]) \right\}
\]

Hence estimate \( \hat{h}_i \) maximizes \( h_i \exp(-h_i \left[ \sum_{k \geq i} \exp(z_k \beta) \right]) \)

Maximizer satisfies \( 1/h_i = \sum_{k \geq i} \exp(z_k \beta) \) after taking logs and differentiating
Continued

1. Pulling all of them inside the big product,

\[
L = \prod_{i \in C} \left\{ \left[ h_i \exp(z_i \beta) \right] \exp\left( -h_i \left[ \sum_{k \geq i} \exp(z_k \beta) \right] \right) \right\}
\]

1. Hence estimate \( \hat{h}_i \) maximizes \( h_i \exp(-h_i \left[ \sum_{k \geq i} \exp(z_k \beta) \right]) \)

2. Maximizer satisfies \( \frac{1}{h_i} = \sum_{k \geq i} \exp(z_k \beta) \) after taking logs and differentiating

2. Profile likelihood is \( L(\beta) = \prod_{i \in C} \frac{\exp(z_i \beta)}{\sum_{k \geq i} \exp(z_k \beta)} \)
Objectives Lecture 06

1. Confidence intervals

Readings: C §3.4, CO §7.2, KM §8.1b, 8.3pn1, 8.3pn2, 8.4, Appendix B
Objectives Lecture 06

1. Confidence intervals
2. Justification of partial likelihood

Readings: C § 3.4, CO § 7.2, KM § 8.1b, 8.3pn1, 8.3pn2, 8.4, Appendix B
Objectives Lecture 06

1. Confidence intervals
2. Justification of partial likelihood
3. Hazard Ratios
Objectives Lecture 06

1. Confidence intervals
2. Justification of partial likelihood
3. Hazard Ratios
4. Computing the estimator

Readings: C §3.4, CO §7.2, KM §8.1b, 8.3pn1, 8.3pn2, 8.4, Appendix B
Objectives Lecture 06

1. Confidence intervals
2. Justification of partial likelihood
3. Hazard Ratios
4. Computing the estimator
5. Infinite Estimators

Readings: C §3.4, CO §7.2, KM §8.1b, 8.3pn1, 8.3pn2, 8.4, Appendix B
Objectives Lecture 06

1. Confidence intervals
2. Justification of partial likelihood
3. Hazard Ratios
4. Computing the estimator
5. Infinite Estimators
6. Ties
Objectives Lecture 06

1. Confidence intervals
2. Justification of partial likelihood
3. Hazard Ratios
4. Computing the estimator
5. Infinite Estimators
6. Ties
7. Likelihood testing
Objectives Lecture 06

1. Confidence intervals
2. Justification of partial likelihood
3. Hazard Ratios
4. Computing the estimator
5. Infinite Estimators
6. Ties
7. Likelihood testing
8. Readings: C §3.4, CO §7.2, KM §8.1b, 8.3pn1, 8.3pn2, 8.4, Appendix B
Alternate Derivation:

1. Let $\mathcal{R}(i)$ be uncensored and unfailed obsn at time $t_i$ (including $i$)
Alternate Derivation:

1. Let \( R(i) \) be uncensored and unfailed obsn at time \( t_i \) (including \( i \))
2. \( P \) [obsn \( i \) fails|obsns \( R(i) \) at risk, fail at time] = \( \frac{\exp(z_i \beta)}{\sum_{k \in R(i)} \exp(z_k \beta)} \)
Alternate Derivation:

1. Let $\mathcal{R}(i)$ be uncensored and unfailed obsn at time $t_i$ (including $i$).
2. $P[\text{obsn } i \text{ fails} | \text{obsns } \mathcal{R}(i) \text{ at risk, fail at time}] = \frac{\exp(z_i \beta)}{\sum_{k \in \mathcal{R}(i)} \exp(z_k \beta)}$
3. Multiply to get Partial likelihood.
Alternate Derivation:

1. Let $R(i)$ be uncensored and unfailed obsn at time $t_i$ (including $i$).
2. $P[\text{obsn } i \text{ fails}| \text{obsns } R(i) \text{ at risk, fail at time}] = \frac{\exp(z_i \beta)}{\sum_{k \in R(i)} \exp(z_k \beta)}$
3. Multiply to get Partial likelihood.
4. Profile likelihood is $L(\beta) = \prod_{i \in C} \frac{\exp(z_i \beta)}{\sum_{k \in R(i)} \exp(z_k \beta)}$
Alternate Derivation:

1. Let $\mathcal{R}(i)$ be uncensored and unfailed obsn at time $t_i$ (including $i$)

2. $\Pr[\text{obsn } i \text{ fails} | \text{obsns } \mathcal{R}(i) \text{ at risk, fail at time}] = \frac{\exp(z_i \beta)}{\sum_{k \in \mathcal{R}(i)} \exp(z_k \beta)}$

3. Multiply to get Partial likelihood.

4. Profile likelihood is $L(\beta) = \prod_{i \in \mathcal{C}} \frac{\exp(z_i \beta)}{\sum_{k \in \mathcal{R}(i)} \exp(z_k \beta)}$

  Definition makes sense for unordered times if risk set is defined to be all those not failed at time $t_i$
Alternate Derivation:

1. Let $\mathcal{R}(i)$ be uncensored and unfailed obsn at time $t_i$ (including $i$).

2. $P[\text{obsn } i \text{ fails} | \text{obsns } \mathcal{R}(i) \text{ at risk, fail at time}] = \frac{\exp(z_i \beta)}{\sum_{k \in \mathcal{R}(i)} \exp(z_k \beta)}$

3. Multiply to get Partial likelihood.

4. Profile likelihood is $L(\beta) = \prod_{i \in C} \frac{\exp(z_i \beta)}{\sum_{k \in \mathcal{R}(i)} \exp(z_k \beta)}$

   - Definition makes sense for unordered times if risk set is defined to be all those not failed at time $t_i$
   - Requires censoring distributions the same for all values of $\beta$
Alternate Derivation:

1. Let $\mathcal{R}(i)$ be uncensored and unfailed obsn at time $t_i$ (including $i$)
2. $P[\text{obsn } i \text{ fails}| \text{obsns } \mathcal{R}(i) \text{ at risk, fail at time}] = \frac{\exp(z_i \beta)}{\sum_{k \in \mathcal{R}(i)} \exp(z_k \beta)}$
3. Multiply to get Partial likelihood.
4. Profile likelihood is $L(\beta) = \prod_{i \in C} \frac{\exp(z_i \beta)}{\sum_{k \in \mathcal{R}(i)} \exp(z_k \beta)}$

- Definition makes sense for unordered times if risk set is defined to be all those not failed at time $t_i$
- Requires censoring distributions the same for all values of $\beta$
- Also is conditional likelihood if censoring happens instantaneously after failure
Alternate Derivation:

1. Let $\mathcal{R}(i)$ be uncensored and unfailed obSN at time $t_i$ (including $i$).
2. $P[\text{obsn } i \text{ fails}|\text{obsns } \mathcal{R}(i) \text{ at risk, fail at time}] = \frac{\exp(z_i \beta)}{\sum_{k \in \mathcal{R}(i)} \exp(z_k \beta)}$
3. Multiply to get *Partial likelihood*.
4. Profile likelihood is $L(\beta) = \prod_{i \in C} \frac{\exp(z_i \beta)}{\sum_{k \in \mathcal{R}(i)} \exp(z_k \beta)}$
   - Definition makes sense for unordered times if risk set is defined to be all those not failed at time $t_i$
   - Requires censoring distributions the same for all values of $\beta$
   - Also is conditional likelihood if censoring happens instantaneously after failure
5. Can show that properties of likelihood extend to partial likelihood
Partial likelihood loses information contained in censoring.

Example: Two data sets, having the same partial likelihood.

Two groups, red ($z = 1$) and blue ($z = 0$).

See Fig. 11.

Fig. 11: Two Data Sets with Identical Partial Likelihood

Data identical except for consecutive censored items swapped.

In fully-parametric model, right panel is stronger evidence favoring red.

Partial likelihood treats them identically:

\[ L(\beta) = \exp(\beta) \exp(0) + \exp(0) \exp(\beta) \times \exp(0) \exp(\beta) = \exp(\beta) \exp(0) + \exp(0) \exp(\beta) \times \exp(\beta) \times \exp(0) \exp(\beta) \]

If no censorings follow the last event, the last factor is always 1.

If we ignored censored observations:

\[ L(\beta) = \exp(\beta) \exp(0) + \exp(0) \exp(\beta) \times \exp(0) \exp(\beta) \]

Regression Models: Proportional Hazards approach Lecture 06
Partial likelihood loses information contained in censoring.

Example: Two data sets, having the same partial likelihood.

Two groups, red \((z = 1)\) and blue \((z = 0)\).
Partial likelihood loses information contained in censoring.

Example: Two data sets, having the same partial likelihood.

- Two groups, red \((z = 1)\) and blue \((z = 0)\).
- See Fig. 11.

**Fig. 11: Two Data Sets with Identical Partial Likelihood**

- Data identical except for consecutive censored items swapped.
- In fully-parametric model, right panel is stronger evidence favoring red.
- Partial likelihood treats them identically:

\[
L(\beta) = \exp(\beta) + \exp(0) + \exp(\beta) \times \exp(0) + \exp(\beta) \times \exp(\beta)
\]

- If no censorings follow the last event, the last factor is always 1.
- If we ignored censored observations:

\[
L(\beta) = \exp(\beta) + \exp(0) + \exp(\beta) \times \exp(0) + \exp(\beta) \times \exp(\beta)
\]
Partial likelihood loses information contained in censoring.

1. Example: Two data sets, having the same partial likelihood.
   - Two groups, red \((z = 1)\) and blue \((z = 0)\).
   - See Fig. 11.

   \[ L(\beta) = \exp(\beta) \exp(0) \exp(\beta) \exp(0) \exp(\beta) \times \exp(0) \exp(0) \exp(\beta) \exp(0) \exp(\beta) \]

2. Data identical except for consecutive censored items swapped.

Fig. 11: Two Data Sets with Identical Partial Likelihood
Partial likelihood loses information contained in censoring.

1 Example: Two data sets, having the same partial likelihood.

   Two groups, red \((z = 1)\) and blue \((z = 0)\).

   See Fig. 11.

   \[ \text{Fig. 11: Two Data Sets with Identical Partial Likelihood} \]

\[ \begin{array}{c}
\chi \quad \chi \\
\chi \quad \chi \\
\chi \quad \chi \\
\chi \quad \chi \\
\chi \quad \chi \\
\chi \quad \chi \\
\chi \quad \chi \\
\chi \quad \chi \\
\chi \quad \chi \\
\chi \quad \chi \\
\chi \quad \chi \\
\chi \quad \chi \\
\chi \quad \chi \\
\end{array} \]

   \text{Event}

\[ \begin{array}{c}
\chi \\
\chi \\
\chi \\
\chi \\
\chi \\
\chi \\
\chi \\
\chi \\
\chi \\
\chi \\
\chi \\
\chi \\
\chi \\
\end{array} \]

   \text{Censored}

2 Data identical except for consecutive censored items swapped.

3 In fully-parametric model, right panel is stronger evidence favoring red.
Partial likelihood loses information contained in censored.

Example: Two data sets, having the same partial likelihood.
- Two groups, red \((z = 1)\) and blue \((z = 0)\).
- See Fig. 11.

Fig. 11: Two Data Sets with Identical Partial Likelihood

- Data identical except for consecutive censored items swapped.
- In fully-parametric model, right panel is stronger evidence favoring red.
- Partial likelihood treats them identically:

\[
L(\beta) = \frac{\exp(\beta)}{\exp(\beta) + \exp(0) + \exp(\beta) + \exp(0) + \exp(\beta)} \times \frac{\exp(0)}{\exp(0) + \exp(\beta) + \exp(0) + \exp(\beta)} \times \frac{\exp(\beta)}{\exp(\beta)}
\]
Partial likelihood loses information contained in censoring.

1 Example: Two data sets, having the same partial likelihood.
   1 Two groups, red ($z = 1$) and blue ($z = 0$).
   1 See Fig. 11.

*Fig. 11: Two Data Sets with Identical Partial Likelihood*

- Data identical except for consecutive censored items swapped.
- In fully-parametric model, right panel is stronger evidence favoring red.
- Partial likelihood treats them identically:

$$L(\beta) = \frac{\exp(\beta)}{\exp(\beta) + \exp(0) + \exp(\beta) + \exp(0) + \exp(\beta)} \times \frac{\exp(0)}{\exp(0) + \exp(\beta) + \exp(0) + \exp(0) + \exp(\beta)} \times \frac{\exp(\beta)}{\exp(\beta)}$$

- If no censorings follow the last event, the last factor is always 1.
Partial likelihood loses information contained in censoring.

1. Example: Two data sets, having the same partial likelihood.
   - Two groups, red ($z = 1$) and blue ($z = 0$).
   - See Fig. 11.

   **Fig. 11: Two Data Sets with Identical Partial Likelihood**

   ![Diagram showing two data sets with identical partial likelihood]

   - Data identical except for consecutive censored items swapped.
   - In fully-parametric model, right panel is stronger evidence favoring red.
   - Partial likelihood treats them identically:

   $$L(\beta) = \frac{\exp(\beta)}{\exp(\beta) + \exp(0) + \exp(\beta) + \exp(0) + \exp(\beta)} \times \frac{\exp(0)}{\exp(0) + \exp(\beta) + \exp(0) + \exp(\beta)} \times \frac{\exp(\beta)}{\exp(\beta)}$$

   - If no censorings follow the last event, the last factor is always 1.
   - If we ignored censored observations:

   $$L(\beta) = \frac{\exp(\beta)}{\exp(\beta) + \exp(0) + \exp(\beta)} \times \frac{\exp(0)}{\exp(0) + \exp(\beta)} \times \frac{\exp(\beta)}{\exp(\beta)}$$
Individual censored first in an interval ought to give evidence that the associated covariate is associated with early failure.
Individual censored first in an interval ought to give evidence that the associated covariate is associated with early failure.

Size of evidence is governed by baseline hazard between censoring points.
Individual censored first in an interval ought to give evidence that the associated covariate is associated with early failure.

Size of evidence is governed by baseline hazard between censoring points

We want to avoid estimating this
Individual censored first in an interval ought to give evidence that the associated covariate is associated with early failure.

1. Size of evidence is governed by baseline hazard between censoring points
2. We want to avoid estimating this
3. Hence we ignore evidence from censoring times
Partial Likelihood derivatives

\[ L(\beta) = \prod_{i \in C} \frac{\exp(z_i \beta)}{\sum_{k \in \mathcal{R}(i)} \exp(z_k \beta)} \]
Partial Likelihood derivatives

1. \( L(\beta) = \prod_{i \in C} \frac{\exp(z_i \beta)}{\sum_{k \in R(i)} \exp(z_k \beta)} \)

2. Let

\[
\pi_{ik} = \frac{\exp(z_k \beta)}{\sum_{m \in R(i)} \exp(z_m \beta)} = \exp(z_m \beta) \left[ \sum_{m \in R(i)} \exp(z_m \beta) \right]^{-1}
\]
Partial Likelihood derivatives

1. \( L(\beta) = \prod_{i \in C} \frac{\exp(z_i \beta)}{\sum_{k \in R(i)} \exp(z_k \beta)} \)

2. Let
   \[ \pi_{ik} = \frac{\exp(z_k \beta)}{\sum_{m \in R(i)} \exp(z_m \beta)} = \exp(z_m \beta) \left[ \sum_{m \in R(i)} \exp(z_m \beta) \right]^{-1} \]

Indices:

Regression Models: Proportional Hazards approach
Partial Likelihood derivatives

1. \( L(\beta) = \prod_{i \in C} \frac{\exp(z_i \beta)}{\sum_{k \in \mathcal{R}(i)} \exp(z_k \beta)} \)

2. Let
\[
\pi_{ik} = \frac{\exp(z_k \beta)}{\sum_{m \in \mathcal{R}(i)} \exp(z_m \beta)} = \exp(z_m \beta) \left[ \sum_{m \in \mathcal{R}(i)} \exp(z_m \beta) \right]^{-1}
\]

Indices:
1. \( i \) represents individual with event, \( k \) represents an individual (including \( i \)) at risk when \( i \) has event,
Partial Likelihood derivatives

1. \[ L(\beta) = \prod_{i \in C} \frac{\exp(z_i \beta)}{\sum_{k \in \mathcal{R}(i)} \exp(z_k \beta)} \]

2. Let

\[ \pi_{ik} = \frac{\exp(z_k \beta)}{\sum_{m \in \mathcal{R}(i)} \exp(z_m \beta)} = \frac{\exp(z_k \beta)}{\sum_{m \in \mathcal{R}(i)} \exp(z_m \beta)} = \exp(z_m \beta) \left[ \sum_{m \in \mathcal{R}(i)} \exp(z_m \beta) \right]^{-1} \]

Indices:

1. \( i \) represents individual with event, \( k \) represents an individual (including \( i \)) at risk when \( i \) has event,

2. and \( j \) will represent a component of the the covariate vector, \( m \) will represent another individual at risk
Then

\[d \beta_j \pi_{ik} = \exp(\sum_{m \in \mathbb{R}} \pi_i \exp(\sum_{m \in \mathbb{R}} \pi_{im} z_{mj}) - 2) + \exp(\sum_{m \in \mathbb{R}} \pi_{im} z_{mj}) - 1 = \pi_{ik} (z_{ij} - \sum_{m \in \mathbb{R}} \pi_{im} z_{mj})\]
Then \( \frac{d}{d\beta_j} \pi_{ik} \) is

\[
\exp(z_k \beta)(-1) \left[\sum_{m \in \mathcal{R}(i)} \exp(z_m \beta) z_m \right] \left[\sum_{m \in \mathcal{R}(i)} \exp(z_m \beta) \right]^{-2} \\
+ \exp(z_k \beta) z_{kj} \left[\sum_{m \in \mathcal{R}(i)} \exp(z_m \beta) \right]^{-1} \\
= \pi_{ik} (z_{ij} - \sum_{m \in \mathcal{R}(i)} \pi_{im} z_m) 
\]
Continued

\[ \pi'_{ik} = \pi_{ik} \left( z_k^\top - \sum_{m \in R} (i)^{\pi_{im}} z_m^\top \right) \]

\[ \ell(\beta) = \sum_{i \in C} z_i \beta - \log(\sum_{k \in R} (i)^{\pi_{ik}} \exp(z_k^\beta)) \]

\[ \ell'(\beta) = \sum_{i \in C} \left[ z_i - \sum_{k \in R} (i)^{\pi_{ik}} z_k \exp(z_k^\beta) \right] \frac{\sum_{k \in R} (i)^{\pi_{ik}} z_k^\top}{\sum_{k \in R} (i)^{\pi_{ik}} z_k^\top \pi_{ik}} \]

\[ \ell''(\beta) = \sum_{i \in C} \left[ -\sum_{k \in R} (i)^{\pi_{ik}} z_k z_k^\top \pi_{ik} + \sum_{k \in R} (i)^{\pi_{ik}} z_k z_k^\top \pi_{ik} \right] \]
Continued

Continued

1. Vector form \( \pi'_{ik} = \pi_{ik}(z_k^\top - \sum_{m \in R(i)} \pi_{im}z_m^\top) \)
Continued

1. Vector form $\pi'_{ik} = \pi_{ik}(z_k^\top - \sum_{m \in R(i)} \pi_{im}z_m^\top)$

2. $\ell(\beta) = \sum_{i \in C} z_i\beta - \log(\sum_{k \in R(i)} \exp(z_k\beta))$
Continued

1. Vector form \( \pi'_{ik} = \pi_{ik}(z_k^\top - \sum_{m \in \mathcal{R}(i)} \pi_{im}z_m^\top) \)

2. \( \ell(\beta) = \sum_{i \in \mathcal{C}} z_i \beta - \log(\sum_{k \in \mathcal{R}(i)} \exp(z_k \beta)) \)

3. \( \ell'(\beta) = \sum_{i \in \mathcal{C}} \left[ z_i - \frac{\sum_{k \in \mathcal{R}(i)} z_k \exp(z_k \beta)}{\sum_{k \in \mathcal{R}(i)} \exp(z_k \beta)} \right] = \sum_{i \in \mathcal{C}} \left[ z_i - \sum_{k \in \mathcal{R}(i)} z_k \pi_{ik} \right] \)
Continued

1. **Vector form** \( \pi'_{ik} = \pi_{ik}(z_k^\top - \sum_{m \in R(i)} \pi_{im} z_m^\top) \)

2. \( \ell(\beta) = \sum_{i \in C} z_i \beta - \log(\sum_{k \in R(i)} \exp(z_k \beta)) \)

3. \( \ell'(\beta) = \sum_{i \in C} \left[ z_i - \frac{\sum_{k \in R(i)} z_k \exp(z_k \beta)}{\sum_{k \in R(i)} \exp(z_k \beta)} \right] = \sum_{i \in C} \left[ z_i - \sum_{k \in R(i)} z_k \pi_{ik} \right] \)

4. \( \ell''(\beta) = \sum_{i \in C} \left[ - \sum_{k \in R(i)} z_k z_k^\top \pi_{ik} + \sum_{k \in R(i)} z_k \pi_{ik} \sum_{k \in R(i)} z_k^\top \pi_{ik} \right] \)
Estimator satisfies $\ell'(\hat{\beta}) = 0$

Estimator satisfies $\ell'(\hat{\beta}) = 0$

   1. Guess $\beta^0$

2. proc phreg
   in SAS,
   coxph
   in R

R Code
SAS Code

Regression Models: Proportional Hazards approach Lecture 06
Estimator satisfies $\ell'(\hat{\beta}) = 0$

   1. Guess $\beta^0$
      1. Generally start at zero.
Estimator satisfies $\ell'(\hat{\beta}) = 0$

   1. Guess $\beta^0$
      1. Generally start at zero.
   2. $0 = \ell'(\hat{\beta}) \approx \ell'(\beta^0) + \ell''(\beta^0)(\hat{\beta} - \beta^0)$
Estimator satisfies $\ell'(\hat{\beta}) = 0$

   1. Guess $\beta^0$
      1. Generally start at zero.
   2. $0 = \ell'(\hat{\beta}) \approx \ell'(\beta^0) + \ell''(\beta^0)(\hat{\beta} - \beta^0)$
   3. Solution $\hat{\beta} = \beta^0 - \ell''(\beta^0)^{-1}\ell'(\beta^0)$
Estimator satisfies $\ell'(\hat{\beta}) = 0$

   1. Guess $\beta^0$
      1. Generally start at zero.
   2. $0 = \ell'(\hat{\beta}) \approx \ell'(\beta^0) + \ell''(\beta^0)(\hat{\beta} - \beta^0)$
   3. Solution $\hat{\beta} = \beta^0 - \ell''(\beta^0)^{-1}\ell'(\beta^0)$
   4. Update using new guess $\hat{\beta} \rightarrow \beta^0$. 

\[ \]
Estimator satisfies $\ell'(\hat{\beta}) = 0$

   1. Guess $\beta^0$
      1. Generally start at zero.
   2. $0 = \ell'(\hat{\beta}) \approx \ell'(\beta^0) + \ell''(\beta^0)(\hat{\beta} - \beta^0)$
   3. Solution $\hat{\beta} = \beta^0 - \ell''(\beta^0)^{-1} \ell'(\beta^0)$
   4. Update using new guess $\hat{\beta} \rightarrow \beta^0$.
   5. Repeat as needed:
Estimator satisfies $\ell'(\hat{\beta}) = 0$

   - Guess $\beta^0$
     - Generally start at zero.
   - $0 = \ell'(\hat{\beta}) \approx \ell'(\beta^0) + \ell''(\beta^0)(\hat{\beta} - \beta^0)$
   - Solution $\hat{\beta} = \beta^0 - \ell''(\beta^0)^{-1} \ell'(\beta^0)$
   - Update using new guess $\hat{\beta} \rightarrow \beta^0$.
   - Repeat as needed:
     - Stop when $\ell'(\beta^0)$ is sufficiently small.
Estimator satisfies $\ell'(\hat{\beta}) = 0$

1. **Solve for ex. using Newton–Raphson method.**
   1. **Guess** $\beta^0$
      1. Generally start at zero.
   2. $0 = \ell'(\hat{\beta}) \approx \ell'(\beta^0) + \ell''(\beta^0)(\hat{\beta} - \beta^0)$
   3. Solution $\hat{\beta} = \beta^0 - \ell''(\beta^0)^{-1}\ell'(\beta^0)$
   4. Update using new guess $\hat{\beta} \rightarrow \beta^0$.
   5. Repeat as needed:
      1. Stop when $\ell'(\beta^0)$ is sufficiently small.
      2. Stop when update $\ell''(\beta^0)^{-1}\ell'(\beta^0)$ is sufficiently small.
Estimator satisfies $\ell'(\hat{\beta}) = 0$

   1. Guess $\beta^0$
      1. Generally start at zero.
   2. $0 = \ell'(\hat{\beta}) \approx \ell'(\beta^0) + \ell''(\beta^0)(\hat{\beta} - \beta^0)$
   3. Solution $\hat{\beta} = \beta^0 - \ell''(\beta^0)^{-1}\ell'(\beta^0)$
   4. Update using new guess $\hat{\beta} \rightarrow \beta^0$.
   5. Repeat as needed:
      1. Stop when $\ell'(\beta^0)$ is sufficiently small.
      2. Stop when update $\ell''(\beta^0)^{-1}\ell'(\beta^0)$ is sufficiently small.

2. `proc phreg` in SAS, `coxph` in R

R Code

SAS Code
What to do about ties?

1. **exact method of Cox**

   - Argue via conditioning
   - If two tied variables covariates $z_i, z_j$,
   - Replace $\exp(z_i \beta) / \sum_{k \in R} \exp(z_k \beta)$ and $\exp(z_j \beta) / \sum_{k \in R} \exp(z_k \beta)$ by $\exp((z_i + z_j) \beta) / \sum_{(k, l) \in P \{i, j\}} \exp((z_k + z_l) \beta)$

   - $P \{i, j\}$ is the set of all pairs of individuals at risk at common failure time of $i$ and $j$

   - Could be lots of them if more than two are tied

---

For more methods and code, see: Regression Models: Proportional Hazards approach Lecture 06 121 / 260
What to do about ties?

1. *exact method of Cox*
   1. Argue via conditioning

2. Breslow’s method
   1. Approximate denominator by usual sum to power of number tied
   2. Unfortunately a bit too big

3. Efron method
   1. Product of sum and some adjustments.

Which one you do is not so important.
What to do about ties?

1. *exact method of Cox*
   1. Argue via conditioning
   2. If two tied variables covariates $z_i, z_j$, 

   
   
   Replace $\exp(z_i \beta) / \sum_{k \in R} \exp(z_k \beta)$ and $\exp(z_j \beta) / \sum_{k \in R} \exp(z_k \beta)$ by $\exp((z_i + z_j) \beta) / \sum_{(k, l) \in P(\{i, j\})} \exp((z_k + z_l) \beta)$

   4. $P(\{i, j\})$ is the set of all pairs of individuals at risk at common failure time of $i$ and $j$

   5. Could be lots of them if more than two are tied

2. *Breslow’s method*
   1. Approximate denominator by usual sum to power of number tied
   2. Unfortunately a bit too big

3. *Efron method*
   1. product of sum and some adjustments.

4. Which one you do is not so important.
What to do about ties?

1. **Exact method of Cox**
   1. Argue via conditioning
   2. If two tied variables covariates \( z_i, z_j \),
   3. Replace \( \exp(z_i \beta) / \sum_{k \in R(i)} \exp(z_k \beta) \) and \( \exp(z_j \beta) / \sum_{k \in R(j)} \exp(z_k \beta) \)
      by \( \exp((z_i + z_j) \beta) / \sum_{(k,l) \in P\{i,j\}} \exp((z_k + z_l) \beta) \)
What to do about ties?

1. **exact method** of Cox
   1. Argue via conditioning
   2. If two tied variables covariates $z_i, z_j$,
   3. Replace $\exp(z_i \beta) / \sum_{k \in R(i)} \exp(z_k \beta)$ and $\exp(z_j \beta) / \sum_{k \in R(j)} \exp(z_k \beta)$ by $\exp((z_i + z_j) \beta) / \sum_{(k,l) \in P\{\{i,j\}\}} \exp((z_k + z_l) \beta)$
   4. $P\{\{i,j\}\}$ is the set of all pairs of individuals at risk at common failure time of $i$ and $j$
What to do about ties?

1. **exact method of Cox**
   1. Argue via conditioning
   2. If two tied variables covariates $z_i, z_j$,
   3. Replace $\exp(z_i/\beta) / \sum_{k \in \mathcal{R}(i)} \exp(z_k/\beta)$ and $\exp(z_j/\beta) / \sum_{k \in \mathcal{R}(j)} \exp(z_k/\beta)$ by $\exp((z_i + z_j)/\beta) / \sum_{(k,l) \in \mathcal{P} \{i,j\}} \exp((z_k + z_l)/\beta)$
   4. $\mathcal{P}(\{i,j\})$ is the set of all pairs of individuals at risk at common failure time of $i$ and $j$
   5. Could be lots of them if more than two are tied
**exact method of Cox**

1. Argue via conditioning
2. If two tied variables covariates $z_i$, $z_j$,
3. Replace $\exp(z_i\beta)/\sum_{k \in R(i)} \exp(z_k \beta)$ and $\exp(z_j\beta)/\sum_{k \in R(j)} \exp(z_k \beta)$ by $\exp((z_i + z_j)\beta)/\sum_{(k,l) \in P\{i,j\}} \exp((z_k + z_l)\beta)$
4. $P\{i,j\}$ is the set of all pairs of individuals at risk at common failure time of $i$ and $j$
5. Could be lots of them if more than two are tied

**Breslow’s method**

1. Approximate denominator by usual sum to power of number tied
2. Unfortunately a bit too big
3. Efron method
4. Product of sum and some adjustments.
5. Which one you do is not so important.
What to do about ties?

1. **exact method** of Cox
   1. Argue via conditioning
   2. If two tied variables covariates $z_i, z_j$,
   3. Replace $\exp(z_i \beta) / \sum_{k \in \mathcal{R}(i)} \exp(z_k \beta)$ and $\exp(z_j \beta) / \sum_{k \in \mathcal{R}(j)} \exp(z_k \beta)$
      by $\exp((z_i + z_j) \beta) / \sum_{(k, l) \in \mathcal{P} \{i, j\}} \exp((z_k + z_l) \beta)$
   4. $\mathcal{P} \{i, j\}$ is the set of all pairs of individuals at risk at common failure time of $i$ and $j$
   5. Could be lots of them if more than two are tied

2. **Breslow’s method**
   1. Approximate denominator by usual sum to power of number tied
What to do about ties?

1 *exact method* of Cox
   1 Argue via conditioning
   2 If two tied variables covariates $z_i$, $z_j$,
   3 Replace $\exp(z_i/\beta) / \sum_{k \in R(i)} \exp(z_k/\beta)$ and $\exp(z_j/\beta) / \sum_{k \in R(j)} \exp(z_k/\beta)$
   by $\exp((z_i + z_j)/\beta) / \sum_{(k,l) \in P\{i,j\}} \exp((z_k + z_l)/\beta)$
   4 $P\{i,j\}$ is the set of all pairs of individuals at risk at common failure time of $i$ and $j$
   5 Could be lots of them if more than two are tied

2 *Breslow’s method*
   1 Approximate denominator by usual sum to power of number tied
   2 Unfortunately a bit too big
What to do about ties?

1. *exact method* of Cox
   1. Argue via conditioning
   2. If two tied variables covariates $z_i, z_j$,
   3. Replace $\exp(z_i \beta)/\sum_{k \in R(i)} \exp(z_k \beta)$ and $\exp(z_j \beta)/\sum_{k \in R(j)} \exp(z_k \beta)$
      by $\exp((z_i + z_j) \beta)/\sum_{(k,l) \in P\{\{i,j\}\}} \exp((z_k + z_l) \beta)$
   4. $P(\{i,j\})$ is the set of all pairs of individuals at risk at common failure time of $i$ and $j$
   5. Could be lots of them if more than two are tied

2. *Breslow’s method*
   1. Approximate denominator by usual sum to power of number tied
   2. Unfortunately a bit too big

3. *Efron method*
What to do about ties?

1. **exact method of Cox**
   1. Argue via conditioning
   2. If two tied variables covariates $z_i$, $z_j$,
   3. Replace $\exp(z_i/\beta) / \sum_{k \in R(i)} \exp(z_k/\beta)$ and $\exp(z_j/\beta) / \sum_{k \in R(j)} \exp(z_k/\beta)$ by $\exp((z_i + z_j)/\beta) / \sum_{(k, l) \in P(\{i, j\})} \exp((z_k + z_l)/\beta)$
   4. $P(\{i, j\})$ is the set of all pairs of individuals at risk at common failure time of $i$ and $j$
   5. Could be lots of them if more than two are tied

2. **Breslow’s method**
   1. Approximate denominator by usual sum to power of number tied
   2. Unfortunately a bit too big

3. **Efron method**
   1. product of sum and some adjustments.
What to do about ties?

1. *exact method* of Cox
   1. Argue via conditioning
   2. If two tied variables covariates $z_i, z_j$,
   3. Replace $\exp(z_i \beta)/\sum_{k \in R(i)} \exp(z_k \beta)$ and $\exp(z_j \beta)/\sum_{k \in R(j)} \exp(z_k \beta)$ by $\exp((z_i + z_j) \beta)/\sum_{(k,l) \in P(\{i,j\})} \exp((z_k + z_l) \beta)$
   4. $P(\{i,j\})$ is the set of all pairs of individuals at risk at common failure time of $i$ and $j$
   5. Could be lots of them if more than two are tied

2. *Breslow’s method*
   1. Approximate denominator by usual sum to power of number tied
   2. Unfortunately a bit too big

3. *Efron method*
   1. product of sum and some adjustments.

4. Which one you do is not so important. R Code  SAS Code
Section: Regression Models

Subsection: Testing
Sampling distribution of the score statistic.

1. Let $U(\beta) = \ell'(\beta)$ be the.
Sampling distribution of the score statistic.

1. Let \( U(\beta) = \ell'(\beta) \) be the .

2. When data are independent observations \( Y_j \),

\[
\ell'(\beta) = \frac{d}{d \beta} \sum_{i=1}^{n} \log(p_j(Y_j, \beta)) = \sum_{i=1}^{n} \frac{d}{d \beta} q_j(Y_j, \beta)
\]
Sampling distribution of the score statistic.

1. Let $U(\beta) = \ell'(\beta)$ be the.

2. When data are independent observations $Y_j$,

$$
\ell'(\beta) = \frac{d}{d\beta} \sum_{i=1}^{n} \log(p_j(Y_j, \beta)) = \sum_{i=1}^{n} \frac{d}{d\beta} q_j(Y_j, \beta)
$$

3. For $q_j(y, \beta) = \log(p_j(y, \beta))$. 

4. As $n$ increases, central limit shows that $U(\beta)$ approximately multivariate normal in identically-distributed case.

More general CLTs imply multivariate normality outside of the non-identically distributed case, under some conditions.
Sampling distribution of the score statistic.

1. Let $U(\beta) = \ell'(\beta)$ be the.

2. When data are independent observations $Y_j$,

$$\ell'(\beta) = \frac{d}{d\beta} \sum_{i=1}^{n} \log(p_j(Y_j, \beta)) = \sum_{i=1}^{n} \frac{d}{d\beta} q_j(Y_j, \beta)$$

for $q_j(y, \beta) = \log(p_j(y, \beta))$.

1. $E\left[ \frac{d}{d\beta_k} \log(p_j(Y_j, \beta)) \right] = \sum_y \left( \frac{\frac{d}{d\beta_k} p_j(y, \beta)}{p_j(y, \beta)} \right) p_j(y, \beta) = \sum_y \frac{d}{d\beta_k} p_j(y, \beta) = \frac{d}{d\beta_k} \sum_y p_j(y, \beta) = \frac{d}{d\beta_k} 1 = 0$
Sampling distribution of the score statistic.

1. Let \( U(\beta) = \ell'(\beta) \) be the .

2. When data are independent observations \( Y_j \),

\[
\ell'(\beta) = \frac{d}{d\beta} \sum_{i=1}^{n} \log(p_j(Y_j, \beta)) = \sum_{i=1}^{n} \frac{d}{d\beta} q_j(Y_j, \beta)
\]

for \( q_j(y, \beta) = \log(p_j(y, \beta)) \).

2. \( \mathbb{E}_\beta \left[ \frac{d}{d\beta_k} \log(p_j(Y_j, \beta)) \right] = \sum_y \left( \frac{\frac{d}{d\beta_k} p_j(y, \beta)}{p_j(y, \beta)} \right) p_j(y, \beta) = \sum_y \frac{d}{d\beta_k} p_j(y, \beta) = \frac{d}{d\beta_k} \sum_y p_j(y, \beta) = \frac{d}{d\beta_k} 1 = 0 \)

3. Implies \( \mathbb{E}_\beta [U] \), under the true distribution.
Sampling distribution of the score statistic.

1. Let $U(\beta) = \ell'(\beta)$ be the.

2. When data are independent observations $Y_j$,

$$
\ell'(\beta) = \frac{d}{d\beta} \sum_{i=1}^{n} \log(p_j(Y_j, \beta)) = \sum_{i=1}^{n} \frac{d}{d\beta} q_j(Y_j, \beta)
$$

for $q_j(y, \beta) = \log(p_j(y, \beta))$.

3. $E\left[ \frac{d}{d\beta_k} \log(p_j(Y_j, \beta)) \right] = \sum_{y} \left( \frac{d}{d\beta_k} \frac{p_j(y, \beta)}{p_j(y, \beta)} \right) p_j(y, \beta) = \sum_{y} \frac{d}{d\beta_k} p_j(y, \beta) = \frac{d}{d\beta_k} \sum_{y} p_j(y, \beta) = \frac{d}{d\beta_k} 1 = 0$

4. Implies $E_{\beta} [U]$, under the true distribution.

5. As $n$ increases, central limit shows that $U(\beta)$ approximately multivariate normal in identically-distributed case.
Sampling distribution of the score statistic.

1. Let $U(\beta) = \ell'(\beta)$ be the.

2. When data are independent observations $Y_j$, 

\[
\ell'(\beta) = \frac{d}{d\beta} \sum_{i=1}^{n} \log(p_j(Y_j, \beta)) = \sum_{i=1}^{n} \frac{d}{d\beta} q_j(Y_j, \beta)
\]

for $q_j(y, \beta) = \log(p_j(y, \beta))$.

3. $E \left[ \frac{d}{d\beta_k} \log(p_j(Y_j, \beta)) \right] = \sum_{y} \left( \frac{d}{d\beta_k} \frac{p_j(y, \beta)}{p_j(y, \beta)} \right) p_j(y, \beta) = \sum_{y} \frac{d}{d\beta_k} p_j(y, \beta) = \frac{d}{d\beta_k} \sum_{y} p_j(y, \beta) = \frac{d}{d\beta_k} 1 = 0$

4. Implies $E_\beta [U]$, under the true distribution.

5. As $n$ increases, central limit shows that $U(\beta)$ approximately multivariate normal in identically-distributed case.

More general CLTs imply multivariate normality outside of the non-identically distributed case, under some conditions.
Continued


1. As in the Product Limit estimator, contributions to score are uncorrelated.
Differentiating again,

\[
0 = \mathbb{E}\left[ \frac{d^2 \log(p_j(Y_j, \beta))}{d\beta_m d\beta_k} \right] = \frac{d}{d\beta_m} \sum_y q_k^m(y, \beta)p_j(y, \beta)
\]

\[
= \sum_y [q_{km}^m(y, \beta)p_j(y, \beta) + q_k^m(y, \beta)q_m^m(y, \beta)p_j(y, \beta)]
\]
Differentiating again,

\[
0 = \mathbb{E} \left[ \frac{d^2 \log(p_j(Y_j, \beta))}{d\beta_m d\beta_k} \right] = \frac{d}{d\beta_m} \sum_y q^k_j(y, \beta)p_j(y, \beta)
\]

\[
= \sum_y [q^{km}_j(y, \beta)p_j(y, \beta) + q^k_j(y, \beta)q^m_j(y, \beta)p_j(y, \beta)]
\]

\[
q^m_j(y, \beta) = \frac{d}{d\beta_k} q_j(y, \beta), \quad q^{km}_j(y, \beta) = \frac{d^2}{d\beta_k \beta_m} q_j(y, \beta),
\]
Differentiating again,

\[ 0 = \mathbb{E} \left[ \frac{d^2 \log(p_j(Y_j, \beta))}{d\beta_m d\beta_k} \right] = \frac{d}{d\beta_m} \sum_y q^k_j(y, \beta)p_j(y, \beta) \]

\[ = \sum_y [q^{km}_j(y, \beta)p_j(y, \beta) + q^k_j(y, \beta)q^m_j(y, \beta)p_j(y, \beta)] \]

1. \( q^m_j(y, \beta) = \frac{d}{d\beta_k} q_j(y, \beta), \ q^{km}_j(y, \beta) = \frac{d^2}{d\beta_k d\beta_m} q_j(y, \beta), \)

2. Independence implies \( \text{Var} [U] = -\mathbb{E} [\ell''(\beta)] \)
Differentiating again,

\[ 0 = \mathbb{E} \left[ \frac{d^2 \log(p_j(Y_j, \beta))}{d\beta_m d\beta_k} \right] = \frac{d}{d\beta_m} \sum_y q_j^k(y, \beta) p_j(y, \beta) \]

\[ = \sum_y \left[ q_j^{km}(y, \beta) p_j(y, \beta) + q_j^k(y, \beta) q_j^m(y, \beta) p_j(y, \beta) \right] \]

1. \( q_j^m(y, \beta) = \frac{d}{d\beta_k} q_j(y, \beta), \quad q_j^{km}(y, \beta) = \frac{d^2}{d\beta_k d\beta_m} q_j(y, \beta) \)

2. Independence implies \( \text{Var}[U] = -\mathbb{E}[\ell''(\beta)] \)

3. Often estimate \( \mathbb{E}[\ell''(\beta)] \) by its observed value.
Differentiating again,

\[ 0 = \mathbb{E} \left[ \frac{d^2 \log(p_j(Y_j, \beta))}{d\beta_m d\beta_k} \right] = \frac{d}{d\beta_m} \sum_y q_j^k(y, \beta)p_j(y, \beta) \]

\[ = \sum_y [q_j^{km}(y, \beta)p_j(y, \beta) + q_j^k(y, \beta)q_j^m(y, \beta)p_j(y, \beta)] \]

1. \[ q_j^m(y, \beta) = \frac{d}{d\beta_k} q_j(y, \beta), \quad q_j^{km}(y, \beta) = \frac{d^2}{d\beta_k \beta_m} q_j(y, \beta), \]

2. Independence implies \( \text{Var} [U] = -\mathbb{E} [\ell''(\beta)] \)

3. Often estimate \( \mathbb{E} [\ell''(\beta)] \) by its observed value.

4. Hence \( U \approx \mathcal{N}(0, -\ell''(\beta)) \).
Sampling Distribution for the estimator

\[ \hat{\beta} \approx \beta - \ell''(\beta)^{-1} \ell'(\beta^0). \]
Sampling Distribution for the estimator

\[ \hat{\beta} \approx \beta - \ell''(\beta)^{-1} \ell'(\beta^0). \]

1. Newton Raphson iterations starting at (unknown) true value.
Sampling Distribution for the estimator

1. $\hat{\beta} \approx \beta - \ell''(\beta)^{-1}\ell'(\beta^0)$.
   - Newton Raphson iterations starting at (unknown) true value.

2. Estimate variance of $\hat{\beta}$ by $(-\ell''(\beta)^{-1})(-\ell''(\beta))(-\ell''(\beta)^{-1}) = -\ell''(\beta)^{-1}$
Sampling Distribution for the estimator

1. \( \hat{\beta} \approx \beta - \ell''(\beta)^{-1} \ell'(\beta^0) \).
   - Newton Raphson iterations starting at (unknown) true value.

2. Estimate variance of \( \hat{\beta} \) by
   \[ (-\ell''(\beta)^{-1})(-\ell''(\beta))(-\ell''(\beta)^{-1}) = -\ell''(\beta)^{-1} \]

3. Hence \( \hat{\beta} \approx N(\beta, -\ell''(\beta)^{-1}) \)
Likelihood Ratio and Wald Test Statistics

1. $H_0 : \beta = \beta^0 \text{ vs } H_A : \beta \neq \beta^0$
Likelihood Ratio and Wald Test Statistics

1. $H_0 : \beta = \beta^0$ vs $H_A : \beta \neq \beta^0$

2. $(\hat{\beta} - \beta^0)^\top[-\ell''(\hat{\beta})](\hat{\beta} - \beta^0)$ is Wald test statistic
Likelihood Ratio and Wald Test Statistics

1. $H_0: \beta = \beta^0$ vs $H_A: \beta \neq \beta^0$

2. $(\hat{\beta} - \beta^0)^\top [-\ell''(\hat{\beta})](\hat{\beta} - \beta^0)$ is Wald test statistic
   
   $\hat{\beta} \approx \mathcal{N}(\beta, \ell''(\hat{\beta})^{-1})$
Likelihood Ratio and Wald Test Statistics

1. \( H_0 : \beta = \beta^0 \) vs \( H_A : \beta \neq \beta^0 \)

2. \((\hat{\beta} - \beta^0)^\top [−\ell''(\hat{\beta})](\hat{\beta} - \beta^0)\) is \textit{Wald test} statistic

1. \( \hat{\beta} \approx \mathcal{N}(\beta, \ell''(\hat{\beta})^{-1}) \)

2. \( \hat{\beta} \) approximately unbiased.
Likelihood Ratio and Wald Test Statistics

1. $H_0 : \beta = \beta^0$ vs $H_A : \beta \neq \beta^0$

2. $(\hat{\beta} - \beta^0)^\top [-\ell''(\hat{\beta})](\hat{\beta} - \beta^0)$ is Wald test statistic
   - $\hat{\beta} \approx N(\beta, \ell''(\hat{\beta})^{-1})$
   - $\hat{\beta}$ approximately unbiased.

3. $2 \times \max \ell$ is (log) likelihood ratio test statistic
Likelihood Ratio and Wald Test Statistics

1. $H_0 : \beta = \beta^0$ vs $H_A : \beta \neq \beta^0$

2. $(\hat{\beta} - \beta^0)^\top[-\ell''(\hat{\beta})](\hat{\beta} - \beta^0)$ is Wald test statistic
   - $\hat{\beta} \approx \mathcal{N}(\beta, \ell''(\hat{\beta})^{-1})$
   - $\hat{\beta}$ approximately unbiased.

3. $2 \times \max \ell$ is (log) likelihood ratio test statistic
   - $-2[\ell(\beta^0) - \ell(\hat{\beta})] \approx -2[\ell'(\hat{\beta})(\beta^0 - \hat{\beta}) + \frac{1}{2}(\beta^0 - \hat{\beta})^\top \ell''(\hat{\beta})(\beta^0 - \hat{\beta})] = (\beta^0 - \hat{\beta})^\top(-\ell''(\hat{\beta}))(\beta^0 - \hat{\beta})$. 

Hence statistic is approximately $\chi^2$ if smaller model is correct.

Degrees of freedom is difference in number of parameters.
Likelihood Ratio and Wald Test Statistics

1. $H_0 : \beta = \beta^0$ vs $H_A : \beta \neq \beta^0$

2. $(\hat{\beta} - \beta^0)^\top [-\ell''(\hat{\beta})](\hat{\beta} - \beta^0)$ is *Wald test* statistic
   - $\hat{\beta} \approx \mathcal{N}(\beta, \ell''(\hat{\beta})^{-1})$
   - $\hat{\beta}$ approximately unbiased.

3. $2 \times \max \ell$ is *(log) likelihood ratio* test statistic
   - $-2[\ell(\beta^0) - \ell(\hat{\beta})] \approx -2[\ell'(\hat{\beta})(\beta^0 - \hat{\beta}) + \frac{1}{2}(\beta^0 - \hat{\beta})^\top \ell''(\hat{\beta})(\beta^0 - \hat{\beta})] = (\beta^0 - \hat{\beta})^\top (-\ell''(\hat{\beta}))(\beta^0 - \hat{\beta})$.
   - Hence statistic is approximately $\chi^2$ if smaller model is correct.
Likelihood Ratio and Wald Test Statistics

1. $H_0 : \beta = \beta^0$ vs $H_A : \beta \neq \beta^0$

2. $(\hat{\beta} - \beta^0)^\top[-\ell''(\hat{\beta})](\hat{\beta} - \beta^0)$ is Wald test statistic
   - $\hat{\beta} \approx \mathcal{N}(\beta, \ell''(\hat{\beta})^{-1})$
   - $\hat{\beta}$ approximately unbiased.

3. $2 \times \max \ell$ is (log) likelihood ratio test statistic
   - $-2[\ell(\beta^0) - \ell(\hat{\beta})] \approx -2[\ell'(\hat{\beta})(\beta^0 - \hat{\beta}) + \frac{1}{2}(\beta^0 - \hat{\beta})^\top\ell''(\hat{\beta})(\beta^0 - \hat{\beta})] = (\beta^0 - \hat{\beta})^\top(-\ell''(\hat{\beta}))(\beta^0 - \hat{\beta})$
   - Hence statistic is approximately $\chi^2$ if smaller model is correct.
   - degrees of freedom is difference in number of parameters
Continued

1. \( U(\beta^0)^\top \left[ -\ell''(\beta^0) \right]^{-1} U(\beta^0) \) for \( U(\beta) = \ell' \) is score test statistic.
Continued

1. $U(\beta^0)^\top [-\ell'''(\beta^0)]^{-1} U(\beta^0)$ for $U(\beta) = \ell'$ is score test statistic.

2. $U(\beta^0) \approx N(0, -\ell'')$
Continued

1. $U(\beta^0)^\top [-\ell''(\beta^0)]^{-1} U(\beta^0)$ for $U(\beta) = \ell'$ is score test statistic.

2. $U(\beta^0) \approx N(0, -\ell'')$

3. Notation a bit abusive, since $\ell''$ is data dependent.
Continued

1. \( U(\beta^0)^\top \left[ -\ell''(\beta^0) \right]^{-1} U(\beta^0) \) for \( U(\beta) = \ell' \) is score test statistic.

2. \( U(\beta^0) \approx \mathcal{N}(0, -\ell'') \)

3. Notation a bit abusive, since \( \ell'' \) is data dependent.

4. See Fig. 12.

Fig. 12: Log Likelihood and Related Tests

Slope = score test (before norming, squaring)

Vertical distance = (log) likelihood ratio test

Horizontal distance = Wald test (before norming, squaring)
Application to $K$ sample problem

1. Let $z_{ij} = \begin{cases} 1 & \text{if item } i \text{ is in group } j \\ 0 & \text{otherwise} \end{cases}$
Application to $K$ sample problem

1. Let $z_{ij} = \begin{cases} 1 & \text{if item } i \text{ is in group } j \\ 0 & \text{otherwise} \end{cases}$

2. $\beta^0 = 0$. 

As with ANOVA via regression, don’t treat the levels as a single variable with ordered categories.
Application to $K$ sample problem

1. Let $z_{ij} = \begin{cases} 1 & \text{if item } i \text{ is in group } j \\ 0 & \text{otherwise} \end{cases}$

2. $\beta^0 = 0$.

3. $U_j(\beta) = \sum_{i \in C} \left[ z_{ij} - \frac{\sum_{k \in \mathcal{R}(i)} z_{kj} \exp(z_{kj} \beta)}{\sum_{k \in \mathcal{R}(i)} \exp(z_{kj} \beta)} \right]$
Application to $K$ sample problem

1. Let $z_{ij} = \begin{cases} 1 & \text{if item } i \text{ is in group } j \\ 0 & \text{otherwise} \end{cases}$

2. $\beta^0 = 0$.

3. $U_j(\beta) = \sum_{i \in C} \left[ z_{ij} - \frac{\sum_{k \in R(i)} z_{kj} \exp(z_{kj} \beta)}{\sum_{k \in R(i)} \exp(z_{kj} \beta)} \right]$

4. $U_j(0) = \sum_{i \in C} \left[ z_{ij} - \frac{\sum_{k \in R(i)} z_{kj}}{\sum_{k \in R(i)} 1} \right] = \sum_{l=1}^{D} [D_{lj} - D_l Y_{lj} / Y_l]$

As with ANOVA via regression, don't treat the levels as a single variable with ordered categories.
Application to $K$ sample problem

1. Let $z_{ij} = \begin{cases} 1 & \text{if item } i \text{ is in group } j \\ 0 & \text{otherwise} \end{cases}$

2. $\beta^0 = 0$.

3. $U_j(\beta) = \sum_{i \in C} \left[ z_{ij} - \frac{\sum_{k \in \mathcal{R}(i)} z_{kj} \exp(z_{kj} \beta)}{\sum_{k \in \mathcal{R}(i)} \exp(z_{kj} \beta)} \right]$

4. $U_j(0) = \sum_{i \in C} \left[ z_{ij} - \frac{\sum_{k \in \mathcal{R}(i)} z_{kj}}{\sum_{k \in \mathcal{R}(i)} 1} \right] = \sum_{l=1}^D \left[ D_{lj} - D_l Y_{lj} / Y_l \right]$

5. Same as log rank statistic
Application to $K$ sample problem

1. Let $z_{ij} = \begin{cases} 1 & \text{if item } i \text{ is in group } j \\ 0 & \text{otherwise} \end{cases}$

2. $\beta^0 = 0$.

3. $U_j(\beta) = \sum_{i \in C} \left[ z_{ij} - \frac{\sum_{k \in R(i)} z_{kj} \exp(z_{kj} \beta)}{\sum_{k \in R(i)} \exp(z_{kj} \beta)} \right]$

4. $U_j(0) = \sum_{i \in C} \left[ z_{ij} - \frac{\sum_{k \in R(i)} z_{kj}}{\sum_{k \in R(i)} 1} \right] = \sum_{l=1}^{D} [D_{lj} - D_l Y_{lj} / Y_l]$

5. Same as log rank statistic
   
   $\sum_{i \in C} z_{ij} = \sum_l D_{lj}$
Application to $K$ sample problem

1. Let $z_{ij} = \begin{cases} 
1 & \text{if item } i \text{ is in group } j \\
0 & \text{otherwise} 
\end{cases}$

2. $\beta^0 = 0$.

3. $U_j(\beta) = \sum_{i \in C} \left[ z_{ij} - \frac{\sum_{k \in R(i)} z_{kj} \exp(z_{kj} \beta)}{\sum_{k \in R(i)} \exp(z_{kj} \beta)} \right]$.

4. $U_j(0) = \sum_{i \in C} \left[ z_{ij} - \frac{\sum_{k \in R(i)} z_{kj}}{\sum_{k \in R(i)} 1} \right] = \sum_{l=1}^D [D_{lj} - D_l Y_{lj} / Y_l]$.

5. Same as log rank statistic

   1. $\sum_{i \in C} z_{ij} = \sum_l D_{lj}$
   2. $\sum_{k \in R(i)} 1 = Y_l$
Application to $K$ sample problem

1. Let $z_{ij} = \begin{cases} 1 & \text{if item } i \text{ is in group } j \\ 0 & \text{otherwise} \end{cases}$

2. $\beta^0 = 0$.

3. $U_j(\beta) = \sum_{i \in C} \left[ z_{ij} - \frac{\sum_{k \in R(i)} z_{kj} \exp(z_{kj} \beta)}{\sum_{k \in R(i)} \exp(z_{kj} \beta)} \right]$.

4. $U_j(0) = \sum_{i \in C} \left[ z_{ij} - \frac{\sum_{k \in R(i)} z_{kj}}{\sum_{k \in R(i)} 1} \right] = \sum_{l=1}^{D} [D_{lj} - D_l Y_{lj} / Y_l]$

5. Same as log rank statistic

   1. $\sum_{i \in C} z_{ij} = \sum_{l} D_{lj}$
   2. $\sum_{k \in R(i)} 1 = Y_l$
   3. $D_l = 1$
Application to $K$ sample problem

1. Let $z_{ij} = \begin{cases} 1 & \text{if item } i \text{ is in group } j \\ 0 & \text{otherwise} \end{cases}

2. $\beta^0 = 0$.

3. $U_j(\beta) = \sum_{i \in C} \left[ z_{ij} - \frac{\sum_{k \in R(i)} z_{kj} \exp(z_{kj} \beta)}{\sum_{k \in R(i)} \exp(z_{kj} \beta)} \right]

4. $U_j(0) = \sum_{i \in C} \left[ z_{ij} - \frac{\sum_{k \in R(i)} z_{kj}}{\sum_{k \in R(i)} 1} \right] = \sum_{l=1}^{D} [D_{lj} - D_l Y_{lj} / Y_l]

5. Same as log rank statistic

   1. $\sum_{i \in C} z_{ij} = \sum_{l} D_{lj}$
   2. $\sum_{k \in R(i)} 1 = Y_l$
   3. $D_l = 1$
   4. $\sum_{k \in R(i)} z_{kj} = Y_{ij}$
Application to $K$ sample problem

1. Let $z_{ij} = \begin{cases} 1 & \text{if item } i \text{ is in group } j \\ 0 & \text{otherwise} \end{cases}$

2. $\beta^0 = 0$.

3. $U_j(\beta) = \sum_{i \in C} \left[ z_{ij} - \frac{\sum_{k \in R(i)} z_{kj} \exp(z_{kj} \beta)}{\sum_{k \in R(i)} \exp(z_{kj} \beta)} \right]$  

4. $U_j(0) = \sum_{i \in C} \left[ z_{ij} - \frac{\sum_{k \in R(i)} z_{kj}}{\sum_{k \in R(i)} 1} \right] = \sum_{l=1}^{D}[D_{lj} - D_l Y_{lj}/Y_l]$  

5. Same as log rank statistic

   1. $\sum_{i \in C} z_{ij} = \sum_l D_{lj}$  
   2. $\sum_{k \in R(i)} 1 = Y_l$  
   3. $D_l = 1$  
   4. $\sum_{k \in R(i)} z_{kj} = Y_{ij}$  

6. As with ANOVA via regression, don’t treat the levels as a single variable with ordered categories
Only $K - 1$ parameters are identifiable.

Suppose $\beta_1, \beta_2, \ldots, \beta_K$ are such that $h_0(t) \exp(\beta_k)$ is the hazard for group $k$. 

Let $\beta^* = \beta_k - \beta_0$.

Then $h_0(t)$ and $\beta$ and $h^*_0(t)$ and $\beta^*$ give the same probabilities.

Often resolved by fixing $\beta_1 = 0$.

Could also be resolved by fixing $\beta_K = 0$.

Group whose parameter is set to zero is called baseline.

Equivalent to dropping column from design matrix.

Could also resolve nonidentifiability by fixing $\sum_{k=1}^{K} \beta_k = 0$.

Less common.

Without this resolution, when fitting such a model via Newton-Raphson, $\ell''$ does not have an inverse.
Only $K - 1$ parameters are identifiable.

1. Suppose $\beta_1, \beta_2, \ldots, \beta_K$ are such that $h_0(t) \exp(\beta_k)$ is the hazard for group $k$.
2. Let $h^*_0(t) = \exp(\beta_1)h_0(t)$
Only \( K - 1 \) parameters are identifiable.

1. Suppose \( \beta_1, \beta_2, \ldots, \beta_K \) are such that \( h_0(t) \exp(\beta_k) \) is the hazard for group \( k \).
2. Let \( h_0^*(t) = \exp(\beta_1) h_0(t) \)
3. Let \( \beta_k^* = \beta_k - \beta_0 \).
Only $K - 1$ parameters are identifiable.

1. Suppose $\beta_1, \beta_2, \ldots, \beta_K$ are such that $h_0(t) \exp(\beta_k)$ is the hazard for group $k$.
2. Let $h_0^*(t) = \exp(\beta_1) h_0(t)$
3. Let $\beta_k^* = \beta_k - \beta_0$.
4. Then $h_0(t)$ and $\beta$ and $h_0^*(t)$ and $\beta^*$ give the same probabilities.
Only $K - 1$ parameters are identifiable.

1. Suppose $\beta_1, \beta_2, \ldots, \beta_K$ are such that $h_0(t) \exp(\beta_k)$ is the hazard for group $k$.
2. Let $h_0^*(t) = \exp(\beta_1) h_0(t)$
3. Let $\beta_k^* = \beta_k - \beta_0$.
4. Then $h_0(t)$ and $\beta$ and $h_0^*(t)$ and $\beta^*$ give the same probabilities.
5. Often resolved by fixing $\beta_1 = 0$. 

- Could also be resolved by fixing $\beta_K = 0$.
- Group whose parameter is set to zero is called baseline.
- Equivalent to dropping column from design matrix.
- Could also resolve nonidentifiability by fixing $\sum_{k=1}^K \beta_k = 0$.
- Less common.
- Without this resolution, when fitting such a model via Newton-Raphson, $\ell''$ does not have an inverse.
Only \( K - 1 \) parameters are identifiable.

1. Suppose \( \beta_1, \beta_2, \ldots, \beta_K \) are such that \( h_0(t) \exp(\beta_k) \) is the hazard for group \( k \).
2. Let \( h_0^*(t) = \exp(\beta_1)h_0(t) \)
3. Let \( \beta_k^* = \beta_k - \beta_0 \).
4. Then \( h_0(t) \) and \( \beta \) and \( h_0^*(t) \) and \( \beta^* \) give the same probabilities.
5. Often resolved by fixing \( \beta_1 = 0 \).
6. Could also be resolved by fixing \( \beta_K = 0 \).
Only $K - 1$ parameters are identifiable.

1. Suppose $\beta_1, \beta_2, \ldots, \beta_K$ are such that $h_0(t) \exp(\beta_k)$ is the hazard for group $k$.
2. Let $h_0^*(t) = \exp(\beta_1) h_0(t)$
3. Let $\beta_k^* = \beta_k - \beta_0$.
4. Then $h_0(t)$ and $\beta$ and $h_0^*(t)$ and $\beta^*$ give the same probabilities.
5. Often resolved by fixing $\beta_1 = 0$.
6. Could also be resolved by fixing $\beta_K = 0$.
7. Group whose parameter is set to zero is called baseline.
Only $K - 1$ parameters are identifiable.

1. Suppose $\beta_1, \beta_2, \ldots, \beta_K$ are such that $h_0(t) \exp(\beta_k)$ is the hazard for group $k$.
2. Let $h_0^*(t) = \exp(\beta_1)h_0(t)$
3. Let $\beta_k^* = \beta_k - \beta_0$.
4. Then $h_0(t)$ and $\beta$ and $h_0^*(t)$ and $\beta^*$ give the same probabilities.
5. Often resolved by fixing $\beta_1 = 0$.
6. Could also be resolved by fixing $\beta_K = 0$.
7. Group whose parameter is set to zero is called baseline.
8. Equivalent to dropping column from design matrix.
Only $K - 1$ parameters are identifiable.

1. Suppose $\beta_1, \beta_2, \ldots, \beta_K$ are such that $h_0(t) \exp(\beta_k)$ is the hazard for group $k$.
2. Let $h_0^*(t) = \exp(\beta_1)h_0(t)$
3. Let $\beta_k^* = \beta_k - \beta_0$.
4. Then $h_0(t)$ and $\beta$ and $h_0^*(t)$ and $\beta^*$ give the same probabilities.
5. Often resolved by fixing $\beta_1 = 0$.
6. Could also be resolved by fixing $\beta_K = 0$.
7. Group whose parameter is set to zero is called baseline.
8. Equivalent to dropping column from design matrix.
9. Could also resolve nonidentifiability by fixing $\sum_{k=1}^{K} \beta_k = 0$. 

Less common.

Without this resolution, when fitting such a model via Newton-Raphson, $\ell''$ does not have an inverse. R Code SAS Code

Regression Models: Testing Lecture 06
Only $K - 1$ parameters are identifiable.

1. Suppose $\beta_1, \beta_2, \ldots, \beta_K$ are such that $h_0(t) \exp(\beta_k)$ is the hazard for group $k$.

2. Let $h_0^*(t) = \exp(\beta_1) h_0(t)$

3. Let $\beta_k^* = \beta_k - \beta_0$.

4. Then $h_0(t)$ and $\beta$ and $h_0^*(t)$ and $\beta^*$ give the same probabilities.

5. Often resolved by fixing $\beta_1 = 0$.

6. Could also be resolved by fixing $\beta_K = 0$.

7. Group whose parameter is set to zero is called baseline.

8. Equivalent to dropping column from design matrix.

9. Could also resolve nonidentifiability by fixing $\sum_{k=1}^{K} \beta_k = 0$.

   - Less common.
Only $K - 1$ parameters are identifiable.

1. Suppose $\beta_1, \beta_2, \ldots, \beta_K$ are such that $h_0(t) \exp(\beta_k)$ is the hazard for group $k$.
2. Let $h_0^*(t) = \exp(\beta_1)h_0(t)$
3. Let $\beta_k^* = \beta_k - \beta_0$.
4. Then $h_0(t)$ and $\beta$ and $h_0^*(t)$ and $\beta^*$ give the same probabilities.
5. Often resolved by fixing $\beta_1 = 0$.
6. Could also be resolved by fixing $\beta_K = 0$.
7. Group whose parameter is set to zero is called baseline.
8. Equivalent to dropping column from design matrix.
9. Could also resolve nonidentifiability by fixing $\sum_{k=1}^{K} \beta_k = 0$.
   - Less common.
10. Without this resolution, when fitting such a model via Newton-Raphson, $\ell''$ does not have an inverse.
Tests for some parameters and not others:

1. Notation: $\beta = (\psi, \phi)$, $\psi$ of interest.

...
Tests for some parameters and not others:

1. Notation: $\beta = (\psi, \phi)$, $\psi$ of interest.

   1. $\phi$ not of interest

2. Likelihood Ratio: $2 \times$ difference in maximized $\ell$.

3. Maximize $\ell$ over $\beta$.

4. Maximize $\ell$ with $\psi = 0$ and $\phi$ unconstrained.

5. Degrees of freedom is length of $\psi$.

6. Wald Test $(\hat{\psi} - \psi_0)^{\top} [I_{11}(\hat{\beta})]^{-1} (\hat{\psi} - \psi_0)$

7. Let $\hat{\beta} = (\hat{\psi}, \hat{\phi})$ maximize $\ell$ over $\beta$.

8. $\text{Var}[\hat{\beta}] \approx I(\hat{\beta}) = [-\ell''(\hat{\beta})]^{-1}$.

9. $\text{Var}[\hat{\psi}] \approx \text{appropriate sub-matrix} I_{11}.$

10. You need to invert twice, and select a submatrix in between.

11. Score Test $\ell_1(0, \tilde{\eta})^{\top} I_{11}(0, \tilde{\eta}) \ell_1(0, \tilde{\eta})$.

12. Let $\tilde{\phi}$ maximize $\ell$ over $\phi$ with $\psi = 0$.

13. Let $\ell_1$ be components of $\ell''$ corresponding to $\psi$.

14. All three have approximate distribution $\sim \chi^2$, degrees of freedom is number of components in $\psi$. 
Tests for some parameters and not others:

1. Notation: $\beta = (\psi, \phi)$, $\psi$ of interest.
   - $\phi$ not of interest

2. Likelihood Ratio: $2 \times$ difference in maximized $\ell$.
Tests for some parameters and not others:

1. Notation: $\beta = (\psi, \phi)$, $\psi$ of interest.
   - $\phi$ not of interest

2. Likelihood Ratio: $2 \times$ difference in maximized $\ell$.
   - Maximize $\ell$ over $\beta$

3. Wald Test: $(\hat{\psi} - \psi_0)^\top [I_{11}(\hat{\beta})]^{-1} (\hat{\psi} - \psi_0)$
   - Let $\hat{\beta} = (\hat{\psi}, \hat{\phi})$ maximize $\ell$ over $\beta$

4. Score Test: $\ell_1(0, \tilde{\eta})^\top I_{11}(0, \tilde{\eta}) \ell_1(0, \tilde{\eta})$
   - Let $\tilde{\phi}$ maximize $\ell$ over $\phi$ with $\psi = 0$

5. All three have approximate distribution $\sim \chi^2$, degrees of freedom is number of components in $\psi$. 
Tests for some parameters and not others:

1. Notation: \( \beta = (\psi, \phi) \), \( \psi \) of interest.
   - \( \phi \) not of interest

2. Likelihood Ratio: \( 2 \times \) difference in maximized \( \ell \).
   - Maximize \( \ell \) over \( \beta \)
   - Maximize \( \ell \) with \( \psi = 0 \) and \( \phi \) unconstrained.
Tests for some parameters and not others:

1. Notation: \( \beta = (\psi, \phi) \), \( \psi \) of interest.
   1. \( \phi \) not of interest

2. Likelihood Ratio: \( 2 \times \) difference in maximized \( \ell \).
   1. Maximize \( \ell \) over \( \beta \)
   2. Maximize \( \ell \) with \( \psi = 0 \) and \( \phi \) unconstrained.
       1. degrees of freedom is length of \( \psi \)
Tests for some parameters and not others:

1. Notation: $\beta = (\psi, \phi)$, $\psi$ of interest.
   - $\phi$ not of interest

2. Likelihood Ratio: $2 \times$ difference in maximized $\ell$.
   - Maximize $\ell$ over $\beta$
   - Maximize $\ell$ with $\psi = 0$ and $\phi$ unconstrained.
   - degrees of freedom is length of $\psi$

3. Wald Test $(\hat{\psi} - \psi^0)^\top [I^{11}(\hat{\beta})]^{-1} (\hat{\psi} - \psi^0)$
Tests for some parameters and not others:

1. Notation: $\beta = (\psi, \phi)$, $\psi$ of interest.
   - $\phi$ not of interest

2. Likelihood Ratio: $2 \times$ difference in maximized $\ell$.
   - Maximize $\ell$ over $\beta$
   - Maximize $\ell$ with $\psi = 0$ and $\phi$ unconstrained.
     - degrees of freedom is length of $\psi$

3. Wald Test $(\hat{\psi} - \psi^0)^\top [I^{11}(\hat{\beta})]^{-1} (\hat{\psi} - \psi^0)$
   - Let $\hat{\beta} = (\hat{\psi}, \hat{\phi})$ maximize $\ell$ over $\beta$
Tests for some parameters and not others:

1. Notation: $\beta = (\psi, \phi)$, $\psi$ of interest.
   - $\phi$ not of interest

2. Likelihood Ratio: $2 \times$ difference in maximized $\ell$.
   - Maximize $\ell$ over $\beta$
   - Maximize $\ell$ with $\psi = 0$ and $\phi$ unconstrained.
   - degrees of freedom is length of $\psi$

3. Wald Test $\left(\hat{\psi} - \psi_0\right)^\top \left[I_{11}(\hat{\beta})\right]^{-1} \left(\hat{\psi} - \psi_0\right)$
   - Let $\hat{\beta} = (\hat{\psi}, \hat{\phi})$ maximize $\ell$ over $\beta$
   - $\text{Var} \left[\hat{\beta}\right] \approx I(\hat{\beta}) = \left[-\ell''(\hat{\beta})\right]^{-1}$
Tests for some parameters and not others:

1. **Notation:** $\beta = (\psi, \phi)$, $\psi$ of interest.
   - $\phi$ not of interest

2. **Likelihood Ratio:** $2 \times$ difference in maximized $\ell$.
   - Maximize $\ell$ over $\beta$
   - Maximize $\ell$ with $\psi = 0$ and $\phi$ unconstrained.
   - degrees of freedom is length of $\psi$

3. **Wald Test** $(\hat{\psi} - \psi^0)^	op [I^{11}(\hat{\beta})]^{-1} (\hat{\psi} - \psi^0)$
   - Let $\hat{\beta} = (\hat{\psi}, \hat{\phi})$ maximize $\ell$ over $\beta$
   - $\text{Var} \begin{bmatrix} \hat{\beta} \end{bmatrix} \approx I(\hat{\beta}) = [-\ell''(\hat{\beta})]^{-1}$
   - $\text{Var} \begin{bmatrix} \hat{\psi} \end{bmatrix} \approx$ appropriate sub-matrix $I^{11}$
Tests for some parameters and not others:

1. Notation: $\beta = (\psi, \phi)$, $\psi$ of interest.
   - $\phi$ not of interest

2. Likelihood Ratio: $2 \times$ difference in maximized $\ell$.
   - Maximize $\ell$ over $\beta$
   - Maximize $\ell$ with $\psi = 0$ and $\phi$ unconstrained.
   - degrees of freedom is length of $\psi$

3. Wald Test $(\hat{\psi} - \psi^0)^\top [I^{11}(\hat{\beta})]^{-1} (\hat{\psi} - \psi^0)$
   - Let $\hat{\beta} = (\hat{\psi}, \hat{\phi})$ maximize $\ell$ over $\beta$
   - $\text{Var} \left[ \hat{\beta} \right] \approx I(\hat{\beta}) = [-\ell''(\hat{\beta})]^{-1}$
   - $\text{Var} \left[ \hat{\psi} \right] \approx$ appropriate sub-matrix $I^{11}$
   - You need to invert twice, and select a submatrix in between.

All three have approximate distribution $\sim \chi^2$, degrees of freedom is number of components in $\psi$. 
Tests for some parameters and not others:

1. **Notation:** $\beta = (\psi, \phi)$, $\psi$ of interest.
   - $\phi$ not of interest
2. **Likelihood Ratio:** $2 \times$ difference in maximized $\ell$.
   - Maximize $\ell$ over $\beta$
   - Maximize $\ell$ with $\psi = 0$ and $\phi$ unconstrained.
   - degrees of freedom is length of $\psi$
3. **Wald Test** $(\hat{\psi} - \psi^0)^\top [I^{11}(\hat{\beta})]^{-1} (\hat{\psi} - \psi^0)$
   - Let $\hat{\beta} = (\hat{\psi}, \hat{\phi})$ maximize $\ell$ over $\beta$
   - $\text{Var} \left[ \hat{\beta} \right] \approx I(\hat{\beta}) = [\ell''(\hat{\beta})]^{-1}$
   - $\text{Var} \left[ \hat{\psi} \right] \approx$ appropriate sub-matrix $I^{11}$
   - You need to invert twice, and select a submatrix in between.
4. **Score Test** $\ell^1(0, \tilde{\eta})^\top I^{11}(0, \tilde{\eta}) \ell^1(0, \tilde{\eta})$
Tests for some parameters and not others:

1. Notation: $\beta = (\psi, \phi)$, $\psi$ of interest.
   - $\phi$ not of interest

2. Likelihood Ratio: $2 \times$ difference in maximized $\ell$.
   - Maximize $\ell$ over $\beta$
   - Maximize $\ell$ with $\psi = 0$ and $\phi$ unconstrained.
   - degrees of freedom is length of $\psi$

3. Wald Test $(\hat{\psi} - \psi^0)^\top [I^{11}(\hat{\beta})]^{-1}(\hat{\psi} - \psi^0)$
   - Let $\hat{\beta} = (\hat{\psi}, \hat{\phi})$ maximize $\ell$ over $\beta$
   - $\text{Var} \left[ \hat{\beta} \right] \approx I(\hat{\beta}) = [-\ell''(\hat{\beta})]^{-1}$
   - $\text{Var} \left[ \hat{\psi} \right] \approx$ appropriate sub-matrix $I^{11}$
   - You need to invert twice, and select a submatrix in between.

4. Score Test $\ell^1(0, \tilde{\eta})^\top I^{11}(0, \tilde{\eta}) \ell^1(0, \tilde{\eta})$
   - Let $\tilde{\phi}$ maximize $\ell$ over $\phi$ with $\psi = 0$
Tests for some parameters and not others:

1. Notation: $\beta = (\psi, \phi)$, $\psi$ of interest.
   - $\phi$ not of interest

2. Likelihood Ratio: $2 \times$ difference in maximized $\ell$.
   - Maximize $\ell$ over $\beta$
   - Maximize $\ell$ with $\psi = 0$ and $\phi$ unconstrained.
   - degrees of freedom is length of $\psi$

3. Wald Test $(\hat{\psi} - \psi^0)^\top [I_{11}(\hat{\beta})]^{-1} (\hat{\psi} - \psi^0)$
   - Let $\hat{\beta} = (\hat{\psi}, \hat{\phi})$ maximize $\ell$ over $\beta$
   - $\text{Var} \left[ \hat{\beta} \right] \approx I(\hat{\beta}) = [-\ell''(\hat{\beta})]^{-1}$
   - $\text{Var} \left[ \hat{\psi} \right] \approx$ appropriate sub-matrix $I_{11}$
   - You need to invert twice, and select a submatrix in between.

4. Score Test $\ell'(0, \tilde{\eta})^\top I_{11}(0, \tilde{\eta}) \ell'(0, \tilde{\eta})$
   - Let $\tilde{\phi}$ maximize $\ell$ over $\phi$ with $\psi = 0$
   - Let $\ell'$ be components of $\ell'$ corresponding to $\psi$
Tests for some parameters and not others:

1. Notation: $\beta = (\psi, \phi)$, $\psi$ of interest.
   - $\phi$ not of interest

2. Likelihood Ratio: $2 \times$ difference in maximized $\ell$.
   - Maximize $\ell$ over $\beta$
   - Maximize $\ell$ with $\psi = 0$ and $\phi$ unconstrained.
   - Degrees of freedom is length of $\psi$

3. Wald Test $(\hat{\psi} - \psi^0)^\top [I_{11}(\hat{\beta})]^{-1} (\hat{\psi} - \psi^0)$
   - Let $\hat{\beta} = (\hat{\psi}, \hat{\phi})$ maximize $\ell$ over $\beta$
   - $\text{Var} \left[ \hat{\beta} \right] \approx I(\hat{\beta}) = [-\ell''(\hat{\beta})]^{-1}$
   - $\text{Var} \left[ \hat{\psi} \right] \approx$ appropriate sub-matrix $I_{11}$
   - You need to invert twice, and select a submatrix in between.

4. Score Test $\ell^1(0, \tilde{\eta})^\top I_{11}(0, \tilde{\eta}) \ell^1(0, \tilde{\eta})$
   - Let $\tilde{\phi}$ maximize $\ell$ over $\phi$ with $\psi = 0$
   - Let $\ell^1$ be components of $\ell'$ corresponding to $\psi$

5. All three have approximate distribution $\sim \chi^2$, degrees of freedom is number of components in $\psi$. 

Regression Models: Testing Lecture 06
Most common application: Test one parameter at a time
Most common application: Test one parameter at a time

In this case, $\psi$ has only one component, and is scalar.
Most common application: Test one parameter at a time

1. In this case, $\psi$ has only one component, and is scalar.
2. Generally do inference via statistic before squaring,
Most common application: Test one parameter at a time

1. In this case, $\psi$ has only one component, and is scalar.
2. Generally do inference via statistic before squaring,
A common model nested in a smaller model: Interactions.

1. Allow the effect of one variable to depend on the effect of another.
A common model nested in a smaller model: Interactions.

1. Allow the effect of one variable to depend on the effect of another.
2. For right now, take both of these variables to be factors.
A common model nested in a smaller model: Interactions.

1. Allow the effect of one variable to depend on the effect of another.
2. For right now, take both of these variables to be factors.
3. Consider two factors, with $K$ and $L$ levels.
A common model nested in a smaller model: Interactions.

1. Allow the effect of one variable to depend on the effect of another.
2. For right now, take both of these variables to be factors.
3. Consider two factors, with $K$ and $L$ levels.
4. These divide the observations into $K \times L$ groups.
A common model nested in a smaller model: Interactions.

1. Allow the effect of one variable to depend on the effect of another.
2. For right now, take both of these variables to be factors.
3. Consider two factors, with \( K \) and \( L \) levels.
4. These divide the observations into \( K \times L \) groups
5. Interactions by definition allow a different behavior in each of the \( K \times L \) groups

In the proportional hazards context, allow different hazard ratio, but same baseline hazard.
This problem behaves like a single factor having \( K \times L \) levels.
Again, only \( K \times L - 1 \) are identifiable.
Usually parameterized in a way that makes sense if interactions are deemed unnecessary.
A common model nested in a smaller model: Interactions.

1. Allow the effect of one variable to depend on the effect of another.
2. For right now, take both of these variables to be factors.
3. Consider two factors, with $K$ and $L$ levels.
4. These divide the observations into $K \times L$ groups.
5. Interactions by definition allow a different behavior in each of the $K \times L$ groups.
   
   - In the proportional hazards context, allow different hazard ratio, but same baseline hazard.
A common model nested in a smaller model: Interactions.

1. Allow the effect of one variable to depend on the effect of another.
2. For right now, take both of these variables to be factors.
3. Consider two factors, with $K$ and $L$ levels.
4. These divide the observations into $K \times L$ groups.
5. Interactions by definition allow a different behavior in each of the $K \times L$ groups.
   1. In the proportional hazards context, allow different hazard ratio, but same baseline hazard.
6. This problem behaves like a single factor having $K \times L$ levels.
A common model nested in a smaller model: Interactions.

1. Allow the effect of one variable to depend on the effect of another.
2. For right now, take both of these variables to be factors.
3. Consider two factors, with $K$ and $L$ levels.
4. These divide the observations into $K \times L$ groups
5. Interactions by definition allow a different behavior in each of the $K \times L$ groups
   - In the proportional hazards context, allow different hazard ratio, but same baseline hazard.
6. This problem behaves like a single factor having $K \times L$ levels.
7. Again, only $K \times L - 1$ are identifiable.
A common model nested in a smaller model: Interactions.

1. Allow the effect of one variable to depend on the effect of another.
2. For right now, take both of these variables to be factors.
3. Consider two factors, with $K$ and $L$ levels.
4. These divide the observations into $K \times L$ groups.
5. Interactions by definition allow a different behavior in each of the $K \times L$ groups.
   - In the proportional hazards context, allow different hazard ratio, but same baseline hazard.
6. This problem behaves like a single factor having $K \times L$ levels.
7. Again, only $K \times L - 1$ are identifiable.
8. Usually parameterized in a way that makes sense if interactions are deemed unnecessary.
Parameterized in terms of main effects,
Parameterized in terms of main effects,
defined as the effect when the other variable is at reference level.
Parameterized in terms of main effects,

1. defined as the effect when the other variable is at reference level.
2. Again, no estimate is made for main effect at reference level.

So there are \( K - 1 + L - 1 \) main effects estimated.

Cannot estimate an interaction when either variable is at reference level.

Hence you can only estimate \((K - 1)(L - 1)\) main effects.

Software generally puts estimates it cannot estimate either to missing or omitted.
Parameterized in terms of main effects,

1. defined as the effect when the other variable is at reference level.
2. Again, no estimate is made for main effect at reference level.
3. So there are $K - 1 + L - 1$ main effects estimated.
Parameterized in terms of main effects,

1 defined as the effect when the other variable is at reference level.
2 Again, no estimate is made for main effect at reference level.
3 So there are $K - 1 + L - 1$ main effects estimated.
4 Cannot estimate an interaction when either variable is at reference level.
Continued

1. Parameterized in terms of main effects,
   1. defined as the effect when the other variable is at reference level.
   2. Again, no estimate is made for main effect at reference level.
   3. So there are $K - 1 + L - 1$ main effects estimated.
   4. Cannot estimate an interaction when either variable is at reference level.
   5. Hence you can only estimate $(K - 1)(L - 1)$ main effects.
Parameterized in terms of main effects,

1. defined as the effect when the other variable is at reference level.
2. Again, no estimate is made for main effect at reference level.
3. So there are $K - 1 + L - 1$ main effects estimated.
4. Cannot estimate an interaction when either variable is at reference level.
5. Hence you can only estimate $(K - 1)(L - 1)$ main effects.

Software generally puts estimates it cannot estimate either to
Parameterized in terms of main effects,

1. defined as the effect when the other variable is at reference level.
2. Again, no estimate is made for main effect at reference level.
3. So there are $K - 1 + L - 1$ main effects estimated.
4. Cannot estimate an interaction when either variable is at reference level.
5. Hence you can only estimate $(K - 1)(L - 1)$ main effects.

Software generally puts estimates it cannot estimate either to

1. Missing
Parameterized in terms of main effects,

1. defined as the effect when the other variable is at reference level.
2. Again, no estimate is made for main effect at reference level.
3. So there are $K - 1 + L - 1$ main effects estimated.
4. Cannot estimate an interaction when either variable is at reference level.
5. Hence you can only estimate $(K - 1)(L - 1)$ main effects.

Software generally puts estimates it cannot estimate either to
1. Missing
2. 0
Parameterized in terms of main effects,
- defined as the effect when the other variable is at reference level.
- Again, no estimate is made for main effect at reference level.
- So there are \( K - 1 + L - 1 \) main effects estimated.
- Cannot estimate an interaction when either variable is at reference level.
- Hence you can only estimate \((K - 1)(L - 1)\) main effects.

Software generally puts estimates it cannot estimate either to
- Missing
- 0
- omitted.

R Code  SAS Code
Section: Regression Models

Subsection: Estimation
Estimation is via maximum likelihood

Estimate is most easily interpreted after exponentiating
Estimation is via maximum likelihood

Estimate is most easily interpreted after exponentiating

Called *risk ratio*
Estimation is via maximum likelihood

1. Estimate is most easily interpreted after exponentiating
   - Called *risk ratio*

2. For indicator variable, gives ratio of hazards in two groups
Estimation is via maximum likelihood

1. Estimate is most easily interpreted after exponentiating
   - Called risk ratio

2. For indicator variable, gives ratio of hazards in two groups

3. For continuous variable, gives ratios of hazards for people identical except for the covariate taking values 1 unit apart
Get confidence intervals for $\beta_j$ as $\hat{\beta}_j \pm z_{\alpha/2} \text{SE} \left[ \hat{\beta}_j \right]$,
Get confidence intervals for $\beta_j$ as $\hat{\beta}_j \pm z_{\alpha/2} \text{SE}\left[\hat{\beta}_j\right]$, for $\text{SE}\left[\hat{\beta}_j\right]$ the same as in Wald test.
Confidence intervals

1. Get confidence intervals for $\beta_j$ as $\hat{\beta}_j \pm z_{\alpha/2} \text{SE}[\hat{\beta}_j]$.

   for $\text{SE}[\hat{\beta}_j]$ the same as in Wald test

2. CI for $\exp(\beta_j)$ may be calculated directly or on log scale

   Using delta method, $\text{SE}[\exp(\hat{\beta}_j)] = \exp(\hat{\beta}_j) \text{SE}[\hat{\beta}_j]$ or $\exp(\hat{\beta}_j \pm z_{\alpha/2} \text{SE}[\hat{\beta}_j])$.

   This one is probably better, since it doesn’t run into end of range.
Confidence intervals

1. Get confidence intervals for $\beta_j$ as $\hat{\beta}_j \pm z_{\alpha/2} SE[\hat{\beta}_j]$,
   for $SE[\hat{\beta}_j]$ the same as in Wald test

2. CI for $\exp(\beta_j)$ may be calculated directly or on log scale
   $\exp(\hat{\beta}_j) \pm SE[\exp(\beta_j)] z_{\alpha/2}$

This one is probably better, since it doesn't run into end of range.
1 Get confidence intervals for $\beta_j$ as $\hat{\beta}_j \pm z_{\alpha/2} \text{SE} \left[ \hat{\beta}_j \right]$, for $\text{SE} \left[ \hat{\beta}_j \right]$ the same as in Wald test.

2 CI for $\exp(\beta_j)$ may be calculated directly or on log scale

   - $\exp(\hat{\beta}_j) \pm \text{SE} \left[ \exp(\beta_j) \right] z_{\alpha/2}$
   - Using delta method, $\text{SE} \left[ \exp(\beta_j) \right] = \exp(\hat{\beta}_j) \text{SE} \left[ \beta_j \right]$
Confidence intervals

1. Get confidence intervals for $\beta_j$ as $\hat{\beta}_j \pm z_{\alpha/2}SE[\hat{\beta}_j]$,
   for $SE[\hat{\beta}_j]$ the same as in Wald test

2. CI for $\exp(\beta_j)$ may be calculated directly or on log scale
   1. $\exp(\hat{\beta}_j) \pm SE[\exp(\beta_j)]z_{\alpha/2}$
      1. Using delta method, $SE[\exp(\beta_j)] = \exp(\hat{\beta}_j)SE[\beta_j]$
   2. or $\exp(\hat{\beta}_j \pm z_{\alpha/2}SE[\beta_j])$
Confidence intervals

1. Get confidence intervals for $\beta_j$ as
   $$\hat{\beta}_j \pm z_{\alpha/2} \text{SE}[\hat{\beta}_j],$$
   for $\text{SE}[\hat{\beta}_j]$ the same as in Wald test.

2. CI for $\exp(\beta_j)$ may be calculated directly or on log scale
   - $\exp(\hat{\beta}_j) \pm \text{SE}[\exp(\beta_j)] z_{\alpha/2}$
     - Using delta method, $\text{SE}[\exp(\beta_j)] = \exp(\hat{\beta}_j) \text{SE}[\beta_j]$ or $\exp(\hat{\beta}_j \pm z_{\alpha/2} \text{SE}[\beta_j])$
   - This one is probably better, since it doesn’t run into end of range.
If values of single covariate are in same order as event times, then estimator of associated $\beta$ is $\pm\infty$.
Infinite Estimates

1. If values of single covariate are in same order as event times, then estimator of associated $\beta$ is $\pm \infty$

2. Each term of $\ell'$ for covariate $j$ is $z_{ij} - \frac{\sum_{k \in \mathcal{R}(i)} z_{kj} \exp(z_k \beta))}{\sum_{k \in \mathcal{R}(i)} \exp(z_k \beta))}$
Infinite Estimates

1. If values of single covariate are in same order as event times, then estimator of associated $\beta$ is $\pm\infty$

4. Each term of $\ell'$ for covariate $j$ is $z_{ij} - \frac{\sum_{k \in R(i)} z_{kj} \exp(z_k \beta))}{\sum_{k \in R(i)} \exp(z_k \beta))}$

2. Second part is weighted average of $z_{kj}$
Infinite Estimates

1. If values of single covariate are in same order as event times, then estimator of associated $\beta$ is $\pm\infty$

2. Each term of $\ell'$ for covariate $j$ is $z_{ij} - \frac{\sum_{k \in \mathcal{R}(i)} z_{kj} \exp(z_k \beta)}{\sum_{k \in \mathcal{R}(i)} \exp(z_k \beta)}$

3. Second part is weighted average of $z_{kj}$

4. In order to make weighted average $= z_{ij}$ all weight must be on $z_{ij}$
Infinite Estimates

1. If values of single covariate are in same order as event times, then estimator of associated $\beta$ is $\pm \infty$

2. Each term of $\ell'$ for covariate $j$ is $z_{ij} - \frac{\sum_{k \in R(i)} z_{kj} \exp(z_k \beta))}{\sum_{k \in R(i)} \exp(z_k \beta))}$

3. Second part is weighted average of $z_{kj}$

4. In order to make weighted average $=z_{ij}$ all weight must be on $z_{ij}$

5. If $z_{kj} \leq z_{ij}$ for $k \in R(i)$ then $\beta_k = \infty$ works
Infinite Estimates

1. If values of single covariate are in same order as event times, then estimator of associated $\beta$ is $\pm \infty$

2. Each term of $\ell'$ for covariate $j$ is $z_{ij} - \frac{\sum_{k \in \mathcal{R}(i)} z_{kj} \exp(z_k \beta))}{\sum_{k \in \mathcal{R}(i)} \exp(z_k \beta))}$

3. Second part is weighted average of $z_{kj}$

4. In order to make weighted average $= z_{ij}$ all weight must be on $z_{ij}$

5. If $z_{kj} \leq z_{ij}$ for $k \in \mathcal{R}(i)$ then $\beta_k = \infty$ works

2. Algorithm can’t converge in standard sense.
Infinite Estimates

1. If values of single covariate are in same order as event times, then estimator of associated $\beta$ is $\pm \infty$

   1. Each term of $\ell'$ for covariate $j$ is $z_{ij} - \frac{\sum_{k \in \mathcal{R}(i)} z_{kj} \exp(z_k \beta)}{\sum_{k \in \mathcal{R}(i)} \exp(z_k \beta)}$

   2. Second part is weighted average of $z_{kj}$

   3. In order to make weighted average $= z_{ij}$ all weight must be on $z_{ij}$

   4. If $z_{kj} \leq z_{ij}$ for $k \in \mathcal{R}(i)$ then $\beta_k = \infty$ works

2. Algorithm can’t converge in standard sense.

3. Diagnose from convergence behavior.
Infinite Estimates

1. If values of single covariate are in same order as event times, then estimator of associated $\beta$ is $\pm \infty$
   
   1. Each term of $\ell'$ for covariate $j$ is $z_{ij} - \frac{\sum_{k \in R(i)} z_{kj} \exp(z_k \beta)}{\sum_{k \in R(i)} \exp(z_k \beta)}$
   
   2. Second part is weighted average of $z_{kj}$
   
   3. In order to make weighted average $= z_{ij}$ all weight must be on $z_{ij}$
   
   4. If $z_{kj} \leq z_{ij}$ for $k \in R(i)$ then $\beta_k = \infty$ works

2. Algorithm can’t converge in standard sense.

3. Diagnose from convergence behavior.
   
   1. Warning message says algorithm hasn’t converged.

diagram
Infinite Estimates

1. If values of single covariate are in same order as event times, then estimator of associated $\beta$ is $\pm \infty$

   1. Each term of $\ell'$ for covariate $j$ is $z_{ij} - \frac{\sum_{k \in R(i)} z_{kj} \exp(z_k \beta)}{\sum_{k \in R(i)} \exp(z_k \beta)}$

   2. Second part is weighted average of $z_{kj}$

   3. In order to make weighted average $= z_{ij}$ all weight must be on $z_{ij}$

   4. If $z_{kj} \leq z_{ij}$ for $k \in R(i)$ then $\beta_k = \infty$ works

2. Algorithm can’t converge in standard sense.

3. Diagnose from convergence behavior.

   1. Warning message says algorithm hasn’t converged.

   2. Or Very large parameter estimates.
Infinite Estimates

1. If values of single covariate are in same order as event times, then estimator of associated $\beta$ is $\pm \infty$

   1. Each term of $\ell'$ for covariate $j$ is $z_{ij} - \frac{\sum_{k \in R(i)} z_{kj} \exp(z_k \beta)}{\sum_{k \in R(i)} \exp(z_k \beta)}$

   2. Second part is weighted average of $z_{kj}$

   3. In order to make weighted average $= z_{ij}$ all weight must be on $z_{ij}$

   4. If $z_{kj} \leq z_{ij}$ for $k \in R(i)$ then $\beta_k = \infty$ works

2. Algorithm can’t converge in standard sense.

3. Diagnose from convergence behavior.

   1. Warning message says algorithm hasn’t converged.

   2. Or Very large parameter estimates.

4. Also can happen with linear combination of $\beta$
Infinite Estimates

1. If values of single covariate are in same order as event times, then estimator of associated $\beta$ is $\pm \infty$

   1. Each term of $\ell'$ for covariate $j$ is $z_{ij} - \frac{\sum_{k \in \mathcal{R}(i)} z_{kj} \exp(z_k \beta)}{\sum_{k \in \mathcal{R}(i)} \exp(z_k \beta)}$

   2. Second part is weighted average of $z_{kj}$

   3. In order to make weighted average $= z_{ij}$ all weight must be on $z_{ij}$

   4. If $z_{kj} \leq z_{ij}$ for $k \in \mathcal{R}(i)$ then $\beta_k = \infty$ works

2. Algorithm can’t converge in standard sense.

3. Diagnose from convergence behavior.

   1. Warning message says algorithm hasn’t converged.

   2. Or Very large parameter estimates.

4. Also can happen with linear combination of $\beta$

   1. Harder to see by looking at data  
   R Code  
   SAS Code
Objectives Lecture 07

1. Model Building

Readings: KM §8.7, 8.8, 9.4
Objectives Lecture 07

1. Model Building
2. Estimating the baseline survival function
Objectives Lecture 07

1. Model Building
2. Estimating the baseline survival function
3. Late Entry
Objectives Lecture 07

1. Model Building
2. Estimating the baseline survival function
3. Late Entry
4. Readings: KM §8.7, 8.8, 9.4
Recall: Regression model has unknown baseline hazard, unknown parameters

Bayesian Paradigm:
1. Treat unknown quantities as random
2. Put prior distribution on these
3. Calculate distribution conditional on data: posterior

Partial likelihood approach: remove baseline hazard via profiling.

We continue this here.

Alternatively, can put a prior on function space.

Let $\varpi(\beta)$ be prior density on parameter space.
Recall: Regression model has unknown baseline hazard, unknown parameters

Bayesian Paradigm:
1. Treat unknown quantities as random
2. Put prior distribution on these
3. Calculate distribution conditional on data: posterior

Partial likelihood approach: remove baseline hazard via profiling. We continue this here. Alternatively, can put a prior on function space.

Let $\varpi(\beta)$ be prior density on parameter space.
Bayesian Inference and Regularization

1. Recall: Regression model has unknown baseline hazard, unknown parameters
2. Bayesian Paradigm:
   1. Treat unknown quantities as random
Recall: Regression model has unknown baseline hazard, unknown parameters

Baysian Paradigm:
1. Treat unknown quantities as random
2. Put prior distribution on these

Alternatively, can put a prior on function space.

Let $\varpi(\beta)$ be prior density on parameter space.
Bayesian Inference and Regularization

1. Recall: Regression model has unknown baseline hazard, unknown parameters

2. Bayesian Paradigm:
   1. Treat unknown quantities as random
   2. Put prior distribution on these
   3. Calculate distribution conditional on data: posterior
Recall: Regression model has unknown baseline hazard, unknown parameters

Baysian Paradigm:
1. Treat unknown quantities as random
2. Put prior distribution on these
3. Calculate distribution conditional on data: posterior

Partial likelihood approach: remove baseline hazard via profiling.
Recall: Regression model has unknown baseline hazard, unknown parameters

Baysian Paradigm:
1. Treat unknown quantities as random
2. Put prior distribution on these
3. Calculate distribution conditional on data: posterior

Partial likelihood approach: remove baseline hazard via profiling.
1. We continue this here.
Recall: Regression model has unknown baseline hazard, unknown parameters

Baysian Paradigm:
1. Treat unknown quantities as random
2. Put prior distribution on these
3. Calculate distribution conditional on data: posterior

Partial likelihood approach: remove baseline hazard via profiling.
1. We continue this here.
2. Alternatively, can put a prior on function space.
1. Recall: Regression model has unknown baseline hazard, unknown parameters

2. Bayesian Paradigm:
   1. Treat unknown quantities as random
   2. Put prior distribution on these
   3. Calculate distribution conditional on data: posterior

3. Partial likelihood approach: remove baseline hazard via profiling.
   1. We continue this here.
   2. Alternatively, can put a prior on function space.
   3. Let \( \varpi(\beta) \) be prior density on parameter space.
The posterior as a regularization method.

Posterior is \( \propto L(\beta) \varpi(\beta) \)

Parameter estimate maximizes posterior.

If \( \lim_{\|\beta\| \to \infty} \varpi(\beta) = 0 \), then the posterior does not have the monotonicity problem that we saw could arise in frequentist approach.

On log scale, log (partial) posterior is \( \ell(\beta) - \varsigma(\beta) + C \) for \( \varsigma(\beta) = -\log(\varpi(\beta)) \)

Equivalent to frequentist technique of regularization

\( \varsigma(\beta) = \lambda \sum_j |\beta_j|^2 \) if \( \beta \) independent \( N(0, 1/\lambda) \)

\( \varsigma(\beta) = \lambda \sum_j |\beta_j| \) if \( \beta \) independent Laplace with scale \( 1/\sqrt{\lambda} \)

We investigated both of these regularizations for multiple regression

\( \ell(\beta) \propto \sum_j (Y_j - \beta z_j)^2 \)

Most common procedure is to use Jeffreys prior. R Code SAS Code
The posterior as a regularization method.

1. Posterior is $\propto L(\beta) \varpi(\beta)$
   - Parameter estimate maximizes posterior.
The posterior as a regularization method.

1. Posterior is $\propto L(\beta) \varpi(\beta)$
   1. Parameter estimate maximizes posterior.
   2. If $\lim_{\|\beta\| \to \infty} \varpi(\beta) = 0$, then the posterior does not have the monotonicity problem that we saw could arise in frequentist approach.
The posterior as a regularization method.

1. Posterior is $\propto L(\beta) \varpi(\beta)$
   - Parameter estimate maximizes posterior.
   - If $\lim_{\|\beta\| \to \infty} \varpi(\beta) = 0$, then the posterior does not have the
     monotonicity problem that we saw could arise in frequentist approach.

2. On log scale, log (partial) posterior is $\ell(\beta) - \varsigma(\beta) + C$ for $\varsigma(\beta) = -\log(\varpi(\beta))$
Posterior is \( \propto L(\beta)\varpi(\beta) \)

1. Parameter estimate maximizes posterior.
2. If \( \lim_{\|\beta\| \to \infty} \varpi(\beta) = 0 \), then the posterior does not have the monotoncity problem that we saw could arise in frequentist approach.

On log scale, log (partial) posterior is \( \ell(\beta) - \varsigma(\beta) + C \) for \( \varsigma(\beta) = -\log(\varpi(\beta)) \)

1. Equivalent to frequentist technique of regularization.
The posterior as a regularization method.

1. Posterior is $\propto L(\beta) \varpi(\beta)$
   1. Parameter estimate maximizes posterior.
   2. If $\lim_{\|\beta\| \to \infty} \varpi(\beta) = 0$, then the posterior does not have the monotonicity problem that we saw could arise in frequentist approach.

2. On log scale, log (partial) posterior is $\ell(\beta) - \varsigma(\beta) + C$ for $\varsigma(\beta) = -\log(\varpi(\beta))$
   1. Equivalent to frequentist technique of regularization
   2. $\varsigma(\beta) = \lambda \sum_j |\beta_j|^2$ if $\beta$ independent $N(0, 1/\lambda)$
The posterior as a regularization method.

1. Posterior is  \( \propto L(\beta) \varpi(\beta) \)
   1. Parameter estimate maximizes posterior.
   2. If \( \lim_{\|\beta\| \to \infty} \varpi(\beta) = 0 \), then the posterior does not have the monotonicity problem that we saw could arise in frequentist approach.

2. On log scale, log (partial) posterior is \( \ell(\beta) - \varsigma(\beta) + C \) for \( \varsigma(\beta) = -\log(\varpi(\beta)) \)
   1. Equivalent to frequentist technique of regularization
   2. \( \varsigma(\beta) = \lambda \sum_j |\beta_j|^2 \) if \( \beta \) independent \( N(0, 1/\lambda) \)
   3. \( \varsigma(\beta) = \lambda \sum_j |\beta_j| \) if \( \beta \) independent Laplace with scale \( 1/\sqrt{\lambda} \)

Regression Models: Estimation Lecture 07
The posterior as a regularization method.

1. Posterior is $\propto L(\beta) \varpi(\beta)$
   - Parameter estimate maximizes posterior.
   - If $\lim_{\|\beta\| \to \infty} \varpi(\beta) = 0$, then the posterior does not have the monotonicity problem that we saw could arise in frequentist approach.

2. On log scale, log (partial) posterior is $\ell(\beta) - \varsigma(\beta) + C$ for $\varsigma(\beta) = -\log(\varpi(\beta))$
   - Equivalent to frequentist technique of regularization
   - $\varsigma(\beta) = \lambda \sum_j |\beta_j|^2$ if $\beta$ independent $N(0, 1/\lambda)$
   - $\varsigma(\beta) = \lambda \sum_j |\beta_j|$ if $\beta$ independent Laplace with scale $1/\sqrt{\lambda}$
   - We investigated both of these regularizations for multiple regression
The posterior as a regularization method.

1. Posterior is \( \propto L(\beta) \varpi(\beta) \)
   - Parameter estimate maximizes posterior.
   - If \( \lim_{\|\beta\| \to \infty} \varpi(\beta) = 0 \), then the posterior does not have the monotonicity problem that we saw could arise in frequentist approach.

2. On log scale, log (partial) posterior is \( \ell(\beta) - \varsigma(\beta) + C \) for \( \varsigma(\beta) = -\log(\varpi(\beta)) \)
   - Equivalent to frequentist technique of regularization
   - \( \varsigma(\beta) = \lambda \sum_j |\beta_j|^2 \) if \( \beta \) independent \( N(0, 1/\lambda) \)
   - \( \varsigma(\beta) = \lambda \sum_j |\beta_j| \) if \( \beta \) independent Laplace with scale \( 1/\sqrt{\lambda} \)
   - We investigated both of these regularizations for multiple regression
     - \( \ell(\beta) \propto \sum_j (Y_j - \beta z_j)^2 \)
The posterior as a regularization method.

1. Posterior is $\propto L(\beta) \varpi(\beta)$
   - Parameter estimate maximizes posterior.
   - If $\lim_{\|\beta\| \to \infty} \varpi(\beta) = 0$, then the posterior does not have the monotonicity problem that we saw could arise in frequentist approach.

2. On log scale, log (partial) posterior is $\ell(\beta) - \varsigma(\beta) + C$ for $\varsigma(\beta) = -\log(\varpi(\beta))$
   - Equivalent to frequentist technique of regularization
   - $\varsigma(\beta) = \lambda \sum_j |\beta_j|^2$ if $\beta$ independent $N(0, 1/\lambda)$
   - $\varsigma(\beta) = \lambda \sum_j |\beta_j|$ if $\beta$ independent Laplace with scale $1/\sqrt{\lambda}$
   - We investigated both of these regularizations for multiple regression
     - $\ell(\beta) \propto \sum_j (Y_j - \beta z_j)^2$

5. Most common procedure is to use Jeffreys prior. R Code  SAS Code
Same regression techniques, constraints

1. Measure quality: AIC, $p$-value.
Same regression techniques, constraints

1. Measure quality: AIC, $p$-value.
   - using Akaike’s Information criterion: $-2\ell + 2p$
Same regression techniques, constraints

1. Measure quality: AIC, \( p \)-value.
   - using *Akaike’s Information criterion*: \(-2\ell + 2p\)
   - for \( p \) the number of parameters
Same regression techniques, constraints

1 Measure quality: AIC, $p$-value.
   1 using Akaike’s Information criterion: $-2\ell + 2p$
     1 for $p$ the number of parameters
     2 Lower is better
Same regression techniques, constraints

1. Measure quality: AIC, $p$-value.
   - using *Akaike’s Information criterion*: $-2\ell + 2p$
     - for $p$ the number of parameters
     - Lower is better
2. Using test $p$-value

Search through models using stepwise:
1. Start with an initial model.
2. Consider models with separate (groups of) parameters added or removed, one at a time.
3. Try nonlinear terms, interactions, etc.
4. Dichotomize continuous variables.
5. Move to model with numerical criteria improved.
7. At each step, one can add or remove variables.
8. Only considering additions: Forward stepwise.
Same regression techniques, constraints

1. Measure quality: AIC, $p$-value.
   - using *Akaike’s Information criterion*: $-2\ell + 2p$
   - for $p$ the number of parameters
   - Lower is better
2. Using test $p$-value
   - Typically set significance higher: 0.15?
Same regression techniques, constraints

1. Measure quality: AIC, $p$-value.
   1. using *Akaike’s Information criterion*: $-2\ell + 2p$
      1. for $p$ the number of parameters
      2. Lower is better
   2. Using test $p$-value
      1. Typically set significance higher: 0.15?
      2. To ensure stability, put level for removal higher than that of inclusion.
Same regression techniques, constraints

1. Measure quality: AIC, $p$-value.
   - using Akaike’s *Information criterion*: $-2\ell + 2p$
     - for $p$ the number of parameters
     - Lower is better
   2. Using test $p$-value
     - Typically set significance higher: 0.15?
     - To ensure stability, put level for removal higher than that of inclusion.

2. Search through models using stepwise:
   - Only considering additions: Forward stepwise.
   - Only considering deletions: Backwards stepwise.
Same regression techniques, constraints

1 Measure quality: AIC, $p$-value.
   - using Akaike’s Information criterion: $-2\ell + 2p$
     - for $p$ the number of parameters
     - Lower is better
   - Using test $p$-value
     - Typically set significance higher: 0.15?
     - To ensure stability, put level for removal higher than that of inclusion.

2 Search through models using stepwise:
   - Start with an initial model.

   - Consider models with separate (groups of) parameters added or removed, one at a time.
   - Try nonlinear terms, interactions, etc
   - Dichotomize continuous variables.
   - Move to model with numerical criteria improved.

   - Only considering additions: Forward stepwise.
   - Only considering deletions: Backwards stepwise.
Same regression techniques, constraints

1. Measure quality: AIC, $p$-value.
   1. using Akaike’s Information criterion: $-2\ell + 2p$
      1. for $p$ the number of parameters
      2. Lower is better
   2. Using test $p$-value
      1. Typically set significance higher: 0.15?
      2. To ensure stability, put level for removal higher than that of inclusion.

2. Search through models using stepwise:
   1. Start with an initial model.
   2. Consider models with separate (groups of) parameters added or removed, one at a time.
Same regression techniques, constraints

1. Measure quality: AIC, \( p \)-value.
   - using *Akaike’s Information criterion*: \(-2\ell + 2p\)
     - for \( p \) the number of parameters
     - Lower is better
   - Using test \( p \)-value
     - Typically set significance higher: 0.15?
     - To ensure stability, put level for removal higher than that of inclusion.

2. Search through models using stepwise:
   - Start with an initial model.
   - Consider models with separate (groups of) parameters added or removed, one at a time.
   - Try nonlinear terms, interactions, etc.
Same regression techniques, constraints

1. Measure quality: AIC, $p$-value.
   - using Akaike’s Information criterion: $-2\ell + 2p$
     - for $p$ the number of parameters
     - Lower is better
   - Using test $p$-value
     - Typically set significance higher: 0.15?
     - To ensure stability, put level for removal higher than that of inclusion.

2. Search through models using stepwise:
   - Start with an initial model.
   - Consider models with separate (groups of) parameters added or removed, one at a time.
   - Try nonlinear terms, interactions, etc
   - Dichotomize continuous variables.
Same regression techniques, constraints

1. Measure quality: AIC, $p$-value.
   - using Akaike’s Information criterion: $-2\ell + 2p$
     - for $p$ the number of parameters
     - Lower is better
   2. Using test $p$-value
     - Typically set significance higher: 0.15?
     - To ensure stability, put level for removal higher than that of inclusion.

2. Search through models using stepwise:
   - Start with an initial model.
   - Consider models with separate (groups of) parameters added or removed, one at a time.
   - Try nonlinear terms, interactions, etc
   - Dichotomize continuous variables.
   - Move to model with numerical criteria improved.
Same regression techniques, constraints

1 Measure quality: AIC, \( p \)-value.
   1 using Akaike's Information criterion: \(-2\ell + 2p\)
     1 for \( p \) the number of parameters
     2 Lower is better
   2 Using test \( p \)-value
     1 Typically set significance higher: 0.15?
     2 To ensure stability, put level for removal higher than that of inclusion.

2 Search through models using stepwise:
   1 Start with an initial model.
   2 Consider models with separate (groups of) parameters added or removed, one at a time.
   3 Try nonlinear terms, interactions, etc
   4 Dichotomize continuous variables.
   5 Move to model with numerical criteria improved.
   6 Gives a local, rather than guaranteed global, optimum.
Same regression techniques, constraints

1. Measure quality: AIC, $p$-value.
   1. using Akaike’s Information criterion: $-2\ell + 2p$
   2. for $p$ the number of parameters
   3. Lower is better

2. Using test $p$-value
   1. Typically set significance higher: 0.15?
   2. To ensure stability, put level for removal higher than that of inclusion.

2. Search through models using stepwise:
   1. Start with an initial model.
   2. Consider models with separate (groups of) parameters added or removed, one at a time.
   3. Try nonlinear terms, interactions, etc
   4. Dichotomize continuous variables.
   5. Move to model with numerical criteria improved.
   7. At each step, one can add or remove variables.
Same regression techniques, constraints

1 Measure quality: AIC, $p$-value.
   - using Akaike's Information criterion: $-2\ell + 2p$
     - for $p$ the number of parameters
     - Lower is better
   - Using test $p$-value
     - Typically set significance higher: 0.15?
     - To ensure stability, put level for removal higher than that of inclusion.

2 Search through models using stepwise:
   - Start with an initial model.
   - Consider models with separate (groups of) parameters added or removed, one at a time.
   - Try nonlinear terms, interactions, etc
   - Dichotomize continuous variables.
   - Move to model with numerical criteria improved.
   - Gives a local, rather than guaranteed global, optimum.
   - At each step, one can add or remove variables.
     - Only considering additions: Forward stepwise.
Same regression techniques, constraints

1. Measure quality: AIC, $p$-value.
   - 1. using Akaike’s Information criterion: $-2\ell + 2p$
     - 2. for $p$ the number of parameters
     - 3. Lower is better
   - 2. Using test $p$-value
     - 1. Typically set significance higher: 0.15?
     - 2. To ensure stability, put level for removal higher than that of inclusion.

2. Search through models using stepwise:
   - 1. Start with an initial model.
   - 2. Consider models with separate (groups of) parameters added or removed, one at a time.
   - 3. Try nonlinear terms, interactions, etc
   - 4. Dichotomize continuous variables.
   - 5. Move to model with numerical criteria improved.
   - 6. Gives a local, rather than guaranteed global, optimum.
   - 7. At each step, one can add or remove variables.
     - 1. Only considering additions: Forward stepwise.
     - 2. Only considering deletions: Backwards stepwise.
Interpretation after selection:

1. Model parameters measure effect of explanatory variable in light of all other variables in model.
Interpretation after selection:

1. Model parameters measure effect of explanatory variable in light of all other variables in model.
2. Hence interpretation of parameter changes as other variables move in and out of the model.
Interpretation after selection:

1. Model parameters measure effect of explanatory variable in light of all other variables in model.

2. Hence interpretation of parameter changes as other variables move in and out of the model.

3. Inference after selection has multiple-comparisons issue.
Interpretation after selection:

1. Model parameters measure effect of explanatory variable in light of all other variables in model.
2. Hence interpretation of parameter changes as other variables move in and out of the model.
3. Inference after selection has multiple-comparisons issue.
   - Effect of variables in a best-fitting model will be exaggerated relative to a model selected a priori.
Interpretation after selection:

1. Model parameters measure effect of explanatory variable in light of all other variables in model.
2. Hence interpretation of parameter changes as other variables move in and out of the model.
3. Inference after selection has multiple-comparisons issue.
   1. Effect of variables in a best-fitting model will be exaggerated relative to a model selected a priori.
   2. One must adjust for this exaggerated effect.
Interpretation after selection:

1. Model parameters measure effect of explanatory variable in light of all other variables in model.

2. Hence interpretation of parameter changes as other variables move in and out of the model.

3. Inference after selection has multiple-comparisons issue.
   1. Effect of variables in a best-fitting model will be exaggerated relative to a model selected a priori.
   2. One must adjust for this exaggerated effect.
   3. Solutions:
Interpretation after selection:

1. Model parameters measure effect of explanatory variable in light of all other variables in model.

2. Hence interpretation of parameter changes as other variables move in and out of the model.

3. Inference after selection has multiple-comparisons issue.
   1. Effect of variables in a best-fitting model will be exaggerated relative to a model selected a priori.
   2. One must adjust for this exaggerated effect.
   3. Solutions:
      1. test and training set.
Interpretation after selection:

1. Model parameters measure effect of explanatory variable in light of all other variables in model.
2. Hence interpretation of parameter changes as other variables move in and out of the model.
3. Inference after selection has multiple-comparisons issue.
   1. Effect of variables in a best-fitting model will be exaggerated relative to a model selected a priori.
   2. One must adjust for this exaggerated effect.
   3. Solutions:
      1. test and training set.
      2. build model without explanatory variable for primary hypothesis.
Interpretation after selection:

1. Model parameters measure effect of explanatory variable in light of all other variables in model.
2. Hence interpretation of parameter changes as other variables move in and out of the model.
3. Inference after selection has multiple-comparisons issue.
   1. Effect of variables in a best-fitting model will be exaggerated relative to a model selected a priori.
   2. One must adjust for this exaggerated effect.
   3. Solutions:
      1. test and training set.
      2. build model without explanatory variable for primary hypothesis.
4. Model selection is impacted by coordinate system for variables.
Interpretation after selection:

1. Model parameters measure effect of explanatory variable in light of all other variables in model.

2. Hence interpretation of parameter changes as other variables move in and out of the model.

3. Inference after selection has multiple-comparisons issue.
   - Effect of variables in a best-fitting model will be exaggerated relative to a model selected a priori.
   - One must adjust for this exaggerated effect.

   Solutions:
   - test and training set.
   - build model without explanatory variable for primary hypothesis.

4. Model selection is impacted by coordinate system for variables.
   - Ex., a model containing baseline value and a change from baseline will be treated differently from a model containing baseline and later value.
Interpretation after selection:

1. Model parameters measure effect of explanatory variable in light of all other variables in model.

2. Hence interpretation of parameter changes as other variables move in and out of the model.

3. Inference after selection has multiple-comparisons issue.
   
   1. Effect of variables in a best-fitting model will be exaggerated relative to a model selected a priori.
   
   2. One must adjust for this exaggerated effect.
   
   3. Solutions:
      
      1. test and training set.
      
      2. build model without explanatory variable for primary hypothesis.

4. Model selection is impacted by coordinate system for variables.
   
   1. Ex., a model containing baseline value and a change from baseline will be treated differently from a model containing baseline and later value.

5. Same warnings and rules for including variables
Interpretation after selection:

1. Model parameters measure effect of explanatory variable in light of all other variables in model.

2. Hence interpretation of parameter changes as other variables move in and out of the model.

3. Inference after selection has multiple-comparisons issue.
   - Effect of variables in a best-fitting model will be exaggerated relative to a model selected a priori.
   - One must adjust for this exaggerated effect.
   - Solutions:
     1. test and training set.
     2. build model without explanatory variable for primary hypothesis.

4. Model selection is impacted by coordinate system for variables.
   - Ex., a model containing baseline value and a change from baseline will be treated differently from a model containing baseline and later value.

5. Same warnings and rules for including variables
   - No interactions without main effects
**Interpretation after selection:**

1. Model parameters measure effect of explanatory variable in light of all other variables in model.
2. Hence interpretation of parameter changes as other variables move in and out of the model.
3. Inference after selection has multiple-comparisons issue.
   1. Effect of variables in a best-fitting model will be exaggerated relative to a model selected a priori.
   2. One must adjust for this exaggerated effect.
   3. Solutions:
      1. Test and training set.
      2. Build model without explanatory variable for primary hypothesis.
4. Model selection is impacted by coordinate system for variables.
   1. Ex., a model containing baseline value and a change from baseline will be treated differently from a model containing baseline and later value.
5. Same warnings and rules for including variables
   1. No interactions without main effects
   2. Watch for multiple comparisons

R Code  SAS Code
Section: Regression Models

Subsection: Estimating nuisance baseline survival function
Before was treated as nuisance parameter
Introduction

1. Before was treated as nuisance parameter
2. Now might be of interest
Introduction

1. Before was treated as nuisance parameter
2. Now might be of interest
   1. Corresponds to person with $z = 0$
Before was treated as nuisance parameter

Now might be of interest

1. Corresponds to person with $z = 0$
2. With suitable redefining, can refer to any fixed $z$
Before was treated as nuisance parameter

Now might be of interest

1. Corresponds to person with $z = 0$
2. With suitable redefining, can refer to any fixed $z$

We will consider only no-ties case
Order event times

Estimate via Cumulative Hazard

1. Order event times

Estimator of baseline hazard at event time \( k \) is

\[
D_k \sum_{j \in R} (k_j) \exp(z_j \beta)
\]

\( \hat{S}(t_i) = \exp\left(-\sum_{k=1}^{d_k} \sum_{j \in R} (i_j) \exp(z_j \beta)\right) \)

With no covariates this corresponds to exponentiated Nelson–Aalen estimator

Can estimate \( S \) at arbitrary \( z \) by \( \hat{S}(t) \exp(z \beta) \)

If \( \beta \) known, can calculate SE just as for Kaplan–Meier

Must be increased for having to estimate \( \beta \)
Estimate via Cumulative Hazard

1. Order event times

2. Estimator of baseline hazard at event time $k$ is

$$\frac{D_k}{\sum_{j \in R(k)} \exp(z_j \beta)}$$
Estimate via Cumulative Hazard

1. Order event times
2. Estimator of baseline hazard at event time \( k \) is
   \[
   D_k \sum_{j \in \mathcal{R}(k)} \exp(z_j \beta)
   \]
3. \( \hat{S}(t_i) = \exp \left( - \sum_{k=1}^{i} \frac{d_k}{\sum_{j \in \mathcal{R}(i)} \exp(z_j \beta)} \right) \)

With no covariates this corresponds to exponentiated Nelson–Aalen estimator. Can estimate \( S \) at arbitrary \( z \) by \( \hat{S}(t_i) \exp(z \beta) \). If \( \beta \) known, can calculate SE just as for Kaplan–Meier.
1 Order event times

2 Estimator of baseline hazard at event time $k$ is

$$D_k \frac{\sum_{j \in \mathcal{R}(k)} \exp(z_j \beta)}{\sum_{j \in \mathcal{R}(i)} \exp(z_j \beta)}$$

3 $\hat{S}(t_i) = \exp \left( - \sum_{k=1}^i \frac{d_k}{\sum_{j \in \mathcal{R}(i)} \exp(z_j \beta)} \right)$

   1 With no covariates this corresponds to exponentiated Nelson–Aalen estimator
Estimate via Cumulative Hazard

1. Order event times
2. Estimator of baseline hazard at event time $k$ is

$$
D_k \frac{\sum_{j \in R(k)} \exp(z_j \beta)}{\sum_{j \in R} \exp(z_j \beta)}
$$

3. $\hat{S}(t_i) = \exp \left( - \sum_{k=1}^{i} \frac{d_k}{\sum_{j \in R(i)} \exp(z_j \beta)} \right)$
   - With no covariates this corresponds to exponentiated Nelson–Aalen estimator
4. Can estimate $S$ at arbitrary $z$ by $\hat{S}(t)^{\exp(z \beta)}$
1. Order event times
2. Estimator of baseline hazard at event time $k$ is

$$D_k \frac{\sum_{j \in R(k)} \exp(z_j \beta)}{\sum_{j \in R(i)} \exp(z_j \beta)}$$

3. $\hat{S}(t_i) = \exp \left( - \sum_{k=1}^i \frac{d_k}{\sum_{j \in R(i)} \exp(z_j \beta)} \right)$

   1. With no covariates this corresponds to exponentiated Nelson–Aalen estimator

4. Can estimate $S$ at arbitrary $z$ by $\hat{S}(t)^{exp(z\beta)}$

5. If $\beta$ known, can calculate SE just as for Kaplan–Meier
Estimate via Cumulative Hazard

1. Order event times
2. Estimator of baseline hazard at event time $k$ is
   \[
   D_k = \frac{\sum_{j \in R(k)} \exp(z_j \beta)}{\sum_{i \in R(i)} \exp(z_i \beta)}
   \]
3. $\hat{S}(t_i) = \exp \left( - \sum_{k=1}^{i} \frac{d_k}{\sum_{j \in R(i)} \exp(z_j \beta)} \right)$
   1. With no covariates this corresponds to exponentiated Nelson–Aalen estimator
4. Can estimate $S$ at arbitrary $z$ by $\hat{S}(t)^{\exp(z \beta)}$
5. If $\beta$ known, can calculate SE just as for Kaplan–Meier
6. Must be increased for having to estimate $\beta$
Alternate estimator:

First estimate survival function $\hat{S}(t_i) = \prod_{k=1}^{i} \left(1 - \frac{d_k}{\sum_{j \in R(i)} \exp(z_j \beta)}\right)$
Alternate estimator:

1. First estimate survival function \( \hat{S}(t_i) = \prod_{k=1}^{i} \left(1 - \frac{d_k}{\sum_{j \in R(i)} \exp(z_j \beta)} \right) \)

1. Made by substituting \( \exp(-x) \approx 1 - x \)
Alternate estimator:

1. First estimate survival function \( \hat{S}(t) = \prod_{k=1}^{i} \left( 1 - \frac{d_k}{\sum_{j \in R(i)} \exp(z_j \beta)} \right) \)

   1. Made by substituting \( \exp(-x) \approx 1 - x \)
   2. Weighted Kaplan–Meier curve
Alternate estimator:

First estimate survival function \( \hat{S}(t_i) = \prod_{k=1}^{i} \left( 1 - \frac{d_k}{\sum_{j \in R(i)} \exp(z_j \beta)} \right) \)

1. Made by substituting \( \exp(-x) \approx 1 - x \)
2. Weighted Kaplan–Meier curve
3. Adjusted for ties if necessary
Alternate estimator:

First estimate survival function \( \hat{S}(t_i) = \prod_{k=1}^{i} \left( 1 - \frac{d_k}{\sum_{j \in R(i)} \exp(z_j \beta)} \right) \)

1. Made by substituting \( \exp(-x) \approx 1 - x \)
2. Weighted Kaplan–Meier curve
3. Adjusted for ties if necessary
4. Both estimators have expressions for standard error

SAS Code

R Code
Section: Regression Models

Subsection: Late Entry
Subject not observed at beginning of study

1. As with Kaplan–Meier, get data set representing survival conditional on having entered

\[ L(\beta) = \exp(\beta) \exp(\beta) + \exp(0) \exp(0) \times \exp(0) \exp(0) + \exp(0) + \exp(\beta) \]

SAS Code

R Code
Subject not observed at beginning of study

1. As with Kaplan–Meier, get data set representing survival conditional on having entered

2. As before, adjust risk set to only contain those who have already entered
Subject not observed at beginning of study

1. As with Kaplan–Meier, get data set representing survival conditional on having entered
2. As before, adjust risk set to only contain those who have already entered
3. Requires entry time to be independent of life time.
Subject not observed at beginning of study

1. As with Kaplan–Meier, get data set representing survival conditional on having entered
2. As before, adjust risk set to only contain those who have already entered
3. Requires entry time to be independent of life time.
4. Treatment is different from adding entry time as covariate
Subject not observed at beginning of study

1. As with Kaplan–Meier, get data set representing survival conditional on having entered
2. As before, adjust risk set to only contain those who have already entered
3. Requires entry time to be independent of life time.
4. Treatment is different from adding entry time as covariate
   
   See Fig. 13.

Fig. 13: Delayed Entry

---

![Diagram showing delayed entry with event and censored observations](image-url)
Subject not observed at beginning of study

1. As with Kaplan–Meier, get data set representing survival conditional on having entered.
2. As before, adjust risk set to only contain those who have already entered.
3. Requires entry time to be independent of life time.
4. Treatment is different from adding entry time as covariate
   - See Fig. 13.

Fig. 13: Delayed Entry

Partial likelihood treats them identically:

\[
L(\beta) = \frac{\exp(\beta)}{\exp(\beta)+\exp(0)+\exp(\beta)+\exp(0)} \times \frac{\exp(0)}{\exp(0)+\exp(\beta)+\exp(0)+\exp(0)} \times \frac{\exp(\beta)}{\exp(\beta)}
\]
Subject removed and returned leaves partial likelihood unchanged

1 Subject now has two lines
Subject removed and returned leaves partial likelihood unchanged

1. Subject now has two lines
   1. First entry censored
Subject removed and returned leaves partial likelihood unchanged

1. Subject now has two lines
   1. First entry censored
   2. Second entry late.
Subject removed and returned leaves partial likelihood unchanged

1. Subject now has two lines
   1. First entry censored
   2. Second entry late.

2. Each risk set under initial structure containing subject now has either first copy or second copy.
Subject removed and returned leaves partial likelihood unchanged

1. Subject now has two lines
   1. First entry censored
   2. Second entry late.

2. Each risk set under initial structure containing subject now has either first copy or second copy.
   1. Hence partial likelihood unchanged.
Subject removed and returned leaves partial likelihood unchanged

1. Subject now has two lines
   1. First entry censored
   2. Second entry late.

2. Each risk set under initial structure containing subject now has either first copy or second copy.
   1. Hence partial likelihood unchanged.

3. Can be repeated to give as many records for a subject as desired. R Code
Objectives Lecture 08

1. Cox Regression diagnostics
Objectives Lecture 08

1. Cox Regression diagnostics
2. Cox and Snell Residuals

Readings: KM § 11.1, 11.2a, 11.2b, 5.3, 9.1–9.2, 9.3
Objectives Lecture 08

1. Cox Regression diagnostics
2. Cox and Snell Residuals
3. Assessing Cox and Snell Residuals

Readings: KM § 11.1, 11.2a, 11.2b, 5.3, 9.1–9.2, 9.3
Objectives Lecture 08

1. Cox Regression diagnostics
2. Cox and Snell Residuals
3. Assessing Cox and Snell Residuals
4. Interpretation

Readings: KM § 11.1, 11.2a, 11.2b, 5.3, 9.1–9.2, 9.3
Objectives Lecture 08

1. Cox Regression diagnostics
2. Cox and Snell Residuals
3. Assessing Cox and Snell Residuals
4. Interpretation
5. Time dependent covariates

Readings: KM § 11.1, 11.2a, 11.2b, 5.3, 9.1–9.2, 9.3
Objectives Lecture 08

1. Cox Regression diagnostics
2. Cox and Snell Residuals
3. Assessing Cox and Snell Residuals
4. Interpretation
5. Time dependent covariates
6. Stratification in Proportional Hazards Estimation
Objectives Lecture 08

1. Cox Regression diagnostics
2. Cox and Snell Residuals
3. Assessing Cox and Snell Residuals
4. Interpretation
5. Time dependent covariates
6. Stratification in Proportional Hazards Estimation
1. Extend model from before:  

\[ h_i(t) = h_0(t) \exp(z_i(t)\beta) \]
Model:

1. Extend model from before: \( h_i(t) = h_0(t) \exp(z_i(t)\beta) \)
2. Hazards are no longer proportional.
Splitting idea above generates separate records where covariate is constant.

Covariate value fixed on each interval.
Splitting idea above generates separate records where covariate is constant.

1. Covariate value fixed on each interval.
   - Ex., remission status
Splitting idea above generates separate records where covariate is constant.

1. Covariate value fixed on each interval.
   1. Ex., remission status
   2. Ex., transplant status
Splitting idea above generates separate records where covariate is constant.

1. Covariate value fixed on each interval.
   1. Ex., remission status
   2. Ex., transplant status

2. Estimating base line hazard rate $h_0$ is very difficult.
Likelihood:

1. Partial log likelihood is still $\sum_{i \in D} (z_i(t) \beta - \log(V_i))$
Likelihood:

1. Partial log likelihood is still \( \sum_{i \in D} (z_i(t) \beta - \log(V_i)) \)

2. for \( V_i = \sum_{j \in R(i)} \exp(z_j(t_i) \beta) \)
Likelihood:

1. Partial log likelihood is still \( \sum_{i \in D} (z_i(t)\beta - \log(V_i)) \)
   for \( V_i = \sum_{j \in R(i)} \exp(z_j(t_i)\beta) \)

2. Math justifying use of this likelihood is harder
Likelihood:

1. Partial log likelihood is still $\sum_{i \in D} (z_i(t) \beta - \log(V_i))$
   for $V_i = \sum_{j \in R(i)} \exp(z_j(t_i) \beta)$

2. Math justifying use of this likelihood is harder

3. Computations are harder
Likelihood:

1. Partial log likelihood is still $\sum_{i \in D} (z_i(t)/\beta - \log(V_i))$
   for $V_i = \sum_{j \in R(i)} \exp(z_j(t_i)/\beta)$

2. Math justifying use of this likelihood is harder

3. Computations are harder
   - Smart way to do fixed covariates
Likelihood:

1. Partial log likelihood is still $\sum_{i \in D} (z_i(t)/\beta - \log(V_i))$
   - for $V_i = \sum_{j \in R(i)} \exp(z_j(t_i)/\beta)$

2. Math justifying use of this likelihood is harder

3. Computations are harder
   - Smart way to do fixed covariates
     - Set $V_{\max(D+1)} = 0$
Likelihood:

1. Partial log likelihood is still $\sum_{i \in D} (z_i(t) \beta - \log(V_i))$
   - for $V_i = \sum_{j \in R(i)} \exp(z_j(t_i) \beta)$
2. Math justifying use of this likelihood is harder
3. Computations are harder
   - Smart way to do fixed covariates
     1. Set $V_{\max(D+1)} = 0$
     2. Cycle through observations backwards
Likelihood:

1. Partial log likelihood is still \( \sum_{i \in D} (z_i(t) \beta - \log(V_i)) \)
   for \( V_i = \sum_{j \in R(i)} \exp(z_j(t_i) \beta) \)

2. Math justifying use of this likelihood is harder

3. Computations are harder
   - Smart way to do fixed covariates
     1. Set \( V_{\text{max}(D+1)} = 0 \)
     2. Cycle through observations backwards
     3. \( V_i = V_{i+1} + \exp(z_j \beta) \)
Likelihood:

1. Partial log likelihood is still \( \sum_{i \in D} (z_i(t)\beta - \log(V_i)) \)
   - for \( V_i = \sum_{j \in R(i)} \exp(z_j(t_i)\beta) \)

2. Math justifying use of this likelihood is harder

3. Computations are harder
   - Smart way to do fixed covariates
     1. Set \( V_{\max(D+1)} = 0 \)
     2. Cycle through observations backwards
     3. \( V_i = V_{i+1} + \exp(z_j\beta) \)
     4. \( \ell(\beta) \leftarrow z_i\beta - \log(V_i) + \ell(\beta) \)
Likelihood:

1. Partial log likelihood is still $\sum_{i \in D} (z_i(t)\beta - \log(V_i))$
   
   for $V_i = \sum_{j \in R(i)} \exp(z_j(t_i)\beta)$

2. Math justifying use of this likelihood is harder

3. Computations are harder
   
   1. Smart way to do fixed covariates
      
      1. Set $V_{\text{max}(D+1)} = 0$
      
      2. Cycle through observations backwards
      
      3. $V_i = V_{i+1} + \exp(z_j\beta)$
      
      4. $\ell(\beta) \leftarrow z_i\beta - \log(V_i) + \ell(\beta)$

   2. With time dependent covariates must calculate $V_i$ all over each time
Likelihood:

1. Partial log likelihood is still $\sum_{i \in D}(z_i(t)\beta - \log(V_i))$
   
   for $V_i = \sum_{j \in R(i)} \exp(z_j(t_i)\beta)$

2. Math justifying use of this likelihood is harder

3. Computations are harder

   1. Smart way to do fixed covariates
      
      1. Set $V_{\text{max}(D+1)} = 0$
      
      2. Cycle through observations backwards
      
      3. $V_i = V_{i+1} + \exp(z_j\beta)$
      
      4. $\ell(\beta) \leftarrow z_i\beta - \log(V_i) + \ell(\beta)$

   2. With time dependent covariates must calculate $V_i$ all over each time
      
      1. Often have to calculate underlying $z$’s as well
Likelihood:

1. Partial log likelihood is still \( \sum_{i \in D} (z_i(t)\beta - \log(V_i)) \)
   for \( V_i = \sum_{j \in \mathcal{R}(i)} \exp(z_j(t_i)\beta) \)
2. Math justifying use of this likelihood is harder
3. Computations are harder
   1. Smart way to do fixed covariates
      1. Set \( V_{\max(D+1)} = 0 \)
      2. Cycle through observations backwards
      3. \( V_i = V_{i+1} + \exp(z_i\beta) \)
      4. \( \ell(\beta) \leftarrow z_i\beta - \log(V_i) + \ell(\beta) \)
   2. With time dependent covariates must calculate \( V_i \) all over each time
      1. Often have to calculate underlying \( z \)'s as well
      2. Can save \( z \) values to reuse in evaluation for new \( \beta \)
Likelihood:

1. Partial log likelihood is still $\sum_{i \in D} (z_i(t) \beta - \log(V_i))$
   1. for $V_i = \sum_{j \in R(i)} \exp(z_j(t_i) \beta)$
2. Math justifying use of this likelihood is harder
3. Computations are harder
   1. Smart way to do fixed covariates
      1. Set $V_{\max(D+1)} = 0$
      2. Cycle through observations backwards
      3. $V_i = V_{i+1} + \exp(z_j \beta)$
      4. $\ell(\beta) \leftarrow z_i \beta - \log(V_i) + \ell(\beta)$
   2. With time dependent covariates must calculate $V_i$ all over each time
      1. Often have to calculate underlying $z$’s as well
      2. Can save $z$ values to reuse in evaluation for new $\beta$
4. Adjustments for ties are as before
Taxonomy

1. *external* or *defined*.
1. *external* or *defined*.

2. Variables whose time course is fixed from the start.

Examples:

1. Study where subjects might switch treatments (cross-over trial)
2. Transplant
3. Treatment group where treatment loses effectiveness
4. Immunization effect might be $\exp(-t \gamma z_i)$ for $z_i$ equal 0 or 1
5. Artificial variables used to assess proportional hazards assumption

Later, to be seen.

1. Ancillary

2. Random variables who change with no causal relation to response.

Refers to a random quantity that one performs analysis conditional on.

Examples:

1. Market rates for replacement mortgages when modeling mortgage prepayments
2. Effect of heat or pollution on mortality
Taxonomy

1. *external* or *defined*.
   1. Variables whose time course is fixed from the start
   2. Examples:
      - Study where subjects might switch treatments
      - Cross over trial
      - Transplant
      - Treatment group where treatment loses effectiveness
        - Immunization effect might be $\exp(-t\gamma z_i)$ for $z_i$ equal 0 or 1
      - Artificial variables used to assess proportional hazards assumption
        - To be seen later

2. *ancillary*
   1. Random variables who change with no causal relation to response
   2. Refers to a random quantity that one performs analysis conditional on
   3. Examples
      - Market rates for replacement mortgages when modeling mortgage prepayments
      - Effect of heat or pollution on mortality
**Taxonomy**

1. *external or defined.*
   1. Variables whose time course is fixed from the start
   2. Examples:
      1. Study where subjects might switch treatments

**Taxonomy**

1. *external* or *defined*.
   1. Variables whose time course is fixed from the start
   2. Examples:
      1. Study where subjects might switch treatments
      2. Cross over trial

Random variables who change with no causal relation to response

Refers to a random quantity that one performs analysis conditional on

Examples

1. market rates for replacement mortgages when modeling mortgage prepayments
2. effect of heat or pollution on mortality
Taxonomy

1. *external* or *defined*.

   1. Variables whose time course is fixed from the start
   2. Examples:
      1. Study where subjects might switch treatments
      2. Cross over trial
      3. Transplant

   Artificial variables used to assess proportional hazards assumption
   To be seen later

2. *ancillary*

   Random variables who change with no causal relation to response
   Refers to a random quantity that one performs analysis conditional on
   Examples
   1. market rates for replacement mortgages when modeling mortgage prepayments
   2. effect of heat or pollution on mortality
Taxonomy

1. *external* or *defined*.
   1. Variables whose time course is fixed from the start
   2. Examples:
      1. Study where subjects might switch treatments
      2. Cross over trial
      3. Transplant
      4. Treatment group where treatment loses effectiveness

Artificial variables used to assess proportional hazards assumption
To be seen later
Taxonomy

1. *external* or *defined*.

1. Variables whose time course is fixed from the start
2. Examples:
   1. Study where subjects might switch treatments
   2. Cross over trial
   3. Transplant
   4. Treatment group where treatment loses effectiveness
   5. Immunization effect might be \( \exp(-t\gamma)z_i \) for \( z_i \) equal 0 or 1
Taxonomy

1. *external or defined.*
   - Variables whose time course is fixed from the start
   - Examples:
     1. Study where subjects might switch treatments
     2. Cross over trial
     3. Transplant
     4. Treatment group where treatment loses effectiveness
     5. Immunization effect might be $\exp(-t\gamma)z_i$ for $z_i$ equal 0 or 1
     6. Artificial variables used to assess proportional hazards assumption
Taxonomy

1. *external or defined.*
   1. Variables whose time course is fixed from the start
   2. Examples:
      1. Study where subjects might switch treatments
      2. Cross over trial
      3. Transplant
      4. Treatment group where treatment loses effectiveness
      5. Immunization effect might be \( \exp(-t\gamma)z_i \) for \( z_i \) equal 0 or 1
      6. Artificial variables used to assess proportional hazards assumption
      7. To be seen later
Taxonomy

1. **external or defined.**
   1. Variables whose time course is fixed from the start
   2. Examples:
      1. Study where subjects might switch treatments
      2. Cross over trial
      3. Transplant
      4. Treatment group where treatment loses effectiveness
      5. Immunization effect might be \( \exp(-t^\gamma)z_i \) for \( z_i \) equal 0 or 1
      6. Artificial variables used to assess proportional hazards assumption
      7. To be seen later

2. **ancillary**
Taxonomy

1. *external* or *defined*.
   - Variables whose time course is fixed from the start
   - Examples:
     1. Study where subjects might switch treatments
     2. Cross over trial
     3. Transplant
     4. Treatment group where treatment loses effectiveness
     5. Immunization effect might be \( \exp(-t\gamma)z_i \) for \( z_i \) equal 0 or 1
     6. Artificial variables used to assess proportional hazards assumption
     7. To be seen later

2. *ancillary*
   - Random variables who change with no causal relation to response
1. *external* or *defined*.
   1. Variables whose time course is fixed from the start
   2. Examples:
      1. Study where subjects might switch treatments
      2. Cross over trial
      3. Transplant
      4. Treatment group where treatment loses effectiveness
      5. Immunization effect might be $\exp(-t\gamma)z_i$ for $z_i$ equal 0 or 1
      6. Artificial variables used to assess proportional hazards assumption
      7. To be seen later

2. *ancillary*
   1. Random variables who change with no causal relation to response
      1. Refers to a random quantity that one performs analysis conditional on
Taxonomy

1. *external* or *defined*.
   - Variables whose time course is fixed from the start
   - Examples:
     1. Study where subjects might switch treatments
     2. Cross over trial
     3. Transplant
     4. Treatment group where treatment loses effectiveness
     5. Immunization effect might be \( \exp(-t \gamma) z_i \) for \( z_i \) equal 0 or 1
     6. Artificial variables used to assess proportional hazards assumption
     7. To be seen later

2. *ancillary*
   - Random variables who change with no causal relation to response
     - Refers to a random quantity that one performs analysis conditional on
   - Examples
Taxonomy

1. *external* or *defined*.
   1. Variables whose time course is fixed from the start
   2. Examples:
      1. Study where subjects might switch treatments
      2. Cross over trial
      3. Transplant
      4. Treatment group where treatment loses effectiveness
      5. Immunization effect might be $\exp(-t\gamma)z_i$ for $z_i$ equal 0 or 1
      6. Artificial variables used to assess proportional hazards assumption
      7. To be seen later

2. *ancillary*
   1. Random variables who change with no causal relation to response
      1. Refers to a random quantity that one performs analysis conditional on
   2. Examples
      1. Market rates for replacement mortgages when modeling mortgage prepayments
**Taxonomy**

1. *external* or *defined*.
   - Variables whose time course is fixed from the start
   - Examples:
     - Study where subjects might switch treatments
     - Cross over trial
     - Transplant
     - Treatment group where treatment loses effectiveness
     - Immunization effect might be \( \exp(-t \gamma)z_i \) for \( z_i \) equal 0 or 1
     - Artificial variables used to assess proportional hazards assumption
     - To be seen later

2. *ancillary*
   - Random variables who change with no causal relation to response
     - Refers to a random quantity that one performs analysis conditional on
   - Examples
     - Market rates for replacement mortgages when modeling mortgage prepayments
     - Effect of heat or pollution on mortality
Continued

1 internal
1. internal

   Variables sharing common cause with mortality
internal

Variables sharing common cause with mortality

Whether disease has progressed to a preliminary stage
Continued

1. *internal*

1. Variables sharing common cause with mortality
   1. Whether disease has progressed to a preliminary stage
   2. Early markers are commonly sought to make studying slowly progressing diseases easier
1. **internal**

1. Variables sharing common cause with mortality
   1. Whether disease has progressed to a preliminary stage
   2. Early markers are commonly sought to make studying slowly progressing diseases easier
   3. In reliability, wear on part as a predictor of failure
Proportional hazards testing (in two sample case)

1. $z_{1i}$ indicates membership in group 2
Proportional hazards testing (in two sample case)

1. $z_{1i}$ indicates membership in group 2
2. Test whether $h_2$ rises relative to $h_1$ by using second covariate $z_{2i}(t) = z_{1i}g(t)$
Proportional hazards testing (in two sample case)

1. $z_{1i}$ indicates membership in group 2
2. Test whether $h_2$ rises relative to $h_1$ by using second covariate
   \[ z_{2i}(t) = z_{1i}g(t) \]
3. $g(t) =$
Proportional hazards testing (in two sample case)

1. $z_{1i}$ indicates membership in group 2
2. Test whether $h_2$ rises relative to $h_1$ by using second covariate $z_{2i}(t) = z_{1i}g(t)$
3. $g(t) =$
   1. $t - \bar{t}$
Proportional hazards testing (in two sample case)

1. \( z_{1i} \) indicates membership in group 2

2. Test whether \( h_2 \) rises relative to \( h_1 \) by using second covariate \( z_{2i}(t) = z_{1i}g(t) \)

3. \( g(t) = \)

   1. \( t - \bar{t} \)
   2. \( \begin{cases} 
     0 & \text{if } t < t_0 \\
     1 & \text{if } t \geq t_0 
   \end{cases} \)

Heaviside function

Might use several for different behaviors in different ranges

\( t_0 \) is called the change point

Regression Models: Time Dependent Covariates

Lecture 08
Proportional hazards testing (in two sample case)

1. $z_{1i}$ indicates membership in group 2

2. Test whether $h_2$ rises relative to $h_1$ by using second covariate $z_2(t) = z_1 g(t)$

3. $g(t) =$
   
   1. $t - \bar{t}$
   2. \[
   \begin{cases} 
   0 & \text{if } t < t_0 \\
   1 & \text{if } t \geq t_0 
   \end{cases}
   \]

   1. Heaviside function

---
Proportional hazards testing (in two sample case)

1. $z_{1i}$ indicates membership in group 2.

2. Test whether $h_2$ rises relative to $h_1$ by using second covariate:
   
   $$z_{2i}(t) = z_{1i}g(t)$$

3. $g(t) =$

   - $t - \bar{t}$
   - $\begin{cases} 
   0 & \text{if } t < t_0 \\
   1 & \text{if } t \geq t_0 
   \end{cases}$

   - 1. Heaviside function
   - 2. Might use several for different behaviors in different ranges
Proportional hazards testing (in two sample case)

1. \( z_{1i} \) indicates membership in group 2
2. Test whether \( h_2 \) rises relative to \( h_1 \) by using second covariate \( z_{2i}(t) = z_{1i}g(t) \)
3. \( g(t) =
   \begin{cases}
   t - t & \text{1 if } t < t_0 \\
   0 & \text{if } t \geq t_0
   \end{cases}
\)
   
   1. Heaviside function
   2. Might use several for different behaviors in different ranges
   3. \( t_0 \) is called the change point
Proportional hazards testing (in two sample case)

1. $z_{1i}$ indicates membership in group 2

2. Test whether $h_2$ rises relative to $h_1$ by using second covariate $z_{2i}(t) = z_{1i}g(t)$

3. $g(t) = \begin{cases} 
  t - \bar{t} & \text{if } t < t_0 \\
  0 & \text{if } t \geq t_0 
\end{cases}$

   1. Heaviside function
   2. Might use several for different behaviors in different ranges
   3. $t_0$ is called the change point

3. $\log(t)$  R Code  SAS Code
Section: Regression Models

Subsection: Causal inference for an internal variable is difficult
Confounding is a problem.

1 Under certain circumstances:
Confounding is a problem.

Under certain circumstances:

1. one covariate is treatment that slows progression to intermediate stage
2. another covariate is progression to intermediate stage
   treatment will show no effect even if effective.
Confounding is a problem.

1. Under certain circumstances:
   1. one covariate is treatment that slows progression to intermediate stage
   2. another covariate is progression to intermediate stage
Confounding is a problem.

1. Under certain circumstances:
   1. one covariate is treatment that slows progression to intermediate stage
   2. another covariate is progression to intermediate stage

2. treatment will show no effect even if effective.
Sometimes surprisingly time dependent covariate makes no difference

1 Suppose we add age as covariate instead of initial age $w$

\[ h_i(t) = h_0(t) \exp(z_i(t) \beta + w_i \gamma) \]

Fixed time hazard

Continuous version

\[ h_i(t) = h_0(t) \exp(z_i(t) \beta + (w_i + t) \gamma) = h_0(t) \exp(t \gamma) \exp(z_i(t) \beta + w_i \gamma) \]

Hence baseline hazard is changed to $h_0(t) \exp(t \gamma)$ but regression is unchanged.
Sometimes surprisingly time dependent covariate makes no difference

1. Suppose we add age as covariate instead of initial age $w$
2. Intuition: individual’s risk increases with time
Sometimes surprisingly time dependent covariate makes no difference

1. Suppose we add age as covariate instead of initial age $w$
2. Intuition: individual’s risk increases with time
   1. Or maybe decreases

\[
h_i(t) = h_0(t) \exp(z_i(t) \beta + w_i \gamma)
\]

\[
h_i(t) = h_0(t) \exp(t \gamma) \exp(z_i(t) \beta + w_i \gamma)
\]

Hence baseline hazard is changed to $h_0(t) \exp(t \gamma)$ but regression is unchanged.
Sometimes surprisingly time dependent covariate makes no difference

1. Suppose we add age as covariate instead of initial age \( w \)
2. Intuition: individual’s risk increases with time
   - Or maybe decreases
3. Effect is even in both groups

\[
h_i(t) = h_0(t) \exp(z_i(t) \beta + w_i \gamma)
\]

Fixed time hazard

\[
h_i(t) = h_0(t) \exp(z_i(t) \beta + \gamma t)
\]

Continuous version

Hence baseline hazard is changed to \( h_0(t) \exp(t \gamma) \) but regression is unchanged. R Code SAS Code
Sometimes surprisingly time dependent covariate makes no difference

1. Suppose we add age as covariate instead of initial age $w$
2. Intuition: individual’s risk increases with time
   1. Or maybe decreases
3. Effect is even in both groups
   1. Fixed time hazard
      \[ h_i(t) = h_0(t) \exp(z_i(t)\beta + w_i\gamma) \]
Sometimes surprisingly time dependent covariate makes no difference

1. Suppose we add age as covariate instead of initial age $w$
2. Intuition: individual’s risk increases with time
   - Or maybe decreases
3. Effect is even in both groups
   - Fixed time hazard $h_i(t) = h_0(t) \exp(z_i(t)\beta + w_i\gamma)$
   - Continuous version

\[
 h_i(t) = h_0(t) \exp(z_i(t)\beta + (w_i + t)\gamma) = h_0(t) \exp(t\gamma) \exp(z_i(t)\beta + w_i\gamma)
\]
Sometimes surprisingly time dependent covariate makes no difference

1. Suppose we add age as covariate instead of initial age $w$.
2. Intuition: individual’s risk increases with time.
   1. Or maybe decreases.
3. Effect is even in both groups.
   1. Fixed time hazard $h_i(t) = h_0(t) \exp(z_i(t)\beta + w_i\gamma)$
   2. Continuous version

$$h_i(t) = h_0(t) \exp(z_i(t)\beta + (w_i + t)\gamma)$$

$$= h_0(t) \exp(t\gamma) \exp(z_i(t)\beta + w_i\gamma)$$

3. Hence baseline hazard is changed to $h_0(t) \exp(t\gamma)$ but regression is unchanged. R Code SAS Code
Section: Regression Models

Subsection: Stratification:
Suppose

1. expect hazards from different strata not to be proportional
Suppose

1. expect hazards from different strata not to be proportional
2. Expect some variable to influence hazard in all strata in the same way
Suppose

1. expect hazards from different strata not to be proportional
2. Expect some variable to influence hazard in all strata in the same way
3. Strata act independently
Then do calculations separately in each strata
Then

1. do calculations separately in each strata
2. Partial log likelihood is sum of individual contributions
Then

1. do calculations separately in each strata
2. Partial log likelihood is sum of individual contributions
3. Estimator maximizes this
Then

1. do calculations separately in each strata
2. Partial log likelihood is sum of individual contributions
3. Estimator maximizes this
4. Derivatives and variances sum
Then

1. do calculations separately in each strata
2. Partial log likelihood is sum of individual contributions
3. Estimator maximizes this
4. Derivatives and variances sum
   1. Testing is straight forward
Then

1. do calculations separately in each strata
2. Partial log likelihood is sum of individual contributions
3. Estimator maximizes this
4. Derivatives and variances sum
   1. Testing is straight forward
5. As alternative to putting interaction with time in for stratification variable.
Advantages and Disadvantages of Stratification

1. Bonuses for stratification:
   
   1. Doesn't depend on correct effect being linear
   2. Computationally easier

Drawbacks:

1. Won't work for continuous covariate
2. Can't simultaneously estimate effect of stratification variable

Loses information

R Code

SAS Code
Advantages and Disadvantages of Stratification

Bonuses for stratification:

1. Doesn’t depend on correct effect being linear
Advantages and Disadvantages of Stratification

1. Bonuses for stratification:
   1. Doesn’t depend on correct effect being linear
   2. Computationally easier
Advantages and Disadvantages of Stratification

1. Bonuses for stratification:
   1. Doesn’t depend on correct effect being linear
   2. Computationally easier

2. Drawbacks

Regression Models: Stratification: Lecture 08
Advantages and Disadvantages of Stratification

1. Bonuses for stratification:
   1. Doesn’t depend on correct effect being linear
   2. Computationally easier

2. Drawbacks
   1. Won’t work for continuous covariate
Advantages and Disadvantages of Stratification

1. Bonuses for stratification:
   1. Doesn’t depend on correct effect being linear
   2. Computationally easier

2. Drawbacks
   1. Won’t work for continuous covariate
   2. Can’t simultaneously estimate effect of stratification variable
Advantages and Disadvantages of Stratification

1. Bonuses for stratification:
   1. Doesn’t depend on correct effect being linear
   2. Computationally easier

2. Drawbacks
   1. Won’t work for continuous covariate
   2. Can’t simultaneously estimate effect of stratification variable
   3. Loses information  R Code  SAS Code
Diagnostic we have seen before:

1. Are cumulative hazard functions parallel?
Diagnostic we have seen before:

1. Are cumulative hazard functions parallel?
2. Testing time dependent covariate  R Code  SAS Code
What we will look for

1. Shape of relationship with covariate
What we will look for

1. Shape of relationship with covariate
   - log scale or original scale
What we will look for

1. **Shape of relationship with covariate**
   1. log scale or original scale
   2. change points
What we will look for

1. Shape of relationship with covariate
   1. log scale or original scale
   2. change points
   3. etc..
What we will look for

1. Shape of relationship with covariate
   1. log scale or original scale
   2. change points
   3. etc..

2. Adequacy of proportional hazards
What we will look for

1. Shape of relationship with covariate
   - log scale or original scale
   - change points
   - etc..

2. Adequacy of proportional hazards

3. Presence of outliers
What we will look for

1. Shape of relationship with covariate
   1. log scale or original scale
   2. change points
   3. etc..

2. Adequacy of proportional hazards

3. Presence of outliers

4. Leverage
Cox and Snell residuals for general regression models:

1. Choose $h$ such that $\epsilon_j = h_j(T_j, \beta)$ are $\approx$ i.i.d.
Cox and Snell residuals for general regression models:

1. Choose $h$ such that $\epsilon_j = h_j(T_j, \beta)$ are $\approx$ i.i.d.
2. Residuals are $R_j = h_j(T_j, \hat{\beta})$
Cox and Snell residuals for general regression models:

1. Choose $h$ such that $\epsilon_j = h_j(T_j, \beta)$ are $\approx$ i.i.d.
2. Residuals are $R_j = h_j(T_j, \hat{\beta})$
3. Examples
Cox and Snell residuals for general regression models:

1. Choose $h$ such that $\epsilon_j = h_j(T_j, \beta)$ are $\approx$ i.i.d.
2. Residuals are $R_j = h_j(T_j, \hat{\beta})$
3. Examples
   1. for normal theory regression $h_j(t, \beta) = t - z_j\beta$
Cox and Snell residuals for general regression models:

1. Choose $h$ such that $\epsilon_j = h_j(T_j, \beta)$ are $\approx$ i.i.d.
2. Residuals are $R_j = h_j(T_j, \hat{\beta})$
3. Examples
   1. For normal theory regression $h_j(t, \beta) = t - z_j \beta$
   2. For logistic regression, $h_j(t, \beta) = (t - n_j \pi_j) / \sqrt{n_j \pi_j (1 - \pi)}$
Cox and Snell residuals for general regression models:

1. Choose $h$ such that $\epsilon_j = h_j(T_j, \beta)$ are $\approx$ i.i.d.
2. Residuals are $R_j = h_j(T_j, \hat{\beta})$
3. Examples
   1. for normal theory regression $h_j(t, \beta) = t - z_j\beta$
   2. For logistic regression, $h_j(t, \beta) = (t - n_j\pi_j)/\sqrt{n_j\pi(1-\pi)}$
4. General definition precedes PH regression.
Cox and Snell residuals for general regression models:

1. Choose $h$ such that $\epsilon_j = h_j(T_j, \beta)$ are $\approx$ i.i.d.
2. Residuals are $R_j = h_j(T_j, \hat{\beta})$
3. Examples
   - For normal theory regression $h_j(t, \beta) = t - z_j\beta$
   - For logistic regression, $h_j(t, \beta) = (t - n_j\pi_j)/\sqrt{n_j\pi(1 - \pi)}$
4. General definition precedes PH regression.
5. Distribution of Hazard at Event times exponential.
Cox and Snell residuals for general regression models:

1. Choose $h$ such that $\epsilon_j = h_j(T_j, \beta)$ are $\approx$ i.i.d.
2. Residuals are $R_j = h_j(T_j, \hat{\beta})$

Examples

1. for normal theory regression $h_j(t, \beta) = t - z_j \beta$
2. For logistic regression, $h_j(t, \beta) = (t - n_j \pi_j) / \sqrt{n_j \pi(1 - \pi)}$

4. General definition precedes PH regression.
5. Distribution of Hazard at Event times exponential.
   1. may be determined through transformation
**Cox and Snell residuals** for general regression models:

1. Choose \( h \) such that \( \epsilon_j = h_j(T_j, \beta) \) are \( \approx \) i.i.d.
2. Residuals are \( R_j = h_j(T_j, \hat{\beta}) \)
3. Examples
   - for normal theory regression \( h_j(t, \beta) = t - z_j \beta \)
   - For logistic regression, \( h_j(t, \beta) = (t - n_j \pi_j) / \sqrt{n_j \pi(1 - \pi)} \)
4. General definition precedes PH regression.
5. Distribution of Hazard at Event times exponential.
   - may be determined through transformaton
   - \( P[T > t] = S(t) \)
Cox and Snell residuals for general regression models:

1. Choose $h$ such that $\epsilon_j = h_j(T_j, \beta)$ are $\approx$ i.i.d.
2. Residuals are $R_j = h_j(T_j, \hat{\beta})$
3. Examples
   1. for normal theory regression $h_j(t, \beta) = t - z_j \beta$
   2. For logistic regression, $h_j(t, \beta) = (t - n_j \pi_j) / \sqrt{n_j \pi_j (1 - \pi)}$
4. General definition precedes PH regression.
5. Distribution of Hazard at Event times exponential.
   1. may be determined through transformaton
   2. $P[T > t] = S(t)$
   3. $P[S(T) < S(t)] = S(t)$
Cox and Snell residuals for general regression models:

1. Choose $h$ such that $\epsilon_j = h_j(T_j, \beta)$ are $\approx$ i.i.d.
2. Residuals are $R_j = h_j(T_j, \hat{\beta})$
3. Examples
   1. For normal theory regression $h_j(t, \beta) = t - z_j \beta$
   2. For logistic regression, $h_j(t, \beta) = (t - n_j \pi_j)/\sqrt{n_j \pi(1 - \pi)}$
4. General definition precedes PH regression.
5. Distribution of Hazard at Event times exponential.
   1. may be determined through transformaton
   2. $P[T > t] = S(t)$
   3. $P[S(T) < S(t)] = S(t)$
   4. $P[-\log(S(T)) > -\log(S(t))] = S(t)$
Cox and Snell residuals for general regression models:

1. Choose $h$ such that $\epsilon_j = h_j(T_j, \beta)$ are $\approx$ i.i.d.
2. Residuals are $R_j = h_j(T_j, \hat{\beta})$
3. Examples
   1. for normal theory regression $h_j(t, \beta) = t - z_j \beta$
   2. For logistic regression, $h_j(t, \beta) = (t - n_j \pi_j)/\sqrt{n_j \pi_j (1 - \pi)}$
4. General definition precedes PH regression.
5. Distribution of Hazard at Event times exponential.
   1. may be determined through transformaton
   2. $P[T > t] = S(t)$
   3. $P[S(T) < S(t)] = S(t)$
   4. $P[-\log(S(T)) > -\log(S(t))] = S(t)$
   5. $P[-\log(S(T)) > -\log(S(t))] = \exp(-[-\log(S(t))])$
Cox and Snell residuals for general regression models:

1. Choose $h$ such that $\epsilon_j = h_j(T_j, \beta)$ are $\approx$ i.i.d.
2. Residuals are $R_j = h_j(T_j, \hat{\beta})$
3. Examples
   1. for normal theory regression $h_j(t, \beta) = t - z_j \beta$
   2. For logistic regression, $h_j(t, \beta) = (t - n_j \pi_j) / \sqrt{n_j \pi_j (1 - \pi)}$
4. General definition precedes PH regression.
5. Distribution of Hazard at Event times exponential.
   1. may be determined through transformation
   2. $P[T > t] = S(t)$
   3. $P[S(T) < S(t)] = S(t)$
   4. $P[-\log(S(T)) > -\log(S(t))] = S(t)$
   5. $P[-\log(S(T)) > -\log(S(t))] = \exp(-[-\log(S(t))])$
   6. $P[-\log(S(T)) > h] = \exp(-h)$
**Cox and Snell residuals** for proportional hazards:

1. Suppose $X_j$ has hazard $h_0(t) \exp(z_j \beta)$
Cox and Snell residuals for proportional hazards:

1. Suppose $X_j$ has hazard $h_0(t) \exp(z_j \beta)$
2. Then survival function is $\exp(-H_0(t) \exp(z_j \beta))$
Cox and Snell residuals for proportional hazards:

1. Suppose $X_j$ has hazard $h_0(t) \exp(z_j \beta)$
   - Then survival function is $\exp(-H_0(t) \exp(z_j \beta))$
2. Then survival function of $\epsilon_j = H_0(X_i) \exp(z_j \beta)$ is $\exp(-\epsilon_j)$
Cox and Snell residuals for proportional hazards:

1. Suppose \( X_j \) has hazard \( h_0(t) \exp(z_j \beta) \)
   - Then survival function is \( \exp(-H_0(t) \exp(z_j \beta)) \)
2. Then survival function of \( \epsilon_j = H_0(X_i) \exp(z_j \beta) \) is \( \exp(-\epsilon_j) \)
3. Let \( R_j = \hat{H}_0(T_j) \exp(z_j \hat{\beta}) \), possibly censored.
Cox and Snell residuals for proportional hazards:

1. Suppose $X_j$ has hazard $h_0(t) \exp(z_j \beta)$
   - Then survival function is $\exp(-H_0(t) \exp(z_j \beta))$
2. Then survival function of $\epsilon_j = H_0(X_i) \exp(z_j \beta)$ is $\exp(-\epsilon_j)$
3. Let $R_j = \hat{H}_0(T_j) \exp(z_j \hat{\beta})$, possibly censored.
4. Analogy with regression:
Cox and Snell residuals for proportional hazards:

1. Suppose $X_j$ has hazard $h_0(t) \exp(z_j \beta)$
   - Then survival function is $\exp(-H_0(t) \exp(z_j \beta))$
2. Then survival function of $\epsilon_j = H_0(X_i) \exp(z_j \beta)$ is $\exp(-\epsilon_j)$
3. Let $R_j = \hat{H}_0(T_j) \exp(z_j \hat{\beta})$, possibly censored.
4. Analogy with regression:
   - For small samples you would want to adjust for negative correlation

Regression Models: Diagnostics Lecture 08 171 / 260
Cox and Snell residuals for proportional hazards:

1. Suppose $X_j$ has hazard $h_0(t) \exp(z_j \beta)$
   - Then survival function is $\exp(-H_0(t) \exp(z_j \beta))$

2. Then survival function of $\epsilon_j = H_0(X_i) \exp(z_j \beta)$ is $\exp(-\epsilon_j)$

3. Let $R_j = \hat{H}_0(T_j) \exp(z_j \hat{\beta})$, possibly censored.

4. Analogy with regression:
   - For small samples you would want to adjust for negative correlation
   - There is no corresponding proportional hazards adjustment, so do not use for small samples.
Cox and Snell residuals for proportional hazards:

1. Suppose $X_j$ has hazard $h_0(t) \exp(z_j \beta)$
   
   Then survival function is $\exp(-H_0(t) \exp(z_j \beta))$

2. Then survival function of $\epsilon_j = H_0(X_i) \exp(z_j \beta)$ is $\exp(-\epsilon_j)$

3. Let $R_j = \hat{H}_0(T_j) \exp(z_j \hat{\beta})$, possibly censored.

4. Analogy with regression:
   
   1. For small samples you would want to adjust for negative correlation
   2. There is no corresponding proportional hazards adjustment, so do not use for small samples.

5. Can also be used for time dependent covariates, if you had $\hat{H}_0$
Cox and Snell residuals for proportional hazards:

1. Suppose $X_j$ has hazard $h_0(t) \exp(z_j \beta)$
   - Then survival function is $\exp(-H_0(t) \exp(z_j \beta))$
2. Then survival function of $\epsilon_j = H_0(X_i) \exp(z_j \beta)$ is $\exp(-\epsilon_j)$
3. Let $R_j = \hat{H}_0(T_j) \exp(z_j \hat{\beta})$, possibly censored.
4. Analogy with regression:
   - For small samples you would want to adjust for negative correlation
   - There is no corresponding proportional hazards adjustment, so do not use for small samples.
5. Can also be used for time dependent covariates, if you had $\hat{H}_0$
6. Can be done within a stratum for a stratified model
Cox and Snell residuals for proportional hazards:

1. Suppose $X_j$ has hazard $h_0(t) \exp(z_j\beta)$

   Then survival function is $\exp(-H_0(t) \exp(z_j\beta))$

2. Then survival function of $\epsilon_j = H_0(X_i) \exp(z_j\beta)$ is $\exp(-\epsilon_j)$

3. Let $R_j = \hat{H}_0(T_j) \exp(z_j\hat{\beta})$, possibly censored.

4. Analogy with regression:
   - For small samples you would want to adjust for negative correlation
   - There is no corresponding proportional hazards adjustment, so do not use for small samples.

5. Can also be used for time dependent covariates, if you had $\hat{H}_0$

6. Can be done within a stratum for a stratified model

7. $\sum_j R_j \approx \sum_j \delta_j$
Cox and Snell residuals for proportional hazards:

1. Suppose $X_j$ has hazard $h_0(t) \exp(z_j \beta)$
   - Then survival function is $\exp(-H_0(t) \exp(z_j \beta))$
2. Then survival function of $\epsilon_j = H_0(X_i) \exp(z_j \beta)$ is $\exp(-\epsilon_j)$
3. Let $R_j = \hat{H}_0(T_j) \exp(z_j \hat{\beta})$, possibly censored.
4. Analogy with regression:
   - For small samples you would want to adjust for negative correlation
   - There is no corresponding proportional hazards adjustment, so do not use for small samples.
5. Can also be used for time dependent covariates, if you had $\hat{H}_0$
6. Can be done within a stratum for a stratified model
7. $\sum_j R_j \approx \sum_j \delta_j$
   - Exact for $\hat{H}(t_i) = \sum_{k=1}^{i} \frac{D_k}{\sum_{j \in R(i)} \exp(z_j \beta)}$
Cox and Snell residuals for proportional hazards:

1. Suppose $X_j$ has hazard $h_0(t) \exp(z_j\beta)$
   - Then survival function is $\exp(-H_0(t) \exp(z_j\beta))$
2. Then survival function of $\epsilon_j = H_0(X_i) \exp(z_j\beta)$ is $\exp(-\epsilon_j)$
3. Let $R_j = \hat{H}_0(T_j) \exp(z_j\hat{\beta})$, possibly censored.
4. Analogy with regression:
   - For small samples you would want to adjust for negative correlation
   - There is no corresponding proportional hazards adjustment, so do not use for small samples.
5. Can also be used for time dependent covariates, if you had $\hat{H}_0$
6. Can be done within a stratum for a stratified model
7. $\sum_j R_j \approx \sum_j \delta_j$
   - Exact for $\hat{H}(t_i) = \sum_{k=1}^i \frac{D_k}{\sum_{j \in R(i)} \exp(z_j\beta)}$
8. See if CDF of $\hat{H}(T_i)$ like a line through 0 with slope 1
Cox and Snell residuals for proportional hazards:

1. Suppose $X_j$ has hazard $h_0(t) \exp(z_j \beta)$
   - Then survival function is $\exp(-H_0(t) \exp(z_j \beta))$

2. Then survival function of $\epsilon_j = H_0(X_i) \exp(z_j \beta)$ is $\exp(-\epsilon_j)$

3. Let $R_j = \hat{H}_0(T_j) \exp(z_j \hat{\beta})$, possibly censored.

4. Analogy with regression:
   - For small samples you would want to adjust for negative correlation
   - There is no corresponding proportional hazards adjustment, so do not use for small samples.

5. Can also be used for time dependent covariates, if you had $\hat{H}_0$

6. Can be done within a stratum for a stratified model

7. $\sum_j R_j \approx \sum_j \delta_j$
   - Exact for $\hat{H}(t_i) = \sum_{k=1}^{i} \frac{D_k}{\sum_{j \in R(i)} \exp(z_j \beta)}$

8. See if CDF of $\hat{H}(T_i)$ like a line through 0 with slope 1
   - Use Nelson–Aalen estimator to account for censoring.

R Code
Objectives Lecture 09

1. Martingale Residuals

Reading: KM §11.3, 11.4, 11.5, 11.6, 12.1, 12.2, 12.3, 12.4, 2.6b, 3.6, §1.3.4-5
Objectives Lecture 09

1. Martingale Residuals
2. Proportional Hazards Plots

Readings: KM §11.3, 11.4, 11.5, 11.6, 12.1, 12.2, 12.3, 12.4, 2.6b, 3.6, L §1.3.4-5
Objectives Lecture 09

1. Martingale Residuals
2. Proportional Hazards Plots
3. Deviance Residuals
Objectives Lecture 09

1. Martingale Residuals
2. Proportional Hazards Plots
3. Deviance Residuals
4. Measuring influence
Objectives Lecture 09

1. Martingale Residuals
2. Proportional Hazards Plots
3. Deviance Residuals
4. Measuring influence
5. Location-Scale family for log lifetime distributions
Objectives Lecture 09

1. Martingale Residuals
2. Proportional Hazards Plots
3. Deviance Residuals
4. Measuring influence
5. Location-Scale family for log lifetime distributions
6. Weibull distribution
Objectives Lecture 09

1. Martingale Residuals
2. Proportional Hazards Plots
3. Deviance Residuals
4. Measuring influence
5. Location-Scale family for log lifetime distributions
6. Weibull distribution
7. Log logistic distribution
Objectives Lecture 09

1. Martingale Residuals
2. Proportional Hazards Plots
3. Deviance Residuals
4. Measuring influence
5. Location-Scale family for log lifetime distributions
6. Weibull distribution
7. Log logistic distribution
8. Log normal and generalized gamma distributions

Readings: KM §11.3, 11.4, 11.5, 11.6, 12.1, 12.2, 12.3, 12.4, 2.6b, 3.6, L §1.3.4-5
Objectives Lecture 09

1. Martingale Residuals
2. Proportional Hazards Plots
3. Deviance Residuals
4. Measuring influence
5. Location-Scale family for log lifetime distributions
6. Weibull distribution
7. Log logistic distribution
8. Log normal and generalized gamma distributions
9. Parametric Failure Time Models

Readings: KM § 11.3, 11.4, 11.5, 11.6, 12.1, 12.2, 12.3, 12.4, 2.6b, 3.6, L § 1.3.4-5
Objectives Lecture 09

1. Martingale Residuals
2. Proportional Hazards Plots
3. Deviance Residuals
4. Measuring influence
5. Location-Scale family for log lifetime distributions
6. Weibull distribution
7. Log logistic distribution
8. Log normal and generalized gamma distributions
9. Parametric Failure Time Models
10. Counting Process
Objectives Lecture 09

1. Martingale Residuals
2. Proportional Hazards Plots
3. Deviance Residuals
4. Measuring influence
5. Location-Scale family for log lifetime distributions
6. Weibull distribution
7. Log logistic distribution
8. Log normal and generalized gamma distributions
9. Parametric Failure Time Models
10. Counting Process
11. Log Normal as a member of the Incomplete Gamma Family
Objectives Lecture 09

1. Martingale Residuals
2. Proportional Hazards Plots
3. Deviance Residuals
4. Measuring influence
5. Location-Scale family for log lifetime distributions
6. Weibull distribution
7. Log logistic distribution
8. Log normal and generalized gamma distributions
9. Parametric Failure Time Models
10. Counting Process
11. Log Normal as a member of the Incomplete Gamma Family
12. Proportional Hazards and the Weibull Distribution
Objectives Lecture 09

1. Martingale Residuals
2. Proportional Hazards Plots
3. Deviance Residuals
4. Measuring influence
5. Location-Scale family for log lifetime distributions
6. Weibull distribution
7. Log logistic distribution
8. Log normal and generalized gamma distributions
9. Parametric Failure Time Models
10. Counting Process
11. Log Normal as a member of the Incomplete Gamma Family
12. Proportional Hazards and the Weibull Distribution
13. Readings: KM §11.3, 11.4, 11.5, 11.6, 12.1, 12.2, 12.3, 12.4, 2.6b, 3.6, L §1.3.4-5
Proportional Hazards Plots

1. Stratify on one of the variables
Proportional Hazards Plots

1. Stratify on one of the variables
   - Perhaps by making continuous variable discrete.

2. Fit Cox model and calculate baseline $\hat{H}$ for each strata

3. Plot quantities involving fitted hazard by strata:
   - $\log \hat{H}_j$ vs time
     - see if they have constant separation
   - $\hat{H}_j$ vs $\hat{H}_1$ and see if it is a straight line through zero
     - Called Andersen plot
   - $\log \hat{H}_j$ vs $\log \hat{H}_1$ and see if it is a straight line
     - R Code

4. For all plots, keep in mind that they are more variable in the tails
Proportional Hazards Plots

1. Stratify on one of the variables
   1. Perhaps by making continuous variable discrete.

2. Fit Cox model and calculate baseline $\hat{H}$ for each strata

   - $\log H_j$ vs time
     - see if they have constant separation
   - $H_j$ vs $H_1$ and see if it is a straight line through zero
     - Called Andersen plot
   - $\log H_j$ vs $\log H_1$ and see if it is a straight line
     1. If convex then $h_j(t)/h_0(t)$ increases with time
     2. If concave then $h_j(t)/h_0(t)$ decreases with time

3. For all plots, keep in mind that they are more variable in the tails
Proportional Hazards Plots

1. Stratify on one of the variables
   - Perhaps by making continuous variable discrete.

2. Fit Cox model and calculate baseline $\hat{H}$ for each strata

3. Plot quantities involving fitted hazard by strata:

   1. $\log\hat{H}_j$ vs time
   2. $\hat{H}_j$ vs $\hat{H}_1$ and see if it is a straight line through zero
   3. $\log\hat{H}_j$ vs $\log\hat{H}_1$ and see if it is a straight line

   Called Andersen plot

   If convex then $h_j(t)/h_0(t)$ increases with time

   If concave then $h_j(t)/h_0(t)$ decreases with time

   For all plots, keep in mind that they are more variable in the tails
Proportional Hazards Plots

1. Stratify on one of the variables
   - Perhaps by making continuous variable discrete.

2. Fit Cox model and calculate baseline $\hat{H}$ for each strata.

3. Plot quantities involving fitted hazard by strata:
   - $\log H_j$ vs time

4. For all plots, keep in mind that they are more variable in the tails.
Proportional Hazards Plots

1. Stratify on one of the variables
   - Perhaps by making continuous variable discrete.
2. Fit Cox model and calculate baseline $\hat{H}$ for each strata
3. Plot quantities involving fitted hazard by strata:
   - $\log H_j$ vs time
     - see if they have constant separation  
   - $\log H_j$ vs $\log H_1$ and see if it is a straight line

   R Code

   Called Andersen plot

   If convex then $h_j(t)/h_0(t)$ increases with time

   If concave then $h_j(t)/h_0(t)$ decreases with time

4. For all plots, keep in mind that they are more variable in the tails
Proportional Hazards Plots

1. Stratify on one of the variables
   - Perhaps by making continuous variable discrete.

2. Fit Cox model and calculate baseline $\hat{H}$ for each strata

3. Plot quantities involving fitted hazard by strata:
   - $\log H_j$ vs time
     - see if they have constant separation R Code
   - $H_j$ vs $H_1$ and see if it is a straight line through zero
Proportional Hazards Plots

1. Stratify on one of the variables
   1. Perhaps by making continuous variable discrete.

2. Fit Cox model and calculate baseline $\hat{H}$ for each strata

3. Plot quantities involving fitted hazard by strata:
   1. $\log H_j$ vs time
      1. see if they have constant separation  
         R Code
   2. $H_j$ vs $H_1$ and see if it is a straight line through zero
      1. Called *Andersen plot*
Proportional Hazards Plots

1. Stratify on one of the variables
   - Perhaps by making continuous variable discrete.
2. Fit Cox model and calculate baseline \( \hat{H} \) for each strata
3. Plot quantities involving fitted hazard by strata:
   - \( \log H_j \) vs time
     - see if they have constant separation  R Code
   - \( H_j \) vs \( H_1 \) and see if it is a straight line through zero
     - Called Andersen plot
   - \( \log H_j \) vs \( \log H_1 \) and see if it is a straight line  R Code
Proportional Hazards Plots

1. Stratify on one of the variables
   1. Perhaps by making continuous variable discrete.

2. Fit Cox model and calculate baseline \( \hat{H} \) for each strata

3. Plot quantities involving fitted hazard by strata:
   1. \( \log H_j \) vs time
      1. see if they have constant separation  
         R Code
   2. \( H_j \) vs \( H_1 \) and see if it is a straight line through zero
      1. Called Andersen plot
   3. \( \log H_j \) vs \( \log H_1 \) and see if it is a straight line  
      R Code
      1. If convex then \( h_j(t)/h_0(t) \) increases with time
Proportional Hazards Plots

1. Stratify on one of the variables
   1. Perhaps by making continuous variable discrete.

2. Fit Cox model and calculate baseline $\hat{H}$ for each strata

3. Plot quantities involving fitted hazard by strata:
   1. $\log H_j$ vs time
      1. see if they have constant separation   R Code
   2. $H_j$ vs $H_1$ and see if it is a straight line through zero
      1. Called *Andersen plot*
   3. $\log H_j$ vs $\log H_1$ and see if it is a straight line   R Code
      1. If convex then $h_j(t)/h_0(t)$ increases with time
      2. If concave then $h_j(t)/h_0(t)$ decreases with time
Proportional Hazards Plots

1. Stratify on one of the variables
   - Perhaps by making continuous variable discrete.
2. Fit Cox model and calculate baseline \( \hat{H} \) for each strata
3. Plot quantities involving fitted hazard by strata:
   - \( \log H_j \) vs time
     - see if they have constant separation \( \text{R Code} \)
   - \( H_j \) vs \( H_1 \) and see if it is a straight line through zero
     - Called Andersen plot
   - \( \log H_j \) vs \( \log H_1 \) and see if it is a straight line \( \text{R Code} \)
     - If convex then \( h_j(t)/h_0(t) \) increases with time
     - If concave then \( h_j(t)/h_0(t) \) decreases with time
4. For all plots, keep in mind that they are more variable in the tails
Construction: Plot cumulative events vs time by group

1. Calculate baseline hazard for a new variable $W$.
2. Stratify the sample based on $W$.
3. For each time point, truncate the population at that time.
4. Sum the Cox and Snell residuals (total amount of cumulative hazard).
5. Calculate the cumulative number of events.

If $W$ is unrelated, the plot should look like a 45° line.
If $W$ is related according to the PH model, the curves are linear with different slope.
If $W$ is related but violates the PH model, the curves are nonlinear.
Arjas Plots

1. Construction: Plot cumulative events vs time by group
   1. Calculate baseline hazard

2. For a new variable $W$, stratify sample based on it
3. For each time point,
   1. Truncate population at that time
   2. Sum the Cox and Snell Residuals (total amount of cumulative hazard)
   3. Calculate the cumulative number of events

2. If $W$ is unrelated, plot should look like $45^\circ$ line
3. If $W$ is related according to PH model, curves are linear with different slope
4. If $W$ is related but violates PH model, curves are nonlinear

R Code
Arjas Plots

Construction: Plot cumulative events vs time by group

1. Calculate baseline hazard
2. For a new variable $W$, stratify sample based on it
Arjas Plots

Construction: Plot cumulative events vs time by group

1. Calculate baseline hazard
2. For a new variable $W$, stratify sample based on it
3. For each time point,

**Arjas Plots**

1. **Construction:** Plot cumulative events vs time by group
   - 1. Calculate baseline hazard
   - 2. For a new variable $W$, stratify sample based on it
   - 3. For each time point,
     - 1. Truncate population at that time

R Code

Regression Models: Diagnostics Lecture 09
**Arjas Plots**

1. **Construction:** Plot cumulative events vs time by group
   1. Calculate baseline hazard
   2. For a new variable $W$, stratify sample based on it
   3. For each time point,
      1. Truncate population at that time
      2. Sum the Cox and Snell Residuals (total amount of cumulative hazard)

   If $W$ is unrelated, plot should look like a $45^\circ$ line
   If $W$ is related according to PH model, curves are linear with different slope
   If $W$ is related but violates PH model, curves are nonlinear

---

R Code

Regression Models: Diagnostics Lecture 09
Construction: Plot cumulative events vs time by group

1. Calculate baseline hazard
2. For a new variable \( W \), stratify sample based on it
3. For each time point,
   1. Truncate population at that time
   2. Sum the Cox and Snell Residuals (total amount of cumulative hazard)
   3. Calculate the cumulative number of events

If \( W \) is unrelated, plot should look like 45\(^\circ\) line

If \( W \) is related according to PH model, curves are linear with different slope

If \( W \) is related but violates PH model, curves are nonlinear

R Code

Regression Models: Diagnostics Lecture 09
Arjas Plots

1. **Construction:** Plot cumulative events vs time by group
   1. Calculate baseline hazard
   2. For a new variable $W$, stratify sample based on it
   3. For each time point,
      1. Truncate population at that time
      2. Sum the Cox and Snell Residuals (total amount of cumulative hazard)
      3. Calculate the cumulative number of events

2. If $W$ is unrelated, plot should look like 45° line
Arjas Plots

1. Construction: Plot cumulative events vs time by group
   1. Calculate baseline hazard
   2. For a new variable $W$, stratify sample based on it
   3. For each time point,
      1. Truncate population at that time
      2. Sum the Cox and Snell Residuals (total amount of cumulative hazard)
      3. Calculate the cumulative number of events

2. If $W$ is unrelated, plot should look like 45° line
   1. If $W$ is related according to PH model, curves are linear with different slope
Arjas Plots

1. Construction: Plot cumulative events vs time by group
   1. Calculate baseline hazard
   2. For a new variable $W$, stratify sample based on it
   3. For each time point,
      1. Truncate population at that time
      2. Sum the Cox and Snell Residuals (total amount of cumulative hazard)
      3. Calculate the cumulative number of events

2. If $W$ is unrelated, plot should look like 45° line
   1. If $W$ is related according to PH model, curves are linear with different slope
   2. If $W$ is related but violates PH model, curves are nonlinear
Counting Process

1. Let $N_i(t)$ represent the number of observations experienced by subject $i$ up to and including time $t$. 

$
N_i(t) \in \{0, 1\}.
$

This is a counting process $N_i(0) = 0$ right-continuous, piecewise constant, with a jump of size 1.

$
\Lambda(t) = \mathbb{E}[N(t) | F_t]
$

for $F_t$ the information available just before $t$, called the cumulative intensity process.
Counting Process

Let $N_i(t)$ represent the number of observations experienced by subject $i$ up to and including time $t$.

$N_i(t) \in \{0, 1\}$. 

This is a counting process: $N_i(0) = 0$ and right-continuous, piecewise constant, with a jump of size 1.

$N(t) = \sum_i N_i(t)$ also a counting process.

$\Lambda(t) = E[N(t)|F_t]$ for $F_t$ the information available just before $t$, called the cumulative intensity process.
Counting Process

1. Let $N_i(t)$ represent the number of observations experienced by subject $i$ up to and including time $t$.
   - $N_i(t) \in \{0, 1\}$.

2. This is a *counting process*
Counting Process

1. Let $N_i(t)$ represent the number of observations experienced by subject $i$ up to and including time $t$.
   - $N_i(t) \in \{0, 1\}$.

2. This is a counting process
   - $N_i(0) = 0$
Counting Process

1. Let $N_i(t)$ represent the number of observations experienced by subject $i$ up to and including time $t$.
   1. $N_i(t) \in \{0, 1\}$.
2. This is a *counting process*
   1. $N_i(0) = 0$
   2. Right-continuous.
Counting Process

1. Let $N_i(t)$ represent the number of observations experienced by subject $i$ up to and including time $t$.
   - $N_i(t) \in \{0, 1\}$.
2. This is a counting process
   - $N_i(0) = 0$
   - Right-continuous.
   - Piecewise constant.

$N_i(t)$ also a counting process.

$\Lambda(t) = E[N(t)|F_t]$ for $F_t = \text{Information available just before } t$ called the cumulative intensity process.
Counting Process

Let $N_i(t)$ represent the number of observations experienced by subject $i$ up to and including time $t$.

1. $N_i(t) \in \{0, 1\}$.

2. This is a *counting process*

   1. $N_i(0) = 0$
   2. Right-continuous.
   3. Piecewise constant.

$N(t) = \sum_i N_i(t)$ also a counting process.

$\Lambda(t) = E[N(t) | F_t]$ for $F_t$ = Information available just before $t$ called the cumulative intensity process.
Counting Process

1. Let $N_i(t)$ represent the number of observations experienced by subject $i$ up to and including time $t$.
   
   1. $N_i(t) \in \{0, 1\}$.

2. This is a counting process
   
   1. $N_i(0) = 0$
   2. Right-continuous.
   3. Piecewise constant.

3. $N(t) = \sum_i N_i(t)$ also a counting process.
Counting Process

1. Let $N_i(t)$ represent the number of observations experienced by subject $i$ up to and including time $t$. 
   - $N_i(t) \in \{0, 1\}$.

2. This is a counting process
   - $N_i(0) = 0$
   - Right-continuous.
   - Piecewise constant.
   - Jump of size 1.

3. $N(t) = \sum_i N_i(t)$ also a counting process.

4. $\Lambda(t) = \mathbb{E} [N(t) | \mathcal{F}_t]$
Counting Process

1. Let $N_i(t)$ represent the number of observations experienced by subject $i$ up to and including time $t$.
   - $N_i(t) \in \{0, 1\}$.
2. This is a counting process
   - $N_i(0) = 0$
   - Right-continuous.
   - Piecewise constant.
   - Jump of size 1.
3. $N(t) = \sum_i N_i(t)$ also a counting process.
4. $\Lambda(t) = \mathbb{E}[N(t)|\mathcal{F}_t]$
   - for $\mathcal{F}_t = $ Information available just before $t$
Counting Process

1. Let \( N_i(t) \) represent the number of observations experienced by subject \( i \) up to and including time \( t \).
   - \( N_i(t) \in \{0, 1\} \).

2. This is a counting process
   - \( N_i(0) = 0 \)
   - Right-continuous.
   - Piecewise constant.
   - Jump of size 1.

3. \( N(t) = \sum_i N_i(t) \) also a counting process.

4. \( \Lambda(t) = \mathbb{E} [N(t) \mid \mathcal{F}_t] \)
   - for \( \mathcal{F}_t = \) Information available just before \( t \)
   - called the cumulative intensity process.
Martingales

Then $M(t) = N(t) - \Lambda(t)$ satisfies $E[M(t)|\mathcal{F}_s] = M(s)$ for $s < t$,
1. Then $M(t) = N(t) - \Lambda(t)$ satisfies $\mathbb{E}[M(t)|F_s] = M(s)$ for $s < t$,
2. Then $M(t)$ is called a Martingale.
Martingales

1. Then $M(t) = N(t) - \Lambda(t)$ satisfies $\mathbb{E}[M(t)|\mathcal{F}_s] = M(s)$ for $s < t$,
2. Then $M(t)$ is called a Martingale.
   1. Expectation always 0, but variance rises with $t$
Martingales

1. Then $M(t) = N(t) - \Lambda(t)$ satisfies $\mathbb{E}[M(t)|\mathcal{F}_s] = M(s)$ for $s < t$.

2. Then $M(t)$ is called a Martingale.

   - Expectation always 0, but variance rises with $t$

   - For $s < t$,

     \[
     \text{Var}[M(t)] = \mathbb{E}[\text{Var}[M(t)|M(s)]] + \text{Var}[\mathbb{E}[M(t)|M(s)]] \\
     = \mathbb{E}[\text{Var}[M(t)|M(s)]] + \text{Var}[M(s)] \\
     \geq \text{Var}[M(s)]
     \]
Martingale Residuals

1. Define as $\hat{M}_j = \delta_j - R_j$
Martingale Residuals

1. Define as $\hat{M}_j = \delta_j - R_j$
2. Sum is $\approx 0$
Martingale Residuals

1. Define as $\hat{M}_j = \delta_j - R_j$
2. Sum is $\approx 0$

   1. Exact for $\hat{H}$ of above
Martingale Residuals

1. Define as $\hat{M}_j = \delta_j - R_j$
2. Sum is $\approx 0$
   1. Exact for $\hat{H}$ of above
3. $E\left[\hat{M}_j\right] = 0$
Martingale Residuals

1. Define as $\hat{M}_j = \delta_j - R_j$
2. Sum is $\approx 0$
   1. Exact for $\hat{H}$ of above
3. $\mathbb{E} \left[ \hat{M}_j \right] = 0$
   1. If subgroup has typically $+$ residuals, more fail than model predicts
Martingale Residuals

1. Define as $\hat{M}_j = \delta_j - R_j$
2. Sum is $\approx 0$
   - Exact for $\hat{H}$ of above
3. $\mathbb{E} \left[ \hat{M}_j \right] = 0$
   - If subgroup has typically $+$ residuals, more fail than model predicts
   - If subgroup has typically $-$ residuals, fewer fail than model predicts

Code
Cox and Snell residuals are approximately linear in proper transformation of missing variable
Cox and Snell residuals are approximately linear in proper transformation of missing variable

Denote the Cox and Snell residuals as

\[ R_j = S_j \exp(-\gamma W_j) / w^* \]

where \( w^* \) is whatever is needed to correct for added stuff in denominators.

Hence \( E[\hat{M}_j] \) is linear in missing variate.

To determine \( f \) that makes \( W = f(V) \) the right scale for some covariate \( V \), plot residuals vs \( V \).
Continued

1. Cox and Snell residuals are approximately linear in proper transformation of missing variable
   - Denote the Cox and Snell residuals as
     - $S_j$ for true model
Cox and Snell residuals are approximately linear in proper transformation of missing variable

Denote the Cox and Snell residuals as

- $S_j$ for true model
- $R_j$ for false model

$$R_j = S_j \exp(-\gamma W_j)/w^*$$

$$E[\hat{M}_j] = E[\delta_j - R_j] = E[\delta_j - S_j] + E[S_j] \approx -E[S_j] (\gamma W_j - \log(w^*))$$

Hence, $E[\hat{M}_j]$ is linear in missing variate

To determine $f$ that makes $W = f(V)$ the right scale for some covariate $V$, plot residuals vs $V$.

Fit nonparametric curve. R Code
Cox and Snell residuals are approximately linear in proper transformation of missing variable

Denote the Cox and Snell residuals as

- $S_j$ for true model
- $R_j$ for false model

\[ R_j = S_j \exp(-\gamma W_j)/w^* \]
Cox and Snell residuals are approximately linear in proper transformation of missing variable.

Denote the Cox and Snell residuals as

1. $S_j$ for true model
2. $R_j$ for false model

$R_j = S_j \exp(-\gamma W_j)/w^*$

$w^*$ is whatever is needed to correct for added stuff in denominators.
Cox and Snell residuals are approximately linear in proper transformation of missing variable

Denote the Cox and Snell residuals as

- $S_j$ for true model
- $R_j$ for false model

\[ R_j = S_j \exp(-\gamma W_j)/w^* \]

- $w^*$ is whatever is needed to correct for added stuff in denominators

\[ E[\hat{M}_j] = E[\hat{\delta}_j - R_j] = E[\hat{\delta}_j - S_j] + E[S_j][1 - \exp(-\gamma W_j)/w^*] \approx -E[S_j][\gamma W_j - \log(w^*)] \]
Cox and Snell residuals are approximately linear in proper transformation of missing variable

1. Denote the Cox and Snell residuals as
   1. \( S_j \) for true model
   2. \( R_j \) for false model

2. \( R_j = S_j \exp(-\gamma W_j) / w^* \)
   1. \( w^* \) is whatever is needed to correct for added stuff in denominators

3. \( \mathbb{E} \left[ \hat{M}_j \right] = \mathbb{E} [\hat{\delta}_j - R_j] = \mathbb{E} [\hat{\delta}_j - S_j] + \mathbb{E} [S_j] \left[ 1 - \exp(-\gamma W_j) / w^* \right] \approx -\mathbb{E} [S_j] [\gamma W_j - \log(w^*)] \)

4. Hence \( \mathbb{E} \left[ \hat{M}_j \right] \) is linear in missing variate
Cox and Snell residuals are approximately linear in proper transformation of missing variable

Denote the Cox and Snell residuals as

- $S_j$ for true model
- $R_j$ for false model

$$R_j = S_j \exp(-\gamma W_j)/w^*$$

$w^*$ is whatever is needed to correct for added stuff in denominators

$$E \left[ \hat{M}_j \right] = E \left[ \delta_j - R_j \right] = E \left[ \delta_j - S_j \right] + E \left[ S_j \right] \left[ 1 - \exp(-\gamma W_j)/w^* \right] \approx -E \left[ S_j \right] \left[ \gamma W_j - \log(w^*) \right]$$

Hence $E \left[ \hat{M}_j \right]$ is linear in missing variate

To determine $f$ that makes $W = f(V)$ the right scale for some covariate $V$, plot residuals vs $V$
Cox and Snell residuals are approximately linear in proper transformation of missing variable

1. Denote the Cox and Snell residuals as
   1. \( S_j \) for true model
   2. \( R_j \) for false model

2. \( R_j = S_j \exp(-\gamma W_j)/w^* \)
   1. \( w^* \) is whatever is needed to correct for added stuff in denominators

3. \( E \left[ \hat{M}_j \right] = E [\delta_j - R_j] = E [\delta_j - S_j] + E [S_j] \left[ 1 - \exp(-\gamma W_j)/w^* \right] \approx -E [S_j] \left[ \gamma W_j - \log(w^*) \right] \)

4. Hence \( E \left[ \hat{M}_j \right] \) is linear in missing variate

5. To determine \( f \) that makes \( W = f(V) \) the right scale for some covariate \( V \), plot residuals vs \( V \)
   1. Fit nonparametric curve. R Code
Deviance Residuals

1. Rescale martingale residuals so that they are closer to normal
Rescale martingale residuals so that they are closer to normal.

Derivation:

Consider letting each observation have its own coefficient vector.
Maximize likelihood under this model. 
Calculate 
\[ -2 \times \text{difference in contribution to } \ell \]
\[ \approx \chi^2 \]
Square root (with same sign as \( \hat{M}_i \))
\[ \approx N(0, 1) \]
Formula is 
\[ D_i = \text{sign}(\hat{M}_i) \sqrt{-2(\hat{M}_i + \delta_i \log(\delta_i - \hat{M}_i))}. \]
Deviance Residuals

1. Rescale martingale residuals so that they are closer to normal
2. Derivation:
   1. Consider letting each observation have its own coefficient vector
Rescale martingale residuals so that they are closer to normal

Derivation:

1. Consider letting each observation have its own coefficient vector
2. Maximize likelihood under this model
Deviance Residuals

1. Rescale martingale residuals so that they are closer to normal

2. Derivation:
   1. Consider letting each observation have its own coefficient vector
   2. Maximize likelihood under this model
   3. Calculate $-2 \times$ difference in contribution to $\ell$ from each observation
      $\approx \chi_1^2$
**Deviance Residuals**

1. Rescale martingale residuals so that they are closer to normal

2. **Derivation:**
   1. Consider letting each observation have its own coefficient vector
   2. Maximize likelihood under this model
   3. Calculate $-2 \times$ difference in contribution to $\ell$ from each observation
      $\approx \chi^2_1$
   4. Square root (with same sign as $\hat{M}_i$) $\approx N(0, 1)$
Deviance Residuals

1. Rescale martingale residuals so that they are closer to normal

2. Derivation:
   1. Consider letting each observation have its own coefficient vector
   2. Maximize likelihood under this model
   3. Calculate $-2 \times$ difference in contribution to $\ell$ from each observation $
   \approx \chi^2_1$
   4. Square root (with same sign as $\hat{M}_i$) \( \approx \mathcal{N}(0, 1) \)

3. Formula is $D_i = \text{sign}(\hat{M}_i) \sqrt{-2(\hat{M}_i + \delta_i \log(\delta_i - \hat{M}_i))}$. 

Plot vs $z_j \hat{\beta}$ 

Better job of assessing relative residual size, since they are closer to normal.
Deviance Residuals

1. Rescale martingale residuals so that they are closer to normal

2. Derivation:
   1. Consider letting each observation have its own coefficient vector
   2. Maximize likelihood under this model
   3. Calculate $-2 \times$ difference in contribution to $\ell$ from each observation
      \[ \approx \chi_1^2 \]
   4. Square root (with same sign as $\hat{M}_i$) \[ \approx N(0, 1) \]

3. Formula is $D_i = \text{sign}(\hat{M}_i) \sqrt{-2(\hat{M}_i + \delta_i \log(\delta_i - \hat{M}_i))}.$

4. Plot vs $z_j\hat{\beta}$
Deviance Residuals

1. Rescale martingale residuals so that they are closer to normal

2. Derivation:
   1. Consider letting each observation have its own coefficient vector
   2. Maximize likelihood under this model
   3. Calculate $-2 \times$ difference in contribution to $\ell$ from each observation
      \[ \approx \chi_1^2 \]
   4. Square root (with same sign as $\hat{M}_i$) \[ \approx N(0, 1) \]

3. Formula is $D_i = \text{sign}(\hat{M}_i) \sqrt{-2(\hat{M}_i + \delta_i \log(\delta_i - \hat{M}_i))}$.

4. Plot vs $z_j \hat{\beta}$
   1. Better job of assessing relative residual size, since they are closer to normal. R Code
Score Residuals

1. Subject’s contribution to derivative of log partial likelihood.

Express partial log likelihood as

$$\ell(\beta) = \sum_j \ell_j(\beta)$$

1. Equals number of events by time \(t\), minus expected number, weighted by covariate

Score Residuals

1. Subject’s contribution to derivative of log partial likelihood.
   - Express partial log likelihood as $\ell(\beta) = \sum_j \ell_j(\beta)$
Score Residuals

1. Subject’s contribution to derivative of log partial likelihood.
   1. Express partial log likelihood as $\ell(\beta) = \sum_j \ell_j(\beta)$
   1. $\ell_j(\beta) = 0$ if subject $j$ censored.
Score Residuals

1. Subject’s contribution to derivative of log partial likelihood.
   1. Express partial log likelihood as \( \ell(\beta) = \sum_j \ell_j(\beta) \)
   2. \( \ell_j(\beta) = 0 \) if subject \( j \) censored.
   2. Define score process \( U(\beta; t) = \frac{d}{d\beta_k} \ell(\beta) \)
Score Residuals

1. Subject’s contribution to derivative of log partial likelihood.
   1. Express partial log likelihood as $\ell(\beta) = \sum_j \ell_j(\beta)$
   2. $\ell_j(\beta) = 0$ if subject $j$ censored.

2. Define score process $U(\beta; t) = \frac{d}{d\beta_k} \ell(\beta)$

3. Equals number of events by time $t$, minus expected number, weighted by covariate
Score Residuals

1. Subject’s contribution to derivative of log partial likelihood.

   1. Express partial log likelihood as $\ell(\beta) = \sum_j \ell_j(\beta)$

   2. $\ell_j(\beta) = 0$ if subject $j$ censored.

2. Define score process $\mathbf{U}(\beta; t) = \frac{d}{d\beta_k} \ell(\beta)$

3. Equals number of events by time $t$, minus expected number, weighted by covariate


R Code
Measuring influence

1. Look for effect of observation $k$ on $\hat{\beta}_j$. 

$$
\sum_{m} \left( Z_{mj} - \sum_{r \in R} Z_{rj} \exp(z_j \beta) \right) \approx \ell''(\beta) \left( \beta - \hat{\beta} \right)
$$

Let $\hat{\beta}$ be estimator for full data set

$\ell$ be log likelihood for full data

$$
\ell'\left( \beta \right) \approx \ell''\left( \beta \right) \left( \beta - \hat{\beta} \right)
$$

Approximate $\ell''\left( \beta \right)$ by $\ell''\left( \hat{\beta} \right)$

Then $\hat{\gamma} - \hat{\beta} \approx \ell''\left( \hat{\beta} \right) - 1 \left[ \ell\left( \beta \right) - \ell\left( \hat{\beta} \right) \right]$

for $\ell$ be log likelihood for reduced data and $\hat{\gamma}$ be estimator with observation $k$ removed

Contributions to $\left[ \ell\left( \beta \right) - \ell\left( \hat{\beta} \right) \right]$ are approximated by score residuals.

Approximation is acceptable for small deviations, and for large deviations we don't really need to know how large. R calls these residuals $\text{dfbeta}$. SAS Code R Code
Measuring influence

1. Look for effect of observation $k$ on $\hat{\beta}_j$

2. \[ \ell^j(\hat{\beta}) = \sum_m \left( Z_{mj} - \frac{\sum_{r \in R} Z_{rj} \exp(z_j \beta)}{\sum_{r \in R} \exp(z_j \beta)} \right). \]
Measuring influence

1. Look for effect of observation $k$ on $\hat{\beta}_j$

$$\ell^j(\hat{\beta}) = \sum_m \left( Z_{mj} - \frac{\sum_{r \in R} Z_{rj} \exp(z_j \beta)}{\sum_{r \in R} \exp(z_j \beta)} \right).$$

2. Let
Measuring influence

1. Look for effect of observation $k$ on $\hat{\beta}_j$

$$\ell^j(\hat{\beta}) = \sum_m \left( Z_{mj} - \frac{\sum_{r \in R} Z_{ij} \exp(z_{ij}\beta)}{\sum_{r \in R} \exp(z_{ij}\beta)} \right).$$

2. Let $\hat{\beta}$ be estimator for full data set

3. Contributions to $[\ell'(\beta) - \ell'(\beta)\dagger]$ are approximated by score residuals.

4. Approximation is acceptable for small deviations, and for large deviations we don't really need to know how large

5. R calls these residuals dfbeta. SAS Code R Code
Measuring influence

1. Look for effect of observation $k$ on $\hat{\beta}_j$

$$\ell^j(\hat{\beta}) = \sum_{m} \left( Z_{mj} - \frac{\sum_{r \in R} Z_{rj} \exp(z_j \beta)}{\sum_{r \in R} \exp(z_j \beta)} \right).$$

2. Let

1. $\hat{\beta}$ be estimator for full data set
2. $\ell$ be log likelihood for full data
Measuring influence

1. Look for effect of observation $k$ on $\hat{\beta}_j$
   \[ \ell^j(\hat{\beta}) = \sum_m \left( Z_{mj} - \frac{\sum_{r \in \mathcal{R}} Z_{rij} \exp(z_j \beta)}{\sum_{r \in \mathcal{R}} \exp(z_j \beta)} \right). \]

2. Let
   - $\hat{\beta}$ be estimator for full data set
   - $\ell$ be log likelihood for full data

3. $\ell'(\beta) \approx \ell''(\beta)(\beta - \hat{\beta})$

4. Contributions to $\ell'(\beta) - \ell'(\hat{\beta})$ are approximated by score residuals.

5. Approximation is acceptable for small deviations, and for large deviations we don't really need to know how large

R calls these residuals $\text{dfbeta}$. SAS Code R Code
Measuring influence

1. Look for effect of observation $k$ on $\hat{\beta}_j$

   $\ell^i(\hat{\beta}) = \sum_m \left( Z_{mj} - \frac{\sum_{r \in R} Z_{rj} \exp(z_{j}\beta)}{\sum_{r \in R} \exp(z_{j}\beta)} \right)$.

2. Let
   - $\hat{\beta}$ be estimator for full data set
   - $\ell$ be log likelihood for full data

3. $\ell'(\beta) \approx \ell''(\beta)(\beta - \hat{\beta})$

4. $\ell'_\dagger(\beta) \approx \ell''_{\dagger}(\beta)(\beta - \hat{\gamma})$
Measuring influence

1. Look for effect of observation $k$ on $\hat{\beta}_j$

\[ \ell^i(\hat{\beta}) = \sum_m \left( Z_{mj} - \frac{\sum_{r \in \mathcal{R}} Z_{rj} \exp(z_j \beta)}{\sum_{r \in \mathcal{R}} \exp(z_j \beta)} \right). \]

2. Let

- $\hat{\beta}$ be estimator for full data set
- $\ell$ be log likelihood for full data

3. $\ell'(\beta) \approx \ell''(\beta)(\beta - \hat{\beta})$

4. $\ell'_\dagger(\beta) \approx \ell''_\dagger(\beta)(\beta - \hat{\gamma})$

5. Approximate $\ell''_\dagger(\beta)$ by $\ell''(\beta)$
Measuring influence

1. Look for effect of observation $k$ on $\hat{\beta}_j$

$$
\ell^j(\hat{\beta}) = \sum_m \left( Z_{mj} - \frac{\sum_{r \in R} Z_{rj} \exp(z_j \beta)}{\sum_{r \in R} \exp(z_j \beta)} \right).
$$

2. Let
   - $\hat{\beta}$ be estimator for full data set
   - $\ell$ be log likelihood for full data

3. $\ell'(\beta) \approx \ell''(\beta)(\beta - \hat{\beta})$

4. $\ell^\dagger'(\beta) \approx \ell^\dagger''(\beta)(\beta - \hat{\gamma})$

5. Approximate $\ell^\dagger''(\beta)$ by $\ell''(\beta)$

2. Then $\hat{\gamma} - \hat{\beta} \approx \ell''(\hat{\beta})^{-1}[\ell'(\beta) - \ell^\dagger'(\beta)]$ for $\ell^\dagger$ be log likelihood for reduced data and $\hat{\gamma}$ be estimator with observation $k$ removed.
Measuring influence

1. Look for effect of observation $k$ on $\hat{\beta}_j$

   $\ell^j(\hat{\beta}) = \sum_m \left(Z_{mj} - \frac{\sum_{r \in R} Z_{rj} \exp(z_{j\beta})}{\sum_{r \in R} \exp(z_{j\beta})}\right)$.

2. Let
   - $\hat{\beta}$ be estimator for full data set
   - $\ell$ be log likelihood for full data

3. $\ell'(\beta) \approx \ell''(\beta)(\beta - \hat{\beta})$

4. $\ell^\dagger'(\beta) \approx \ell''(\beta)(\beta - \hat{\gamma})$

5. Approximate $\ell''(\beta)$ by $\ell''(\beta)$

Then $\hat{\gamma} - \hat{\beta} \approx \ell''(\hat{\beta})^{-1}[\ell'(\beta) - \ell^\dagger'(\beta)]$ for $\ell^\dagger$ be log likelihood for reduced data and $\hat{\gamma}$ be estimator with observation $k$ removed

Contributions to $[\ell'(\beta) - \ell^\dagger'(\beta)]$ are approximated by score residuals.
Measuring influence

1. Look for effect of observation $k$ on $\hat{\beta}_j$
\[
\ell^i(\hat{\beta}) = \sum_m \left( Z_{mj} - \frac{\sum_{r \in \mathcal{R}} Z_{rj} \exp(z_j \beta)}{\sum_{r \in \mathcal{R}} \exp(z_j \beta)} \right).
\]

2. Let
   1. $\hat{\beta}$ be estimator for full data set
   2. $\ell$ be log likelihood for full data
\[
\ell'(\beta) \approx \ell''(\beta)(\beta - \hat{\beta})
\]
\[
\ell'(\beta) \approx \ell''(\beta)(\beta - \hat{\gamma})
\]

3. Approximate $\ell''(\beta)$ by $\ell''(\beta)$

4. Then $\hat{\gamma} - \hat{\beta} \approx \ell''(\hat{\beta})^{-1}[\ell'(\beta) - \ell'(\beta)]$ for $\ell$ be log likelihood for reduced data and $\hat{\gamma}$ be estimator with observation $k$ removed

5. Contributions to $[\ell'(\beta) - \ell'(\beta)]$ are approximated by score residuals.

Approximation is
Measuring influence

1. Look for effect of observation $k$ on $\hat{\beta}_j$

$$\ell^j(\hat{\beta}) = \sum_m \left( Z_{mj} - \frac{\sum_{r \in R} Z_{rj} \exp(z_j \beta)}{\sum_{r \in R} \exp(z_j \beta)} \right).$$

2. Let

   - $\hat{\beta}$ be estimator for full data set
   - $\ell$ be log likelihood for full data

3. $\ell'(\beta) \approx \ell''(\beta)(\beta - \hat{\beta})$

4. $\ell'^{\dagger}(\beta) \approx \ell''^{\dagger}(\beta)(\beta - \hat{\gamma})$

5. Approximate $\ell''^{\dagger}(\beta)$ by $\ell''(\beta)$

2. Then $\hat{\gamma} - \hat{\beta} \approx \ell''(\hat{\beta})^{-1}[\ell'(\beta) - \ell'^{\dagger}(\beta)]$ for $\ell^{\dagger}$ be log likelihood for reduced data and $\hat{\gamma}$ be estimator with observation $k$ removed

3. Contributions to $[\ell'(\beta) - \ell'^{\dagger}(\beta)]$ are approximated by score residuals.

4. Approximation is

   - acceptable for small deviations,
Measuring influence

1. Look for effect of observation $k$ on $\hat{\beta}_j$

\[
\ell^j(\hat{\beta}) = \sum_m \left( Z_{mj} - \frac{\sum_{r \in R} Z_{rj} \exp(z_j \beta)}{\sum_{r \in R} \exp(z_j \beta)} \right).
\]

2. Let

- $\hat{\beta}$ be estimator for full data set
- $\ell$ be log likelihood for full data

3. $\ell'(\beta) \approx \ell''(\beta)(\beta - \hat{\beta})$

4. $\ell'_+(\beta) \approx \ell''_+(\beta)(\beta - \hat{\gamma})$

5. Approximate $\ell''_+(\beta)$ by $\ell''(\beta)$

Then $\hat{\gamma} - \hat{\beta} \approx \ell''(\hat{\beta})^{-1}[\ell'(\beta) - \ell'_+(\beta)]$ for $\ell_+$ be log likelihood for reduced data and $\hat{\gamma}$ be estimator with observation $k$ removed.

Contributions to $[\ell'(\beta) - \ell'_+(\beta)]$ are approximated by score residuals.

Approximation is

- acceptable for small deviations,
- and for large deviations we don’t really need to know how large
Measuring influence

1. Look for effect of observation $k$ on $\hat{\beta}_j$

$$\ell_j(\hat{\beta}) = \sum_m \left( Z_{mj} - \frac{\sum_{r \in R} Z_{rj} \exp(z_j \beta)}{\sum_{r \in R} \exp(z_j \beta)} \right).$$

2. Let
   - $\hat{\beta}$ be estimator for full data set
   - $\ell$ be log likelihood for full data

3. $\ell'(\beta) \approx \ell''(\beta)(\beta - \hat{\beta})$
4. $\ell_\uparrow'(\beta) \approx \ell''_\uparrow(\beta)(\beta - \hat{\gamma})$
5. Approximate $\ell''_\uparrow(\beta)$ by $\ell''(\beta)$

2. Then $\hat{\gamma} - \hat{\beta} \approx \ell''(\hat{\beta})^{-1}[\ell'(\beta) - \ell'_\uparrow(\beta)]$ for $\ell_\uparrow$ be log likelihood for reduced data and $\hat{\gamma}$ be estimator with observation $k$ removed

3. Contributions to $[\ell'(\beta) - \ell'_\uparrow(\beta)]$ are approximated by score residuals.

4. Approximation is
   - acceptable for small deviations,
   - and for large deviations we don’t really need to know how large

5. R calls these residuals $dfbeta$. SAS Code R Code
Section: Regression Models

Subsection: Parametric Failure Time Models:
Transformation and regression approach

1. Transform times so distn is on $(-\infty, \infty)$
Transformation and regression approach

1. Transform times so distn is on \((-\infty, \infty)\)
2. Generally using log
Use linear model for covariates

1 \[ \log(T_i) = \alpha + z_i \beta + \sigma U_i, \ U_i \text{ i.i.d.} \]
Use linear model for covariates

1. \( \log(T_i) = \alpha + z_i \beta + \sigma U_i, \ U_i \text{ i.i.d.} \)
2. \( T_i = \exp(\alpha + z_i \beta) \exp(U_i)^\sigma, \ U_i \text{ i.i.d.} \)
Use linear model for covariates

1. \( \log(T_i) = \alpha + z_i \beta + \sigma U_i, \ U_i \text{ i.i.d.} \)
2. \( T_i = \exp(\alpha + z_i \beta) \exp(U_i)^\sigma, \ U_i \text{ i.i.d.} \)
3. Let \( S_0 \) be survival function for \( \exp(U) \)
Use linear model for covariates

1. \( \log(T_i) = \alpha + z_i \beta + \sigma U_i, \ U_i \ \text{i.i.d.} \)

2. \( T_i = \exp(\alpha + z_i \beta) \exp(U_i)\sigma, \ U_i \ \text{i.i.d.} \)

3. Let \( S_0 \) be survival function for \( \exp(U) \)

4. Survival function for obsn \( i \) is

\[
S_i(t) = P[T_i > t] = P[\alpha + z_i \beta + \sigma U_i > \log(t)].
\]
Use linear model for covariates

1. \( \log(T_i) = \alpha + z_i \beta + \sigma U_i, \ U_i \text{ i.i.d.} \)

2. \( T_i = \exp(\alpha + z_i \beta) \exp(U_i)^\sigma, \ U_i \text{ i.i.d.} \)

Let \( S_0 \) be survival function for \( \exp(U) \)

Survival function for obsn \( i \) is

\[
S_i(t) = P[T_i > t] = P[\alpha + z_i \beta + \sigma U_i > \log(t)].
\]

Express in terms of \( U_i \):

\[
S_i(t) = P[\sigma U_i > \log(t) - z_i \beta - \alpha]
\]
Use linear model for covariates

1. \[ \log(T_i) = \alpha + z_i\beta + \sigma U_i, \ U_i \text{ i.i.d.}. \]
2. \[ T_i = \exp(\alpha + z_i\beta) \exp(U_i)^\sigma, \ U_i \text{ i.i.d.}. \]

Let \( S_0 \) be survival function for \( \exp(U) \)

Survival function for obsn \( i \) is

\[ S_i(t) = P[T_i > t] = P[\alpha + z_i\beta + \sigma U_i > \log(t)]. \]

1. Express in terms of \( U_i \):

\[ S_i(t) = P[\sigma U_i > \log(t) - z_i\beta - \alpha] \]

2. Express in terms of \( \exp(U_i) \):

\[ S_i(t) = P\left[\exp(U_i) > t^{1/\sigma} \exp(-z_i\beta/\sigma - \alpha/\sigma)\right] \]
Use linear model for covariates

1. \(\log(T_i) = \alpha + z_i \beta + \sigma U_i,\ U_i\ \text{i.i.d.}..\)

2. \(T_i = \exp(\alpha + z_i \beta) \exp(U_i)\sigma,\ U_i\ \text{i.i.d.}..\)

3. Let \(S_0\) be survival function for \(\exp(U)\)

4. Survival function for obsn \(i\) is

\[
S_i(t) = \mathbb{P}[T_i > t] = \mathbb{P}[\alpha + z_i \beta + \sigma U_i > \log(t)].
\]

1. Express in terms of \(U_i\):

\[
S_i(t) = \mathbb{P}[\sigma U_i > \log(t) - z_i \beta - \alpha]
\]

2. Express in terms of \(\exp(U_i)\):

\[
S_i(t) = \mathbb{P}\left[\exp(U_i) > t^{1/\sigma} \exp(-z_i \beta / \sigma - \alpha / \sigma)\right]
\]

3. Express in terms of \(S_0\):

\[
S_i(t) = S_0([t \exp(-\alpha - z_i \beta)]^{1/\sigma})
\]
Time scale for person with $z_j$ is on a time scale $\exp(-[z_j - z_k] \beta)$ times that of $z_k$: acceleration factor.
Model Interpretation

1. Time scale for person with $z_j$ is on a time scale $\exp(-[z_j - z_k]/\beta)$ times that of $z_k$: *acceleration factor*.

2. Estimating $\alpha$ and $\sigma$ usually equivalent to estimating which member $S_0$ is in a parametric family.
Time scale for person with \( z_j \) is on a time scale \( \exp(-[z_j - z_k]/\beta) \) times that of \( z_k \): *acceleration factor*.

Estimating \( \alpha \) and \( \sigma \) usually equivalent to estimating which member \( S_0 \) is in a parametric family

This model has an intercept
Model Interpretation

1. Time scale for person with $z_j$ is on a time scale $\exp(-[z_j - z_k]/\beta)$ times that of $z_k$: *acceleration factor*.

2. Estimating $\alpha$ and $\sigma$ usually equivalent to estimating which member $S_0$ is in a parametric family.

3. This model has an intercept
   - like normal theory, logistic regression, etc.
Model Interpretation

1. Time scale for person with \( z_j \) is on a time scale \( \exp(-[z_j - z_k]/\beta) \) times that of \( z_k \): *acceleration factor*.

2. Estimating \( \alpha \) and \( \sigma \) usually equivalent to estimating which member \( S_0 \) is in a parametric family.

3. This model has an intercept
   - like normal theory, logistic regression, etc.
   - unlike Cox regression
Model Interpretation

1. Time scale for person with $z_j$ is on a time scale $\exp(-[z_j - z_k]/\beta)$ times that of $z_k$: acceleration factor.

2. Estimating $\alpha$ and $\sigma$ usually equivalent to estimating which member $S_0$ is in a parametric family.

3. This model has an intercept
   1. like normal theory, logistic regression, etc.
   2. unlike Cox regression

4. Larger value of linear predictor $\Rightarrow$ longer life
Model Interpretation

1. Time scale for person with $z_j$ is on a time scale $\exp(-[z_j - z_k] \beta)$ times that of $z_k$: *acceleration factor*.

2. Estimating $\alpha$ and $\sigma$ usually equivalent to estimating which member $S_0$ is in a parametric family.

3. This model has an intercept
   1. like normal theory, logistic regression, etc.
   2. unlike Cox regression

4. Larger value of linear predictor $\Rightarrow$ longer life
   1. Recall relation for Cox model opposite.
1. Time scale for person with \( z_j \) is on a time scale \( \exp(-[z_j - z_k]/\beta) \) times that of \( z_k \): *acceleration factor*.

2. Estimating \( \alpha \) and \( \sigma \) usually equivalent to estimating which member \( S_0 \) is in a parametric family.

3. This model has an intercept
   - like normal theory, logistic regression, etc.
   - unlike Cox regression

4. Larger value of linear predictor \( \Rightarrow \) longer life
   - Recall relation for Cox model opposite.

5. Hence effect of covariates is to accelerate (or decelerate) the life: *accelerated failure model*.
Section: Regression Models

Subsection: Parametric Failure Time Distributions
Next step is to specify distribution for $U_i$. 

Potential distributions for $\log(T)$ will allow for shift and rescaling.
Overview of Parametric Failure Time Distributions

1. Next step is to specify distribution for $U_i$
2. Generally iid

Potential distributions for $\log(T)$ will allow for shift and rescaling.

Definition: Set of distns $G$ is a location and scale family if $\psi X + \phi \in G$ whenever $X \in G$ and $\psi > 0$ and $\phi \in \mathbb{R}$

Associated family turns a baseline $S_0$ into a location and scale family $G = \{S_0((x - \phi)/\psi) | \psi > 0\}$

Can estimate $\alpha, \sigma$ only if error distribution fixed to standard.

Pick one member of family to be the standard one

Some of our distributions don't have means and variances

Median 0, IQR 1?

Some simple choice of two parameters?

Even if we select standard value for location and scale, there might be other parameters in model

Analogue to least squares: skewness

Hence typically, we estimate family member rather than $\alpha, \sigma$.

Family member is the one with $\beta = 0$. 
Overview of Parametric Failure Time Distributions

1. Next step is to specify distribution for $U_i$:
   1. Generally iid
   2. Normal is possible
Overview of Parametric Failure Time Distributions

1. Next step is to specify distribution for $U_i$

   1. Generally iid
   2. Normal is possible
   3. We explore this, and alternatives.
Overview of Parametric Failure Time Distributions

1. Next step is to specify distribution for $U_i$
   - Generally iid
   - Normal is possible
   - We explore this, and alternatives.

2. Potential distributions for $\log(T)$ will allow for shift and rescaling.
Next step is to specify distribution for $U_i$:

1. Generally iid
2. Normal is possible
3. We explore this, and alternatives.

Potential distributions for $\log(T)$ will allow for shift and rescaling.

Definition: Set of distns $\mathcal{G}$ is a *location and scale family* if $\psi X + \phi \in \mathcal{G}$ whenever $X \in \mathcal{G}$ and $\psi > 0$ and $\phi \in \mathbb{R}$.

Can estimate $\alpha, \sigma$ only if error distribution fixed to standard.

Pick one member of family to be the standard one.

Some of our distributions don't have means and variances.

Median 0, IQR 1?

Some simple choice of two parameters?

Even if we select standard value for location and scale, there might be other parameters in model.

Analogue to least squares: skewness.

Hence typically, we estimate family member rather than $\alpha, \sigma$.

Family member is the one with $\beta = 0$. 

Overview of Parametric Failure Time Distributions

1. Next step is to specify distribution for $U_i$
   1. Generally iid
   2. Normal is possible
   3. We explore this, and alternatives.

2. Potential distributions for $\log(T)$ will allow for shift and rescaling.
   1. Definition: Set of distns $\mathcal{G}$ is a location and scale family if $\psi X + \phi \in \mathcal{G}$ whenever $X \in \mathcal{G}$ and $\psi > 0$ and $\phi \in \mathbb{R}$
   2. Associated family turns a baseline $S_0$ into a location and scale family
      $\mathcal{G} = \{S_0((x - \phi)/\psi)|\psi > 0\}$

Can estimate $\alpha, \sigma$ only if error distribution fixed to standard.

Pick one member of family to be the standard one

Some of our distributions don’t have means and variances

Median 0, IQR 1?

Even if we select standard value for location and scale, there might be other parameters in model

Analogue to least squares: skewness

Hence typically, we estimate family member rather than $\alpha, \sigma$.

Family member is the one with $\beta = 0$. 
Overview of Parametric Failure Time Distributions

1. Next step is to specify distribution for \( U_i \):
   1. Generally iid
   2. Normal is possible
   3. We explore this, and alternatives.

2. Potential distributions for \( \log(T) \) will allow for shift and rescaling.
   1. Definition: Set of distns \( \mathcal{G} \) is a \textit{location and scale family} if \( \psi X + \phi \in \mathcal{G} \)
      whenever \( X \in \mathcal{G} \) and \( \psi > 0 \) and \( \phi \in \mathbb{R} \)
   2. Associated family turns a baseline \( S_0 \) into a location and scale family
      \( \mathcal{G} = \{ S_0((x - \phi)/\psi)|\psi > 0 \} \)

3. Can estimate \( \alpha, \sigma \) only if error distribution fixed to standard.
Overview of Parametric Failure Time Distributions

1. Next step is to specify distribution for $U_i$
   1. Generally iid
   2. Normal is possible
   3. We explore this, and alternatives.

2. Potential distributions for $\log(T)$ will allow for shift and rescaling.
   1. Definition: Set of distns $\mathcal{G}$ is a location and scale family if $\psi X + \phi \in \mathcal{G}$ whenever $X \in \mathcal{G}$ and $\psi > 0$ and $\phi \in \mathbb{R}$
   2. Associated family turns a baseline $S_0$ into a location and scale family
      $\mathcal{G} = \{S_0((x - \phi)/\psi)|\psi > 0\}$

3. Can estimate $\alpha, \sigma$ only if error distribution fixed to standard.
   1. Pick one member of family to be the standard one
Overview of Parametric Failure Time Distributions

1 Next step is to specify distribution for $U_i$
   1 Generally iid
   2 Normal is possible
   3 We explore this, and alternatives.

2 Potential distributions for $\log(T)$ will allow for shift and rescaling.
   1 Definition: Set of distns $\mathcal{G}$ is a location and scale family if $\psi X + \phi \in \mathcal{G}$ whenever $X \in \mathcal{G}$ and $\psi > 0$ and $\phi \in \mathbb{R}$
   2 Associated family turns a baseline $S_0$ into a location and scale family $\mathcal{G} = \{S_0((x - \phi)/\psi) | \psi > 0\}$

3 Can estimate $\alpha, \sigma$ only if error distribution fixed to standard.
   1 Pick one member of family to be the standard one
      1 Some of our distributions don’t have means and variances
Overview of Parametric Failure Time Distributions

1. Next step is to specify distribution for $U_i$;
   1. Generally iid
   2. Normal is possible
   3. We explore this, and alternatives.

2. Potential distributions for $\log(T)$ will allow for shift and rescaling.
   1. Definition: Set of distns $\mathcal{G}$ is a location and scale family if $\psi X + \phi \in \mathcal{G}$ whenever $X \in \mathcal{G}$ and $\psi > 0$ and $\phi \in \mathbb{R}$
   2. Associated family turns a baseline $S_0$ into a location and scale family $\mathcal{G} = \{S_0((x - \phi)/\psi)|\psi > 0\}$

3. Can estimate $\alpha, \sigma$ only if error distribution fixed to standard.
   1. Pick one member of family to be the standard one
      1. Some of our distributions don’t have means and variances
      2. Median 0, IQR 1?
Overview of Parametric Failure Time Distributions

1. Next step is to specify distribution for $U_i$
   - Generally iid
   - Normal is possible
   - We explore this, and alternatives.

2. Potential distributions for $\log(T)$ will allow for shift and rescaling.
   - Definition: Set of distns $\mathcal{G}$ is a location and scale family if $\psi X + \phi \in \mathcal{G}$ whenever $X \in \mathcal{G}$ and $\psi > 0$ and $\phi \in \mathbb{R}$
   - Associated family turns a baseline $S_0$ into a location and scale family $\mathcal{G} = \{S_0((x - \phi)/\psi)|\psi > 0\}$

3. Can estimate $\alpha, \sigma$ only if error distribution fixed to standard.
   - Pick one member of family to be the standard one
     - Some of our distributions don’t have means and variances
     - Median 0, IQR 1?
     - Some simple choice of two parameters?

Even if we select standard value for location and scale, there might be other parameters in model.

Analogue to least squares: skewness

Hence typically, we estimate family member rather than $\alpha, \sigma$.

Family member is the one with $\beta = 0$. 
Overview of Parametric Failure Time Distributions

Next step is to specify distribution for \( U_i \)
1. Generally iid
2. Normal is possible
3. We explore this, and alternatives.

Potential distributions for \( \log(T) \) will allow for shift and rescaling.
1. Definition: Set of distns \( \mathcal{G} \) is a *location and scale family* if \( \psi X + \phi \in \mathcal{G} \) whenever \( X \in \mathcal{G} \) and \( \psi > 0 \) and \( \phi \in \mathbb{R} \)
2. Associated family turns a baseline \( S_0 \) into a location and scale family \( \mathcal{G} = \{ S_0((x - \phi)/\psi)|\psi > 0 \} \)

Can estimate \( \alpha, \sigma \) only if error distribution fixed to standard.
1. Pick one member of family to be the standard one
   1. Some of our distributions don’t have means and variances
   2. Median 0, IQR 1?
   3. Some simple choice of two parameters?
2. Even if we select standard value for location and scale, there might be other parameters in model
Next step is to specify distribution for $U_i$:

1. Generally iid
2. Normal is possible
3. We explore this, and alternatives.

Potential distributions for $\log(T)$ will allow for shift and rescaling.

1. Definition: Set of distns $\mathcal{G}$ is a location and scale family if $\psi X + \phi \in \mathcal{G}$ whenever $X \in \mathcal{G}$ and $\psi > 0$ and $\phi \in \mathbb{R}$
2. Associated family turns a baseline $S_0$ into a location and scale family $\mathcal{G} = \{S_0((x - \phi)/\psi)|\psi > 0\}$

Can estimate $\alpha, \sigma$ only if error distribution fixed to standard.

1. Pick one member of family to be the standard one
   1. Some of our distributions don’t have means and variances
   2. Median 0, IQR 1?
   3. Some simple choice of two parameters?
2. Even if we select standard value for location and scale, there might be other parameters in model
   1. Analogue to least squares: skewness
Overview of Parametric Failure Time Distributions

1. Next step is to specify distribution for $U_i$:
   1. Generally iid
   2. Normal is possible
   3. We explore this, and alternatives.

2. Potential distributions for $\log(T)$ will allow for shift and rescaling.
   1. Definition: Set of distns $\mathcal{G}$ is a *location and scale family* if $\psi X + \phi \in \mathcal{G}$ whenever $X \in \mathcal{G}$ and $\psi > 0$ and $\phi \in \mathbb{R}$
   2. Associated family turns a baseline $S_0$ into a location and scale family $\mathcal{G} = \{S_0((x - \phi)/\psi) | \psi > 0\}$

3. Can estimate $\alpha, \sigma$ only if error distribution fixed to standard.
   1. Pick one member of family to be the standard one
      1. Some of our distributions don’t have means and variances
      2. Median 0, IQR 1?
      3. Some simple choice of two parameters?
   2. Even if we select standard value for location and scale, there might be other parameters in model
      1. Analogue to least squares: skewness
      3. Hence typically, we estimate family member rather than $\alpha, \sigma$. 

Regression Models: Parametric Failure Time Distributions
Next step is to specify distribution for $U_i$:

1. Generally iid
2. Normal is possible
3. We explore this, and alternatives.

Potential distributions for $\log(T)$ will allow for shift and rescaling.

1. Definition: Set of distns $\mathcal{G}$ is a location and scale family if $\psi X + \phi \in \mathcal{G}$ whenever $X \in \mathcal{G}$ and $\psi > 0$ and $\phi \in \mathbb{R}$
2. Associated family turns a baseline $S_0$ into a location and scale family $\mathcal{G} = \{S_0((x - \phi)/\psi)|\psi > 0\}$

Can estimate $\alpha, \sigma$ only if error distribution fixed to standard.

1. Pick one member of family to be the standard one
   - Some of our distributions don’t have means and variances
   - Median 0, IQR 1?
   - Some simple choice of two parameters?
2. Even if we select standard value for location and scale, there might be other parameters in model
   - Analogue to least squares: skewness
3. Hence typically, we estimate family member rather than $\alpha, \sigma$.
4. Family member is the one with $\beta = 0$. 
Weibull

1. $W_j$ has survival function $\exp(-\lambda w^\eta) = \exp(-\exp(\log(w)\eta + \log(\lambda)))$

Regression Models: Parametric Failure Time Distributions
Weibull

1. $W_j$ has survival function $\exp(-\lambda w^{\eta}) = \exp(- \exp(\log(w)\eta + \log(\lambda)))$

2. Hence $\eta$ and $\log(\lambda_j)$ are scale and location parameters for $\log(W_j)$. 

---

Regression Models: Parametric Failure Time Distributions
Weibull

1. $W_j$ has survival function $\exp(-\lambda w^\eta) = \exp(-\exp(\log(w)\eta + \log(\lambda)))$

2. Hence $\eta$ and $\log(\lambda_j)$ are scale and location parameters for $\log(W_j)$.

3. $W_j$ has cumulative hazard function $\lambda w^\eta$. 

4. If $\sigma = 1$ and hence $\eta = 1$, $U_j = \log(W_j)$ has survival function $\exp(-\lambda \exp(u)\eta)$.

5. Standard $\text{distn}$ with $\eta = 1$ and $\lambda = 1$ is $f_{U}(u) = \exp(u - \exp(u \eta + \log(\lambda)))$.

$\eta$ is scale parameter

$\log(\lambda)$ is location parameter

6. $T_i = \exp(\alpha + z_i \beta + \sigma U_i)$

7. $\mathbb{P}[T_i > t] = \exp(-\exp((\log(t) - \alpha - z_i \beta)/\sigma))$.

8. Hence $T_i \sim \text{Weibull}$, with $\eta_i = 1/\sigma$ and $\lambda = \exp(-\alpha/\sigma - z_i \beta/\sigma)$.

9. $\log(T_i) - z_i \beta$ has Extreme value $\text{distn}$ or Fisher–Tippett Distribution.
Weibull

1. \( W_j \) has survival function \( \exp(-\lambda w^\eta) = \exp(- \exp(\log(w)\eta + \log(\lambda))) \)

1. Hence \( \eta \) and \( \log(\lambda_j) \) are scale and location parameters for \( \log(W_j) \).

2. \( W_j \) has cumulative hazard function \( \lambda w^\eta \).

4. Exponential Distribution if \( \sigma = 1 \) and hence \( \eta = 1 \)

Regression Models: Parametric Failure Time Distributions
Weibull

1. $W_j$ has survival function $\exp(-\lambda w^\eta) = \exp(-\exp(\log(w)\eta + \log(\lambda)))$

   Hence $\eta$ and $\log(\lambda_j)$ are scale and location parameters for $\log(W_j)$.

2. $W_j$ has cumulative hazard function $\lambda w^\eta$.
   - Exponential Distribution if $\sigma = 1$ and hence $\eta = 1$

3. $U_j = \log(W_j)$ has survival function $\exp(-\lambda \exp(u)^\eta)$

Regression Models: Parametric Failure Time Distributions Lecture 09 188 / 260
Weibull

1. \( W_j \) has survival function \( \exp(-\lambda w^\eta) = \exp(-\exp(\log(w)\eta + \log(\lambda))) \)

   Hence \( \eta \) and \( \log(\lambda_j) \) are scale and location parameters for \( \log(W_j) \).

2. \( W_j \) has cumulative hazard function \( \lambda w^\eta \).

   Exponential Distribution if \( \sigma = 1 \) and hence \( \eta = 1 \)

3. \( U_j = \log(W_j) \) has survival function \( \exp(-\lambda \exp(u)^\eta) \)

   Standard distn with \( \eta = 1 \) and \( \lambda = 1 \) is \( f_U(u) = \exp(u - \exp(u)) \):
Weibull

1. $W_j$ has survival function $\exp(-\lambda w^\eta) = \exp(- \exp(\log(w)\eta + \log(\lambda)))$

   Hence $\eta$ and $\log(\lambda_j)$ are scale and location parameters for $\log(W_j)$.

2. $W_j$ has cumulative hazard function $\lambda w^\eta$.

   Exponential Distribution if $\sigma = 1$ and hence $\eta = 1$

3. $U_j = \log(W_j)$ has survival function $\exp(-\lambda \exp(u)^\eta)$

   Standard distn with $\eta = 1$ and $\lambda = 1$ is $f_U(u) = \exp(u - \exp(u))$:

   Density on $U$ scale is $\exp(- \exp(u\eta + \log(\lambda)))$
Weibull

1. $W_j$ has survival function $\exp(-\lambda w^\eta) = \exp(-\exp(\log(w)\eta + \log(\lambda)))$

Hence $\eta$ and $\log(\lambda_j)$ are scale and location parameters for $\log(W_j)$.

2. $W_j$ has cumulative hazard function $\lambda w^\eta$.

3. Exponential Distribution if $\sigma = 1$ and hence $\eta = 1$

4. $U_j = \log(W_j)$ has survival function $\exp(-\lambda \exp(u)^\eta)$

   1. Standard distn with $\eta = 1$ and $\lambda = 1$ is $f_U(u) = \exp(u - \exp(u))$: 
   2. Density on $U$ scale is $\exp(-\exp(u\eta + \log(\lambda)))$
   3. $\eta$ is scale parameter
Weibull

1. $W_j$ has survival function $\exp(-\lambda w^\eta) = \exp(-\exp(\log(w)\eta + \log(\lambda)))$

   - Hence $\eta$ and $\log(\lambda_j)$ are scale and location parameters for $\log(W_j)$.

2. $W_j$ has cumulative hazard function $\lambda w^\eta$.

   - Exponential Distribution if $\sigma = 1$ and hence $\eta = 1$

3. $U_j = \log(W_j)$ has survival function $\exp(-\lambda \exp(u)^\eta)$

   - Standard distn with $\eta = 1$ and $\lambda = 1$ is $f_u(u) = \exp(u - \exp(u))$:
   - Density on $U$ scale is $\exp(-\exp(u\eta + \log(\lambda)))$
   - $\eta$ is scale parameter
   - $\log(\lambda)$ is location parameter
Weibull

1. $W_j$ has survival function $\exp(-\lambda w^\eta) = \exp(-\exp(\log(w)\eta + \log(\lambda)))$  
   Hence $\eta$ and $\log(\lambda_j)$ are scale and location parameters for $\log(W_j)$.

2. $W_j$ has cumulative hazard function $\lambda w^\eta$.

3. $U_j = \log(W_j)$ has survival function $\exp(-\lambda \exp(u)^\eta)$
   1. Standard distn with $\eta = 1$ and $\lambda = 1$ is $f_U(u) = \exp(u - \exp(u))$:  
   2. Density on $U$ scale is $\exp(-\exp(u\eta + \log(\lambda)))$
   3. $\eta$ is scale parameter
   4. $\log(\lambda)$ is location parameter

4. $T_i = \exp(\alpha + z_i\beta + \sigma U_i)$
Weibull

1. $W_j$ has survival function $\exp(-\lambda w^\eta) = \exp(-\exp(\log(w)\eta + \log(\lambda)))$

   Hence $\eta$ and $\log(\lambda_j)$ are scale and location parameters for $\log(W_j)$.

2. $W_j$ has cumulative hazard function $\lambda w^\eta$.

3. $U_j = \log(W_j)$ has survival function $\exp(-\lambda \exp(u)^\eta)$

   a. Exponential Distribution if $\sigma = 1$ and hence $\eta = 1$

4. $U_j = \log(W_j)$ has survival function $\exp(-\lambda \exp(u)^\eta)$

   a. Standard distn with $\eta = 1$ and $\lambda = 1$ is $f_U(u) = \exp(u - \exp(u))$:

   b. Density on $U$ scale is $\exp(-\exp(u\eta + \log(\lambda)))$

   c. $\eta$ is scale parameter

   d. $\log(\lambda)$ is location parameter

4. $T_i = \exp(\alpha + z_i\beta + \sigma U_i)$

   a. $P[T_i > t] = \exp(-\exp((\log(t) - \alpha - z_i\beta)/\sigma))$. 
Weibull

1. $W_j$ has survival function \( \exp(-\lambda w^\eta) = \exp(-\exp(\log(w)\eta + \log(\lambda))) \)

   Hence $\eta$ and $\log(\lambda_j)$ are scale and location parameters for $\log(W_j)$.

2. $W_j$ has cumulative hazard function $\lambda w^\eta$.

3. $W_j$ has survival function $\exp(-\lambda \exp(u^\eta))$.

4. $U_j = \log(W_j)$ has survival function $\exp(-\lambda \exp(u^\eta))$.

   - Exponential Distribution if $\sigma = 1$ and hence $\eta = 1$

5. $U_j = \log(W_j)$ has survival function $\exp(-\lambda \exp(u^\eta))$.

   - Standard distn with $\eta = 1$ and $\lambda = 1$ is $f_U(u) = \exp(u - \exp(u))$:
     - Density on $U$ scale is $\exp(-\exp(u\eta + \log(\lambda)))$
     - $\eta$ is scale parameter
     - $\log(\lambda)$ is location parameter

4. $T_i = \exp(\alpha + z_i\beta + \sigma U_i)$

   - $P[T_i > t] = \exp(-\exp((\log(t) - \alpha - z_i\beta)/\sigma))$.
   - Hence $T_i \sim$ Weibull, with $\eta_i = 1/\sigma$ and $\lambda = \exp(-\alpha/\sigma - z_i\beta/\sigma)$. 
Weibull

1. $W_j$ has survival function $\exp(-\lambda w^\eta) = \exp(- \exp(\log(w)\eta + \log(\lambda)))$

   Hence $\eta$ and $\log(\lambda_j)$ are scale and location parameters for $\log(W_j)$.

2. $W_j$ has cumulative hazard function $\lambda w^\eta$.

3. $U_j = \log(W_j)$ has survival function $\exp(-\lambda \exp(u)^\eta)$

   Standard distn with $\eta = 1$ and $\lambda = 1$ is $f_U(u) = \exp(u - \exp(u))$:  
   
   Density on $U$ scale is $\exp(- \exp(u\eta + \log(\lambda)))$

   $\eta$ is scale parameter

   $\log(\lambda)$ is location parameter

4. $T_i = \exp(\alpha + z_i\beta + \sigma U_i)$

   1. $P[T_i > t] = \exp(- \exp((\log(t) - \alpha - z_i\beta)/\sigma))$.

   2. Hence $T_i \sim$ Weibull, with $\eta_i = 1/\sigma$ and $\lambda = \exp(-\alpha/\sigma - z_i\beta/\sigma)$.

   3. $\log(T_i) - z_i\beta$ has Extreme value distn or Fisher–Tippett Distribution.
Weibull Hazards are Proportional
Weibull Hazards are Proportional

\[ h(w) = \frac{d}{dw} \lambda w^{\eta} = \eta \lambda w^{\eta-1} \]
Weibull Hazards are Proportional

1. $h(w) = \frac{d}{dw} \lambda w^{\eta} = \eta \lambda w^{\eta-1}$

2. Ratio of hazards with $\lambda_1$ and $\lambda_2$: $(\eta \lambda_1 w^{\eta-1})/(\eta \lambda_2 w^{\eta-1}) = \lambda_1/\lambda_2$:

   Proportional hazards holds
Weibull Hazards are Proportional

1. $h(w) = \frac{d}{dw} \lambda w^\eta = \eta \lambda w^{\eta-1}$
2. Ratio of hazards with $\lambda_1$ and $\lambda_2$: $(\eta \lambda_1 w^{\eta-1})/(\eta \lambda_2 w^{\eta-1}) = \lambda_1/\lambda_2$:
   Proportional hazards holds
3. Risk for individual with covariates $z_j$ relative to one with $z_k$ is
   $\exp((z_k - z_j)\beta/\sigma)$. 

Regression Models: Parametric Failure Time Distributions Lecture 09
Weibull Hazards are Proportional

1. \[ h(w) = \frac{d}{dw} \lambda w^\eta = \eta \lambda w^{\eta-1} \]
2. Ratio of hazards with \( \lambda_1 \) and \( \lambda_2 \):
   \[ \frac{(\eta \lambda_1 w^{\eta-1})}{(\eta \lambda_2 w^{\eta-1})} = \frac{\lambda_1}{\lambda_2} \]
   Proportional hazards holds
3. Risk for individual with covariates \( z_j \) relative to one with \( z_k \) is
   \[ \exp((z_k - z_j)\beta/\sigma) \]
4. Can use delta method to give CI
Weibull Hazards are Proportional

1. \( h(w) = \frac{d}{dw} \lambda w^{\eta} = \eta \lambda w^{\eta-1} \)

2. Ratio of hazards with \( \lambda_1 \) and \( \lambda_2 \): 
   \( \frac{\eta \lambda_1 w^{\eta-1}}{\eta \lambda_2 w^{\eta-1}} = \frac{\lambda_1}{\lambda_2} \) 
   Proportional hazards holds

3. Risk for individual with covariates \( z_j \) relative to one with \( z_k \) is 
   \( \exp((z_k - z_j)\beta/\sigma) \).

4. Can use delta method to give CI

Lemma: For any real function \( g \) such that \( g(\tau + \nu) = g(\tau) + g(\nu) \) \( \forall \tau, \nu \), then \( g(\tau) = \tau g(1) \)
Continued

1. Weibull Hazards are Proportional
   - \( h(w) = \frac{d}{dw} \lambda w^n = \eta \lambda w^{n-1} \)
   - Ratio of hazards with \( \lambda_1 \) and \( \lambda_2 \): \( \frac{\eta \lambda_1 w^{n-1}}{\eta \lambda_2 w^{n-1}} \) = \( \frac{\lambda_1}{\lambda_2} \):
     Proportional hazards holds
   - Risk for individual with covariates \( z_j \) relative to one with \( z_k \) is
     \( \exp((z_k - z_j)\beta/\sigma) \).
   - Can use delta method to give CI

2. Lemma: For any real function \( g \) such that \( g(\tau + \nu) = g(\tau) + g(\nu) \) for all \( \tau, \nu \), then \( g(\tau) = \tau g(1) \)
   - The condition implies that for any \( m \), then \( g(m\tau) = mg(\tau) \).
Weibull Hazards are Proportional

1. \( h(w) = \frac{d}{dw} \lambda w^{\eta} = \eta \lambda w^{\eta-1} \)

2. Ratio of hazards with \( \lambda_1 \) and \( \lambda_2 \): \( (\eta \lambda_1 w^{\eta-1})/(\eta \lambda_2 w^{\eta-1}) = \lambda_1/\lambda_2 \):
   Proportional hazards holds

3. Risk for individual with covariates \( z_j \) relative to one with \( z_k \) is
   \( \exp((z_k - z_j)\beta/\sigma) \).

   Can use delta method to give CI

Lemma: For any real function \( g \) such that \( g(\tau + \nu) = g(\tau) + g(\nu) \forall \tau, \nu \), then \( g(\tau) = \tau g(1) \)

1. The condition implies that for any \( m \), then \( g(m\tau) = mg(\tau) \).
2. The condition implies that for any \( n \), then \( g(\tau/n) = g(\tau)/n \).
Continued

1. **Weibull Hazards are Proportional**
   1. \( h(w) = \frac{d}{dw} \lambda w^\eta = \eta \lambda w^{\eta-1} \)
   2. Ratio of hazards with \( \lambda_1 \) and \( \lambda_2 \):
      \[ \frac{\eta \lambda_1 w^{\eta-1}}{\eta \lambda_2 w^{\eta-1}} = \frac{\lambda_1}{\lambda_2} \]
      Proportional hazards holds
   3. Risk for individual with covariates \( z_j \) relative to one with \( z_k \) is
      \[ \exp((z_k - z_j) \beta / \sigma) \]
      Can use delta method to give CI

2. **Lemma:** For any real function \( g \) such that \( g(\tau + \nu) = g(\tau) + g(\nu) \) for all \( \tau, \nu \), then \( g(\tau) = \tau g(1) \)
   1. The condition implies that for any \( m \), then \( g(m\tau) = mg(\tau) \).
   2. The condition implies that for any \( n \), then \( g(\tau/n) = g(\tau)/n. \)
   3. Then \( g(m/n) = g(1)m/n. \)
Weibull Hazards are Proportional

1. \( h(w) = \frac{d}{dw} \lambda w^n = \eta \lambda w^{n-1} \)
2. Ratio of hazards with \( \lambda_1 \) and \( \lambda_2 \): \((\eta \lambda_1 w^{n-1})/(\eta \lambda_2 w^{n-1}) = \lambda_1/\lambda_2\): Proportional hazards holds
3. Risk for individual with covariates \( z_j \) relative to one with \( z_k \) is \( \exp((z_k - z_j)\beta/\sigma) \).
4. Can use delta method to give CI

Lemma: For any real function \( g \) such that \( g(\tau + \nu) = g(\tau) + g(\nu) \) for all \( \tau, \nu \), then \( g(\tau) = \tau g(1) \)

1. The condition implies that for any \( m \), then \( g(m\tau) = mg(\tau) \).
2. The condition implies that for any \( n \), then \( g(\tau/n) = g(\tau)/n \).
3. Then \( g(m/n) = g(1)m/n \).
4. By continuity from right, \( g(\tau) = \tau g(1) \) for every \( \tau \).
Weibull is only distribution giving proportional hazards:
Weibull is only distribution giving proportional hazards:

Accelerated life model equivalent to $H_\nu(t) = H(t \exp(\nu))$
Weibull is only distribution giving proportional hazards:

1. Accelerated life model equivalent to $H_\nu(t) = H(t \exp(\nu))$
2. Proportional hazards $\iff H_\nu(t) = H(t) \exp(c(\nu))$
Weibull is only distribution giving proportional hazards:

1. Accellerated life model equivalent to $H_\nu(t) = H(t \exp(\nu))$
2. Proportional hazards $\iff H_\nu(t) = H(t) \exp(c(\nu))$
3. Both $\iff H(\exp(\tau) \exp(\nu)) = H(\exp(\tau)) \exp(c(\nu)) \forall \tau$
Weibull is the only distribution giving proportional hazards:

1. Accellerated life model equivalent to \( H_\nu(t) = H(t \exp(\nu)) \)
2. Proportional hazards \( \iff H_\nu(t) = H(t) \exp(c(\nu)) \)
3. Both \( \iff H(\exp(\tau) \exp(\nu)) = H(\exp(\tau)) \exp(c(\nu)) \forall \tau \)
4. Let \( g(\tau) = \log\left(\frac{H(\exp(\tau))}{H(1)}\right) \) (and hence \( g(0) = 0 \))
Weibull is only distribution giving proportional hazards:

1. Accellerated life model equivalent to \( H_\nu(t) = H(t \exp(\nu)) \)
2. Proportional hazards \( \Leftrightarrow H_\nu(t) = H(t) \exp(c(\nu)) \)
3. Both \( \Leftrightarrow H(\exp(\tau) \exp(\nu)) = H(\exp(\tau)) \exp(c(\nu)) \forall \tau \)
4. Let \( g(\tau) = \log(H(\exp(\tau))/H(1))) \) (and hence \( g(0) = 0 \))
5. Then \( g(\tau + \nu) = g(\tau) + c(\nu) \forall \tau, \nu \).
Weibull is only distribution giving proportional hazards:

1. Accellerated life model equivalent to $H_\nu(t) = H(t \exp(\nu))$
2. Proportional hazards $\iff H_\nu(t) = H(t) \exp(c(\nu))$
3. Both $\iff H(\exp(\tau) \exp(\nu)) = H(\exp(\tau)) \exp(c(\nu)) \forall \tau$
4. Let $g(\tau) = \log(H(\exp(\tau))/H(1)))$ (and hence $g(0) = 0$)
5. Then $g(\tau + \nu) = g(\tau) + c(\nu) \forall \tau, \nu$.
6. Setting $\tau = 0$, then $c(\nu) = g(\nu)$. 
Weibull is only distribution giving proportional hazards:

1. Accellerated life model equivalent to $H_\nu(t) = H(t \exp(\nu))$
2. Proportional hazards $\iff H_\nu(t) = H(t) \exp(c(\nu))$
3. Both $\iff H(\exp(\tau) \exp(\nu)) = H(\exp(\tau)) \exp(c(\nu)) \forall \tau$
4. Let $g(\tau) = \log(H(\exp(\tau))/H(1)))$ (and hence $g(0) = 0$)
5. Then $g(\tau + \nu) = g(\tau) + c(\nu) \forall \tau, \nu$.
6. Setting $\tau = 0$, then $c(\nu) = g(\nu)$.
7. Then $g(\tau + \nu) = g(\tau) + g(\nu) \forall \tau, \nu$. 

So $H(t) = H(1) t^{\frac{g(1)}{g(0)}}$. 

By lemma $H(\exp(\tau)) = g(1) \exp(\tau)$.
Weibull is only distribution giving proportional hazards:

1. Accellerated life model equivalent to $H_{\nu}(t) = H(t \exp(\nu))$
2. Proportional hazards $\Leftrightarrow H_{\nu}(t) = H(t) \exp(c(\nu))$
3. Both $\Leftrightarrow H(\exp(\tau) \exp(\nu)) = H(\exp(\tau)) \exp(c(\nu)) \forall \tau$
4. Let $g(\tau) = \log(H(\exp(\tau))/H(1)))$ (and hence $g(0) = 0$)
5. Then $g(\tau + \nu) = g(\tau) + c(\nu) \forall \tau, \nu$
6. Setting $\tau = 0$, then $c(\nu) = g(\nu)$.
7. Then $g(\tau + \nu) = g(\tau) + g(\nu) \forall \tau, \nu$

By lemma $H(\exp(\tau)) = g(1) \exp(\tau)^{H(1)}$. 

---

**Continued**
Weibull is only distribution giving proportional hazards:

1. Accellerated life model equivalent to \( H_{\nu}(t) = H(t \exp(\nu)) \)
2. Proportional hazards \( \iff H_{\nu}(t) = H(t) \exp(c(\nu)) \)
3. Both \( \iff H(\exp(\tau) \exp(\nu)) = H(\exp(\tau)) \exp(c(\nu)) \forall \tau \)
4. Let \( g(\tau) = \log(H(\exp(\tau))/H(1))) \) (and hence \( g(0) = 0 \))
5. Then \( g(\tau + \nu) = g(\tau) + c(\nu) \forall \tau, \nu. \)
6. Setting \( \tau = 0 \), then \( c(\nu) = g(\nu). \)
7. Then \( g(\tau + \nu) = g(\tau) + g(\nu) \forall \tau, \nu. \)

2. By lemma \( H(\exp(\tau)) = g(1) \exp(\tau)^{H(1)}. \)
1. So \( H(t) = H(1)t^{g(1)} \)
Log logistic distn

1. Gives proportional odds model

\[ \text{Survival function} = \frac{1}{1 + \exp(-u)} \]

\[ \text{Odds of survival} = \exp(-u) \]

\[ \text{Odds ratio of survival for individual with } \eta^\dagger \text{ and } \eta^\ddagger \text{ is } \exp(\eta^\dagger - \eta^\ddagger) \]

\[ \text{Define variable } W = \exp(U) \]

\[ \text{Survival function} = \frac{1}{1 + \lambda w^\alpha} \]

\[ \eta = \log(\lambda) \]
Log logistic distn

1. Gives proportional odds model
   - Take logistic random variable $U$

Survival function: $\frac{1}{1 + \exp(-U)}$

Odds of survival: $\exp(-U)$

Make into location scale family

Survival function: $\frac{1}{1 + \exp(\alpha U + \eta)}$

Odds of survival: $\exp(-\alpha U - \eta)$

Odds ratio of survival for individual with $\eta$ of $\eta^*$ and $\eta^\dagger$ is $\exp(\eta^* - \eta^\dagger)$

Define variable $W = \exp(U)$

Survival function: $\frac{1}{1 + \lambda W^\alpha}$

$\eta = \log(\lambda)$
Log logistic distn

1. Gives proportional odds model
   1. Take logistic random variable $U$
      1. Survival function $1/(1 + \exp(u))$

2. Odds of survival $\exp(-u)$

3. Make into location scale family
   1. Survival function $1/(1 + \exp(\alpha u + \eta))$
   2. Odds of survival are $\exp(-\alpha u - \eta)$

3. Odds ratio of survival for individual with $\eta$ of $\eta^*$ and $\eta^*$ is $\exp(\eta^* - \eta^*)$

4. Define variable $W = \exp(U)$
   1. Survival function is $1/(1 + \lambda w \alpha)$
   2. $\eta = \log(\lambda)$
Log logistic distn

1. Gives proportional odds model
   1. Take logistic random variable $U$
   2. Survival function $1/(1 + \exp(u))$
   3. Odds of survival $\exp(-u)$

Regression Models: Parametric Failure Time Distributions
Log logistic distn

1. Gives proportional odds model
   1. Take logistic random variable $U$
      1. Survival function $1/(1 + \exp(u))$
      2. Odds of survival $\exp(-u)$
   2. Make into location scale family
Log logistic distn

Gives proportional odds model

1. Take logistic random variable $U$
   - Survival function $1/(1 + \exp(u))$
   - Odds of survival $\exp(-u)$

2. Make into location scale family
   - Survival function $1/(1 + \exp(\alpha u + \eta))$

Odds ratio of survival for individual with $\eta^\dagger$ and $\eta^\ddagger$ is $\exp(\eta^\dagger - \eta^\ddagger)$

Define variable $W = \exp(U)$

Survival function is $1/(1 + \lambda w^{\alpha \eta})$
Log logistic distn

1. Gives proportional odds model
   1. Take logistic random variable $U$
      1. Survival function $1/(1 + \exp(u))$
      2. Odds of survival $\exp(-u)$
   2. Make into location scale family
      1. Survival function $1/(1 + \exp(\alpha u + \eta))$
      2. Odds of survival are $\exp(-\alpha u - \eta)$
Log logistic distn

1. Gives proportional odds model
   1. Take logistic random variable $U$
      1. Survival function $1/(1 + \exp(u))$
      2. Odds of survival $\exp(-u)$
   2. Make into location scale family
      1. Survival function $1/(1 + \exp(\alpha u + \eta))$
      2. Odds of survival are $\exp(-\alpha u - \eta)$
      3. Odds ratio of survival for individual with $\eta$ of $\eta^\dagger$ and $\eta^\ddagger$ is $\exp(\eta^\dagger - \eta^\ddagger)$

Define variable $W = \exp(U)$

Survival function is $1/(1 + \lambda w^\alpha)$

$\eta = \log(\lambda)$
Log logistic distn

1. Gives proportional odds model
   a. Take logistic random variable $U$
   b. Survival function $1/(1 + \exp(u))$
   c. Odds of survival $\exp(-u)$
2. Make into location scale family
   a. Survival function $1/(1 + \exp(\alpha u + \eta))$
   b. Odds of survival are $\exp(-\alpha u - \eta)$
   c. Odds ratio of survival for individual with $\eta$ of $\eta^{\dagger}$ and $\eta^{\ddagger}$ is $\exp(\eta^{\dagger} - \eta^{\ddagger})$
3. Define variable $W = \exp(U)$
Log logistic distn

1. **Gives proportional odds model**
   1. Take logistic random variable $U$
      1. Survival function $1/(1 + \exp(u))$
      2. Odds of survival $\exp(-u)$
   2. Make into location scale family
      1. Survival function $1/(1 + \exp(\alpha u + \eta))$
      2. Odds of survival are $\exp(-\alpha u - \eta)$
      3. Odds ratio of survival for individual with $\eta$ of $\eta^\dagger$ and $\eta^\ddagger$ is $\exp(\eta^\dagger - \eta^\ddagger)$
   3. Define variable $W = \exp(U)$
      1. Survival function is $1/(1 + \lambda w^\alpha)$
Log logistic distn

1. Gives proportional odds model
   1. Take logistic random variable $U$
      1. Survival function $1/(1 + \exp(u))$
      2. Odds of survival $\exp(-u)$
   2. Make into location scale family
      1. Survival function $1/(1 + \exp(\alpha u + \eta))$
      2. Odds of survival are $\exp(-\alpha u - \eta)$
      3. Odds ratio of survival for individual with $\eta$ of $\eta^\dagger$ and $\eta^\ddagger$ is $\exp(\eta^\dagger - \eta^\ddagger)$
   3. Define variable $W = \exp(U)$
      1. Survival function is $1/(1 + \lambda w^\alpha)$
      2. $\eta = \log(\lambda)$
No other family gives proportional odds model
No other family gives proportional odds model

Let $LO(x) = \log(S(x)) - \log(1 - S(x))$. 
No other family gives proportional odds model

Let $LO(x) = \log(S(x)) - \log(1 - S(x))$.

$S(x) = \exp(LO(x))/(1 + \exp(LO(x)))$. 

No other family gives proportional odds model

1. Let $LO(x) = \log(S(x)) - \log(1 - S(x))$.
2. $S(x) = \exp(LO(x))/(1 + \exp(LO(x)))$.
3. Suppose $LO(x \exp(\mu)) - LO(x) = c(\mu)$ for some function $c$ and every $x, \mu$. 

Continued
No other family gives proportional odds model

1. Let $LO(x) = \log(S(x)) - \log(1 - S(x))$.
   
2. $S(x) = \exp(LO(x))/(1 + \exp(LO(x)))$.

3. Suppose $LO(x \exp(\mu)) - LO(x) = c(\mu)$ for some function $c$ and every $x, \mu$.

4. Let $g(\xi) = LO(\exp(\xi)) - LO(1)$
No other family gives proportional odds model

1. Let \( LO(x) = \log(S(x)) - \log(1 - S(x)) \).

2. \( S(x) = \exp(LO(x))/(1 + \exp(LO(x))) \).

3. Suppose \( LO(x \exp(\mu)) - LO(x) = c(\mu) \) for some function \( c \) and every \( x, \mu \).

4. Let \( g(\xi) = LO(\exp(\xi)) - LO(1) \)

5. \( g(\xi + \mu) - g(\xi) = c(\mu) \),
No other family gives proportional odds model

1. Let \( LO(x) = \log(S(x)) - \log(1 - S(x)) \).

1. \( S(x) = \exp(LO(x))/(1 + \exp(LO(x))) \).

2. Suppose \( LO(x \exp(\mu)) - LO(x) = c(\mu) \) for some function \( c \) and every \( x, \mu \).

3. Let \( g(\xi) = LO(\exp(\xi)) - LO(1) \)

4. \( g(\xi + \mu) - g(\xi) = c(\mu) \),

5. Setting \( \xi = 0 \) to see \( g(\mu) = c(\mu) \).
No other family gives proportional odds model

Let \( LO(x) = \log(S(x)) - \log(1 - S(x)) \).

\[ S(x) = \frac{\exp(LO(x))}{1 + \exp(LO(x))}. \]

Suppose \( LO(x \exp(\mu)) - LO(x) = c(\mu) \) for some function \( c \) and every \( x, \mu \).

Let \( g(\xi) = LO(\exp(\xi)) - LO(1) \)

\[ g(\xi + \mu) - g(\xi) = c(\mu), \]

Setting \( \xi = 0 \) to see \( g(\mu) = c(\mu) \).

So \( g(\xi + \mu) = g(\mu) + g(\xi) \forall \mu, \xi \).
No other family gives proportional odds model

1. Let $LO(x) = \log(S(x)) - \log(1 - S(x))$.
2. $S(x) = \exp(LO(x))/(1 + \exp(LO(x)))$.
3. Suppose $LO(x \exp(\mu)) - LO(x) = c(\mu)$ for some function $c$ and every $x, \mu$.
4. Let $g(\xi) = LO(\exp(\xi)) - LO(1)$
5. $g(\xi + \mu) - g(\xi) = c(\mu)$,
6. Setting $\xi = 0$ to see $g(\mu) = c(\mu)$.
7. So $g(\xi + \mu) = g(\mu) + g(\xi)\forall \mu, \xi$,
8. By the lemma, $g(\xi) = \xi \alpha$ for $\alpha = g(1)$. 

Regression Models: Parametric Failure Time Distributions

Lecture 09
No other family gives proportional odds model

Let \( LO(x) = \log(S(x)) - \log(1 - S(x)) \).

\( S(x) = \frac{\exp(LO(x))}{1 + \exp(LO(x))} \).

Suppose \( LO(x \exp(\mu)) - LO(x) = c(\mu) \) for some function \( c \) and every \( x, \mu \).

Let \( g(\xi) = LO(\exp(\xi)) - LO(1) \)

\( g(\xi + \mu) - g(\xi) = c(\mu) \),

Setting \( \xi = 0 \) to see \( g(\mu) = c(\mu) \).

So \( g(\xi + \mu) = g(\mu) + g(\xi) \forall \mu, \xi \),

By the lemma, \( g(\xi) = \xi \alpha \) for \( \alpha = g(1) \).

Then \( LO(x) = \log(\beta) + \alpha \log(x) \) for \( \beta = \exp(LO(1)) \).
No other family gives proportional odds model

1. Let $LO(x) = \log(S(x)) - \log(1 - S(x))$.
   
   $S(x) = \exp(LO(x))/(1 + \exp(LO(x)))$.

2. Suppose $LO(x \exp(\mu)) - LO(x) = c(\mu)$ for some function $c$ and every $x, \mu$.

3. Let $g(\xi) = LO(\exp(\xi)) - LO(1)$

4. $g(\xi + \mu) - g(\xi) = c(\mu)$,

5. Setting $\xi = 0$ to see $g(\mu) = c(\mu)$.

6. So $g(\xi + \mu) = g(\mu) + g(\xi) \forall \mu, \xi$,

7. By the lemma, $g(\xi) = \xi \alpha$ for $\alpha = g(1)$.

8. Then $LO(x) = \log(\beta) + \alpha \log(x)$ for $\beta = \exp(LO(1))$.

9. Then $S(x) = \beta x^{\alpha}/(1 + \beta x^{\alpha})$
Log normal

We have seen this one before
Generalized gamma

1 Idea: $\Gamma(\gamma, x)$ is defined as $\int_0^x x^{\gamma-1} \exp(-x) \, dx$ and is called the incomplete gamma function.

2 $\Gamma(\gamma, x)$ is increasing in $x$.

3 $\lim_{x \to \infty} \Gamma(\gamma, x)$ is called the Gamma function.

Then

\[ \frac{\Gamma(\gamma, x)}{\Gamma(\gamma)} \quad \text{if} \quad x \geq 0 \]
\[ 0 \quad \text{otherwise} \]

is a distribution function

Non-decreasing, Limits of 1 and 0 as $x \to \pm \infty$.

Generalized gamma CDF has variable transformed by raising to a power and multiplying:

\[ F(x) = \frac{\Gamma(\gamma, \lambda x^\alpha)}{\Gamma(\gamma)} \]

That is, if $X \sim \Gamma(\gamma)$, $Y = (X/\lambda)^{1/\alpha}$. then $Y$ has generalized gamma distribution with parameters $\alpha$, $\lambda$, $\gamma$. Argument to the Gamma CDF is the inverse transformation.

Text gives two different density parameterizations, including $f(x) = \frac{\alpha \lambda^{\gamma}}{\Gamma(\gamma)} x^{\alpha \gamma-1} \exp(-\lambda x^{\alpha})$.

Using $u = \lambda y^{1/\alpha}$, $y = (u/\lambda)^{1/\alpha}$, $dy = \left(\frac{u}{\lambda}\right)^{1/\alpha - 1}/(\alpha \lambda) \, du$. 

Regression Models: Parametric Failure Time Distributions Lecture 09
Generalized gamma

Idea: \( \Gamma(\gamma, x) \) is defined as \( \int_0^x x^{\gamma-1} \exp(-x) \, dx \) and is called the *incomplete gamma function*

\( \Gamma(\gamma, x) \) is increasing in \( x \).
Generalized gamma

1 Idea: $\Gamma(\gamma, x)$ is defined as $\int_0^x x^{\gamma-1} \exp(-x) \, dx$ and is called the \textit{incomplete gamma function}

1 $\Gamma(\gamma, x)$ is increasing in $x$.
2 $\lim_{x \to \infty} \Gamma(\gamma, x)$ is called the \textit{Gamma function}
Generalized gamma

1. Idea: $\Gamma(\gamma, x)$ is defined as $\int_0^x x^{\gamma-1} \exp(-x) \, dx$ and is called the *incomplete gamma function*

1. $\Gamma(\gamma, x)$ is increasing in $x$.

2. $\lim_{x \to \infty} \Gamma(\gamma, x)$ is called the *Gamma function*

3. Then $\begin{cases} \frac{\Gamma(\gamma, x)}{\Gamma(\gamma)} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$ is a distribution function
Generalized gamma

1. Idea: $\Gamma(\gamma, x)$ is defined as $\int_0^x x^{\gamma-1} \exp(-x) \, dx$ and is called the *incomplete gamma function*

2. $\Gamma(\gamma, x)$ is increasing in $x$.

3. $\lim_{x \to \infty} \Gamma(\gamma, x)$ is called the *Gamma function*

4. Then \[
\begin{cases} 
\Gamma(\gamma, x)/\Gamma(\gamma) & \text{if } x \geq 0 \\
0 & \text{otherwise}
\end{cases}
\]

5. Non-decreasing.

**Regression Models: Parametric Failure Time Distributions**
Generalized gamma

1. Idea: $\Gamma(\gamma, x)$ is defined as $\int_0^x x^{\gamma-1} \exp(-x) \, dx$ and is called the *incomplete gamma function*

   1. $\Gamma(\gamma, x)$ is increasing in $x$.
   2. $\lim_{x \to \infty} \Gamma(\gamma, x)$ is called the *Gamma function*.

3. Then

   \[
   \begin{cases} 
   \frac{\Gamma(\gamma, x)}{\Gamma(\gamma)} & \text{if } x \geq 0 \\
   0 & \text{otherwise}
   \end{cases}
   \]

   1. Non-decreasing,
   2. Limits of 1 and 0 as $x \to \pm \infty$. 

Regression Models: Parametric Failure Time Distributions  Lecture 09 194 / 260
Generalized gamma

1. Idea: \( \Gamma(\gamma, x) \) is defined as \( \int_0^x x^{\gamma-1} \exp(-x) \, dx \) and is called the *incomplete gamma function*

   - \( \Gamma(\gamma, x) \) is increasing in \( x \).
   - \( \lim_{x \to \infty} \Gamma(\gamma, x) \) is called the *Gamma function*.

2. Then \( \begin{cases} \frac{\Gamma(\gamma, x)}{\Gamma(\gamma)} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \) is a distribution function

   - Non-decreasing,
   - Limits of 1 and 0 as \( x \to \pm \infty \).

2. Generalized gamma CDF has variable transformed by raising to a power and multiplying: \( F(x) = \frac{\Gamma(\gamma, \lambda x^\alpha)}{\Gamma(\gamma)} \)

   - Argument to the Gamma CDF is the inverse transformation.

   - Text gives two different density parameterizations, including \( f(x) = \frac{\alpha \lambda^{\gamma}}{\gamma^{\alpha}} \left( \frac{x}{\lambda} \right)^{\alpha(\gamma-1)} \exp(-\lambda x^\alpha) / \Gamma(\gamma) \).

   - Using \( u = \lambda y^\alpha \), \( y = \left( \frac{u}{\lambda} \right)^{1/\alpha} \), \( dy = \left( \frac{u}{\lambda} \right)^{1-1/\alpha} du / (\alpha \lambda) \).
Generalized gamma

1. Idea: $\Gamma(\gamma, x)$ is defined as $\int_0^x x^{\gamma-1} \exp(-x) \, dx$ and is called the *incomplete gamma function*
   - $\Gamma(\gamma, x)$ is increasing in $x$.
   - $\lim_{x \to \infty} \Gamma(\gamma, x)$ is called the *Gamma function*

2. Then
   
   \[
   \begin{cases}
   \frac{\Gamma(\gamma, x)}{\Gamma(\gamma)} & \text{if } x \geq 0 \\
   0 & \text{otherwise}
   \end{cases}
   \]

   is a distribution function
   - Non-decreasing,
   - Limits of 1 and 0 as $x \to \pm \infty$.

3. Generalized gamma CDF has variable transformed by raising to a power and multiplying: $F(x) = \frac{\Gamma(\gamma, \lambda x^\alpha)}{\Gamma(\gamma)}$
   - That is, if $X \sim \Gamma(\gamma)$, $Y = (X/\lambda)^{1/\alpha}$. then $Y$ has generalized gamma distribution with parameters $\alpha$, $\lambda$, $\gamma$. 

Regression Models: Parametric Failure Time Distributions Lecture 09 194 / 260
Generalized gamma

Idea: \( \Gamma(\gamma, x) \) is defined as \( \int_0^x x^{\gamma-1} \exp(-x) \, dx \) and is called the *incomplete gamma function*

1. \( \Gamma(\gamma, x) \) is increasing in \( x \).
2. \( \lim_{x \to \infty} \Gamma(\gamma, x) \) is called the *Gamma function*.
3. Then \( \left\{ \begin{array}{ll} \Gamma(\gamma, x)/\Gamma(\gamma) & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{array} \right. \) is a distribution function.

- Non-decreasing,
- Limits of 1 and 0 as \( x \to \pm \infty \).

Generalized gamma CDF has variable transformed by raising to a power and multiplying: \( F(x) = \Gamma(\gamma, \lambda x^{\alpha})/\Gamma(\gamma) \)

1. That is, if \( X \sim \Gamma(\gamma) \), \( Y = (X/\lambda)^{1/\alpha} \). then \( Y \) has generalized gamma distribution with parameters \( \alpha, \lambda, \gamma \).
2. Argument to the Gamma CDF is the inverse transformation.
Generalized gamma

1. Idea: $\Gamma(\gamma, x)$ is defined as $\int_0^x x^{\gamma-1} \exp(-x) \, dx$ and is called the *incomplete gamma function*

   - $\Gamma(\gamma, x)$ is increasing in $x$.
   - $\lim_{x \to \infty} \Gamma(\gamma, x)$ is called the *Gamma function*

2. Then
   \[
   \begin{cases} 
   \frac{\Gamma(\gamma, x)}{\Gamma(\gamma)} & \text{if } x \geq 0 \\
   0 & \text{otherwise} 
   \end{cases}
   \]
   is a distribution function

   - Non-decreasing,
   - Limits of 1 and 0 as $x \to \pm\infty$.

3. Generalized gamma CDF has variable transformed by raising to a power and multiplying: $F(x) = \frac{\Gamma(\gamma, \lambda x^{\alpha})}{\Gamma(\gamma)}$

   - That is, if $X \sim \Gamma(\gamma)$, $Y = (X/\lambda)^{1/\alpha}$. then $Y$ has generalized gamma distribution with parameters $\alpha, \lambda, \gamma$.
   - Argument to the Gamma CDF is the inverse transformation.

4. Text gives two different density parameterizations, including

   \[
   f(x) = \alpha \lambda^\gamma x^{\alpha \gamma - 1} \exp(-\lambda x^\alpha)/\Gamma(\gamma)
   \]
Generalized gamma

1 Idea: $\Gamma(\gamma, x)$ is defined as $\int_0^x x^{\gamma-1} \exp(-x) \, dx$ and is called the **incomplete gamma function**

1 $\Gamma(\gamma, x)$ is increasing in $x$.

2 $\lim_{x \to \infty} \Gamma(\gamma, x)$ is called the **Gamma function**

3 Then \[ \begin{cases} \frac{\Gamma(\gamma, x)}{\Gamma(\gamma)} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \] is a distribution function

1 Non-decreasing,

2 Limits of 1 and 0 as $x \to \pm \infty$.

2 Generalized gamma CDF has variable transformed by raising to a power and multiplying: $F(x) = \Gamma(\gamma, \lambda x^\alpha)/\Gamma(\gamma)$

1 That is, if $X \sim \Gamma(\gamma)$, $Y = (X/\lambda)^{1/\alpha}$. then $Y$ has generalized gamma distribution with parameters $\alpha, \lambda, \gamma$.

2 Argument to the Gamma CDF is the inverse transformation.

3 Text gives two different density parameterizations, including

$$f(x) = \alpha \lambda^\gamma x^{\alpha \gamma - 1} \exp(-\lambda x^\alpha)/\Gamma(\gamma)$$

1 using $u = \lambda y^\alpha$, $y = (u/\lambda)^{1/\alpha}$, $dy = (u/\lambda)^{1/\alpha - 1}/(\alpha \lambda) \, du$. 

Regression Models: Parametric Failure Time Distributions Lecture 09
Generalized Gamma Special Cases

1. Weibull if $\gamma = 1$
Generalized Gamma Special Cases

1. Weibull if $\gamma = 1$
2. Gamma if $\alpha = 1$. 
Generalized Gamma Special Cases

1. Weibull if $\gamma = 1$
2. Gamma if $\alpha = 1$.
3. Log normal if $\gamma \to \infty$

CDF of $W = \log(X)$ is $\Gamma(\gamma, \exp(\alpha(w + \log(\lambda)/\alpha)))/\Gamma(\gamma)$.

Need to make $\alpha, \lambda$ depend on $\gamma$.

Use the fact that $\Gamma(\gamma, (x - \gamma)/\sqrt{\gamma})/\Gamma(\gamma) \to \Phi(x)$ by CLT.

Hence generalized gamma distribution close to normal for large $\gamma$, so long as $\alpha$ and $\lambda$ are allowed to vary appropriately with $\gamma$. 
Generalized Gamma Special Cases

1. Weibull if $\gamma = 1$
2. Gamma if $\alpha = 1$.
3. Log normal if $\gamma \to \infty$
   1. CDF of $W = \log(X)$ is $\Gamma(\gamma, \exp(\alpha(w + \log(\lambda)/\alpha)))/\Gamma(\gamma)$
Generalized Gamma Special Cases

1. Weibull if $\gamma = 1$
2. Gamma if $\alpha = 1$.
3. Log normal if $\gamma \to \infty$
   1. CDF of $W = \log(X)$ is $\Gamma(\gamma, \exp(\alpha(w + \log(\lambda)/\alpha))/\Gamma(\gamma)$
   2. Need to make $\alpha, \lambda$ depend on $\gamma$. 
Generalized Gamma Special Cases

1. Weibull if $\gamma = 1$
2. Gamma if $\alpha = 1$.
3. Log normal if $\gamma \to \infty$

- CDF of $W = \log(X)$ is $\Gamma(\gamma, \exp(\alpha(w + \log(\lambda)/\alpha))/\Gamma(\gamma)$
- Need to make $\alpha, \lambda$ depend on $\gamma$.
- Use the fact that $\Gamma(\gamma, (x - \gamma)/\sqrt{\gamma})/\Gamma(\gamma) \to \Phi(x)$ by CLT
Generalized Gamma Special Cases

1. Weibull if $\gamma = 1$
2. Gamma if $\alpha = 1$.
3. Log normal if $\gamma \to \infty$

   1. CDF of $W = \log(X)$ is $\Gamma(\gamma, \exp(\alpha(w + \log(\lambda)/\alpha))/\Gamma(\gamma)$
   2. Need to make $\alpha$, $\lambda$ depend on $\gamma$.
   3. Use the fact that $\Gamma(\gamma, (x - \gamma)/\sqrt{\gamma})/\Gamma(\gamma) \to \Phi(x)$ by CLT
   4. Hence generalized gamma distribution close to normal for large $\gamma$, so long as $\alpha$ and $\lambda$ are allowed to vary appropriately with $\gamma$. 
Objectives Lecture 10

1. Fitting Models via Maximum Likelihood
Objectives Lecture 10

1. Fitting Models via Maximum Likelihood
2. Accelerated failure diagnostics
Objectives Lecture 10

1. Fitting Models via Maximum Likelihood
2. Accelerated failure diagnostics
3. Readings: KM §12.2b, 12.5
Section: Regression Models

Subsection: Estimation, Confidence Intervals for Various Quantities
Estimate Parameters and Standard Errors Via Likelihood

1. Likelihood is

\[ L(\beta, \alpha, \sigma) = \prod_{j=1}^{n} f_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{\delta_j} \sigma^{-\delta_j} S_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{1-\delta_j} \]
Likelihood is
\[ L(\beta, \alpha, \sigma) = \prod_{j=1}^{n} f_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{\delta_j} \sigma^{-\delta_j} S_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{1-\delta_j} \]

Maximizer gives estimate
Likelihood is \( L(\beta, \alpha, \sigma) = \prod_{j=1}^{n} f_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{\delta_j} \sigma^{-\delta_j} S_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{1-\delta_j} \)

Maximizer gives estimate

Simplest example is Weibull
Likelihood is \( L(\beta, \alpha, \sigma) = \prod_{j=1}^{n} f_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{\delta_j} \sigma^{-\delta_j} S_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{1-\delta_j} \)

Maximizer gives estimate

Simplest example is Weibull

1. \( W \) has unit exponential distribution.
Estimate Parameters and Standard Errors Via Likelihood

1. Likelihood is \( L(\beta, \alpha, \sigma) = \prod_{j=1}^{n} f_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{\delta_j} \sigma^{-\delta_j} S_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{1-\delta_j} \)

2. Maximizer gives estimate

3. Simplest example is Weibull
   - \( W \) has unit exponential distribution.
   - \( S_W(w) = \exp(-w) \)
Likelihood is \( L(\beta, \alpha, \sigma) = \prod_{j=1}^{n} f_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{\delta_j} \sigma^{-\delta_j} S_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{1-\delta_j} \)

Maximizer gives estimate

Simplest example is Weibull

1. \( W \) has unit exponential distribution.
2. \( S_W(w) = \exp(-w) \)
3. \( S_U(u) = P[U > u] = P[W > \exp(u)] = \exp(-\exp(u)) \).

For location and scale families, can fit regression model with only intercept to estimate \( \eta \) by residual variation, \( \lambda \) as exponential of intercept.
Likelihood is \( L(\beta, \alpha, \sigma) = \prod_{j=1}^{n} f_{U}\left(\frac{\log(T_j) - \alpha - \beta z_j}{\sigma}\right)^{\delta_j} \sigma^{-\delta_j} S_{U}\left(\frac{\log(T_j) - \alpha - \beta z_j}{\sigma}\right)^{1-\delta_j}\)

Maximizer gives estimate

Simplest example is Weibull

1. \( W \) has unit exponential distribution.
2. \( S_{W}(w) = \exp(-w) \)
3. \( S_{U}(u) = P[U > u] = P[W > \exp(u)] = \exp(-\exp(u)) \).
4. \( f_{U}(u) = \exp(-\exp(u)) \exp(u) \).
Likelihood is \( L(\beta, \alpha, \sigma) = \prod_{j=1}^{n} f_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{\delta_j} \sigma^{-\delta_j} S_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{1-\delta_j} \)

Maximizer gives estimate

Simplest example is Weibull

1. \( W \) has unit exponential distribution.
2. \( S_W(w) = \exp(-w) \)
3. \( S_U(u) = P[U > u] = P[W > \exp(u)] = \exp(-\exp(u)). \)
4. \( f_U(u) = \exp(-\exp(u)) \exp(u). \)

For location and scale families, can fit regression model with only intercept to estimate \( \eta \) by residual variation, \( \lambda \) as exponential of intercept.
Estimate Parameters and Standard Errors Via Likelihood

1. Likelihood is
   \[ L(\beta, \alpha, \sigma) = \prod_{j=1}^{n} f_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{\delta_j} \sigma^{-\delta_j} S_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{1-\delta_j} \]

2. Maximizer gives estimate

3. Simplest example is Weibull
   1. \( W \) has unit exponential distribution.
   2. \( S_W(w) = \exp(-w) \)
   3. \( S_U(u) = P[U > u] = P[W > \exp(u)] = \exp(-\exp(u)) \).
   4. \( f_U(u) = \exp(-\exp(u)) \exp(u) \).

4. For location and scale families, can fit regression model with only intercept to estimate \( \eta \) by residual variation, \( \lambda \) as exponential of intercept.
   1. ex. Weibull, Log Normal.
Likelihood is \( L(\beta, \alpha, \sigma) = \prod_{j=1}^{n} f_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{\delta_j} \sigma^{-\delta_j} S_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{1-\delta_j} \)

Maximizer gives estimate

Simplest example is Weibull

- \( W \) has unit exponential distribution.
- \( S_W(w) = \exp(-w) \)
- \( S_U(u) = P[U > u] = P[W > \exp(u)] = \exp(-\exp(u)). \)
- \( f_U(u) = \exp(-\exp(u)) \exp(u). \)

For location and scale families, can fit regression model with only intercept to estimate \( \eta \) by residual variation, \( \lambda \) as exponential of intercept.

- ex. Weibull, Log Normal.

Some distributions have additional parameters estimable similarly.
Estimate Parameters and Standard Errors Via Likelihood

1. Likelihood is \( L(\beta, \alpha, \sigma) = \prod_{j=1}^{n} f_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{\delta_j} \sigma^{-\delta_j} S_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{1-\delta_j} \)

2. Maximizer gives estimate

3. Simplest example is Weibull

   1. \( W \) has unit exponential distribution.
   2. \( S_W(w) = \exp(-w) \)
   3. \( S_U(u) = P[U > u] = P[W > \exp(u)] = \exp(-\exp(u)) \).
   4. \( f_u(u) = \exp(-\exp(u)) \exp(u) \).

4. For location and scale families, can fit regression model with only intercept to estimate \( \eta \) by residual variation, \( \lambda \) as exponential of intercept.

   1. ex. Weibull, Log Normal.

5. Some distributions have additional parameters estimable similarly.

   1. ex. generalized gamma
Likelihood is $L(\beta, \alpha, \sigma) =$
\[
\prod_{j=1}^{n} f_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{\delta_j} \sigma^{-\delta_j} S_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{1-\delta_j}
\]

Maximizer gives estimate

Simplest example is Weibull

$W$ has unit exponential distribution.

$S_W(w) = \exp(-w)$

$S_U(u) = \Pr[U > u] = \Pr[W > \exp(u)] = \exp(-\exp(u))$.

$f_U(u) = \exp(-\exp(u)) \exp(u)$.

For location and scale families, can fit regression model with only intercept to estimate $\eta$ by residual variation, $\lambda$ as exponential of intercept.

ex. Weibull, Log Normal.

Some distributions have additional parameters estimable similarly.

ex. generalized gamma

Get approximate covariance matrix $\Sigma$ of $(\hat{\alpha}, \hat{\beta}, \hat{\sigma})$ using $\ell''$
Likelihood is \( L(\beta, \alpha, \sigma) = \prod_{j=1}^{n} f_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{\delta_j} \sigma^{-\delta_j} S_U \left( \frac{\log(T_j) - \alpha - \beta z_j}{\sigma} \right)^{1-\delta_j} \)

Maximizer gives estimate

Simplest example is Weibull

1. \( W \) has unit exponential distribution.
2. \( S_W(w) = \exp(-w) \)
3. \( S_U(u) = P[U > u] = P[W > \exp(u)] = \exp(-\exp(u)) \).
4. \( f_U(u) = \exp(-\exp(u)) \exp(u) \).

For location and scale families, can fit regression model with only intercept to estimate \( \eta \) by residual variation, \( \lambda \) as exponential of intercept.

1. ex. Weibull, Log Normal.

Some distributions have additional parameters estimable similarly.

1. ex. generalized gamma

Get approximate covariance matrix \( \Sigma \) of \((\hat{\alpha}, \hat{\beta}, \hat{\sigma})\) using \( \ell'' \)

\[ \Sigma = \text{Var} \left[ (\hat{\alpha}, \hat{\beta}, \hat{\sigma}, \ldots) \right] \approx [\ell'']^{-1} \]

SAS Code R Code
Estimate derived quantities via maximum likelihood

1. $\hat{S}(t)$ with covariates $z$ is $\hat{p} = S_U([\log(t) - \hat{\alpha} - \hat{\beta} z]/\hat{\sigma})$
Estimate derived quantities via maximum likelihood

\( \hat{S}(t) \) with covariates \( z \) is \( \hat{p} = S_U([\log(t) - \hat{\alpha} - \hat{\beta}z]/\hat{\sigma}) \)

\[ \frac{d\hat{p}}{d\hat{\alpha}} = -f_U(\cdots)/\hat{\sigma}, \]

\[ \frac{d\hat{p}}{d\hat{\beta}} = -f_U(\cdots)/\hat{\sigma}, \]

\[ \frac{d\hat{p}}{d\hat{\sigma}} = f_U(\cdots)/\hat{\sigma}^2, \]
Estimate derived quantities via maximum likelihood

1. $\hat{S}(t)$ with covariates $z$ is $\hat{p} = S_U([\log(t) - \hat{\alpha} - \hat{\beta}z]/\hat{\sigma})$

   1. $\frac{d\hat{p}}{d\hat{\alpha}} = -f_U(\cdots)/\hat{\sigma}$,

   2. $\frac{d\hat{p}}{d\hat{\beta}} = -f_U(\cdots)z/\hat{\sigma}$
Estimate derived quantities via maximum likelihood

1. $\hat{S}(t)$ with covariates $z$ is $\hat{p} = S_U([\log(t) - \hat{\alpha} - \hat{\beta}z]/\hat{\sigma})$

1. $\frac{d\hat{p}}{d\hat{\alpha}} = -f_U(\cdots)/\hat{\sigma}$,

2. $\frac{d\hat{p}}{d\hat{\beta}} = -f_U(\cdots)z/\hat{\sigma}$

3. $\frac{d\hat{p}}{d\hat{\sigma}} = f_U(\cdots)/\hat{\sigma}^2$
Estimate derived quantities via maximum likelihood

1. $\hat{S}(t)$ with covariates $z$ is $\hat{p} = S_U([\log(t) - \hat{\alpha} - \hat{\beta}z]/\hat{\sigma})$

1. $\frac{d\hat{p}}{d\hat{\alpha}} = -f_U(\cdots)/\hat{\sigma}$,
2. $\frac{d\hat{p}}{d\hat{\beta}} = -f_U(\cdots)z/\hat{\sigma}$
3. $\frac{d\hat{p}}{d\hat{\sigma}} = f_U(\cdots)/\hat{\sigma}^2$
4. Let $\mathbf{v} = (\frac{d\hat{p}}{d\hat{\alpha}}, \frac{d\hat{p}}{d\hat{\beta}}, \frac{d\hat{p}}{d\hat{\sigma}})$
Estimate derived quantities via maximum likelihood

1. \( \hat{S}(t) \) with covariates \( z \) is \( \hat{p} = S_U([\log(t) - \hat{\alpha} - \hat{\beta}z]/\hat{\sigma}) \)

   - \( \frac{d\hat{p}}{d\hat{\alpha}} = -f_U(\cdots)/\hat{\sigma} \),
   - \( \frac{d\hat{p}}{d\hat{\beta}} = -f_U(\cdots)z/\hat{\sigma} \)
   - \( \frac{d\hat{p}}{d\hat{\sigma}} = f_U(\cdots)/\hat{\sigma}^2 \)

2. Let \( \mathbf{v} = (\frac{d\hat{p}}{d\hat{\alpha}}, \frac{d\hat{p}}{d\hat{\beta}}, \frac{d\hat{p}}{d\hat{\sigma}}) \)

3. \[ \text{Var} \left[ g(\hat{\alpha}, \hat{\beta}, \hat{\sigma}, \ldots) \right] = (g')^\top \text{Var} \left[ (\hat{\alpha}, \hat{\beta}, \hat{\sigma}, \ldots) \right] g' \]
Estimate derived quantities via maximum likelihood

\[ \hat{S}(t) \text{ with covariates } z \text{ is } \hat{p} = S_U([\log(t) - \hat{\alpha} - \hat{\beta}z]/\hat{\sigma}) \]

1. \[ \frac{d\hat{p}}{d\hat{\alpha}} = -f_U(\cdots)/\hat{\sigma}, \]
2. \[ \frac{d\hat{p}}{d\hat{\beta}} = -f_U(\cdots)z/\hat{\sigma} \]
3. \[ \frac{d\hat{p}}{d\hat{\sigma}} = f_U(\cdots)/\hat{\sigma}^2 \]
4. Let \( \mathbf{v} = (\frac{d\hat{p}}{d\hat{\alpha}}, \frac{d\hat{p}}{d\hat{\beta}}, \frac{d\hat{p}}{d\hat{\sigma}}) \)
5. \[ \text{Var} \left[ g(\hat{\alpha}, \hat{\beta}, \hat{\sigma}, \ldots) \right] = (g')^\top \text{Var} \left[ (\hat{\alpha}, \hat{\beta}, \hat{\sigma}, \ldots) \right] g' \]

\text{Delta method, lecture 2.}
Estimate derived quantities via maximum likelihood

$\hat{S}(t)$ with covariates $z$ is $\hat{p} = S_U([\log(t) - \hat{\alpha} - \hat{\beta}z]/\hat{\sigma})$

1. $\frac{d\hat{p}}{d\hat{\alpha}} = -f_U(\cdots)/\hat{\sigma}$
2. $\frac{d\hat{p}}{d\hat{\beta}} = -f_U(\cdots)z/\hat{\sigma}$
3. $\frac{d\hat{p}}{d\hat{\sigma}} = f_U(\cdots)/\hat{\sigma}^2$

4. Let $v = (\frac{d\hat{p}}{d\hat{\alpha}}, \frac{d\hat{p}}{d\hat{\beta}}, \frac{d\hat{p}}{d\hat{\sigma}})$

5. $\text{Var} \left[ g(\hat{\alpha}, \hat{\beta}, \hat{\sigma}, \ldots) \right] = (g')^\top \text{Var} \left[ (\hat{\alpha}, \hat{\beta}, \hat{\sigma}, \ldots) \right] g'$

6. $\text{SE} [\hat{p}] = \sqrt{v^\top \Sigma v}$
Estimate derived quantities via maximum likelihood

$\hat{S}(t)$ with covariates $\mathbf{z}$ is $\hat{p} = S_U([\log(t) - \hat{\alpha} - \hat{\beta}\mathbf{z}] / \hat{\sigma})$

1. $\frac{d\hat{p}}{d\hat{\alpha}} = -f_U(\cdots) / \hat{\sigma}$,
2. $\frac{d\hat{p}}{d\hat{\beta}} = -f_U(\cdots) \mathbf{z} / \hat{\sigma}$
3. $\frac{d\hat{p}}{d\hat{\sigma}} = f_U(\cdots) / \hat{\sigma}^2$
4. Let $\mathbf{v} = (\frac{d\hat{p}}{d\hat{\alpha}}, \frac{d\hat{p}}{d\hat{\beta}}, \frac{d\hat{p}}{d\hat{\sigma}})$
5. $\text{Var} \left[ g(\hat{\alpha}, \hat{\beta}, \hat{\sigma}, \ldots) \right] = (g')^\top \text{Var} \left[ (\hat{\alpha}, \hat{\beta}, \hat{\sigma}, \ldots) \right] g'$

Delta method, lecture 2.

6. $\text{SE} [\hat{\rho}] = \sqrt{\mathbf{v}^\top \Sigma \mathbf{v}}$
7. $\text{SE} [\log(\hat{\rho})] = \text{SE} [\hat{\rho}] / \hat{\rho}$
acceleration factor $\exp((z_k - z_j) \hat{\beta})$
acceleration factor \( \exp((z_k - z_j)\hat{\beta}) \)

Let \( \mathbf{v} = (0, z_k - z_j, 0) \)
acceleration factor $\exp((z_k - z_j)\hat{\beta})$

1. Let $v = (0, z_k - z_j, 0)$
2. SE on log scale is $\sqrt{v^T \Sigma v}$
acceleration factor \( \exp((z_k - z_j)\hat{\beta}) \)

1. Let \( \mathbf{v} = (0, z_k - z_j, 0) \)

2. SE on log scale is \( \sqrt{\mathbf{v}^\top \Sigma \mathbf{v}} \)

relative risk \( \exp((z_k - z_j)\hat{\beta}/\hat{\sigma}) \)

1. Let \( \mathbf{v} = (0, (z_k - z_j)/\hat{\sigma}, -1/\hat{\sigma}) \)

2. SE on log scale is \( \sqrt{\mathbf{v}^\top \Sigma \mathbf{v}} \)
acceleration factor $\exp((z_k - z_j)\hat{\beta})$ 

Let $\mathbf{v} = (0, z_k - z_j, 0)$

SE on log scale is $\sqrt{\mathbf{v}^\top \Sigma \mathbf{v}}$

relative risk $\exp((z_k - z_j)\hat{\beta}/\hat{\sigma})$ 

Let $\mathbf{v} = (0, (z_k - z_j)/\hat{\sigma}, -1/\hat{\sigma}^2)$
acceleration factor $\exp((z_k - z_j)\hat{\beta})$

1. Let $\mathbf{v} = (0, z_k - z_j, 0)$
2. SE on log scale is $\sqrt{\mathbf{v}^\top \Sigma \mathbf{v}}$

relative risk $\exp((z_k - z_j)\hat{\beta}/\hat{\sigma})$

1. Let $\mathbf{v} = (0, (z_k - z_j)/\hat{\sigma}, -1/\hat{\sigma}^2)$
2. SE on log scale is $\sqrt{\mathbf{v}^\top \Sigma \mathbf{v}}$
acceleration factor \( \exp((z_k - z_j)\hat{\beta}) \)

1. Let \( v = (0, z_k - z_j, 0) \)
2. SE on log scale is \( \sqrt{v^\top \Sigma v} \)

relative risk \( \exp((z_k - z_j)\hat{\beta}/\hat{\sigma}) \)

1. Let \( v = (0, (z_k - z_j)/\hat{\sigma}, -1/\hat{\sigma}^2) \)
2. SE on log scale is \( \sqrt{v^\top \Sigma v} \)

CI’s best on log scale  R Code  SAS Code
acceleration factor \( \exp((z_k - z_j)\hat{\beta}) \)

1. Let \( \mathbf{v} = (0, z_k - z_j, 0) \)
2. SE on log scale is \( \sqrt{\mathbf{v}^\top \Sigma \mathbf{v}} \)

relative risk \( \exp((z_k - z_j)\hat{\beta}/\hat{\sigma}) \)

1. Let \( \mathbf{v} = (0, (z_k - z_j)/\hat{\sigma}, -1/\hat{\sigma}^2) \)
2. SE on log scale is \( \sqrt{\mathbf{v}^\top \Sigma \mathbf{v}} \)

Cl’s best on log scale  R Code  SAS Code

Can fit hazard ratios and acceleration factors.  SAS Code  R Code
Interpretation:

1. Scale gives shape parameter for exponential family
Interpretation:

1. Scale gives shape parameter for exponential family
2. Intercept gives location parameter (sort of).
One-Sample:

1. If you want to diagnose family with specific member unspecified,

$$S(t) = S_0(\exp((\log(t) - \psi)/\phi))$$ for $$S_0$$ known, $$\psi$$, $$\phi$$ unknown,

Cumulative hazard estimated without using model should agree with model

$$H(t) = H_0(\exp((\log(t) - \psi)/\phi))$$ for $$H_0 = -\log(S_0)$$

$$\log(H_0^{-1}(\hat{H}(t))) \approx \text{linear in } \log(t)$$.

$$\hat{H}$$ is Nelson-Allen estimator

Slope is dispersion parameter
One-Sample:

1. If you want to diagnose family with specific member unspecified,

\[ S(t) = S_0\left(\exp\left(\frac{\log(t) - \psi}{\phi}\right)\right) \]

for

\[ H(t) = H_0\left(\exp\left(\frac{\log(t) - \psi}{\phi}\right)\right) \]

Hence \[ \log(H^{-1}(\hat{H}(t))) \] approx. linear in \[ \log(t) \].

\[ \hat{H} \] is Nelson-Allen estimator

slope is dispersion parameter
One-Sample:

1. If you want to diagnose family with specific member unspecified,
   1. \( S(t) = S_0(\exp((\log(t) - \psi)/\phi)) \) for
   2. \( S_0 \) known
If you want to diagnose family with specific member unspecified,

1. $S(t) = S_0(\exp((\log(t) - \psi)/\phi))$ for
2. $S_0$ known
3. $\psi, \phi$ unknown,
One-Sample:

1. If you want to diagnose family with specific member unspecified,
   a. \( \hat{S}(t) = S_0(\exp((\log(t) - \psi)/\phi)) \) for
   b. \( S_0 \) known
   c. \( \psi, \phi \) unknown,

2. Cumulative hazard estimated without using model should agree with model

\( \hat{\psi} \) is Nelson-Allen estimator

Slope is dispersion parameter
One-Sample:

1. If you want to diagnose family with specific member unspecified,
   
   \[ S(t) = S_0(\exp((\log(t) - \psi)/\phi)) \] for

   2. \( S_0 \) known

   3. \( \psi, \phi \) unknown,

2. Cumulative hazard estimated without using model should agree with model

   \[ H(t) = H_0(\exp((\log(t) - \psi)/\phi)) \] for \( H_0 = -\log(S_0) \)
One-Sample:

1. If you want to diagnose family with specific member unspecified,
   - \[ S(t) = S_0(\exp((\log(t) - \psi)/\phi)) \]
   1. \( S_0 \) known
   2. \( \psi, \phi \) unknown,

2. Cumulative hazard estimated without using model should agree with model
   - \[ H(t) = H_0(\exp((\log(t) - \psi)/\phi)) \]
   - \( H_0 = -\log(S_0) \)
   1. \( \log(H_0^{-1}(H(t))) = (\log(t) - \psi)/\phi \), linear in \( \log(t) \).
One-Sample:

1. If you want to diagnose family with specific member unspecified,
   \[ S(t) = S_0(\exp((\log(t) - \psi)/\phi)) \] for
2. \( S_0 \) known
3. \( \psi, \phi \) unknown,

Cumulative hazard estimated without using model should agree with model

1. \( H(t) = H_0(\exp((\log(t) - \psi)/\phi)) \) for \( H_0 = -\log(S_0) \)
2. \( \log(H_0^{-1}(H(t))) = (\log(t) - \psi)/\phi \), linear in \( \log(t) \)
3. Hence \( \log(H_0^{-1}({\hat H}(t))) \) approx. linear in \( \log(t) \).
One-Sample:

1. If you want to diagnose family with specific member unspecified, 
   \[ S(t) = S_0(\exp((\log(t) - \psi)/\phi)) \text{ for } \]
   \[ S_0 \text{ known} \]
   \[ \psi, \phi \text{ unknown}, \]

2. Cumulative hazard estimated without using model should agree with model 
   \[ H(t) = H_0(\exp((\log(t) - \psi)/\phi)) \text{ for } H_0 = -\log(S_0) \]
   \[ \log(H_0^{-1}(H(t))) = (\log(t) - \psi)/\phi, \text{ linear in } \log(t) \]
   \[ \text{Hence } \log(H_0^{-1}(\hat{H}(t))) \text{ approx. linear in } \log(t). \]
   \[ \hat{H} \text{ is Nelson-Allen estimator} \]
One-Sample:

1. If you want to diagnose family with specific member unspecified,
   \[ S(t) = S_0(\exp((\log(t) - \psi)/\phi)) \] for
   1. \( S_0 \) known
   2. \( \psi, \phi \) unknown,

2. Cumulative hazard estimated without using model should agree with model
   \[ H(t) = H_0(\exp((\log(t) - \psi)/\phi)) \] for \( H_0 = -\log(S_0) \)
   1. \( \log(H_0^{-1}(H(t))) = (\log(t) - \psi)/\phi \), linear in \( \log(t) \)
   2. Hence \( \log(H_0^{-1}(\hat{H}(t))) \) approx. linear in \( \log(t) \).
   3. \( \hat{H} \) is Nelson-Allen estimator
   4. Slope is dispersion parameter
Form depends on specific family:

- **Weibull:**
  - Easiest member is exponential
  - \( \log(H - 10(s)) = \log(s) \)
  - Plot \( \log(\hat{H}(t)) \) vs \( \log(t) \)

- **Log normal:**
  - Easiest member is standard normal
  - \( H_0(t) = -\log(\Phi(-\log(t))) \)
  - \( H^{-1}_0(s) = \exp(-\Phi^{-1}(\exp(-s))) \)
  - Plot \( -\Phi^{-1}(\exp(-\hat{H}(t))) \) vs \( \log(t) \)

- **Log logistic:**
  - \( H_0(t) = \log(1 + t) \)
  - \( H^{-1}_0(s) = \exp(s - 1) \)
  - \( \log(H^{-1}_0(s)) = \log(\exp(s) - 1) \)
  - Plot \( \log(\exp(\hat{H}(t)) - 1) \) vs \( \log(t) \)
Form depends on specific family:

1. **Weibull:**
   - Easiest member is exponential
   - \( \log(H - 10(s)) = \log(s) \)
   - Plot \( \log(\hat{H}(t)) \) vs \( \log(t) \)
   - Exponential if slope is 1

2. **Log normal:**
   - Easiest member is standard normal
   - \( H_0(t) = -\log(\Phi(-\log(t))) \)
   - \( H^{-10}(s) = \exp(-\Phi^{-1}(\exp(-s))) \)
   - Plot \( -\Phi^{-1}(\exp(-\hat{H}(t))) \) vs \( \log(t) \)

3. **Log logistic:**
   - \( H_0(t) = \log(1 + t) \)
   - \( H^{-10}(s) = \exp(s - 1) \)
   - \( \log(H^{-10}(s)) = \log(\exp(s) - 1) \)
   - Plot \( \log(\exp(\hat{H}(t)) - 1) \) vs \( \log(t) \)
Form depends on specific family:

- Weibull:
  - Easiest member is exponential

Lognormal:

\[
H(0)(t) = -\log(\Phi(-\log(t)))
\]

\[
H^{-1}(0)(s) = \exp(-\Phi^{-1}(\exp(-s)))
\]

Log logistic:

\[
H(0)(t) = \log(1 + t)
\]

\[
H^{-1}(0)(s) = \exp(s) - 1
\]

\[
\log(H^{-1}(0)(s)) = \log(\exp(s) - 1)
\]
Form depends on specific family:

- **Weibull:**
  1. Easiest member is exponential
  2. \[ \log(H_{0}^{-1}(s)) = \log(s) \]
Form depends on specific family:

1. **Weibull:**
   1. Easiest member is exponential
   2. \( \log(H_0^{-1}(s)) = \log(s) \)
   3. Plot \( \log(\hat{H}(t)) \) vs \( \log(t) \)
Form depends on specific family:

Weibull:

- Easiest member is exponential
- $\log(H^{-1}(s)) = \log(s)$
- Plot $\log(\hat{H}(t))$ vs $\log(t)$
- Exponential if slope is 1
Form depends on specific family:

1. **Weibull:**
   1. Easiest member is exponential
   2. \( \log(H^{-1}(s)) = \log(s) \)
   3. Plot \( \log(H(t)) \) vs \( \log(t) \)
   4. Exponential if slope is 1

2. **Log normal:**
Form depends on specific family:

1. **Weibull:**
   1. Easiest member is exponential
   2. \( \log(H_0^{-1}(s)) = \log(s) \)
   3. Plot \( \log(H(t)) \) vs \( \log(t) \)
   4. Exponential if slope is 1

2. **Log normal:**
   1. Easiest member is standard normal
Form depends on specific family:

1. **Weibull:**
   1. Easiest member is exponential
   2. \( \log(H_0^{-1}(s)) = \log(s) \)
   3. Plot \( \log(\hat{H}(t)) \) vs \( \log(t) \)
   4. Exponential if slope is 1

2. **Log normal:**
   1. Easiest member is standard normal
   2. \( H_0(t) = -\log(\Phi(-\log(t))) \)
Form depends on specific family:

1. Weibull:
   1. Easiest member is exponential
   2. \( \log(H_0^{-1}(s)) = \log(s) \)
   3. Plot \( \log(\hat{H}(t)) \) vs \( \log(t) \)
   4. Exponential if slope is 1

2. Log normal:
   1. Easiest member is standard normal
   2. \( H_0(t) = -\log(\Phi(-\log(t))) \)
   3. \( H_0^{-1}(s) = \exp(-\Phi^{-1}(\exp(-s))) \)
Form depends on specific family:

1. **Weibull:**
   1. Easiest member is exponential
   2. \( \log(H_0^{-1}(s)) = \log(s) \)
   3. Plot \( \log(\hat{H}(t)) \) vs \( \log(t) \)
   4. Exponential if slope is 1

2. **Log normal:**
   1. Easiest member is standard normal
   2. \( H_0(t) = -\log(\Phi(-\log(t))) \)
   3. \( H_0^{-1}(s) = \exp(-\Phi^{-1}(\exp(-s))) \)
   4. Plot \( -\Phi^{-1}(\exp(-\hat{H}(t))) \) vs \( \log(t) \)
Form depends on specific family:

1. **Weibull:**
   - Easiest member is exponential
   - \( \log(H_0^{-1}(s)) = \log(s) \)
   - Plot \( \log(\hat{H}(t)) \) vs \( \log(t) \)
   - Exponential if slope is 1

2. **Log normal:**
   - Easiest member is standard normal
   - \( H_0(t) = -\log(\Phi(-\log(t))) \)
   - \( H_0^{-1}(s) = \exp(-\Phi^{-1}(\exp(-s))) \)
   - Plot \(-\Phi^{-1}(\exp(-\hat{H}(t)))\) vs \( \log(t) \)

3. **Log logistic:**
Form depends on specific family:

1. **Weibull:**
   - Easiest member is exponential
   - \( \log(H_0^{-1}(s)) = \log(s) \)
   - Plot \( \log(\hat{H}(t)) \) vs \( \log(t) \)
   - Exponential if slope is 1

2. **Log normal:**
   - Easiest member is standard normal
   - \( H_0(t) = -\log(\Phi(-\log(t))) \)
   - \( H_0^{-1}(s) = \exp(-\Phi^{-1}(\exp(-s))) \)
   - Plot \( -\Phi^{-1}(\exp(-\hat{H}(t))) \) vs \( \log(t) \)

3. **Log logistic:**
   - \( H_0(t) = \log(1 + t) \)
Form depends on specific family:

**Weibull:**
1. Easiest member is exponential
2. \( \log(H^{-1}_0(s)) = \log(s) \)
3. Plot \( \log(\hat{H}(t)) \) vs \( \log(t) \)
4. Exponential if slope is 1

**Log normal:**
1. Easiest member is standard normal
2. \( H_0(t) = -\log(\Phi(-\log(t))) \)
3. \( H^{-1}_0(s) = \exp(-\Phi^{-1}(\exp(-s))) \)
4. Plot \( -\Phi^{-1}(\exp(-\hat{H}(t))) \) vs \( \log(t) \)

**Log logistic:**
1. \( H_0(t) = \log(1 + t) \)
2. \( H^{-1}_0(s) = \exp(s) - 1 \)
Form depends on specific family:

1. **Weibull:**
   1. Easiest member is exponential
   2. \( \log(H_0^{-1}(s)) = \log(s) \)
   3. Plot \( \log(\hat{H}(t)) \) vs \( \log(t) \)
   4. Exponential if slope is 1

2. **Log normal:**
   1. Easiest member is standard normal
   2. \( H_0(t) = -\log(\Phi(-\log(t))) \)
   3. \( H_0^{-1}(s) = \exp(-\Phi^{-1}(\exp(-s))) \)
   4. Plot \( -\Phi^{-1}(\exp(-\hat{H}(t))) \) vs \( \log(t) \)

3. **Log logistic:**
   1. \( H_0(t) = \log(1 + t) \)
   2. \( H_0^{-1}(s) = \exp(s) - 1 \)
   3. \( \log(H_0^{-1}(s)) = \log(\exp(s) - 1) \)
Form depends on specific family:

1. **Weibull:**
   - Easiest member is exponential
   - \( \log(H_0^{-1}(s)) = \log(s) \)
   - Plot \( \log(\hat{H}(t)) \) vs \( \log(t) \)
   - Exponential if slope is 1

2. **Log normal:**
   - Easiest member is standard normal
   - \( H_0(t) = -\log(\Phi(-\log(t))) \)
   - \( H_0^{-1}(s) = \exp(-\Phi^{-1}(\exp(-s))) \)
   - Plot \( -\Phi^{-1}(\exp(-\hat{H}(t))) \) vs \( \log(t) \)

3. **Log logistic:**
   - \( H_0(t) = \log(1 + t) \)
   - \( H_0^{-1}(s) = \exp(s) - 1 \)
   - \( \log(H_0^{-1}(s)) = \log(\exp(s) - 1) \)
   - Plot \( \log(\exp(\hat{H}(t)) - 1) \) vs \( \log(t) \)
Multiple Sample:

Counterpart to Anderson Plots:
Multiple Sample:

1. Counterpart to Anderson Plots:
2. Plot $\hat{S}^{-1}(p)$ computed group-wise against each other
Multiple Sample:

1. Counterpart to Anderson Plots:
2. Plot $\hat{S}^{-1}(p)$ computed group-wise against each other
3. $\hat{S}$ computed nonparametrically R Code
Regression Case:

1. Calculate Cox and Snell Residuals
Regression Case:

1. Calculate Cox and Snell Residuals
   1. This time \( \hat{H} \) computed parametrically

\[ P[\text{\textit{T}}_i > \text{\textit{t}}] = \exp(-\exp((\log(\text{\textit{t}}) - \alpha - z_i \beta) / \sigma)). \]

Cumulative hazard
\[ -\log(P[\text{\textit{T}}_i > \text{\textit{t}}]) = \exp((\log(\text{\textit{t}}) - \alpha - z_i \beta) / \sigma)). \]

CS residual is fitted CH \[ -\log(\bar{\Phi}((\log(\text{\textit{t}}) - \hat{\alpha} - z_i \hat{\beta}) / \hat{\sigma})). \]

\[ P[\text{\textit{T}}_i > \text{\textit{t}}] = \frac{1}{1 + \exp((\log(\text{\textit{t}}) - \alpha/\sigma - z_i \beta) / \sigma)). \]

Cumulative hazard
\[ -\log(1 + \exp((\log(\text{\textit{t}}) - \alpha/\sigma - z_i \beta) / \sigma))). \]

CS residual is fitted CH
\[ -\log(\bar{\Phi}((\log(\text{\textit{t}}) - \hat{\alpha} - z_i \hat{\beta}) / \hat{\sigma})). \]

\&H for residuals should be approximately 45° line

Simulation shows that this technique can identify correct model.

Regression Case:

1. Calculate Cox and Snell Residuals
   1. This time $\hat{H}$ computed parametrically
   2. Weibull:

   $P[T_i > t] = \exp\left(-\exp\left((\log(t) - \alpha - z_i \beta)/\sigma\right)\right)$.

   Cumulative hazard $-\log(P[T_i > t]) = \exp\left((\log(t) - \alpha - z_i \beta)/\sigma\right)$.

   CS residual is fitted CH $-\log\left(\bar{\Phi}\left((\log(t) - \hat{\alpha} - z_i \hat{\beta})/\hat{\sigma}\right)\right)$.

2. Log normal:

   $P[T_i > t] = \bar{\Phi}\left((\log(t) - \alpha/\sigma - z_i \beta)/\sigma\right)$.

   Cumulative hazard $-\log(\bar{\Phi}\left((\log(t) - \alpha - z_i \beta)/\sigma\right))$.

   CS residual is fitted CH $-\log\left(\bar{\Phi}\left((\log(t) - \hat{\alpha} - z_i \hat{\beta})/\hat{\sigma}\right)\right)$.

3. Log logistic:

   $P[T_i > t] = \frac{1}{1 + \exp\left((\log(t) - \alpha/\sigma - z_i \beta)/\sigma\right)}$.

   Cumulative hazard $\log\left(1 + \exp\left((\log(t) - \alpha/\sigma - z_i \beta)/\sigma\right)\right)$.

   CS residual is fitted CH $\log\left(1 + \exp\left((\log(t) - \hat{\alpha} - z_i \hat{\beta})/\hat{\sigma}\right)\right)$.

   $\hat{H}$ for residuals should be approximately 45° line.

Simulation shows that this technique can identify correct model.
Regression Case:

1. Calculate Cox and Snell Residuals
   1. This time $\hat{H}$ computed parametrically
   2. Weibull:
      1. $P[T_i > t] = \exp(-\exp((\log(t) - \alpha - z_i\beta)/\sigma))$. 

Regression Case:

1. Calculate Cox and Snell Residuals
   1. This time \( \hat{H} \) computed parametrically
   2. Weibull:
      1. \( P[T_i > t] = \exp(-\exp((\log(t) - \alpha - z_i \beta)/\sigma)). \)
      2. Cumulative hazard \( -\log(P[T_i > t]) = \exp((\log(t) - \alpha - z_i \beta)/\sigma)). \)

\( \hat{H} \) for residuals should be approximately 45\( ^\circ \) line

Simulation shows that this technique can identify correct model.
Regression Case:

1. Calculate Cox and Snell Residuals
   1. This time \( \hat{H} \) computed parametrically
   2. Weibull:
      1. \( P[T_i > t] = \exp(-\exp((\log(t) - \alpha - z_i \beta)/\sigma)). \)
      2. Cumulative hazard - \( \log(P[T_i > t]) = \exp((\log(t) - \alpha - z_i \beta)/\sigma). \)
      3. CS residual is fitted CH \( \exp((\log(t) - \hat{\alpha} - z_i \hat{\beta})/\hat{\sigma}). \)
Regression Case:

1. Calculate Cox and Snell Residuals
   1. This time $\hat{H}$ computed parametrically
   2. Weibull:
      1. $P[T_i > t] = \exp(-\exp((\log(t) - \alpha - z_i\beta)/\sigma))$.
      2. Cumulative hazard $-\log(P[T_i > t]) = \exp((\log(t) - \alpha - z_i\beta)/\sigma)$.
      3. CS residual is fitted CH $\exp((\log(t) - \hat{\alpha} - z_i\hat{\beta})/\hat{\sigma})$.

3. Log normal:

$\hat{H}$ for residuals should be approximately 45° line

Simulation shows that this technique can identify correct model.
Regression Case:

1. Calculate Cox and Snell Residuals
   1. This time $\hat{H}$ computed parametrically
   2. Weibull:
      1. $P[T_i > t] = \exp(-\exp((\log(t) - \alpha - z_i\beta)/\sigma))$.
      2. Cumulative hazard $- \log(P[T_i > t]) = \exp((\log(t) - \alpha - z_i\beta)/\sigma)$.
      3. CS residual is fitted CH $\exp((\log(t) - \hat{\alpha} - z_i\hat{\beta})/\hat{\sigma})$.
   3. Log normal:
      1. $P[T_i > t] = \Phi((\log(t) - \alpha/\sigma - z_i\beta)/\sigma)$.
Regression Case:

1. Calculate Cox and Snell Residuals
   1. This time $\hat{H}$ computed parametrically
   2. Weibull:
      1. $P[T_i > t] = \exp(-\exp((\log(t) - \alpha - z_i\beta)/\sigma))$.
      2. Cumulative hazard $-\log(P[T_i > t]) = \exp((\log(t) - \alpha - z_i\beta)/\sigma)$.
      3. CS residual is fitted $\text{CH} \exp((\log(t) - \hat{\alpha} - z_i\hat{\beta})/\hat{\sigma})$.
   3. Log normal:
      1. $P[T_i > t] = \Phi((\log(t) - \alpha/\sigma - z_i\beta)/\sigma)$.
      2. Cumulative hazard $-\log(\Phi((\log(t) - \alpha - z_i\beta)/\sigma))$.
Regression Case:

1. Calculate Cox and Snell Residuals
   1. This time $\hat{H}$ computed parametrically
   2. Weibull:
      1. $P[T_i > t] = \exp(-\exp((\log(t) - \alpha - z_i\beta)/\sigma))$.
      2. Cumulative hazard $-\log(P[T_i > t]) = \exp((\log(t) - \alpha - z_i\beta)/\sigma)$.
      3. CS residual is fitted CH $\exp((\log(t) - \hat{\alpha} - z_i\hat{\beta})/\hat{\sigma})$.
   3. Log normal:
      1. $P[T_i > t] = \Phi((\log(t) - \alpha/\sigma - z_i\beta)/\sigma)$.
      2. Cumulative hazard $-\log(\Phi((\log(t) - \alpha - z_i\beta)/\sigma))$.
      3. CS residual is fitted CH $-\log(\Phi((\log(t) - \hat{\alpha} - z_i\hat{\beta})/\hat{\sigma}))$.
Regression Case:

1. Calculate Cox and Snell Residuals
   1. This time $\hat{H}$ computed parametrically
   2. Weibull:
      1. $P[T_i > t] = \exp(-\exp((\log(t) - \alpha - z_i\beta)/\sigma))$.
      2. Cumulative hazard $-\log(P[T_i > t]) = \exp((\log(t) - \alpha - z_i\beta)/\sigma)$.
      3. CS residual is fitted CH $\exp((\log(t) - \hat{\alpha} - z_i\hat{\beta})/\hat{\sigma})$.
   3. Log normal:
      1. $P[T_i > t] = \Phi((\log(t) - \alpha/\sigma - z_i\beta)/\sigma)$.
      2. Cumulative hazard $-\log(\Phi((\log(t) - \alpha - z_i\beta)/\sigma))$.
      3. CS residual is fitted CH $-\log(\Phi((\log(t) - \hat{\alpha} - z_i\hat{\beta})/\hat{\sigma}))$.
   4. Log logistic:
Regression Case:

1. Calculate Cox and Snell Residuals
   1. This time $\hat{H}$ computed parametrically
   2. Weibull:
      1. $P[T_i > t] = \exp(-\exp((\log(t) - \alpha - z_i\beta)/\sigma))$.
      2. Cumulative hazard $-\log(P[T_i > t]) = \exp((\log(t) - \alpha - z_i\beta)/\sigma)$.
      3. CS residual is fitted CH $\exp((\log(t) - \hat{\alpha} - z_i\hat{\beta})/\hat{\sigma})$.
   3. Log normal:
      1. $P[T_i > t] = \Phi((\log(t) - \alpha/\sigma - z_i\beta)/\sigma)$.
      2. Cumulative hazard $-\log(\Phi((\log(t) - \alpha - z_i\beta)/\sigma))$.
      3. CS residual is fitted CH $-\log(\Phi((\log(t) - \hat{\alpha} - z_i\hat{\beta})/\hat{\sigma}))$.
   4. Log logistic:
      1. $P[T_i > t] = 1/(1 + \exp((\log(t) - \alpha/\sigma - z_i\beta)/\sigma))$.
Regression Case:

1. Calculate Cox and Snell Residuals
   1. This time $\hat{H}$ computed parametrically
   2. Weibull:
      1. $P[T_i > t] = \exp(-\exp((\log(t) - \alpha - z_i \beta)/\sigma))$.
      2. Cumulative hazard $-\log(P[T_i > t]) = \exp((\log(t) - \alpha - z_i \beta)/\sigma)$.
      3. CS residual is fitted CH $\exp((\log(t) - \hat{\alpha} - z_i \hat{\beta})/\hat{\sigma})$.
   3. Log normal:
      1. $P[T_i > t] = \Phi((\log(t) - \alpha/\sigma - z_i \beta)/\sigma)$.
      2. Cumulative hazard $-\log(\Phi((\log(t) - \alpha - z_i \beta)/\sigma))$.
      3. CS residual is fitted CH $-\log(\Phi((\log(t) - \hat{\alpha} - z_i \hat{\beta})/\hat{\sigma}))$.
   4. Log logistic:
      1. $P[T_i > t] = 1/(1 + \exp((\log(t) - \alpha/\sigma - z_i \beta)/\sigma))$.
      2. Cumulative hazard $\log(1 + \exp((\log(t) - \alpha/\sigma - z_i \beta)/\sigma))$
Calculate Cox and Snell Residuals

This time \( \hat{H} \) computed parametrically

Weibull:
1. \( P[T_i > t] = \exp(-\exp((\log(t) - \alpha - z_i\beta)/\sigma)) \).
2. Cumulative hazard \(-\log(P[T_i > t]) = \exp((\log(t) - \alpha - z_i\beta)/\sigma)\).
3. CS residual is fitted CH \( \exp((\log(t) - \hat{\alpha} - z_i\hat{\beta})/\hat{\sigma}) \).

Log normal:
1. \( P[T_i > t] = \Phi((\log(t) - \alpha/\sigma - z_i\beta)/\sigma) \).
2. Cumulative hazard \(-\log(\Phi((\log(t) - \alpha - z_i\beta)/\sigma)) \).
3. CS residual is fitted CH \(-\log(\Phi((\log(t) - \hat{\alpha} - z_i\hat{\beta})/\hat{\sigma})) \).

Log logistic:
1. \( P[T_i > t] = 1/(1 + \exp((\log(t) - \alpha/\sigma - z_i\beta)/\sigma)) \).
2. Cumulative hazard \( \log(1 + \exp((\log(t) - \alpha/\sigma - z_i\beta)/\sigma)) \).
3. CS residual is fitted CH \( \log(1 + \exp((\log(t) - \hat{\alpha} - z_i\hat{\beta})/\hat{\sigma})) \).
Regression Case:

1. Calculate Cox and Snell Residuals
   1. This time $\hat{H}$ computed parametrically
   2. Weibull:
      1. $P[T_i > t] = \exp(-\exp((\log(t) - \alpha - z_i\beta)/\sigma))$.
      2. Cumulative hazard $-\log(P[T_i > t]) = \exp((\log(t) - \alpha - z_i\beta)/\sigma)$.
      3. CS residual is fitted CH $\exp((\log(t) - \hat{\alpha} - z_i\hat{\beta})/\hat{\sigma})$.
   3. Log normal:
      1. $P[T_i > t] = \Phi((\log(t) - \alpha/\sigma - z_i\beta)/\sigma)$.
      2. Cumulative hazard $-\log(\Phi((\log(t) - \alpha - z_i\beta)/\sigma))$.
      3. CS residual is fitted CH $-\log(\Phi((\log(t) - \hat{\alpha} - z_i\hat{\beta})/\hat{\sigma}))$.
   4. Log logistic:
      1. $P[T_i > t] = 1/(1 + \exp((\log(t) - \alpha/\sigma - z_i\beta)/\sigma))$.
      2. Cumulative hazard $\log(1 + \exp((\log(t) - \alpha/\sigma - z_i\beta)/\sigma))$.
      3. CS residual is fitted CH $\log(1 + \exp((\log(t) - \hat{\alpha} - z_i\hat{\beta})/\hat{\sigma}))$.

2. $\hat{H}$ for residuals should be approximately 45° line
Regression Case:

1. Calculate Cox and Snell Residuals
   1. This time \( \hat{H} \) computed parametrically
   2. Weibull:
      1. \( P[T_i > t] = \exp(- \exp((\log(t) - \alpha - z_i\beta)/\sigma)). \)
      2. Cumulative hazard \( - \log(P[T_i > t]) = \exp((\log(t) - \alpha - z_i\beta)/\sigma)). \)
      3. CS residual is fitted CH \( \exp((\log(t) - \hat{\alpha} - z_i\hat{\beta})/\hat{\sigma})). \)
   3. Log normal:
      1. \( P[T_i > t] = \Phi((\log(t) - \alpha/\sigma - z_i\beta)/\sigma)). \)
      2. Cumulative hazard \( - \log(\Phi((\log(t) - \alpha - z_i\beta)/\sigma)). \)
      3. CS residual is fitted CH \( - \log(\Phi((\log(t) - \hat{\alpha} - z_i\hat{\beta})/\hat{\sigma})). \)
   4. Log logistic:
      1. \( P[T_i > t] = 1/(1 + \exp((\log(t) - \alpha/\sigma - z_i\beta)/\sigma)). \)
      2. Cumulative hazard \( \log(1 + \exp((\log(t) - \alpha/\sigma - z_i\beta)/\sigma)). \)
      3. CS residual is fitted CH \( \log(1 + \exp((\log(t) - \hat{\alpha} - z_i\hat{\beta})/\hat{\sigma})). \)

2. \( \hat{H} \) for residuals should be approximately 45° line

3. Simulation shows that this technique can identify correct model.
Comparison of interpretations of different models:

<table>
<thead>
<tr>
<th>Event</th>
<th>Model</th>
<th>Interp. of $z_j \hat{\beta} &gt; 0$ (compared to baseline)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>Proportional Hazards</td>
<td>Better</td>
</tr>
<tr>
<td>Bad</td>
<td>Proportional Hazards</td>
<td>Worse</td>
</tr>
<tr>
<td>Good</td>
<td>Accelerated Failure</td>
<td>Worse</td>
</tr>
<tr>
<td>Bad</td>
<td>Accelerated Failure</td>
<td>Better</td>
</tr>
</tbody>
</table>
Section: Regression Models

Subsection: Estimation with interval censoring:
Notation for interval censoring:

1. Patients are screened at fixed intervals $t_0, t_1, t_2, \ldots, t_J$ for $t_{J+1} = \infty$.

2. Knowing that event happened within $(L_i, R_i]$ for $L_i, R_i \in \{t_1, \ldots, t_J\}$,

   - $R_i = \infty$ reflects right censoring.
   - $L_i = R_i$ reflects observation without censoring (notwithstanding $(L_i, R_i]$ is empty if $L_i = R_i$).

3. Because times are fixed, the profiling argument putting all weight on times doesn't apply.

4. Data format similar to life-table approach.
Notation for interval censoring:

1. Patients are screened at fixed intervals $t_0, t_1, t_2, \ldots, t_J$ for $t_{J+1} = \infty$.

2. Know that event happened within $(L_i, R_i]$ for $L_i, R_i \in \{t_1, \ldots, t_J\}$.

3. Because times are fixed, the profiling argument putting all weight on times doesn't apply.

4. Data format similar to life-table approach.
Notation for interval censoring:

1. Patients are screened at fixed intervals $t_0, t_1, t_2, \ldots, t_J$ for $t_{J+1} = \infty$.

2. Know that event happened within $(L_i, R_i]$ for $L_i, R_i \in \{t_1, \ldots, t_J\}$.

   - $R_i = \infty$ reflects right censoring.
Notation for interval censoring:

1. Patients are screened at fixed intervals $t_0, t_1, t_2, \ldots, t_J$ for $t_{J+1} = \infty$.
2. Know that event happened within $(L_i, R_i]$ for $L_i, R_i \in \{t_1, \ldots, t_J\}$.
   1. $R_i = \infty$ reflects right censoring.
   2. $L_i = R_i$ reflects observation without censoring.
Notation for interval censoring:

1. Patients are screened at fixed intervals $t_0, t_1, t_2, \ldots, t_J$ for $t_{J+1} = \infty$.

2. Know that event happened within $(L_i, R_i]$ for $L_i, R_i \in \{t_1, \ldots, t_J\}$.
   - $R_i = \infty$ reflects right censoring.
   - $L_i = R_i$ reflects observation without censoring
     - (not withstanding $(L_i, R_i]$ is empty if $L_i = R_i$).
Notation for interval censoring:

1. Patients are screened at fixed intervals \( t_0, t_1, t_2, \ldots, t_J \) for \( t_{J+1} = \infty \).
2. Know that event happened within \((L_i, R_i]\) for \( L_i, R_i \in \{t_1, \ldots, t_J\} \).
   - \( R_i = \infty \) reflects right censoring.
   - \( L_i = R_i \) reflects observation without censoring
     - (not withstanding \((L_i, R_i]\) is empty if \( L_i = R_i \)).
3. Because times are fixed, the profiling argument putting all weight on times doesn’t apply.
1. Patients are screened at fixed intervals $t_0, t_1, t_2, \ldots, t_J$ for $t_{J+1} = \infty$.

2. Know that event happened within $(L_i, R_i]$ for $L_i, R_i \in \{t_1, \ldots, t_J\}$.
   - $R_i = \infty$ reflects right censoring.
   - $L_i = R_i$ reflects observation without censoring (notwithstanding $(L_i, R_i]$ is empty if $L_i = R_i$).

3. Because times are fixed, the profiling argument putting all weight on times doesn’t apply.

4. Data format similar to life-table approach.
Likelihood for interval censoring

1. Likelihood \( L = \prod_{i=1}^{n} [S_i(L_i) - S_i(R_i)]. \)
Likelihood for interval censoring

1. Likelihood $L = \prod_{i=1}^{n} [S_i(L_i) - S_i(R_i)]$.
   
   1. $n$ is the number of observations.
Likelihood for interval censoring

**Likelihood** $L = \prod_{i=1}^{n} [S_i(L_i) - S_i(R_i)]$.

1. $n$ is the number of observations.
2. Assume observations are independent, and so contributions multiply.

Ex., if potential visits are Jan, Apr, Jul, Oct, but visits are 6 mo apart, then we need to split an event that happens between Apr and Oct to Apr-Jul or Aug-Oct.
Likelihood for interval censoring

1. Likelihood \( L = \prod_{i=1}^{n} [S_i(L_i) - S_i(R_i)] \).
   - \( n \) is the number of observations.
   - Assume observations are independent, and so contributions multiply.
   - \( S_i(t) = S_0(t)^{\exp(z_i \beta)} \)
Likelihood for interval censoring

1. Likelihood $L = \prod_{i=1}^{n} [S_i(L_i) - S_i(R_i)]$.
   - $n$ is the number of observations.
   - Assume observations are independent, and so contributions multiply.
   - $S_i(t) = S_0(t)^{\exp(z_i\beta)}$
   - Estimate $\beta$ and $S_0(t_j)$ for all $j$. 

Ex., if potential visits are Jan, Apr, Jul, Oct, but visits are 6 mo apart, then we need to split an event that happens between Apr and Oct to Apr-Jul or Aug-Oct.
Likelihood for interval censoring

1. Likelihood \( L = \prod_{i=1}^{n} [S_i(L_i) - S_i(R_i)] \).
   1. \( n \) is the number of observations.
   2. Assume observations are independent, and so contributions multiply.
   3. \( S_i(t) = S_0(t)^{\exp(z_i \beta)} \)
   4. Estimate \( \beta \) and \( S_0(t_j) \) for all \( j \).

2. If \( L_i \) and \( R_i \) are all consecutive times from \( t_0, t_1, t_2, \ldots, t_J \) for \( t_{J+1} = \infty \).
Likelihood for interval censoring

1. Likelihood \( L = \prod_{i=1}^{n} [S_i(L_i) - S_i(R_i)] \).
   - \( n \) is the number of observations.
   - Assume observations are independent, and so contributions multiply.
   - \( S_i(t) = S_0(t)^{\exp(z_i\beta)} \)
   - Estimate \( \beta \) and \( S_0(t_j) \) for all \( j \).

2. If \( L_i \) and \( R_i \) are all consecutive times from \( t_0, t_1, t_2, \ldots, t_J \) for \( t_{J+1} = \infty \).
   - \( L = \prod_{j=1}^{J+1} [S_i(t_j) - S_i(t_{j-1})]_j^N \)
Likelihood for interval censoring

1. Likelihood \( L = \prod_{i=1}^{n} [S_i(L_i) - S_i(R_i)] \).
   1. \( n \) is the number of observations.
   2. Assume observations are independent, and so contributions multiply.
   3. \( S_i(t) = S_0(t)^{ \exp(z_i \beta)} \)
   4. Estimate \( \beta \) and \( S_0(t_j) \) for all \( j \).

2. If \( L_i \) and \( R_i \) are all consecutive times from \( t_0, t_1, t_2, \ldots, t_J \) for \( t_{J+1} = \infty \).
   1. \( L = \prod_{j=1}^{J+1} [S_i(t_j) - S_i(t_{j-1})]^{N_j} \)
   2. \( N_j \) are the number of observations in interval \((t_{j-1}, t_j]\).
Likelihood for interval censoring

1. Likelihood  \( L = \prod_{i=1}^{n} [S_i(L_i) - S_i(R_i)] \).
   
   1. \( n \) is the number of observations.
   
   2. Assume observations are independent, and so contributions multiply.
   
   3. \( S_i(t) = S_0(t)^{\exp(z_i\beta)} \)
   
   4. Estimate \( \beta \) and \( S_0(t_j) \) for all \( j \).

2. If \( L_i \) and \( R_i \) are all consecutive times from \( t_0, t_1, t_2, \ldots, t_J \) for \( t_{J+1} = \infty \).
   
   1. \( L = \prod_{j=1}^{J+1} [S_i(t_j) - S_i(t_{j-1})] \)
   
   2. \( N_j \) are the number of observations in interval \( (t_{j-1}, t_j] \).

3. More complicated if \( L_i \) and \( R_i \) are not consecutive items from \( t_0, t_1, t_2, \ldots, t_J \) for \( t_{J+1} = \infty \).
Likelihood for interval censoring

1. Likelihood \( L = \prod_{i=1}^{n} [S_i(L_i) - S_i(R_i)]. \)
   - \( n \) is the number of observations.
   - Assume observations are independent, and so contributions multiply.
   - \( S_i(t) = S_0(t)^{\exp(z_i\beta)} \)
   - Estimate \( \beta \) and \( S_0(t_j) \) for all \( j \).

2. If \( L_i \) and \( R_i \) are all consecutive times from \( t_0, t_1, t_2, \ldots, t_J \) for \( t_{J+1} = \infty \).
   - \( L = \prod_{j=1}^{J+1} [S_i(t_j) - S_i(t_{j-1})]^N \)
   - \( N_j \) are the number of observations in interval \((t_{j-1}, t_j]\).

3. More complicated if \( L_i \) and \( R_i \) are not consecutive items from \( t_0, t_1, t_2, \ldots, t_J \) for \( t_{J+1} = \infty \).
   - Ex., if potential visits are Jan, Apr, Jul, Oct, but visits are 6 mo apart, then we need to split an event that happens between Apr and Oct to Apr-Jul or Aug-Oct.
Likelihood for interval censoring

1. Likelihood \( L = \prod_{i=1}^{n} [S_i(L_i) - S_i(R_i)] \).
   1. \( n \) is the number of observations.
   2. Assume observations are independent, and so contributions multiply.
   3. \( S_i(t) = S_0(t)^{\exp(z_i\beta)} \)
   4. Estimate \( \beta \) and \( S_0(t_j) \) for all \( j \).

2. If \( L_i \) and \( R_i \) are all consecutive times from \( t_0, t_1, t_2, \ldots, t_J \) for \( t_{J+1} = \infty \).
   1. \( L = \prod_{j=1}^{J+1} [S_i(t_j) - S_i(t_{j-1})]_j^N \)
   2. \( N_j \) are the number of observations in interval \( (t_{j-1}, t_j] \).

3. More complicated if \( L_i \) and \( R_i \) are not consecutive times from \( t_0, t_1, t_2, \ldots, t_J \) for \( t_{J+1} = \infty \).
   1. Ex., if potential visits are Jan, Apr, Jul, Oct, but visits are 6 mo apart, then we need to split an event that happens between Apr and Oct to Apr-Jul or Aug-Oct.
   2. Iterative procedure for splitting events over sub-intervals.
Likelihood for interval censoring

1. Likelihood \( L = \prod_{i=1}^{n} [S_i(L_i) - S_i(R_i)] \).
   - \( n \) is the number of observations.
   - Assume observations are independent, and so contributions multiply.
   - \( S_i(t) = S_0(t)^{\exp(z_i \beta)} \)
   - Estimate \( \beta \) and \( S_0(t) \) for all \( j \).

2. If \( L_i \) and \( R_i \) are all consecutive times from \( t_0, t_1, t_2, \ldots, t_J \) for \( t_{J+1} = \infty \).
   - \( L = \prod_{j=1}^{J+1} [S_i(t_j) - S_i(t_{j-1})]_j^N \)
   - \( N_j \) are the number of observations in interval \((t_{j-1}, t_j]\).

3. More complicated if \( L_i \) and \( R_i \) are not consecutive times from \( t_0, t_1, t_2, \ldots, t_J \) for \( t_{J+1} = \infty \).
   - Ex., if potential visits are Jan, Apr, Jul, Oct, but visits are 6 mo apart, then we need to split an event that happens between Apr and Oct to Apr-Jul or Aug-Oct.
   - Iterative procedure for splitting events over sub-intervals.
   - Issue does not arise in the fully-parametric case.  R Code  SAS Code
Case with all subjects having the same assessment times

Let \( \pi_{ij} = P [T_i \leq t_j | T_i > t_{j-1}] \)
Case with all subjects having the same assessment times

1. Let $\pi_{ij} = P[T_i \leq t_j | T_i > t_{j-1}]$

2. $P[T_i > t_j] = \prod_{l=1}^{j}(1 - \pi_{il})$
Case with all subjects having the same assessment times

1. Let $\pi_{ij} = P[T_i \leq t_j | T_i > t_{j-1}]
2. \quad P[T_i > t_j] = \prod_{l=1}^{j}(1 - \pi_{il})
3. Let $J_i$ be index such that $T_i \in (t_{J_i}, t_{J_i+1}].$
Case with all subjects having the same assessment times

1. Let $\pi_{ij} = P[T_i \leq t_j | T_i > t_{j-1}]$
2. $P[T_i > t_j] = \prod_{l=1}^{j}(1 - \pi_{il})$
3. Let $J_i$ be index such that $T_i \in (t_{J_i}, t_{J_i+1}]$.
4. $Y_{ij}$ indicate which interval subject $i$ has event in.
Case with all subjects having the same assessment times

1. Let \( \pi_{ij} = P[T_i \leq t_j | T_i > t_{j-1}] \)
2. \( P[T_i > t_j] = \prod_{l=1}^{j} (1 - \pi_{il}) \)
3. Let \( J_i \) be index such that \( T_i \in (t_{J_i}, t_{J_i+1}] \).
4. \( Y_{ij} \) indicate which interval subject \( i \) has event in.
   - 1 if subject \( i \) had the event in interval \( (t_j, t_{j+1}] \),
Case with all subjects having the same assessment times

1. Let $\pi_{ij} = P[T_i \leq t_j | T_i > t_{j-1}]$
2. $P[T_i > t_j] = \prod_{l=1}^{j} (1 - \pi_{il})$
3. Let $J_i$ be index such that $T_i \in (t_{J_i}, t_{J_i+1}]$.
4. $Y_{ij}$ indicate which interval subject $i$ has event in.
   - 1 if subject $i$ had the event in interval $(t_j, t_{j+1}]$,
   - $Y_{i,J+1} = 1$ if item not observed to fail,
Case with all subjects having the same assessment times

1. Let $\pi_{ij} = P[T_i \leq t_j|T_i > t_{j-1}]$
2. $P[T_i > t_j] = \prod_{l=1}^{j}(1 - \pi_{il})$
3. Let $J_i$ be index such that $T_i \in (t_{J_i}, t_{J_i+1}]$.
4. $Y_{ij}$ indicate which interval subject $i$ has event in.
   1. 1 if subject $i$ had the event in interval $(t_j, t_{j+1}]$,
   2. $Y_{i,J+1} = 1$ if item not observed to fail,
   3. 0 otherwise.
Case with all subjects having the same assessment times

1. Let \( \pi_{ij} = P[T_i \leq t_j | T_i > t_{j-1}] \)

2. \( P[T_i > t_j] = \prod_{l=1}^{j}(1 - \pi_{il}) \)

3. Let \( J_i \) be index such that \( T_i \in (t_{J_i}, t_{J_i+1}] \).

4. \( Y_{ij} \) indicate which interval subject \( i \) has event in.
   1. 1 if subject \( i \) had the event in interval \((t_j, t_{j+1}],\)
   2. \( Y_{iJ_{i+1}} = 1 \) if item not observed to fail,
   3. 0 otherwise.

5. Likelihood is \( \prod_{i=1}^{n} \pi_{iJ_i} \prod_{l=1}^{J_i-1}(1 - \pi_{il}) = \prod_{i=1}^{n} \pi_{iJ_i} Y_{iJ_i} \prod_{l=1}^{J_i-1}(1 - \pi_{il})^{1-Y_{il}} \),
Case with all subjects having the same assessment times

1. Let $\pi_{ij} = P[T_i \leq t_j | T_i > t_{j-1}]$
2. $P[T_i > t_j] = \prod_{l=1}^{j}(1 - \pi_{il})$
3. Let $J_i$ be index such that $T_i \in (t_{J_i}, t_{J_i+1}]$.
4. $Y_{ij}$ indicate which interval subject $i$ has event in.
   - 1 if subject $i$ had the event in interval $(t_j, t_{j+1}]$,
   - $Y_{i,J+1} = 1$ if item not observed to fail,
   - 0 otherwise.
5. Likelihood is $\prod_{i=1}^{n} \pi_{iJ_i} \prod_{l=1}^{J_i-1}(1 - \pi_{il}) = \prod_{i=1}^{n} Y_{iJ_i}^{Y_{iJ_i}} \prod_{l=1}^{J_i-1}(1 - \pi_{il})^{1-Y_{il}}$,

   Likelihood for Bernoulli trials $Y_{ij}$ with success probabilities $\pi_{ij}$.
Case with all subjects having the same assessment times

1. Let \( \pi_{ij} = P \left( T_i \leq t_j \mid T_i > t_{j-1} \right) \)
2. \( P \left( T_i > t_j \right) = \prod_{l=1}^{j} (1 - \pi_{il}) \)
3. Let \( J_i \) be index such that \( T_i \in (t_{J_i}, t_{J_i+1}] \).
4. \( Y_{ij} \) indicate which interval subject \( i \) has event in.
    1. 1 if subject \( i \) had the event in interval \((t_j, t_{j+1}]\),
    2. \( Y_{i,J+1} = 1 \) if item not observed to fail,
    3. 0 otherwise.
5. Likelihood is \( \prod_{i=1}^{n} \pi_{iJ_i} \prod_{l=1}^{J_i-1} (1 - \pi_{il}) = \prod_{i=1}^{n} Y_{iJ_i} \prod_{l=1}^{J_i-1} (1 - \pi_{il})^{1-Y_{il}} \),

- Likelihood for Bernoulli trials \( Y_{ij} \) with success probabilities \( \pi_{ij} \)
- Likelihood contributions multiply through conditioning rather than through independence.
Proportional hazards

\[ P(T_i > t_j) = P(T_m > t_j)^{\exp((z_i - z_m)\beta)} \quad \forall i, j, m \]
Proportional hazards

1. \( P [T_i > t_j] = P [T_m > t_j] \exp((z_i - z_m)\beta) \ \forall i, j, m \)

2. \( \prod_{l=1}^{j} (1 - \pi_{il}) = \prod_{l=1}^{j} (1 - \pi_{ml}) \exp((z_i - z_m)\beta) \ \forall i, j, m \)
Proportional hazards

1. \( P[T_i > t_j] = P[T_m > t_j] \exp((z_i - z_m)\beta) \quad \forall i, j, m \)

2. \( \prod_{l=1}^{j} (1 - \pi_{il}) = \prod_{l=1}^{j} (1 - \pi_{ml}) \exp((z_i - z_m)\beta) \quad \forall i, j, m \)

3. \( (1 - \pi_{ij}) = (1 - \pi_{mj}) \exp((z_i - z_m)\beta) \quad \forall i, j, m \)
Proportional hazards

1. \( P[T_i > t_j] = P[T_m > t_j]^{\exp((z_i - z_m)\beta)} \quad \forall i, j, m \)

2. \( \prod_{l=1}^{j}(1 - \pi_{il}) = \prod_{l=1}^{j}(1 - \pi_{ml})^{\exp((z_i - z_m)\beta)} \quad \forall i, j, m \)

3. \( (1 - \pi_{ij}) = (1 - \pi_{mj})^{\exp((z_i - z_m)\beta)} \quad \forall i, j, m \)

4. \( \log(1 - \pi_{ij}) = \exp(\alpha_j + z_i\beta) \quad \forall i, j \)
Proportional hazards

1. \( P [ T_i > t_j ] = P [ T_m > t_j ] \exp((z_i - z_m) \beta) \quad \forall i, j, m \)
2. \( \prod_{l=1}^{j} (1 - \pi_{il}) = \prod_{l=1}^{j} (1 - \pi_{ml}) \exp((z_i - z_m) \beta) \quad \forall i, j, m \)
3. \( (1 - \pi_{ij}) = (1 - \pi_{mj}) \exp((z_i - z_m) \beta) \quad \forall i, j, m \)
4. \( \log(1 - \pi_{ij}) = \exp(\alpha_j + z_i \beta) \quad \forall i, j \)
5. \( \log(\log(1 - \pi_{ij})) = \alpha_j + z_i \beta \quad \forall i, j \)
Proportional hazards

1. \[ P[T_i > t_j] = P[T_m > t_j]^{\exp((z_i - z_m)\beta)} \forall i, j, m \]
2. \[ \prod_{l=1}^{j}(1 - \pi_{il}) = \prod_{l=1}^{j}(1 - \pi_{ml})^{\exp((z_i - z_m)\beta)} \forall i, j, m \]
3. \[ (1 - \pi_{ij}) = (1 - \pi_{mj})^{\exp((z_i - z_m)\beta)} \forall i, j, m \]
4. \[ \log(1 - \pi_{ij}) = \exp(\alpha_j + z_i\beta) \forall i, j \]
5. \[ \log(\log(1 - \pi_{ij})) = \alpha_j + z_i\beta \forall i, j \]
6. Gives *complimentary log log link* for regression model for \( \pi_{ij} \)

R Code

SAS Code

Regression Models: Estimation with interval censoring: Lecture 10
Objectives Lecture 11

1. Dependent survival times
Objectives Lecture 11

1. Dependent survival times
2. Multiple events of same type per individual
Objectives Lecture 11

1. Dependent survival times
2. Multiple events of same type per individual
3. Competing Risk
Objectives Lecture 11

1. Dependent survival times
2. Multiple events of same type per individual
3. Competing Risk
4. Threshold Models
Objectives Lecture 11

1. Dependent survival times
2. Multiple events of same type per individual
3. Competing Risk
4. Threshold Models
Section: Regression Models

Subsection: Alternative Use of Parametric Survival Regression
Suppose a linear regression model.
Suppose a linear regression model.

\[ V_i = \alpha + z_i \beta + \sigma U_i \]
Setup

1 Suppose a linear regression model.

1 $V_i$ (not necessarily log times) satisfy $V_i = \alpha + z_i\beta + \sigma U_i$

2 $V_i$ independent
Setup

1. Suppose a linear regression model.
   1. $V_i$ (not necessarily log times) satisfy $V_i = \alpha + z_i \beta + \sigma U_i$
   2. $V_i$ independent
   3. $V_i$ not observed if it falls below a threshold
Setup

1. Suppose a linear regression model.
   1. $V_i$ (not necessarily log times) satisfy $V_i = \alpha + z_i \beta + \sigma U_i$
   2. $V_i$ independent
   3. $V_i$ not observed if it falls below a threshold

2. Examples in which the response is (a transformation of)
Setup

1. Suppose a linear regression model.
   1. $V_i$ (not necessarily log times) satisfy $V_i = \alpha + z_i \beta + \sigma U_i$
   2. $V_i$ independent
   3. $V_i$ not observed if it falls below a threshold

2. Examples in which the response is (a transformation of)
   1. chemical concentration, with threshold is lowest detectable value
Suppose a linear regression model.

1. $V_i$ (not necessarily log times) satisfy $V_i = \alpha + z_i \beta + \sigma U_i$
2. $V_i$ independent
3. $V_i$ not observed if it falls below a threshold

Examples in which the response is (a transformation of)

1. chemical concentration, with threshold is lowest detectable value
2. sale price for a security, with 0 the lower bound
Then $\exp(V_i)$ follows accelerated life model

Typically need to make censoring on right
Then $\exp(V_i)$ follows accelerated life model

1. Typically need to make censoring on right
   - By flipping scale if necessary.
Related model: If $U_i$ normal, and threshold is 0, model (on log($V_i$) scale) is called *Tobit model*.  

1 If $U_i$ normal, and threshold is 0, model (on log($V_i$) scale) is called *Tobit model*.  
R Code  SAS Code
Section: More complicated Situations

Subsection: Non-Independent Survival Times
Examples of Non-Independent Survival Times

1. Times till event in correlated individuals

More complicated Situations: Non-Independent Survival Times
Examples of Non-Independent Survival Times

1. Times till event in correlated individuals
   1. litters of rats
Examples of Non-Independent Survival Times

Times till event in correlated individuals

1. litters of rats
2. parts from same batch

More complicated Situations: Non-Independent Survival Times
Examples of Non-Independent Survival Times

1. Times till event in correlated individuals
   1. litters of rats
   2. parts from same batch

2. Times till consecutive events in same individual
Examples of Non-Independent Survival Times

1. Times till event in correlated individuals
   - litters of rats
   - parts from same batch

2. Times till consecutive events in same individual

3. Times till different kinds of events in same individual
Examples of Non-Independent Survival Times

1. Times till event in correlated individuals
   1. litters of rats
   2. parts from same batch

2. Times till consecutive events in same individual

3. Times till different kinds of events in same individual
   1. different types of GVHD for leukemia patients
Analogy with regression models:

1. $Y_{ij}$ is measurement $i$ from cluster $j$

3. Key component is $W_j$:
   1. effect of cluster $j$
   2. Not directly observable
   3. Standard analyses average this out.
   4. Must specify distn (Ex., standard normal)
   5. Called a random effects model.

4. Strategies for Estimation in Random Effects Models
   1. $L(\alpha, \beta, \sigma, \tau; Y) = E_{W}[L(\alpha, \beta, \sigma, \tau; Y, W)]$
   2. As integral, $= \int L(\alpha, \beta, \sigma, \tau; Y, w)f_W(w)dw$
   4. More complicated cases (Ex., logistic regression, other generalized linear models)
   5. Numerical integration
   6. Monte Carlo integration
   7. Use missing data techniques to impute values of $W$.

More complicated Situations: Non-Independent Survival Times

Lecture 11
Analogy with regression models:

1. $Y_{ij}$ is measurement $i$ from cluster $j$
   1. Clusters are people, animal litters, manufacturing lots, etc
Analogy with regression models:

1. \( Y_{ij} \) is measurement \( i \) from cluster \( j \)
   - Clusters are people, animal litters, manufacturing lots, etc

2. \[ Y_{ij} = \alpha + z_{ij} \beta + \sigma W_j + \tau \epsilon_{ij} \]
Analogy with regression models:

1. $Y_{ij}$ is measurement $i$ from cluster $j$
   - Clusters are people, animal litters, manufacturing lots, etc
2. $Y_{ij} = \alpha + z_{ij}\beta + \sigma W_j + \tau \epsilon_{ij}$
3. Key component is $W_j$:

More complicated Situations: Non-Independent Survival Times
Analogy with regression models:

1. $Y_{ij}$ is measurement $i$ from cluster $j$
   - Clusters are people, animal litters, manufacturing lots, etc

2. $Y_{ij} = \alpha + z_{ij} \beta + \sigma W_j + \tau \epsilon_{ij}$

3. Key component is $W_j$:
   - effect of cluster $j$

More complicated Situations: Non-Independent Survival Times
Analogy with regression models:

1. $Y_{ij}$ is measurement $i$ from cluster $j$
   - Clusters are people, animal litters, manufacturing lots, etc

2. $Y_{ij} = \alpha + z_{ij}\beta + \sigma W_j + \tau \epsilon_{ij}$

3. Key component is $W_j$:
   - Effect of cluster $j$
   - Not directly observable

More complicated Situations: Non-Independent Survival Times
Analogy with regression models:

1. $Y_{ij}$ is measurement $i$ from cluster $j$
   - Clusters are people, animal litters, manufacturing lots, etc
2. $Y_{ij} = \alpha + z_{ij}\beta + \sigma W_j + \tau \epsilon_{ij}$
3. Key component is $W_j$:
   - Effect of cluster $j$
   - Not directly observable
   - Standard analyses average this out.
Analogy with regression models:

1. $Y_{ij}$ is measurement $i$ from cluster $j$
   - Clusters are people, animal litters, manufacturing lots, etc

2. $Y_{ij} = \alpha + z_{ij}\beta + \sigma W_j + \tau \epsilon_{ij}$

3. Key component is $W_j$:
   - effect of cluster $j$
   - Not directly observable
   - Standard analyses average this out.
   - Must specify distn

More complicated Situations: Non-Independent Survival Times
Analogy with regression models:

1. $Y_{ij}$ is measurement $i$ from cluster $j$
   - Clusters are people, animal litters, manufacturing lots, etc

2. $Y_{ij} = \alpha + z_{ij}\beta + \sigma W_j + \tau \epsilon_{ij}$

3. Key component is $W_j$:
   - effect of cluster $j$
   - Not directly observable
   - Standard analyses average this out.
   - Must specify distn
     - Ex, standard normal

More complicated Situations: Non-Independent Survival Times
Analogy with regression models:

1. $Y_{ij}$ is measurement $i$ from cluster $j$
   - Clusters are people, animal litters, manufacturing lots, etc
2. $Y_{ij} = \alpha + z_{ij}\beta + \sigma W_j + \tau \epsilon_{ij}$
3. Key component is $W_j$:
   - Effect of cluster $j$
   - Not directly observable
   - Standard analyses average this out.
   - Must specify distn
   - Ex, standard normal
4. Called a random effects model.

More complicated Situations: Non-Independent Survival Times
Analogy with regression models:

1. $Y_{ij}$ is measurement $i$ from cluster $j$
   - Clusters are people, animal litters, manufacturing lots, etc

2. $Y_{ij} = \alpha + z_{ij}\beta + \sigma W_j + \tau \epsilon_{ij}$

3. Key component is $W_j$:
   - Effect of cluster $j$
   - Not directly observable
   - Standard analyses average this out.
   - Must specify distn
   - Ex., standard normal

4. Called a *random effects model*.

5. Strategies for Estimation in Random Effects Models
Analogy with regression models:

1. \( Y_{ij} \) is measurement \( i \) from cluster \( j \)
   - Clusters are people, animal litters, manufacturing lots, etc

2. \( Y_{ij} = \alpha + z_{ij}\beta + \sigma W_j + \tau \epsilon_{ij} \)

3. Key component is \( W_j \):
   - Effect of cluster \( j \)
   - Not directly observable
   - Standard analyses average this out.
   - Must specify distn
     - Ex, standard normal

4. Called a *random effects model*.

5. Strategies for Estimation in Random Effects Models
   - \( L(\alpha, \beta, \sigma, \tau; \mathbf{Y}) = \mathbb{E}_W[L(\alpha, \beta, \sigma, \tau; \mathbf{Y}, \mathbf{W})] \)
Analogy with regression models:

1. $Y_{ij}$ is measurement $i$ from cluster $j$
   - Clusters are people, animal litters, manufacturing lots, etc

2. $Y_{ij} = \alpha + z_{ij}\beta + \sigma W_j + \tau \epsilon_{ij}$

3. Key component is $W_j$:
   - Effect of cluster $j$
   - Not directly observable
   - Standard analyses average this out.
   - Must specify distn
     - Ex, standard normal

4. Called a random effects model.

5. Strategies for Estimation in Random Effects Models
   - $L(\alpha, \beta, \sigma, \tau; Y) = \mathbb{E}_W [L(\alpha, \beta, \sigma, \tau; Y, W)]$
     - As integral, $= \int L(\alpha, \beta, \sigma, \tau; Y, w)f_W(w) \, dw$
Analogy with regression models:

1. \( Y_{ij} \) is measurement \( i \) from cluster \( j \)
   - Clusters are people, animal litters, manufacturing lots, etc

2. \( Y_{ij} = \alpha + z_{ij}\beta + \sigma W_j + \tau \epsilon_{ij} \)

3. Key component is \( W_j \):
   - Effect of cluster \( j \)
   - Not directly observable
   - Standard analyses average this out.
   - Must specify distn
     - Ex., standard normal

4. Called a \textit{random effects model}.

5. Strategies for Estimation in Random Effects Models
   - \( L(\alpha, \beta, \sigma, \tau; Y) = E_W \left[ L(\alpha, \beta, \sigma, \tau; Y, W) \right] \)
     - As integral, \( = \int L(\alpha, \beta, \sigma, \tau; Y, w) f_W(w) \, dw \)
   - Standard normal case: can often do integral in closed form.

More complicated Situations: Non-Independent Survival Times
Analogy with regression models:

1. $Y_{ij}$ is measurement $i$ from cluster $j$
   - Clusters are people, animal litters, manufacturing lots, etc
2. $Y_{ij} = \alpha + z_{ij}\beta + \sigma W_j + \tau \varepsilon_{ij}$
3. Key component is $W_j$:
   - effect of cluster $j$
   - Not directly observable
   - Standard analyses average this out.
   - Must specify distn
     - Ex, standard normal
4. Called a *random effects model*.
5. Strategies for Estimation in Random Effects Models
   - $L(\alpha, \beta, \sigma, \tau; Y) = E_W [L(\alpha, \beta, \sigma, \tau; Y, W)]$
     - As integral, $= \int L(\alpha, \beta, \sigma, \tau; Y, w)f_W(w) \, dw$
   - Standard normal case: can often do integral in closed form.
   - More complicated cases
Analogy with regression models:

1. $Y_{ij}$ is measurement $i$ from cluster $j$
   - Clusters are people, animal litters, manufacturing lots, etc
2. $Y_{ij} = \alpha + z_{ij}\beta + \sigma W_j + \tau \epsilon_{ij}$
3. Key component is $W_j$:
   - effect of cluster $j$
   - Not directly observable
   - Standard analyses average this out.
   - Must specify distn
   - Ex, standard normal
4. Called a **random effects model**.

Strategies for Estimation in Random Effects Models

1. $L(\alpha, \beta, \sigma, \tau; Y) = \mathbb{E}_W [L(\alpha, \beta, \sigma, \tau; Y, W)]$
   - As integral, $= \int L(\alpha, \beta, \sigma, \tau; Y, w) f_W(w) \, dw$
2. Standard normal case: can often do integral in closed form.
3. More complicated cases
   - Ex., logistic regression, other generalized linear models
Analogy with regression models:

1. $Y_{ij}$ is measurement $i$ from cluster $j$
   - Clusters are people, animal litters, manufacturing lots, etc
2. $Y_{ij} = \alpha + z_{ij}\beta + \sigma W_j + \tau \epsilon_{ij}$
3. Key component is $W_j$:
   - effect of cluster $j$
   - Not directly observable
   - Standard analyses average this out.
   - Must specify distn
     - Ex, standard normal
4. Called a random effects model.

Strategies for Estimation in Random Effects Models

1. $L(\alpha, \beta, \sigma, \tau; Y) = E_W [L(\alpha, \beta, \sigma, \tau; Y, W)]$
   - As integral, $= \int L(\alpha, \beta, \sigma, \tau; Y, w)f_W(w) \, dw$
2. Standard normal case: can often do integral in closed form.
3. More complicated cases
   - Ex., logistic regression, other generalized linear models
   - Numerical integration
Analogy with regression models:

1. $Y_{ij}$ is measurement $i$ from cluster $j$
   - Clusters are people, animal litters, manufacturing lots, etc
2. $Y_{ij} = \alpha + z_{ij}\beta + \sigma W_j + \tau \epsilon_{ij}$
3. Key component is $W_j$:
   - Effect of cluster $j$
   - Not directly observable
   - Standard analyses average this out.
   - Must specify distn
     - Ex. standard normal
4. Called a *random effects model*.

Strategies for Estimation in Random Effects Models

1. $L(\alpha, \beta, \sigma, \tau; Y) = E_W [L(\alpha, \beta, \sigma, \tau; Y, W)]$
   - As integral, $= \int L(\alpha, \beta, \sigma, \tau; Y, w)f_W(w) \, dw$
2. Standard normal case: can often do integral in closed form.
3. More complicated cases
   - Ex., logistic regression, other generalized linear models
   - Numerical integration
   - Monte Carlo integration
Analogy with regression models:

1. $Y_{ij}$ is measurement $i$ from cluster $j$
   - Clusters are people, animal litters, manufacturing lots, etc

2. $Y_{ij} = \alpha + z_{ij}\beta + \sigma W_j + \tau \epsilon_{ij}$

3. Key component is $W_j$:
   - effect of cluster $j$
   - Not directly observable
   - Standard analyses average this out.
   - Must specify distn
     - Ex., standard normal

4. Called a **random effects model**.

5. Strategies for Estimation in Random Effects Models
   - $L(\alpha, \beta, \sigma, \tau; Y) = E_W[L(\alpha, \beta, \sigma, \tau; Y, W)]$
     - As integral, $= \int L(\alpha, \beta, \sigma, \tau; Y, w)f_W(w) \, dw$
   - Standard normal case: can often do integral in closed form.
   - More complicated cases
     - Ex., logistic regression, other generalized linear models
     - Numerical integration
     - Monte Carlo integration
     - Use missing data techniques to impute values of $W$. 

Examples illustrating effect of various clustering models

1. Setup: Regression model with

- 10 clusters,
- 4 obs/cluster,
- 1 covariate of interest
- 1 nuisance covariate

Alternatives:
- Ignoring clusters: 37 df.
- Stratifying with different intercept and nuisance parameter value for each cluster: 40-10-10-1 = 19 df.
- Fixed effect model: 40-10-1-1 = 28 df.
- Random effects model: 36 df.

For g.l.m.s, distn of $Y_{ij}$ given by

$$
\eta_{ij} = \alpha + z_{ij} \beta + \tau W_j + \sigma \epsilon_{ij}
$$

Note $\text{Var}[Y_{ij}] = \text{E}[\text{Var}[Y_{ij} | W_j]] + \text{Var}[\text{E}[Y_{ij} | W_j]]$

Second term makes variance larger than would be expected for fixed effect model

Phenomenon is called over-dispersion.
Examples illustrating effect of various clustering models

Setup: Regression model with

- 10 clusters,

Alternatives:

1. Ignoring clusters: 37 df.
2. Stratifying with different intercept and nuisance parameter value for each cluster: 40-10-10-1= 19 df.
3. Fixed effect model: 40-10-1-1=28 df.
4. Random effects model: 36 df.

For g.l.m.s, distn of $Y_{ij}$ given by

$$
\eta_{ij} = \alpha + z_{ij} \beta + \tau_{Wj} + \sigma \epsilon_{ij}
$$

Note $\text{Var}[Y_{ij}] = E[\text{Var}[Y_{ij}|W_j]] + \text{Var}[E[Y_{ij}|W_j]]$

Second term makes variance larger than would be expected for fixed effect model. Phenomenon is called over-dispersion.

More complicated Situations: Non-Independent Survival Times
Examples illustrating effect of various clustering models

1 Setup: Regression model with
   1 10 clusters,
   2 4 obs/cluster,
Examples illustrating effect of various clustering models

1. Setup: Regression model with
   1. 10 clusters,
   2. 4 obs/cluster,
   3. 1 covariate of interest

2. Alternatives:
   1. Ignoring clusters: 37 df.
   2. Stratifying with different intercept and nuisance parameter value for each cluster: 40-10-10-1 = 19 df.
   3. Fixed effect model: 40-10-1-1 = 28 df.
   4. Random effects model: 36 df.

3. For g.l.m.s, distn of $Y_{ij}$ given by
   
   \[ \eta_{ij} = \alpha + z_{ij} \beta + \tau W_j + \sigma \epsilon_{ij} \]

4. Note \( \text{Var}[Y_{ij}] = E[\text{Var}[Y_{ij}|W_j]] + \text{Var}[E[Y_{ij}|W_j]] \)

   Second term makes variance larger than would be expected for fixed effect model

   Phenomenon is called over-dispersion.

More complicated Situations: Non-Independent Survival Times Lecture 11 221 / 260
Examples illustrating effect of various clustering models

1. Setup: Regression model with
   1. 10 clusters,
   2. 4 obs/cluster,
   3. 1 covariate of interest
   4. 1 nuisance covariate

Alternatives:
1. Ignoring clusters: 37 df.
2. Stratifying with different intercept and nuisance parameter value for each cluster: 40-10-10-1= 19 df.
3. Fixed effect model: 40-10-1-1=28 df.
4. Random effects model: 36 df.

For g.l.m.s, distn of $Y_{ij}$ given by

\[ \eta_{ij} = \alpha + z_{ij} \beta + \tau W_j + \sigma \epsilon_{ij} \]

Note \[ \text{Var}[Y_{ij}] = E[\text{Var}[Y_{ij}|W_j]] + \text{Var}[E[Y_{ij}|W_j]] \]

Second term makes variance larger than would be expected for fixed effect model
Phenomenon is called over-dispersion.
Examples illustrating effect of various clustering models

1. Setup: Regression model with
   - 10 clusters,
   - 4 obs/cluster,
   - 1 covariate of interest
   - 1 nuisance covariate

2. Alternatives:
   - Ignoring clusters: 37 df.
   - Stratifying with different intercept and nuisance parameter value for each cluster: 40-10-10-1 = 19 df.
   - Fixed effect model: 40-10-1-1 = 28 df.
   - Random effects model: 36 df.

For g.l.m.s, distn of $Y_{ij}$ given by

\[ \eta_{ij} = \alpha + z_{ij} \beta + \tau W_j + \sigma \epsilon_{ij} \]

Note $\text{Var}[Y_{ij}] = E[\text{Var}[Y_{ij} | W_j]] + \text{Var}[E[Y_{ij} | W_j]]$

Second term makes variance larger than would be expected for fixed effect model.

Phenomenon is called over-dispersion.
Examples illustrating effect of various clustering models

1. Setup: Regression model with
   - 10 clusters,
   - 4 obs/cluster,
   - 1 covariate of interest
   - 1 nuisance covariate

2. Alternatives:
   - Ignoring clusters: 37 df.
Examples illustrating effect of various clustering models

1. Setup: Regression model with
   1. 10 clusters,
   2. 4 obs/cluster,
   3. 1 covariate of interest
   4. 1 nuisance covariate

2. Alternatives:
   1. Ignoring clusters: 37 df.
   2. Stratifying with different intercept and nuisance parameter value for each cluster: 40-10-10-1 = 19 df.

For g.l.m.s, distn of \( Y_{ij} \) given by

\[ \eta_{ij} = \alpha + z_{ij} \beta + \tau W_j + \sigma \epsilon_{ij} \]

Note \( \text{Var}[Y_{ij}] = E[\text{Var}[Y_{ij}|W_j]] + \text{Var}[E[Y_{ij}|W_j]] \)

Second term makes variance larger than would be expected for fixed effect model.

Phenomenon is called over-dispersion.
Examples illustrating effect of various clustering models

1. Setup: Regression model with
   - 10 clusters,
   - 4 obs/cluster,
   - 1 covariate of interest
   - 1 nuisance covariate

2. Alternatives:
   - Ignoring clusters: 37 df.
   - Stratifying with different intercept and nuisance parameter value for each cluster: 40-10-10-1 = 19 df.
   - Fixed effect model: 40-10-1-1 = 28 df.
Examples illustrating effect of various clustering models

1. Setup: Regression model with
   - 10 clusters,
   - 4 obs/cluster,
   - 1 covariate of interest
   - 1 nuisance covariate

2. Alternatives:
   - Ignoring clusters: 37 df.
   - Stratifying with different intercept and nuisance parameter value for each cluster: 40-10-10-1 = 19 df.
   - Fixed effect model: 40-10-1-1 = 28 df.
   - Random effects model: 36 df.
Examples illustrating effect of various clustering models

1. Setup: Regression model with
   1. 10 clusters,
   2. 4 obs/cluster,
   3. 1 covariate of interest
   4. 1 nuisance covariate

2. Alternatives:
   1. Ignoring clusters: 37 df.
   2. Stratifying with different intercept and nuisance parameter value for each cluster: 40-10-10-1 = 19 df.
   3. Fixed effect model: 40-10-1-1 = 28 df.
   4. Random effects model: 36 df.

3. For g.l.m.s, distn of $Y_{ij}$ given by $f_Y(y; \eta_{ij})$
Examples illustrating effect of various clustering models

1. Setup: Regression model with
   1. 10 clusters,
   2. 4 obs/cluster,
   3. 1 covariate of interest
   4. 1 nuisance covariate

2. Alternatives:
   1. Ignoring clusters: 37 df.
   2. Stratifying with different intercept and nuisance parameter value for each cluster: 40-10-10-1 = 19 df.
   3. Fixed effect model: 40-10-1-1 = 28 df.
   4. Random effects model: 36 df.

3. For g.l.m.s, distn of $Y_{ij}$ given by $f_Y(y; \eta_{ij})$
   
   $\eta_{ij} = \alpha + z_{ij}\beta + \tau W_j + \sigma \epsilon_{ij}$

More complicated Situations: Non-Independent Survival Times
Examples illustrating effect of various clustering models

1 Setup: Regression model with
   1 10 clusters,
   2 4 obs/cluster,
   3 1 covariate of interest
   4 1 nuisance covariate

2 Alternatives:
   1 Ignoring clusters: 37 df.
   2 Stratifying with different intercept and nuisance parameter value for each cluster: 40-10-10-1= 19 df.
   3 Fixed effect model: 40-10-1-1=28 df.
   4 Random effects model: 36 df.

3 For g.l.m.s, distn of $Y_{ij}$ given by $f_Y(y; \eta_{ij})$
   1 $\eta_{ij} = \alpha + z_{ij}\beta + \tau W_j + \sigma \epsilon_{ij}$

4 Note $\text{Var}[Y_{ij}] = \mathbb{E}[\text{Var}[Y_{ij}|W_j]] + \text{Var}[\mathbb{E}[Y_{ij}|W_j]]$
Examples illustrating effect of various clustering models

1. Setup: Regression model with
   1. 10 clusters,
   2. 4 obs/cluster,
   3. 1 covariate of interest
   4. 1 nuisance covariate

2. Alternatives:
   1. Ignoring clusters: 37 df.
   2. Stratifying with different intercept and nuisance parameter value for each cluster: 40-10-10-1= 19 df.
   3. Fixed effect model: 40-10-1-1=28 df.
   4. Random effects model: 36 df.

3. For g.l.m.s, distn of \( Y_{ij} \) given by \( f_Y(y; \eta_{ij}) \)
   \[ \eta_{ij} = \alpha + z_{ij}\beta + \tau W_j + \sigma \epsilon_{ij} \]

4. Note \( \text{Var}[Y_{ij}] = \mathbb{E}[\text{Var}[Y_{ij}|W_j]] + \text{Var}[\mathbb{E}[Y_{ij}|W_j]] \)
   Second term makes variance larger than would be expected for fixed effect model
Examples illustrating effect of various clustering models

1. Setup: Regression model with
   - 10 clusters,
   - 4 obs/cluster,
   - 1 covariate of interest
   - 1 nuisance covariate

2. Alternatives:
   - Ignoring clusters: 37 df.
   - Stratifying with different intercept and nuisance parameter value for each cluster: 40-10-10-1 = 19 df.
   - Fixed effect model: 40-10-1-1 = 28 df.
   - Random effects model: 36 df.

3. For g.l.m.s, distn of $Y_{ij}$ given by $f_Y(y; \eta_{ij})$
   - $\eta_{ij} = \alpha + z_{ij}\beta + \tau W_j + \sigma \epsilon_{ij}$

4. Note $\text{Var}[Y_{ij}] = \mathbb{E}[\text{Var}[Y_{ij} \mid W_j]] + \text{Var}[\mathbb{E}[Y_{ij} \mid W_j]]$
   - Second term makes variance larger than would be expected for fixed effect model
   - Phenomenon is called over-dispersion.

More complicated Situations: Non-Independent Survival Times
Life Data Random Effect Model

1. Add subject–specific random effect

Called frailty model because it allows for some clusters to be more frail than others. Can be applied to Proportional Hazards

\[ h_{ij}(t) = h_0(t) \exp(z_{ij}\beta + \tau W_j) \]

If we knew \( W_j \), would treat as Cox model with the partial likelihood \( L(\beta,\tau; Y, W) \).

Can be applied to Accelerated Failure

\[ \log(T_{ij}) = \alpha + z_{ij}\beta + \tau W_j + \sigma\epsilon_{ij} \]

If we knew \( W_j \), would treat as AFT model with the likelihood \( L(\alpha, \beta, \tau, \sigma; Y, W) \).

More complicated Situations: Non-Independent Survival Times
Life Data Random Effect Model

1. Add subject-specific random effect
2. Called *frailty model*.

More complicated Situations: Non-Independent Survival Times
Life Data Random Effect Model

1. Add subject–specific random effect
2. Called *frailty model*. 
   - because it allows for some clusters to be more frail than others
Add subject–specific random effect

Called *frailty model.*

because it allows for some clusters to be more frail than others

Can be applied to Proportional Hazards
Add subject–specific random effect

Called *frailty model*.

because it allows for some clusters to be more frail than others

Can be applied to Proportional Hazards

\[ h_{ij}(t) = h_0(t) \exp(z_{ij}\beta + \tau W_j) \]
Add subject–specific random effect

Called \textit{frailty model}.

because it allows for some clusters to be more frail than others

Can be applied to Proportional Hazards

\begin{align*}
  h_{ij}(t) &= h_0(t) \exp(z_{ij} \beta + \tau W_j) \\
  \text{If we knew } W, \text{ would treat as Cox model with the partial likelihood } & L(\beta, \tau; Y, W)
\end{align*}
Life Data Random Effect Model

1. Add subject–specific random effect
2. Called *frailty model.*

   1. because it allows for some clusters to be more frail than others

3. Can be applied to Proportional Hazards

   1. \( h_{ij}(t) = h_0(t) \exp(z_{ij}\beta + \tau W_j) \)
   2. If we knew \( W \), would treat as Cox model with the partial likelihood \( L(\beta, \tau; Y, W) \)

4. Can be applied to Accelerated Failure

More complicated Situations: Non-Independent Survival Times
Add subject–specific random effect

Called *frailty model*. because it allows for some clusters to be more frail than others

Can be applied to Proportional Hazards

\[ h_{ij}(t) = h_0(t) \exp(z_{ij}\beta + \tau W_j) \]

If we knew \( W \), would treat as Cox model with the partial likelihood \( L(\beta, \tau; Y, W) \)

Can be applied to Accelerated Failure

\[ \log(T_{ij}) = \alpha + z_{ij}\beta + \tau W_j + \sigma \epsilon_{ij} \]
Life Data Random Effect Model

1. Add subject–specific random effect
2. Called \textit{frailty model}.
   - because it allows for some clusters to be more frail than others
3. Can be applied to Proportional Hazards
   1. \( h_{ij}(t) = h_0(t) \exp(z_{ij} \beta + \tau W_j) \)
   2. If we knew \( W \), would treat as Cox model with the partial likelihood \( L(\beta, \tau; Y, W) \)
4. Can be applied to Accelerated Failure
   1. \( \log(T_{ij}) = \alpha + z_{ij} \beta + \tau W_j + \sigma \epsilon_{ij} \)
   2. If we knew \( W \), would treat as AFT model with the likelihood \( L(\alpha, \beta, \tau, \sigma; Y, W) \)

More complicated Situations: Non-Independent Survival Times
Integrate to remove unobserved common factor

$L(\alpha, \beta, \tau, \sigma; Y) = E_{\tau} [L(\alpha, \beta, \tau, \sigma; Y, W)]$
Integrate to remove unobserved common factor

\[ L(\alpha, \beta, \tau, \sigma; Y) = E_\tau [L(\alpha, \beta, \tau, \sigma; Y, W)] \]

1. Expectation calculation is difficult

More complicated Situations: Non-Independent Survival Times
Integrate to remove unobserved common factor

1. \[ L(\alpha, \beta, \tau, \sigma; Y) = E_\tau [L(\alpha, \beta, \tau, \sigma; Y, W)] \]

2. Expectation calculation is difficult

3. May be done:

   1. via Monte Carlo
   2. via E-M
   3. via approximation

More complicated Situations: Non-Independent Survival Times
Integrate to remove unobserved common factor

\[ L(\alpha, \beta, \tau, \sigma; \mathbf{Y}) = \mathbb{E}_\tau [L(\alpha, \beta, \tau, \sigma; \mathbf{Y}, \mathbf{W})] \]

Expectation calculation is difficult

May be done:
1. via Monte Carlo
Integrate to remove unobserved common factor

1. $L(\alpha, \beta, \tau, \sigma; Y) = E_\tau [L(\alpha, \beta, \tau, \sigma; Y, W)]$
2. Expectation calculation is difficult
3. May be done:
   1. via Monte Carlo
   2. via E-M
Integrate to remove unobserved common factor

1. \( L(\alpha, \beta, \tau, \sigma; Y) = \mathbb{E}_\tau [L(\alpha, \beta, \tau, \sigma; Y, W)] \)

2. Expectation calculation is difficult

3. May be done:
   1. via Monte Carlo
   2. via E-M
     1. Unobserved frailties are missing data
Integrate to remove unobserved common factor

1. \[ L(\alpha, \beta, \tau, \sigma; Y) = E_\tau [L(\alpha, \beta, \tau, \sigma; Y, W)] \]

2. Expectation calculation is difficult

3. May be done:
   - via Monte Carlo
   - via E-M

   1. Unobserved frailties are missing data
   2. Estimate them and coefficient of distn simultaneously

More complicated Situations: Non-Independent Survival Times
Integrate to remove unobserved common factor

1. \( L(\alpha, \beta, \tau, \sigma; Y) = E_{\tau} [L(\alpha, \beta, \tau, \sigma; Y, W)] \)

2. Expectation calculation is difficult

3. May be done:
   1. via Monte Carlo
   2. via E-M
      1. Unobserved frailties are missing data
      2. Estimate them and coefficient of distn simultaneously
      3. Text web site gives SAS macro
Integrate to remove unobserved common factor

\[ L(\alpha, \beta, \tau, \sigma; Y) = E_\tau [L(\alpha, \beta, \tau, \sigma; Y, W)] \]

1. Expectation calculation is difficult
2. May be done:
   1. via Monte Carlo
   2. via E-M
      1. Unobserved frailties are missing data
      2. Estimate them and coefficient of distn simultaneously
      3. Text web site gives SAS macro
   3. via approximation

More complicated Situations: Non-Independent Survival Times
Often parameterize in terms of exponentiated error

Typically want $E[U_j] = 1$ for $U_j = \exp(\sigma W_j)$
Often parameterize in terms of exponentiated error

1. Typically want $E[U_j] = 1$ for $U_j = \exp(\sigma W_j)$
2. Not same as $E[W_j] = 0$
Often parameterize in terms of exponentiated error

1. Typically want \( \mathbb{E}[U_j] = 1 \) for \( U_j = \exp(\sigma W_j) \)
2. Not same as \( \mathbb{E}[W_j] = 0 \)
3. Can’t work for various values of \( \sigma \)
Often parameterize in terms of exponentiated error

1. Typically want $\mathbb{E}[U_j] = 1$ for $U_j = \exp(\sigma W_j)$
2. Not same as $\mathbb{E}[W_j] = 0$
3. Can’t work for various values of $\sigma$
4. Often $U_j$ take as gamma
Often parameterize in terms of exponentiated error

1. Typically want $E[U_j] = 1$ for $U_j = \exp(\sigma W_j)$
2. Not same as $E[W_j] = 0$
3. Can’t work for various values of $\sigma$
4. Often $U_j$ take as gamma
   - with density $u^{(1/\theta-1)} \exp(-u/\theta)\theta^{-1/\theta} / \Gamma(1/\theta)$

More complicated Situations: Non-Independent Survival Times
Often parameterize in terms of exponentiated error

1. Typically want $E[U_j] = 1$ for $U_j = \exp(\sigma W_j)$
2. Not same as $E[W_j] = 0$
3. Can’t work for various values of $\sigma$
4. Often $U_j$ take as gamma
   1. with density $u^{(1/\theta - 1)} \exp(-u/\theta) \theta^{-1/\theta} / \Gamma(1/\theta)$
   2. Same scale and shape parameter $1/\theta$
Often parameterize in terms of exponentiated error

1. Typically want $E[U_j] = 1$ for $U_j = \exp(\sigma W_j)$
2. Not same as $E[W_j] = 0$
3. Can’t work for various values of $\sigma$
4. Often $U_j$ take as gamma
   1. with density $u^{(1/\theta-1)} \exp(-u/\theta)\theta^{-1/\theta}/\Gamma(1/\theta)$
   2. Same scale and shape parameter $1/\theta$
   3. Independence corresponds to $\theta = 0$
Section: Multiple types of event per individual

Subsection: Repeated Event Modeling
Multiple events of the same type per individual

1. Can treat intervals between observations as separate events

Idea:
- Individual might contract recurrent disease
- Repeated repairs to machinery

Model:
- Times between events as independent (conditional on individual)
- Let mean time depend on how many recurrences have already happened (at least).
- Time to events treated as independent conditional on subject.

Possible extensions:
- Censoring tied to event times?
- Use of frailties that depend on time?

R Code
SAS Code
Multiple events of the same type per individual

1. Can treat intervals between observations as separate events
2. Handle dependence using frailty models
Multiple events of the same type per individual

1. Can treat intervals between observations as separate events
2. Handle dependence using frailty models
3. Idea:

   - Individual might contract recurrent disease
   - Repeated repairs to machinery

4. Model:
   - Times between events as independent (conditional on individual)
   - Let mean time depend on how many recurrences have already happened (at least).
   - Time to events treated as independent conditional on subject.

5. Possible extensions:
   - Censoring tied to event times?
   - Use of frailties that depend on time? R Code SAS Code
Multiple events of the same type per individual

1. Can treat intervals between observations as separate events
2. Handle dependence using frailty models
3. Idea:
   1. Individual might contract recurrent disease
Multiple events of the same type per individual

1. Can treat intervals between observations as separate events
2. Handle dependence using frailty models
3. Idea:
   1. Individual might contract recurrent disease
   2. Repeated repairs to machinery
Multiple events of the same type per individual

1. Can treat intervals between observations as separate events
2. Handle dependence using frailty models
3. Idea:
   1. Individual might contract recurrent disease
   2. Repeated repairs to machinery
4. Model:
Multiple events of the same type per individual

1. Can treat intervals between observations as separate events
2. Handle dependence using frailty models
3. Idea:
   1. Individual might contract recurrent disease
   2. Repeated repairs to machinery
4. Model:
   1. Times between events as independent (conditional on individual)
Multiple events of the same type per individual

Can treat intervals between observations as separate events

Handle dependence using frailty models

Idea:

- Individual might contract recurrent disease
- Repeated repairs to machinery

Model:

- Times between events as independent (conditional on individual)
- Let mean time depend on how many recurrences have already happened (at least).
Multiple events of the same type per individual

1. Can treat intervals between observations as separate events
2. Handle dependence using frailty models
3. Idea:
   1. Individual might contract recurrent disease
   2. Repeated repairs to machinery
4. Model:
   1. Times between events as independent (conditional on individual)
   2. Let mean time depend on how many recurrences have already happened (at least).
   3. Time to events treated as independent independent conditional on subject.

Possible extensions:
1. Censoring tied to event times?
2. Use of frailties that depend on time?
Multiple events of the same type per individual

1. Can treat intervals between observations as separate events
2. Handle dependence using frailty models
3. Idea:
   1. Individual might contract recurrent disease
   2. Repeated repairs to machinery
4. Model:
   1. Times between events as independent (conditional on individual)
   2. Let mean time depend on how many recurrences have already happened (at least).
   3. Time to events treated as independent independent conditional on subject.
5. Possible extensions:
Multiple events of the same type per individual

1. Can treat intervals between observations as separate events
2. Handle dependence using frailty models
3. Idea:
   1. Individual might contract recurrent disease
   2. Repeated repairs to machinery
4. Model:
   1. Times between events as independent (conditional on individual)
   2. Let mean time depend on how many recurrences have already happened (at least).
   3. Time to events treated as independent conditional on subject.
5. Possible extensions:
   1. Censoring tied to event times?
Multiple events of the same type per individual

1. Can treat intervals between observations as separate events
2. Handle dependence using frailty models
3. Idea:
   1. Individual might contract recurrent disease
   2. Repeated repairs to machinery
4. Model:
   1. Times between events as independent (conditional on individual)
   2. Let mean time depend on how many recurrences have already happened (at least).
   3. Time to events treated as independent independent conditional on subject.
5. Possible extensions:
   1. Censoring tied to event times?
   2. Use of frailties that depend on time?  R Code  SAS Code
Section: Multiple types of event per individual

Subsection: When only the first is seen:
Examples of Competing Risks

1. Car failure due to

2. Death due to various disease causes

Multiple types of event per individual: When only the first is seen:
Examples of Competing Risks

1. Car failure due to
   1. engine failure

Multiple types of event per individual: When only the first is seen:
Examples of Competing Risks

1. Car failure due to
   1. engine failure
   2. rusting out
Examples of Competing Risks

1. Car failure due to
   1. engine failure
   2. rusting out
   3. censoring is accident
Examples of Competing Risks

1. Car failure due to
   1. engine failure
   2. rusting out
   3. censoring is accident

2. Death due to
Examples of Competing Risks

1. Car failure due to
   1. engine failure
   2. rusting out
   3. censoring is accident

2. Death due to
   1. various disease causes
Examples of Competing Risks

1. Car failure due to
   - 1. engine failure
   - 2. rusting out
   - 3. censoring is accident

2. Death due to
   - 1. various disease causes
   - 2. censoring is loss to follow up

Multiple types of event per individual: When only the first is seen:
Model is called *competing risk*.

Then $T = \min(T^1, \ldots, T^k)$
Model is called *competing risk*.

1. Then $T = \min(T^1, \ldots, T^k)$
2. Let $V$ denote type of event we see
Model is called *competing risk*. 

1. Then $T = \min(T^1, \ldots, T^k)$

2. Let $V$ denote type of event we see

3. Cause-specific hazard for type $\nu$ failure is

   $$h_\nu(t) = P \left[ T \in (t, t + \delta), V = \nu | T \geq t \right] / \delta.$$
Model is called \textit{competing risk}.

1. Then $T = \min(T^1, \ldots, T^k)$
2. Let $V$ denote type of event we see
3. Cause-specific hazard for type $v$ failure is
   \[
   h_v(t) = P \left[ T \in (t, t+\delta), V = v \mid T \geq t \right] / \delta.
   \]
4. Note nonstandard conditioning argument

\[1\] More random variables on left of bar than on right.
\[5\] Aggregate hazard for $T$ is $\sum_{v=1}^{k} h_v(t)$
\[6\] Cause-specific hazard reduces to usual definition if $T_v$ independent.
\[7\] Independence not checkable from data, since any distribution of $(T, V)$ is obtainable from independence model
\[8\] See Tsiadis (1975). Argument involves solving a system of equations for the hazard function.
\[9\] He\'ckman and Honor\'e (1989) cite also Cox (1962), Objective is generally estimating some notion of probability of event from specific cause.
Model is called \textit{competing risk}.

1. Then $T = \min(T^1, \ldots, T^k)$
2. Let $V$ denote type of event we see
3. Cause-specific hazard for type $v$ failure is
   \[ h_v(t) = P[T \in (t, t+\delta), V = v | T \geq t] / \delta. \]
4. Note nonstandard conditioning argument
   1. More random variables on left of bar than on right.

Aggregated hazard for $T$ is $\sum_{v=1}^{k} h_v(t)$.

Cause-specific hazard reduces to usual definition if $T^v$ independent.

Independence not checkable from data, since any distn of $(T, V)$ is obtainable from independence model. See Tsiadis (1975).

Heckman and Honoré (1989) cite also Cox (1962).

Objective is generally estimating some notion of probability of event from specific cause.
Model is called *competing risk*.

1. Then \( T = \min(T^1, \ldots, T^k) \)
2. Let \( V \) denote type of event we see
3. Cause-specific hazard for type \( v \) failure is

\[
h_v(t) = P [ T \in (t, t+\delta), V = v | T \geq t ] / \delta.
\]

4. Note nonstandard conditioning argument
   - More random variables on left of bar than on right.
5. Aggregate hazard for \( T \) is \( \sum_{v=1}^{k} h_v(t) \)

Multiple types of event per individual: When only the first is seen:
Model is called *competing risk*.

1. Then $T = \min(T^1, \ldots, T^k)$
2. Let $V$ denote type of event we see
3. Cause-specific hazard for type $\nu$ failure is

\[ h_{\nu}(t) = P[T \in (t, t+\delta), V = \nu | T \geq t] / \delta. \]

4. Note nonstandard conditioning argument
   - More random variables on left of bar than on right.
5. Aggregate hazard for $T$ is $\sum_{\nu=1}^{k} h_{\nu}(t)$
6. Cause-specific hazard reduces to usual definition if $T^\nu$ independent.
Model is called *competing risk*.

1. Then $T = \min(T^1, \ldots, T^k)$
2. Let $V$ denote type of event we see
3. Cause-specific hazard for type $\nu$ failure is
   
   $$h_\nu(t) = P[T \in (t, t+\delta), V = \nu | T \geq t] / \delta.$$  

4. Note nonstandard conditioning argument
   
   - More random variables on left of bar than on right.

5. Aggregate hazard for $T$ is $\sum_{\nu=1}^{k} h_\nu(t)$
6. Cause-specific hazard reduces to usual definition if $T^\nu$ independent.
   
   - Usually this not even plausible, much less certain.

Multiple types of event per individual: When only the first is seen:
Model is called *competing risk*.

1. Then \( T = \min(T^1, \ldots, T^k) \)
2. Let \( V \) denote type of event we see
3. Cause-specific hazard for type \( v \) failure is
   \[
   h_v(t) = P[T \in (t, t + \delta), V = v | T \geq t] / \delta.
   \]
4. Note nonstandard conditioning argument
   - More random variables on left of bar than on right.
5. Aggregate hazard for \( T \) is \( \sum_{v=1}^{k} h_v(t) \)
6. Cause-specific hazard reduces to usual definition if \( T^v \) independent.
   - Usually this not even plausible, much less certain.
7. Independence not checkable from data,
Model is called *competing risk*.

1. Then $T = \min(T^1, \ldots, T^k)$
2. Let $V$ denote type of event we see
3. Cause-specific hazard for type $v$ failure is

   $$h_v(t) = P[T \in (t, t + \delta), V = v | T \geq t] / \delta.$$ 

4. Note nonstandard conditioning argument
   - More random variables on left of bar than on right.
5. Aggregate hazard for $T$ is $\sum_{v=1}^{k} h_v(t)$
6. Cause-specific hazard reduces to usual definition if $T^v$ independent.
   - Usually this not even plausible, much less certain.
7. Independence not checkable from data,
   - since any distn of $(T, V)$ is obtainable from independence model
Model is called *competing risk*.

1. Then \( T = \min(T^1, \ldots, T^k) \)
2. Let \( V \) denote type of event we see
3. Cause-specific hazard for type \( v \) failure is
   \[
   h_v(t) = P\left[T \in (t, t + \delta), V = v \mid T \geq t\right] / \delta.
   \]
4. Note nonstandard conditioning argument
   1. More random variables on left of bar than on right.
5. Aggregate hazard for \( T \) is \( \sum_{v=1}^{k} h_v(t) \)
6. Cause-specific hazard reduces to usual definition if \( T^v \) independent.
   1. Usually this not even plausible, much less certain.
7. Independence not checkable from data,
   1. since any distn of \( (T, V) \) is obtainable from independence model
   2. See Tsiadis (1975).
Model is called *competing risk*.

1. Then \( T = \min(T^1, \ldots, T^k) \)
2. Let \( V \) denote type of event we see
3. Cause-specific hazard for type \( v \) failure is

\[
h_v(t) = P[T \in (t, t+\delta), V = v|T \geq t]/\delta.
\]

4. Note nonstandard conditioning argument
   1. More random variables on left of bar than on right.
5. Aggregate hazard for \( T \) is \( \sum_{v=1}^{k} h_v(t) \)
6. Cause-specific hazard reduces to usual definition if \( T^v \) independent.
   1. Usually this not even plausible, much less certain.
7. Independence not checkable from data,
   1. since any distn of \((T, V)\) is obtainable from independence model
   2. See Tsiadis (1975).
   3. Argument involves solving a system of equations for the hazard function.
Model is called **competing risk**.

1. Then $T = \min(T^1, \ldots, T^k)$

2. Let $V$ denote type of event we see

3. Cause-specific hazard for type $v$ failure is

   $$h_v(t) = P[T \in (t, t+\delta), V = v|T \geq t] / \delta.$$  

4. Note nonstandard conditioning argument
   1. More random variables on left of bar than on right.

5. Aggregate hazard for $T$ is $\sum_{v=1}^{k} h_v(t)$  

6. Cause-specific hazard reduces to usual definition if $T^v$ independent.
   1. Usually this not even plausible, much less certain.

7. Independence not checkable from data,
   1. since any distn of $(T, V)$ is obtainable from independence model
   2. See Tsiadis (1975).
   3. Argument involves solving a system of equations for the hazard function.
   4. Heckman and Honoré (1989) cite also Cox (1962),

Multiple types of event per individual: When only the first is seen:
Model is called **competing risk**.

1. Then \( T = \min(T_1, \ldots, T^k) \)
2. Let \( V \) denote type of event we see
3. Cause-specific hazard for type \( v \) failure is

\[
h_v(t) = \frac{P\left[T \in (t, t+\delta), V = v \mid T \geq t\right]}{\delta}.
\]

4. Note nonstandard conditioning argument
   - More random variables on left of bar than on right.
5. Aggregate hazard for \( T \) is \( \sum_{v=1}^{k} h_v(t) \)
6. Cause-specific hazard reduces to usual definition if \( T^v \) independent.
   - Usually this not even plausible, much less certain.
7. Independence not checkable from data,
   - since any distn of \( (T, V) \) is obtainable from independence model
   - See Tsiadis (1975).
   - Argument involves solving a system of equations for the hazard function.
   - Heckman and Honoré (1989) cite also Cox (1962),
8. Objective is generally estimating some notion of probability of event from specific cause.
Taxonomy of Summaries

- Crude: treat other failures as precluding failure of interest.

  - Subjects lost to followup are treated as censored.
  - Subjects whose progress is not tracked for other causes are treated as never having event.
  - Ex. car scrapped because transmission goes out is treated as never having engine failure.
  - car removed from data set because owner stops responding is treated as censored.

- Report using cumulative incidence function $P[T \leq t, V = v]$.

- Net: treat other failures as censored.

  - Estimate net survival using Kaplan–Meier.
  - Treat all failures not of type $v$ as censored.
  - Only corresponds to survival function for process $v$ acting alone under independence.
Taxonomy of Summaries

1. Crude: treat other failures as precluding failure of interest.
   - Subjects lost to followup are treated as censored.
Crude: treat other failures as precluding failure of interest.

1. Subjects lost to followup are treated as censored.
2. Subjects whose progress is not tracked for other causes are treated as never having event.

Ex., car scrapped because transmission goes out is treated as never having engine failure.
Car removed from data set because owner stops responding is treated as censored.

Report using cumulative incidence function $P[T \leq t, V = v]$.

Net: treat other failures as censored.

Estimate net survival using Kaplan–Meier.
Taxonomy of Summaries

1. Crude: treat other failures as precluding failure of interest.
   1. Subjects lost to followup are treated as censored.
   2. Subjects whose progress is not tracked for other causes are treated as never having event.
   3. Ex.,

Report using cumulative incidence function $P[T \leq t, V = v]$. 

2. Net: treat other failures as censored.

Estimate net survival using Kaplan–Meier.

Treat all failures not of type $v$ as censored.

Only corresponds to survival function for process $v$ acting alone under independence.

Multiple types of event per individual: When only the first is seen:
Taxonomy of Summaries

1. **Crude**: treat other failures as precluding failure of interest.
   - 1. Subjects lost to followup are treated as censored.
   - 2. Subjects whose progress is not tracked for other causes are treated as never having event.
   - 3. Ex.,
     - 1. Car scrapped because transmission goes out is treated as never having engine failure

2. **Net**: treat other failures as censored.

   - Estimate net survival using Kaplan–Meier
   - Treat all failures not of type $v$ as censored
   - Only corresponds to survival function for process $v$ acting alone under independence
Taxonomy of Summaries

1. Crude: treat other failures as precluding failure of interest.
   1. Subjects lost to followup are treated as censored.
   2. Subjects whose progress is not tracked for other causes are treated as never having event.

3. Ex.,
   1. car scrapped because transmission goes out is treated as never having engine failure
   2. car removed from data set because owner stops responding is treated as censored.

Report using cumulative incidence function $P[T \leq t, V = v]$.

2. Net: treat other failures as censored.

   Estimate net survival using Kaplan–Meier

   Treat all failures not of type $v$ as censored

   Only corresponds to survival function for process $v$ acting alone under independence
Taxonomy of Summaries

1. **Crude**: treat other failures as precluding failure of interest.
   1. Subjects lost to followup are treated as censored.
   2. Subjects whose progress is not tracked for other causes are treated as never having event.
   3. Ex.,
      1. car scrapped because transmission goes out is treated as never having engine failure
      2. car removed from data set because owner stops responding is treated as censored.
   4. Report using *cumulative incidence function* \( P [ T \leq t, V = v ] \).

Multiple types of event per individual: When only the first is seen:
Taxonomy of Summaries

1. Crude: treat other failures as precluding failure of interest.
   1. Subjects lost to followup are treated as censored.
   2. Subjects whose progress is not tracked for other causes are treated as never having event.
   3. Ex.,
      1. car scrapped because transmission goes out is treated as never having engine failure
      2. car removed from data set because owner stops responding is treated as censored.
   4. Report using cumulative incidence function $\mathbb{P}[T \leq t, V = v]$.

2. Net: treat other failures as censored.
Taxonomy of Summaries

1. **Crude**: treat other failures as precluding failure of interest.
   - Subjects lost to followup are treated as censored.
   - Subjects whose progress is not tracked for other causes are treated as never having event.
   - Ex., car scrapped because transmission goes out is treated as never having engine failure.
   - Ex., car removed from data set because owner stops responding is treated as censored.
   - Report using cumulative incidence function $P[T \leq t, V = v]$.

2. **Net**: treat other failures as censored.
   - Estimate net survival using Kaplan–Meier
Taxonomy of Summaries

1. **Crude**: treat other failures as precluding failure of interest.
   1. Subjects lost to followup are treated as censored.
   2. Subjects whose progress is not tracked for other causes are treated as never having event.
   3. Ex.,
      1. car scrapped because transmission goes out is treated as never having engine failure
      2. car removed from data set because owner stops responding is treated as censored.
   4. Report using cumulative incidence function \( P[T \leq t, V = v] \).

2. **Net**: treat other failures as censored.
   1. Estimate net survival using Kaplan–Meier
   2. Treat all failures not of type \( v \) as censored

Multiple types of event per individual: When only the first is seen:
Taxonomy of Summaries

1. Crude: treat other failures as precluding failure of interest.
   1. Subjects lost to followup are treated as censored.
   2. Subjects whose progress is not tracked for other causes are treated as never having event.
   3. Ex.,
      1. car scrapped because transmission goes out is treated as never having engine failure
      2. car removed from data set because owner stops responding is treated as censored.

2. Net: treat other failures as censored.
   1. Estimate net survival using Kaplan–Meier
   2. Treat all failures not of type $v$ as censored
   3. Only corresponds to survival function for process $v$ acting alone under independence

Multiple types of event per individual: When only the first is seen:
Partial Crude: treat some endpoints as ruling out endpoint of intest.
Partial Crude: treat some endpoints as ruling out endpoint of interest.

Select some other causes to treat as censored, and some to treat as removing potential for event.
Partial Crude: treat some endpoints as ruling out endpoint of interest.

1. Select some other causes to treat as censored, and some to treat as removing potential for event.
2. Ex., car that gets scrapped because transmission goes out is treated as never having engine failure, but car removed from data set because of accident is treated as censored.
Aalen-Johansen: Modify Kaplan-Meier idea to capture cause-specific hazard at a point.
Aalen-Johansen: Modify Kaplan-Meier idea to capture cause-specific hazard at a point.

Aalen-Johansen: Modify Kaplan-Meier idea to capture cause-specific hazard at a point.


Probability of moving from censored to event \( \nu \) by time \( t \) is
\[
\int_0^t S(s-)h_\nu(s) \, ds = \int_0^t S(s-)dH_\nu(s).
\]
Aalen-Johansen: Modify Kaplan-Meier idea to capture cause-specific hazard at a point.


Probability of moving from censored to event \( v \) by time \( t \) is
\[
\int_0^t S(s-) h_v(s) \, ds = \int_0^t S(s-) dH_v(s).
\]

\( S(t-) \) is probability of lasting until at least \( t \).
Continued

1. Aalen-Johansen: Modify Kaplan-Meier idea to capture cause-specific hazard at a point.


2. Probability of moving from censored to event \( \nu \) by time \( t \) is
   \[
   \int_0^t S(s-)h_\nu(s) \, ds = \int_0^t S(s-)dH_\nu(s).
   \]

3. \( S(t-) \) is probability of lasting until at least \( t \)

4. Estimate via Stieltjes integral of empirical quantities

Multiple types of event per individual: When only the first is seen:
Aalen-Johansen: Modify Kaplan-Meier idea to capture cause-specific hazard at a point.


Probability of moving from censored to event $v$ by time $t$ is

\[ \int_0^t S(s-)h_v(s) \, ds = \int_0^t S(s-)dH_v(s). \]

$S(t-)$ is probability of lasting until at least $t$

Estimate via *Stieltjes integral* of empirical quantities

For any increasing function $H$, \( \int f(s) \, dH(s) \) is defined to be
Aalen-Johansen: Modify Kaplan-Meier idea to capture cause-specific hazard at a point.


Probability of moving from censored to event \( \nu \) by time \( t \) is
\[
\int_0^t S(s-)h_\nu(s) \, ds = \int_0^t S(s-)dH_\nu(s).
\]

\( S(t-) \) is probability of lasting until at least \( t \)

Estimate via Stieltjes integral of empirical quantities

For any increasing function \( H \), \( \int f(s) \, dH(s) \) is defined to be
\[
\int f(s) \, dh(s) \, ds \text{ if } H \text{ has derivative } h
\]
Aalen-Johansen: Modify Kaplan-Meier idea to capture cause-specific hazard at a point.


Probability of moving from censored to event \( \nu \) by time \( t \) is
\[
\int_0^t S(s-) h_{\nu}(s) \, ds = \int_0^t S(s-) dH_{\nu}(s).
\]

\( S(t-) \) is probability of lasting until at least \( t \)

Estimate via Stieltjes integral of empirical quantities

For any increasing function \( H \), \( \int f(s) \, dH(s) \) is defined to be
\[
\int f(s) \, dh(s) \, ds \quad \text{if} \quad H \text{ has derivative } h
\]
\[
\sum_j f(s_j)(H(s_j+) - H(s_j-)) \quad \text{for points } \{s_j\} \text{ where } H \text{ has jumps, if all variation in } H \text{ comes as jumps at discrete places.} \]
Aalen-Johansen: Modify Kaplan-Meier idea to capture cause-specific hazard at a point.


2. Probability of moving from censored to event \( \nu \) by time \( t \) is

\[
\int_0^t S(s-)h_\nu(s) \, ds = \int_0^t S(s-)dH_\nu(s).
\]

3. \( S(t-) \) is probability of lasting until at least \( t \)

4. Estimate via *Stieltjes integral* of empirical quantities

5. For any increasing function \( H \), \( \int f(s) \, dH(s) \) is defined to be

\[
\int f(s) \, dh(s) \, ds \quad \text{if } H \text{ has derivative } h
\]

6. \[
\sum_j f(s_j)(H(s_j+) - H(s_j-)) \quad \text{for points } \{s_j\} \text{ where } H \text{ has jumps, if all variation in } H \text{ comes as jumps at discrete places.}
\]

7. If \( H \) is a CDF, \( H(s_j+) = H(s_j) \).
1. Aalen-Johansen: Modify Kaplan-Meier idea to capture cause-specific hazard at a point.


2. Probability of moving from censored to event $\nu$ by time $t$ is

$$\int_0^t S(s-)h_\nu(s) \, ds = \int_0^t S(s-)dH_\nu(s).$$

1. $S(t-) \text{ is probability of lasting until at least } t$

3. Estimate via *Stieltjes integral* of empirical quantities

1. For any increasing function $H$, $\int f(s) \, dH(s)$ is defined to be
2. $\int f(s) \, dh(s) \, ds$ if $H$ has derivative $h$
3. $\sum_j f(s_j)(H(s_j+) - H(s_j-))$ for points $\{s_j\}$ where $H$ has jumps, if all variation in $H$ comes as jumps at discrete places.
4. If $H$ is a CDF, $H(s_j+) = H(s_j)$.
5. $d\hat{H}_\nu(s)$ has jumps $N_\nu(s)/R(s)$ at event times.
Aalen-Johansen: Modify Kaplan-Meier idea to capture cause-specific hazard at a point.


Probability of moving from censored to event $\nu$ by time $t$ is
\[
\int_0^t S(s-) h_\nu(s) \, ds = \int_0^t S(s-) dH_\nu(s).
\]

$S(t-)$ is probability of lasting until at least $t$

Estimate via *Stieltjes integral* of empirical quantities

For any increasing function $H$, $\int f(s) \, dH(s)$ is defined to be
\[
\int f(s) \, dh(s) \, ds \text{ if } H \text{ has derivative } h
\]
\[
\sum_j f(s_j)(H(s_j+) - H(s_j-)) \text{ for points } \{s_j\} \text{ where } H \text{ has jumps, if all variation in } H \text{ comes as jumps at discrete places.}
\]
If $H$ is a CDF, $H(s_j+) = H(s_j)$.

$d\hat{H}_\nu(s)$ has jumps $N_\nu(s)/R(s)$ at event times.

$\hat{S}(t_j-)$ = $\hat{S}(t_{j-1})$
Aalen-Johansen: Modify Kaplan-Meier idea to capture cause-specific hazard at a point.


Probability of moving from censored to event $\nu$ by time $t$ is

$$\int_0^t S(s-) h_\nu(s) \, ds = \int_0^t S(s-) dH_\nu(s).$$

$S(t-)$ is probability of lasting until at least $t$

Estimate via *Stieltjes integral* of empirical quantities

For any increasing function $H$, $\int f(s) \, dH(s)$ is defined to be

$$\int f(s) \, dh(s) \, ds$$

if $H$ has derivative $h$

$$\sum_j f(s_j)(H(s_j+) - H(s_j-))$$

for points $\{s_j\}$ where $H$ has jumps, if all variation in $H$ comes as jumps at discrete places.

If $H$ is a CDF, $H(s_j+) = H(s_j)$.

$d\hat{H}_\nu(s)$ has jumps $N_\nu(s)/R(s)$ at event times.

$\hat{S}(t_j-) = \hat{S}(t_{j-1})$

Estimator is $\sum_{j \geq t_j < t} \hat{S}(t_{j-1}) d_j, \nu/R_j$  

R Code

Multiple types of event per individual: When only the first is seen:
Objectives Lecture 12

1. Bayesian Methods

Multiple types of event per individual: When only the first is seen:
Objectives Lecture 12

1. Bayesian Methods
2. Multi-State Models

Multiple types of event per individual: When only the first is seen:
Objectives Lecture 12

1. Bayesian Methods
2. Multi-State Models
3. Readings: KM §6.4, 9.5
Section: What are people working on now?

Subsection: Multistage Modeling
Idea:

1. Individuals go through stages in disease
Idea:

1. Individuals go through stages in disease
   1. Ex. Cancer: Active, Remission, Death
Idea:

1. Individuals go through stages in disease
   - Ex. Cancer: Active, Remission, Death
2. Data sets typically start with all participants in an initial stage
   - Transition to some other state is what was referred to in earlier examples as aggregate survival.
   - Might jump back and forth
3. Model position in network as function of time
   - Times for passing between stages
   - Process that determines which stage is next
Idea:

1. Individuals go through stages in disease
   - Ex. Cancer: Active, Remission, Death
2. Data sets typically start with all participants in an initial stage
   - Initial stage might represent a variety of conditions
Idea:

1. Individuals go through stages in disease
   - Ex. Cancer: Active, Remission, Death

2. Data sets typically start with all participants in an initial stage
   - Initial stage might represent a variety of conditions
     - Healthy

What are people working on now?: Multistage Modeling
Idea:

1. Individuals go through stages in disease
   1. Ex. Cancer: Active, Remission, Death

2. Data sets typically start with all participants in an initial stage
   1. Initial stage might represent a variety of conditions
      1. Healthy
      2. Initially diagnosed
Idea:

1. Individuals go through stages in disease
   - Ex. Cancer: Active, Remission, Death
2. Data sets typically start with all participants in an initial stage
   - Initial stage might represent a variety of conditions
     - Healthy
     - Initially diagnosed
     - Entered clinical trial
Idea:

1. **Individuals go through stages in disease**
   - Ex. Cancer: Active, Remission, Death

2. **Data sets typically start with all participants in an initial stage**
   - Initial stage might represent a variety of conditions
     - Healthy
     - Initially diagnosed
     - Entered clinical trial
     - etc.
Idea:

1. Individuals go through stages in disease
   - Ex. Cancer: Active, Remission, Death

2. Data sets typically start with all participants in an initial stage
   - Initial stage might represent a variety of conditions
     - Healthy
     - Initially diagnosed
     - Entered clinical trial
     - etc.

3. Transition to some other state is what was referred to in earlier examples as aggregate survival.
Idea:

1. Individuals go through stages in disease
   - Ex. Cancer: Active, Remission, Death

2. Data sets typically start with all participants in an initial stage
   - Initial stage might represent a variety of conditions
     - Healthy
     - Initially diagnosed
     - Entered clinical trial
     - etc.

3. Transition to some other state is what was referred to in earlier examples as aggregate survival.
   - Might jump back and forth
Idea:

1. Individuals go through stages in disease
   - Ex. Cancer: Active, Remission, Death

2. Data sets typically start with all participants in an initial stage
   - Initial stage might represent a variety of conditions
     - Healthy
     - Initially diagnosed
     - Entered clinical trial
     - etc.

3. Transition to some other state is what was referred to in earlier examples as aggregate survival.
   - Might jump back and forth

4. Model position in network as function of time
Idea:

1. Individuals go through stages in disease
   - Ex. Cancer: Active, Remission, Death

2. Data sets typically start with all participants in an initial stage
   - Initial stage might represent a variety of conditions
     - Healthy
     - Initially diagnosed
     - Entered clinical trial
     - etc.

3. Transition to some other state is what was referred to in earlier examples as aggregate survival.
   - Might jump back and forth

4. Model position in network as function of time
   - Times for passing between stages

What are people working on now?: Multistage Modeling
Idea:

1. Individuals go through stages in disease
   - Ex. Cancer: Active, Remission, Death
2. Data sets typically start with all participants in an initial stage
   - Initial stage might represent a variety of conditions
     - Healthy
     - Initially diagnosed
     - Entered clinical trial
     - etc.
3. Transition to some other state is what was referred to in earlier examples as aggregate survival.
   - Might jump back and forth
4. Model position in network as function of time
   - Times for passing between stages
   - Process that determines which stage is next
Some cases:

1. Trivial case: Transition from alive to dead.

If censoring is independent of transition to final stage, adjust for it in modeling transition. Otherwise ignore.

Competing Risk model is a slightly less trivial case.

Competing risk model uses this machinery R Code.
Some cases:

1. Trivial case: Transition from alive to dead.
   - See Fig. 15.

Fig. 15: Trivial Transitions

Entry State —> Censored —> Dead, Cause 1

Entry State —> Censored —> Dead, Cause 2
Some cases:

1. Trivial case: Transition from alive to dead.
   - See Fig. 15.

   *Fig. 15: Trivial Transitions*

   - Entry State
   - Censored
   - Dead

2. If censoring is independent of transition to final stage,

   Competing Risk model uses this machinery
Some cases:

1. Trivial case: Transition from alive to dead.
   - See Fig. 15.

   \textbf{Fig. 15: Trivial Transitions}

   \begin{itemize}
   \item Entry State
   \item Censored
   \item Dead
   \end{itemize}

2. If censoring is independent of transition to final stage,
   - Adjust for it in modeling transition
Some cases:

1. Trivial case: Transition from alive to dead.
   - See Fig. 15.

   \[ \text{Entry State} \rightarrow \text{Censored} \rightarrow \text{Dead} \]

2. If censoring is independent of transition to final stage,
   - Adjust for it in modeling transition
   - Otherwise ignore.

\[ \text{Fig. 15: Trivial Transitions} \]
Some cases:

1. Trivial case: Transition from alive to dead.
   1. See Fig. 15.

   ![Fig. 15: Trivial Transitions](image)

   - Entry State
   - Censored
   - Dead

2. If censoring is independent of transition to final stage,
   1. Adjust for it in modeling transition
   2. Otherwise ignore.

3. Competing Risk model is a slightly less trivial case.
Some cases:

1. Trivial case: Transition from alive to dead.
   - See Fig. 15.

   *Fig. 15: Trivial Transitions*

   - Entry State
   - Censored
   - Dead

   If censoring is independent of transition to final stage,
   - Adjust for it in modeling transition
   - Otherwise ignore.

2. Competing Risk model is a slightly less trivial case.
   - See Fig. 15.

   *Fig. 15: Transitions with Two Causes*

   - Entry State
   - Censored
   - Dead, Cause 1
   - Dead, Cause 2

   Competing Risk model uses this machinery

What are people working on now?: Multistage Modeling
Some cases:

1. Trivial case: Transition from alive to dead.
   - See Fig. 15.

   \[ Fig. 15: Trivial Transitions \]

   ![Diagram of trivial transitions](image)

2. If censoring is independent of transition to final stage,
   - Adjust for it in modeling transition
   - Otherwise ignore.

2. Competing Risk model is a slightly less trivial case.
   - See Fig. 15.

   \[ Fig. 15: Transitions with Two Causes \]

   ![Diagram of transitions with two causes](image)

3. Competing risk model uses this machinery

   \[ R \] Code
Multi-state model

1. Individuals start in initial stage
Multi-state model

1. Individuals start in initial stage
2. Stages from which one cannot leave are called *absorbing*.
Multi-state model

1. Individuals start in initial stage
2. Stages from which one cannot leave are called *absorbing*.
3. Some states other than initial are not absorbing.
Multi-state model

1. Individuals start in initial stage.
2. Stages from which one cannot leave are called *absorbing*.
3. Some states other than initial are not absorbing.
   - See Fig. 16.

*Fig. 16: Still Transitions with an Absorbing State*

- Entry State
- Dead
- Disease Progression

Censoring not pictured
Multi-state model

1. Individuals start in initial stage
2. Stages from which one cannot leave are called absorbing.
3. Some states other than initial are not absorbing.
   - See Fig. 16.

   **Fig. 16: Still Transitions with an Absorbing State**

   - Entry State
   - Dead
   - Disease Progression

   Censoring not pictured

4. Need to model entry and exit times
Model as *Markov*:

1. Let $S(t)$ be the random state at time $t$.
2. Probability of future conditional on past depends on past only through present.
3. Then for $t_1 < t_2 < \cdots < t_k < t$, require $P[S(t) = x | S(t_1) = x_1, \ldots, S(t_k) = x_k] = P[S(t) = s | S(t_k) = x_k]$.
4. Recall notation from before: $F_t = \text{Information available just before } t$.
5. Then Markov means for $s < t$, $P[S(t) = x | F_s] = P[S(t) = x | S(s)]$.
6. If $P[S(t) = x | S(s)]$ depends on time only through the time difference $t - s$ then the process is time-homogeneous.
7. Otherwise the process is time-inhomogeneous.
Model as Markov:

Let $S(t)$ be the random state at time $t$. 

Probability of future conditional on past depends on past only through present.

Then for $t_1 < t_2 < \cdots < t_k < t$, require

$$P[S(t) = x | S(t_1) = x_1, \ldots, S(t_k) = x_k] = P[S(t) = s | S(t_k) = x_k].$$

Recall notation from before:

$F_t =$ Information available just before $t$

Then Markov means for $s < t$,

$$P[S(t) = x | F_s] = P[S(t) = x | S(s)].$$

If $P[S(t) = x | S(s)]$ depends on time only through the time difference $t - s$ then the process is time-homogeneous.

Otherwise the process is time-inhomogeneous.
Model as *Markov*:

1. Let $S(t)$ be the random state at time $t$.
2. Probability of future conditional on past depends on past only through present.
1 Model as *Markov*:

1. Let $S(t)$ be the random state at time $t$.
2. Probability of future conditional on past depends on past only through present.
3. Then for $t_1 < t_2 < \cdots < t_k < t$, require

$$P[S(t) = x | S(t_1) = x_1, \ldots, S(t_k) = x_k] = P[S(t) = s | S(t_k) = x_k].$$
Model as *Markov*:

1. Let $S(t)$ be the random state at time $t$.
2. Probability of future conditional on past depends on past only through present.
3. Then for $t_1 < t_2 < \cdots < t_k < t$, require
   
   $$P [S(t) = x | S(t_1) = x_1, \ldots, S(t_k) = x_k] = P [S(t) = s | S(t_k) = x_k].$$

4. Recall notation from before: $\mathcal{F}_t =$ Information available just before $t$. 

What are people working on now?: Multistage Modeling Lecture 12
Model as *Markov*:

1. Let $S(t)$ be the random state at time $t$.
2. Probability of future conditional on past depends on past only through present.
3. Then for $t_1 < t_2 < \cdots < t_k < t$, require

   $$P[S(t) = x | S(t_1) = x_1, \ldots, S(t_k) = x_k] = P[S(t) = s | S(t_k) = x_k].$$

4. Recall notation from before: $\mathcal{F}_t = \text{Information available just before } t$
5. Then Markov means for $s < t$, $P[S(t) = x | \mathcal{F}_s] = P[S(t) = x | S(s)].$
Model as Markov:

1. Let \( S(t) \) be the random state at time \( t \).
2. Probability of future conditional on past depends on past only through present.
3. Then for \( t_1 < t_2 < \cdots < t_k < t \), require

\[
P[S(t) = x | S(t_1) = x_1, \ldots, S(t_k) = x_k] = P[S(t) = s | S(t_k) = x_k].
\]

4. Recall notation from before: \( \mathcal{F}_t = \text{Information available just before } t \)
5. Then Markov means for \( s < t \), \( P[S(t) = x | \mathcal{F}_s] = P[S(t) = x | S(s)] \).
6. If \( P[S(t) = x | S(s)] \) depends on time only through the time difference \( t - s \) then the process is time-homogeneous.
Model as *Markov*:

1. Let $S(t)$ be the random state at time $t$.
2. Probability of future conditional on past depends on past only through present.
3. Then for $t_1 < t_2 < \cdots < t_k < t$, require

$$P[S(t) = x | S(t_1) = x_1, \ldots, S(t_k) = x_k] = P[S(t) = s | S(t_k) = x_k].$$

4. Recall notation from before: $\mathcal{F}_t =$ Information available just before $t$
5. Then Markov means for $s < t$, $P[S(t) = x | \mathcal{F}_s] = P[S(t) = x | S(s)]$.
6. If $P[S(t) = x | S(s)]$ depends on time only through the time difference $t - s$ then the process is *time-homogeneous*.
7. Otherwise the process is *time-inhomogeneous*. 
Report probability of each stage at each time.

1. Recall survival function tells about leaving entry state.
Report probability of each stage at each time.

1. Recall survival function tells about leaving entry state.
2. This model must be more complicated.

What are people working on now?: Multistage Modeling
Report probability of each stage at each time.

1. Recall survival function tells about leaving entry state.
2. This model must be more complicated.
   - See Fig. 17.

*Fig. 17: Most complicated Transitions*

Diagram:

- Pre-disease
- Active Disease
- Remission
- Dead

Censoring not pictured
Mathematical Details

1. Calculations for discrete times

Let \( q_k; i = \text{P} \text{[State } k \text{ at time } t_i \text{].} \)

Let \( p_{jk}; i = \text{P} \text{[State } k \text{ at time } t_{i+1} \mid \text{State } j \text{ at time } t_i \]. \)

Suppose everyone starts out in state 1 at time 0.

Then \( q_{j,1} = p_{1,j}, 0 \).

Then \( q_{j,2} = \sum_k p_{jk}, 1 q_{k,1} = \sum_k p_{jk}, 1 p_{1,k}, 0 \).

This is the same mathematical operation as matrix multiplication.

Let \( P_i \) be matrix with \( p_{jk}, i \) in row \( j \) column \( k \).

Let \( q_i \) be the vector with \( q_{j,i} \) in position \( j \).

Then \( q_2 = P_1 q_1 \).

Similarly, \( q_3 = P_2 P_1 q_1 \).

Similarly, \( q_i = P_i \cdots P_2 P_1 q_1 \).
Calculations for discrete times

Discrete times: transitions can only happen at times $t_1, \ldots, t_i, \ldots$
Mathematical Details

1 Calculations for discrete times
   1 Discrete times: transitions can only happen at times $t_1, \ldots, t_i, \ldots$
   2 Let $q_{k;i} = P \text{[State } k \text{ at time } t_i\text{]}$. 

Let $P_i$ be matrix with $p_{jk}, i$ in row $j$ column $k$. Let $q_i$ be the vector with $q_{j,i}$ in position $j$. Then $q_2 = P_1 q_1$. Similarly, $q_3 = P_2 P_1 q_1$. Similarly, $q_i = P_i \cdots P_2 P_1 q_1$. 

What are people working on now?:  Multistage Modeling
Mathematical Details

1. Calculations for discrete times
   1. Discrete times: transitions can only happen at times $t_1, \ldots, t_i, \ldots$
   2. Let $q_{k;i} = P$ [State $k$ at time $t_i$].
   3. Let $p_{jk;i} = P$ [State $k$ at time $t_{i+1}$|State $j$ at time $t_i$].

Suppose everyone starts out in state 1 at time 0.

Then $q_{j,1} = p_{1,j}, 0$.

Then $q_{j,2} = \sum_k p_{kj,1} q_{k,1} = \sum_k p_{kj,1} p_{1,k}, 0$.

This is the same mathematical operation as matrix multiplication.

Let $P_i$ be matrix with $p_{jk,i}$ in row $j$ column $k$.

Let $q_i$ be the vector with $q_{j,i}$ in position $j$.

Then $q_2 = P_1 q_1$.

Similarly, $q_3 = P_2 P_1 q_1$.

Similarly, $q_i = P_i \cdots P_2 P_1 q_1$. 

What are people working on now?: Multistage Modeling
Mathematical Details

1 Calculations for discrete times

1 Discrete times: transitions can only happen at times $t_1, \ldots, t_i, \ldots$

2 Let $q_{k;i} = P [\text{State } k \text{ at time } t_i]$.

3 Let $p_{jk;i} = P [\text{State } k \text{ at time } t_{i+1} | \text{State } j \text{ at time } t_i]$.

4 Suppose everyone starts out in state 1 at time 0
Mathematical Details

1. Calculations for discrete times
   1. Discrete times: transitions can only happen at times $t_1, \ldots, t_i, \ldots$
   2. Let $q_{k;i} = P[\text{State } k \text{ at time } t_i]$.
   3. Let $p_{jk;i} = P[\text{State } k \text{ at time } t_{i+1} | \text{State } j \text{ at time } t_i]$.
   4. Suppose everyone starts out in state 1 at time 0
   5. Then $q_{j,1} = p_{1j,0}$.
Calculations for discrete times

1. Discrete times: transitions can only happen at times $t_1, \ldots, t_i, \ldots$
2. Let $q_{k;i} = P [\text{State } k \text{ at time } t_i]$
3. Let $p_{jk;i} = P [\text{State } k \text{ at time } t_{i+1}|\text{State } j \text{ at time } t_i]$
4. Suppose everyone starts out in state 1 at time 0
5. Then $q_{j,1} = p_{1j,0}$
6. Then $q_{j,2} = \sum_k p_{kj,1} q_{k,1} = \sum_k p_{kj,1} p_{1k,0}$

Mathematical Details

1. **Calculations for discrete times**

1. **Discrete times**: transitions can only happen at times \( t_1, \ldots, t_i, \ldots \)
2. Let \( q_{k,i} = P [\text{State } k \text{ at time } t_i] \).
3. Let \( p_{jk;i} = P [\text{State } k \text{ at time } t_{i+1} | \text{State } j \text{ at time } t_i] \).
4. Suppose everyone starts out in state 1 at time 0
5. Then \( q_{j,1} = p_{1j,0} \).
6. Then \( q_{j,2} = \sum_k p_{kj,1} q_{k,1} = \sum_k p_{kj,1} p_{1k,0} \).
7. This is the same mathematical operation as matrix multiplication.
Calculations for discrete times

1. Discrete times: transitions can only happen at times $t_1, \ldots, t_i, \ldots$
2. Let $q_{k;i} = \mathbb{P} [\text{State } k \text{ at time } t_i]$.
3. Let $p_{jk;i} = \mathbb{P} [\text{State } k \text{ at time } t_{i+1}| \text{State } j \text{ at time } t_i]$.
4. Suppose everyone starts out in state 1 at time 0
5. Then $q_{j,1} = p_{1j,0}$.
6. Then $q_{j,2} = \sum_k p_{kj,1} q_{k,1} = \sum_k p_{kj,1} p_{1k,0}$.
7. This is the same mathematical operation as matrix multiplication.

Let $P_i$ be matrix with $p_{jk,i}$ in row $j$ column $k$
Calculations for discrete times

1. Discrete times: transitions can only happen at times $t_1, \ldots, t_i, \ldots$
2. Let $q_{k,i} = \mathbb{P} [\text{State } k \text{ at time } t_i]$.
3. Let $p_{jk;i} = \mathbb{P} [\text{State } k \text{ at time } t_{i+1} | \text{State } j \text{ at time } t_i]$.
4. Suppose everyone starts out in state 1 at time 0
5. Then $q_{j,1} = p_{1j,0}$.
6. Then $q_{j,2} = \sum_k p_{kj,1} q_{k,1} = \sum_k p_{kj,1} p_{1k,0}$.
7. This is the same mathematical operation as matrix multiplication.

1. Let $P_i$ be matrix with $p_{jk,i}$ in row $j$ column $k$
2. Let $q_i$ be the vector with $q_{j,i}$ in position $j$. 
Calculations for discrete times

1. Discrete times: transitions can only happen at times $t_1, \ldots, t_i, \ldots$
2. Let $q_{k,i} = P \{ \text{State } k \text{ at time } t_i \}$.
3. Let $p_{jk;i} = P \{ \text{State } k \text{ at time } t_{i+1} | \text{State } j \text{ at time } t_i \}$.
4. Suppose everyone starts out in state 1 at time 0
5. Then $q_{j,1} = p_{1j,0}$.
6. Then $q_{j,2} = \sum_k p_{kj,1} q_{k,1} = \sum_k p_{kj,1} p_{1k,0}$.
7. This is the same mathematical operation as matrix multiplication.

1. Let $P_i$ be matrix with $p_{jk,i}$ in row $j$ column $k$
2. Let $q_i$ be the vector with $q_{j,i}$ in position $j$.
3. Then $q_2 = P_1 q_1$. 
Mathematical Details

Calculations for discrete times

1. Discrete times: transitions can only happen at times $t_1, \ldots, t_i, \ldots$
2. Let $q_{k;i} = \Pr \{\text{State } k \text{ at time } t_i\}$.
3. Let $p_{jk;i} = \Pr \{\text{State } k \text{ at time } t_{i+1} | \text{State } j \text{ at time } t_i\}$.
4. Suppose everyone starts out in state 1 at time 0
5. Then $q_{j,1} = p_{1j,0}$.
6. Then $q_{j,2} = \sum_k p_{kj,1} q_{k,1} = \sum_k p_{kj,1} p_{1k,0}$.
7. This is the same mathematical operation as matrix multiplication.

- Let $P_i$ be matrix with $p_{jk,i}$ in row $j$ column $k$
- Let $q_i$ be the vector with $q_{j,i}$ in position $j$.
- Then $q_2 = P_1 q_1$.
- Similarly, $q_3 = P_2 P_1 q_1$. 
Mathematical Details

1. Calculations for discrete times
   1. Discrete times: transitions can only happen at times \( t_1, \ldots, t_i, \ldots \)
   2. Let \( q_{k;i} = P[\text{State } k \text{ at time } t_i] \).
   3. Let \( p_{jk;i} = P[\text{State } k \text{ at time } t_{i+1} | \text{State } j \text{ at time } t_i] \).
   4. Suppose everyone starts out in state 1 at time 0
   5. Then \( q_{j,1} = p_{1j,0} \).
   6. Then \( q_{j,2} = \sum_k p_{kj,1} q_{k,1} = \sum_k p_{kj,1} p_{1k,0} \).
   7. This is the same mathematical operation as matrix multiplication.
      1. Let \( P_i \) be matrix with \( p_{jk,i} \) in row \( j \) column \( k \)
      2. Let \( q_i \) be the vector with \( q_{j,i} \) in position \( j \).
      3. Then \( q_2 = P_1 q_1 \).
      4. Similarly, \( q_3 = P_2 P_1 q_1 \).
      5. Similarly, \( q_i = P_i \cdots P_2 P_1 q_1 \).

What are people working on now?: Multistage Modeling
Homogeneous case, plus some other conditions
Continued

1. Homogeneous case, plus some other conditions
   - $P_i$ does not depend on $i$

Other conditions requiring process not be periodic

Counter example: process where you deterministically swap between two stages.

Can prove $q_i$ converges as $i$ increases

If $P_i$ does not depend on $i$

$q$ satisfies $q = Pq$

$q$ is eigenvector, satisfies $\sum q_j = 1$.

$q$ is called a fixed point

If process cannot segregate itself into separate subprocesses, the limit does not depend on the initial state.

Tool is often used to probabilistically approximate things too hard to calculate deterministically.

Application to disease progression data

Processes are generally not homogeneous

Can be used for quickly-changing diseases: ex. flu.
Homogeneous case, plus some other conditions

1. \( P_i \) does not depend on \( i \)
2. Other conditions requiring process not be periodic
Homogeneous case, plus some other conditions

1. $P_i$ does not depend on $i$
2. Other conditions requiring process not be periodic

Counter example: process where you deterministically swap between two stages.

Can prove $q_i$ converges as $i$ increases

If $P_i$ does not depend on $i$

$q$ satisfies $q = Pq$

$q$ is eigenvector, satisfies $\sum q_j = 1$.

$q$ is called a fixed point

If process cannot segregate itself into separate subprocesses, the limit does not depend on the initial state.

Tool is often used to probabilistically approximate things too hard to calculate deterministically.

Application to disease progression data

Processes are generally not homogeneous

Can be used for quickly-changing diseases: ex. flu.
Continued

1 Homogeneous case, plus some other conditions
   1 $P_i$ does not depend on $i$
   2 Other conditions requiring process not be periodic
      1 Counter example: process where you deterministically swap between two stages.
   3 Can prove $q_i$ converges as $i$ increases
Continued

1. Homogeneous case, plus some other conditions
   1. $P_i$ does not depend on $i$
   2. Other conditions requiring process not be periodic
      1. Counter example: process where you deterministically swap between two stages.
   3. Can prove $q_i$ converges as $i$ increases
      1. If $P$ does not depend on $i$
1. Homogeneous case, plus some other conditions
   1. $P_i$ does not depend on $i$
   2. Other conditions requiring process not be periodic
      1. Counter example: process where you deterministically swap between two stages.
   3. Can prove $q_i$ converges as $i$ increases
      1. If $P$ does not depend on $i$
      2. $q$ satisfies $q = Pq$
Homogeneous case, plus some other conditions

1. \( P_i \) does not depend on \( i \)
2. Other conditions requiring process not be periodic
   - Counter example: process where you deterministically swap between two stages.
3. Can prove \( q_i \) converges as \( i \) increases
   - If \( P \) does not depend on \( i \)
   - \( q \) satisfies \( q = Pq \)
   - \( q \) is eigenvector, satisfies \( \sum q_j = 1 \).
Homogeneous case, plus some other conditions

1. $P_i$ does not depend on $i$
2. Other conditions requiring process not be periodic
   - Counter example: process where you deterministically swap between two stages.
3. Can prove $q_i$ converges as $i$ increases
   - If $P$ does not depend on $i$
   - $q$ satisfies $q = Pq$
   - $q$ is eigenvector, satisfies $\sum q_j = 1$.
   - $q$ is called a fixed point
Homogeneous case, plus some other conditions

1. $P_i$ does not depend on $i$
2. Other conditions requiring process not be periodic
   - Counter example: process where you deterministically swap between two stages.
3. Can prove $q_i$ converges as $i$ increases
   - If $P$ does not depend on $i$
   - $q$ satisfies $q = Pq$
   - $q$ is eigenvector, satisfies $\sum q_j = 1.$
4. $q$ is called a fixed point
5. If process cannot segregate itself into separate subprocesses, the limit does not depend on the initial state.
Continued

1. Homogeneous case, plus some other conditions
   1. $P_i$ does not depend on $i$
   2. Other conditions requiring process not be periodic
      1. Counter example: process where you deterministically swap between two stages.
   3. Can prove $q_i$ converges as $i$ increases
      1. If $P$ does not depend on $i$
      2. $q$ satisfies $q = Pq$
      3. $q$ is eigenvector, satisfies $\sum q_j = 1$.
      4. $q$ is called a fixed point
      5. If process cannot segregate itself into separate subprocesses, the limit does not depend on the initial state.
      6. Tool is often used to probabilistically approximate things too hard to calculate deterministically.
Homogeneous case, plus some other conditions

1. $P_i$ does not depend on $i$
2. Other conditions requiring process not be periodic
   1. Counter example: process where you deterministically swap between two stages.
3. Can prove $q_i$ converges as $i$ increases
   1. If $P$ does not depend on $i$
   2. $q$ satisfies $q = Pq$
   3. $q$ is eigenvector, satisfies $\sum q_j = 1$.
   4. $q$ is called a fixed point
   5. If process cannot segregate itself into separate subprocesses, the limit does not depend on the initial state.
   6. Tool is often used to probabilistically approximate things too hard to calculate deterministically.

4. Application to disease progression data
Continued

1 Homogeneous case, plus some other conditions
   1 $P_i$ does not depend on $i$
   2 Other conditions requiring process not be periodic
      1 Counter example: process where you deterministically swap between two stages.
   3 Can prove $q_i$ converges as $i$ increases
      1 If $P$ does not depend on $i$
      2 $q$ satisfies $q = Pq$
      3 $q$ is eigenvector, satisfies $\sum q_j = 1$.
      4 $q$ is called a fixed point
      5 If process cannot segregate itself into separate subprocesses, the limit does not depend on the initial state.
      6 Tool is often used to probabilistically approximate things too hard to calculate deterministically.

4 Application to disease progression data
   1 Processes are generally not homogeneous
Homogeneous case, plus some other conditions

- $P_i$ does not depend on $i$
- Other conditions requiring process not be periodic
  - Counter example: process where you deterministically swap between two stages.
- Can prove $q_i$ converges as $i$ increases
  - If $P$ does not depend on $i$
  - $q$ satisfies $q = Pq$
  - $q$ is eigenvector, satisfies $\sum q_j = 1$.
  - $q$ is called a fixed point
  - If process cannot segregate itself into separate subprocesses, the limit does not depend on the initial state.
  - Tool is often used to probabilistically approximate things too hard to calculate deterministically.

Application to disease progression data

- Processes are generally not homogeneous
- Can be used for quickly-changing diseases: ex. flu.
Move to continuous time:

1. Approximate on discrete time for times very close together
Move to continuous time:

1. Approximate on discrete time for times very close together
2. As time interval shrinks, for almost all times, the probability of staying in the same state must be close to 1.

\[ \int_{0}^{t} S(s) - h_k(s) \, ds = \int_{0}^{t} S(s) \, dH_k(s) \]

1. General version of Aalen-Johansen equation
2. Estimation via substitution of empirical estimates.
3. Curve is step function
4. Neither consistently increasing nor decreasing.

Limitations:
1. No adjustment for interval censoring.
2. Can't tolerate censoring in middle.
Move to continuous time:

1. Approximate on discrete time for times very close together.

2. As time interval shrinks, for almost all times, the probability of staying in the same state must be close to 1.

3. Conditioning argument gives probability of moving to state $k$ by time $t$ is

$$
\int_0^t S(s-)h_k(s) \, ds = \int_0^t S(s-)dH_k(s):
$$

General version of Aalen-Johansen equation.

4. Estimation via substitution of empirical estimates.

5. Curve is step function.

Neither consistently increasing nor decreasing. R Code

Limitations:

1. No adjustment for interval censoring.

2. Can’t tolerate censoring in middle.
Move to continuous time:

1. Approximate on discrete time for times very close together
2. As time interval shrinks, for almost all times, the probability of staying in the same state must be close to 1.
3. Conditioning argument gives probability of moving to state \( k \) by time \( t \) is
   \[
   \int_0^t S(s-)h_k(s) \, ds = \int_0^t S(s-)dH_k(s):
   \]
   1. general version of Aalen-Johansen equation

Estimation via substitution of empirical estimates.

Limitations:
1. No adjustment for interval censoring.
2. Can't tolerate censoring in middle.
Move to continuous time:

1. Approximate on discrete time for times very close together
2. As time interval shrinks, for almost all times, the probability of staying in the same state must be close to 1.
3. Conditioning argument gives probability of moving to state $k$ by time $t$ is $\int_0^t S(s-)h_k(s) \, ds = \int_0^t S(s-)dH_k(s)$:
   - general version of Aalen-Johansen equation
4. Estimation via substitution of empirical estimates.
Move to continuous time:

1. Approximate on discrete time for times very close together
2. As time interval shrinks, for almost all times, the probability of staying in the same state must be close to 1.
3. Conditioning argument gives probability of moving to state $k$ by time $t$ is $\int_0^t S(s-)h_k(s) \, ds = \int_0^t S(s-)dH_k(s)$:
   - general version of Aalen-Johansen equation
4. Estimation via substitution of empirical estimates.
   - Curve is step function
Move to continuous time:

1. Approximate on discrete time for times very close together
2. As time interval shrinks, for almost all times, the probability of staying in the same state must be close to 1.
3. Conditioning argument gives probability of moving to state $k$ by time $t$ is $\int_0^t S(s-) h_k(s) \, ds = \int_0^t S(s-) dH_k(s)$:
   - general version of Aalen-Johansen equation
4. Estimation via substitution of empirical estimates.
   - Curve is step function
   - Neither consistently increasing nor decreasing. R Code
Move to continuous time:

1. Approximate on discrete time for times very close together
2. As time interval shrinks, for almost all times, the probability of staying in the same state must be close to 1.
3. Conditioning argument gives probability of moving to state $k$ by time $t$ is $\int_0^t S(s-)h_k(s)\,ds = \int_0^t S(s-)dH_k(s)$:
   - General version of Aalen-Johansen equation
4. Estimation via substitution of empirical estimates.
   - Curve is step function
   - Neither consistently increasing nor decreasing. R Code
5. Limitations:
Move to continuous time:

1. Approximate on discrete time for times very close together.
2. As time interval shrinks, for almost all times, the probability of staying in the same state must be close to 1.
3. Conditioning argument gives probability of moving to state $k$ by time $t$ is
\[
\int_0^t S(s-)h_k(s) \, ds = \int_0^t S(s-)dH_k(s):
\]
   - General version of Aalen-Johansen equation.
4. Estimation via substitution of empirical estimates.
   - Curve is step function.
   - Neither consistently increasing nor decreasing. R Code.
5. Limitations:
   - No adjustment for interval censoring.
Move to continuous time:

1. Approximate on discrete time for times very close together
2. As time interval shrinks, for almost all times, the probability of staying in the same state must be close to 1.
3. Conditioning argument gives probability of moving to state $k$ by time $t$ is
   \[ \int_0^t S(s-)h_k(s) \, ds = \int_0^t S(s-)dH_k(s) : \]
   - general version of Aalen-Johansen equation
4. Estimation via substitution of empirical estimates.
   - Curve is step function
   - Neither consistently increasing nor decreasing.  
5. Limitations:
   - No adjustment for interval censoring.
   - Can’t tolerate censoring in middle.
Can build a Proportional Hazards regression model.

1. Model each observed transition separately.

Profile likelihood is still used to remove baseline hazard.

\[
L(\beta) = \prod_{i \in C} \exp(z_i \beta) \sum_{k \geq i} \exp(z_k \beta)
\]

Denominator must be expanded to include potential transitions for individuals at risk.

Some models feature coefficients shared across transitions. I can’t think of a plausible scenario in which you’d want to do this.

Some models ignore timing of transitions between subjects and consider only transitions within intervals.
Can build a Proportional Hazards regression model.

1. Model each observed transition separately.
2. Profile likelihood is still used to remove baseline hazard.
Can build a Proportional Hazards regression model.

1. Model each observed transition separately.
2. Profile likelihood is still used to remove baseline hazard.
   - Profile likelihood was $L(\beta) = \prod_{i \in C} \frac{\exp(z_i \beta)}{\sum_{k \geq i} \exp(z_k \beta)}$

Denominator must be expanded to include transitions potential transitions for individuals at risk.

Some models feature coefficients shared across transitions. I can't think of a plausible scenario in which you'd want to do this.

Some models ignore timing of transitions between subjects and consider only transitions within intervals.
Can build a Proportional Hazards regression model.

1. Model each observed transition separately.
2. Profile likelihood is still used to remove baseline hazard.
   - Profile likelihood was $L(\beta) = \prod_{i \in C} \frac{\exp(z_i \beta)}{\sum_{k \geq i} \exp(z_k \beta)}$
   - Denominator must be expanded to include transitions potential transitions for individuals at risk.
Can build a Proportional Hazards regression model.

1. Model each observed transition separately.
2. Profile likelihood is still used to remove baseline hazard.
   1. Profile likelihood was \( L(\beta) = \prod_{i \in C} \frac{\exp(z_i \beta)}{\sum_{k \geq i} \exp(z_k \beta)} \)
   2. Denominator must be expanded to include transitions potential transitions for individuals at risk.
Can build a Proportional Hazards regression model.

1. Model each observed transition separately.
2. Profile likelihood is still used to remove baseline hazard.
   1. Profile likelihood was \( L(\beta) = \prod_{i \in C} \frac{\exp(z_i \beta)}{\sum_{k \geq i} \exp(z_k \beta)} \)
   2. Denominator must be expanded to include transitions potential transitions for individuals at risk.
   1. I can’t think of a plausible scenario in which you’d want to do this.
Can build a Proportional Hazards regression model.

1. Model each observed transition separately.
2. Profile likelihood is still used to remove baseline hazard.
   - Profile likelihood was $L(\beta) = \prod_{i \in C} \frac{\exp(z_i \beta)}{\sum_{k \geq i} \exp(z_k \beta)}$
   - Denominator must be expanded to include transitions potential transitions for individuals at risk.
   - I can’t think of a plausible scenario in which you’d want to do this.
4. Some models ignore timing of transitions between subjects and consider only transitions within intervals.

R Code
Section: What are people working on now?
Subsection: Covariate Measurement error
Analogous with Standard Regression Models

1 Suppose one observes covariates with error

\[ Y_j = \alpha + \beta X_j + \epsilon_j \]

You observe \[ Z_j = X_j + \xi_j = \epsilon_j \]

\[ \epsilon_j \sim N(0, \sigma^2) \text{ i.i.d.} \]

\[ \xi_j \sim N(0, \tau^2) \text{ i.i.d.} \]
Analogous with Standard Regression Models

1. Suppose one observes covariates with error
2. \( Y_j = \alpha + \beta X_j + \epsilon_j \)
Analogous with Standard Regression Models

1. Suppose one observes covariates with error
2. $Y_j = \alpha + \beta X_j + \epsilon_j$
3. You observe $Z_j = X_j + \xi_j$
Suppose one observes covariates with error.

You observe $Z_j = X_j + \xi_j$

$\epsilon_j \sim N(0, \sigma^2)$ i.i.d.
Analogous with Standard Regression Models

1. Suppose one observes covariates with error
2. \( Y_j = \alpha + \beta X_j + \epsilon_j \)
3. You observe \( Z_j = X_j + \xi_j \)
4. \( \epsilon_j \sim N(0, \sigma^2) \) i.i.d.
5. \( \xi_j \sim N(0, \tau^2) \) i.i.d.
Then parameter estimates are systematically too small.

\[ \hat{\beta} = \frac{\sum_j (Z_j - \bar{Z}) Y_j}{\sum_j (Z_j - \bar{Z})^2} \]
Then parameter estimates are systematically too small.

\[ \tilde{\beta} = \frac{\sum_j (Z_j - \bar{Z}) Y_j}{\sum_j (Z_j - \bar{Z})^2} \]

\[ \mathbb{E} \left[ \tilde{\beta} \right] = \beta \frac{\sum_j (Z_j - \bar{Z}) X_j}{\sum_j (Z_j - \bar{Z})^2} \]
Then parameter estimates are systematically too small.

\[ \tilde{\beta} = \frac{\sum_j (Z_j - \bar{Z}) Y_j}{\sum_j (Z_j - \bar{Z})^2} \]

\[ \mathbb{E} \left[ \tilde{\beta} \right] = \frac{\beta \sum_j (Z_j - \bar{Z}) X_j}{\sum_j (Z_j - \bar{Z})^2} \]

\[ |\mathbb{E} \left[ \tilde{\beta} \right] | \epsilon \leq |\beta| \sqrt{\frac{\sum_j (X_j - \bar{X})^2}{\sum_j (Z_j - \bar{Z})^2}} \]
Then parameter estimates are systematically too small.

1. \[ \hat{\beta} = \frac{\sum_j (Z_j - \bar{Z}) Y_j}{\sum_j (Z_j - \bar{Z})^2} \]

2. \[ \mathbb{E} \left[ \hat{\beta} \right] = \beta \frac{\sum_j (Z_j - \bar{Z}) X_j}{\sum_j (Z_j - \bar{Z})^2} \]

3. \[ \left| \mathbb{E} \left[ \hat{\beta} \right] \right| \leq |\beta| \sqrt{\frac{\sum_j (X_j - \bar{X})^2}{\sum_j (Z_j - \bar{Z})^2}} \]

4. Bottom tends to be bigger than top
Then parameter estimates are systematically too small.

\[ \tilde{\beta} = \frac{\sum_j (Z_j - \bar{Z}) Y_j}{\sum_j (Z_j - \bar{Z})^2} \]

\[ \mathbb{E} [\tilde{\beta}] = \frac{\beta \sum_j (Z_j - \bar{Z}) X_j}{\sum_j (Z_j - \bar{Z})^2} \]

\[ |\mathbb{E} [\tilde{\beta}]| \epsilon \leq |\beta| \sqrt{\frac{\sum_j (X_j - \bar{X})^2}{\sum_j (Z_j - \bar{Z})^2}} \]

4. Bottom tends to be bigger than top

5. Expectation less than one
Section: What are people working on now?

Subsection: Bayesian methods
Recall the paradigm:

1. Want to make probabilistic statements about model parameters, conditional on data

\[ P[\theta | \text{data}] \]

\[ \propto P[\text{data} | \theta] \times \pi(\theta) \]

\( \pi(\theta) \) = prior density for \( \theta \).

Need to have prior concept of distribution of \( \theta \).

Techniques that will give you the probabilities of sets will also give the proportionality constant.

What are people working on now?: Bayesian methods
Recall the paradigm:

1. Want to make probabilistic statements about model parameters, conditional on data
   - That is, $P[\theta|\text{data}]$ for model parameter $\theta$?
Recall the paradigm:

1. Want to make probabilistic statements about model parameters, conditional on data
   - That is, $P[\theta|\text{data}]$ for model parameter $\theta$?
   - $\propto P[\text{data}|\theta] \varpi(\theta)$

What are people working on now?: Bayesian methods
Recall the paradigm:

1. Want to make probabilistic statements about model parameters, conditional on data
   a. That is, $P[\theta|\text{data}]$ for model parameter $\theta$?
   b. $\propto P[\text{data}|\theta] \varpi(\theta)$
   c. $\varpi(\theta) =$ prior density for $\theta$.
Recall the paradigm:

1. Want to make probabilistic statements about model parameters, conditional on data
   1. That is, $P[\theta|\text{data}]$ for model parameter $\theta$?
   2. $\propto P[\text{data}|\theta] \varpi(\theta)$
   3. $\varpi(\theta) = \text{prior density for } \theta$.
2. Need to have prior concept of distribution of $\theta$.

What are people working on now?: Bayesian methods
Recall the paradigm:

1. Want to make probabilistic statements about model parameters, conditional on data
   1. That is, \( P[\theta|\text{data}] \) for model parameter \( \theta \)?
   2. \( \propto P[\text{data}|\theta] \varpi(\theta) \)
      1. \( \varpi(\theta) = \) prior density for \( \theta \).
3. Need to have prior concept of distribution of \( \theta \).
4. Techniques that will give you the probabilities of sets will also give the proportionality constant.

What are people working on now?: Bayesian methods
For survival analysis regression,

1. **Accelerated failure model (fully parametric):**
   
   $\theta$ are linear parameters, intercept (location parameter for family on the log scale) and scale on the log scale.

   Analysis proceeds as usual.

2. **Proportional Hazards model (semi-parametric):**
   
   $\theta$ is linear parameters, and the baseline survival function. Hence it is infinite-dimensional.

   Prior on baseline is generally Dirichlet prior.

   Constructed by splitting range into many small bins.
   Putting a generalization of the beta distribution on each bin.
   Or by putting a spline on the curve and putting priors on spline coefficients.
For survival analysis regression,

1. Accelerated failure model (fully parametric):
   - \( \theta \) are linear parameters, intercept (location parameter for family on the log scale) and scale on the log scale.

2. Proportional Hazards model (semi-parametric):
   - \( \theta \) is linear parameters, and the baseline survival function.
   - Hence it is infinite-dimensional.
   - Prior on baseline is generally Dirichlet prior.

1. Constructed by splitting range into many small bins
2. Putting a generalization of the beta distribution on each bin
3. Or by putting a spline on the curve and putting priors on spline coefficients.

R Code

What are people working on now?: Bayesian methods
For survival analysis regression,

1. **Accelerated failure model (fully parametric):**
   1. $\theta$ are linear parameters, intercept (location parameter for family on the log scale) and scale on the log scale.
   2. Analysis proceeds as usual.
For survival analysis regression,

1. Accelerated failure model (fully parametric):
   1. $\theta$ are linear parameters, intercept (location parameter for family on the log scale) and scale on the log scale.
   2. Analysis proceeds as usual.

2. Proportional Hazards model (semi-parametric)
For survival analysis regression,

1. Accelerated failure model (fully parametric):
   1. \( \theta \) are linear parameters, intercept (location parameter for family on the log scale) and scale on the log scale.
   2. Analysis proceeds as usual.

2. Proportional Hazards model (semi-parametric)
   1. \( \theta \) is linear parameters, and the baseline survival function.
For survival analysis regression,

1. **Accelerated failure model (fully parametric):**
   - $\theta$ are linear parameters, intercept (location parameter for family on the log scale) and scale on the log scale.
   - Analysis proceeds as usual.

2. **Proportional Hazards model (semi-parametric)**
   - $\theta$ is linear parameters, and the baseline survival function.
   - Hence it is infinite-dimensional
For survival analysis regression,

1. Accelerated failure model (fully parametric):
   - $\theta$ are linear parameters, intercept (location parameter for family on the
     log scale) and scale on the log scale.
   - Analysis proceeds as usual.

2. Proportional Hazards model (semi-parametric)
   - $\theta$ is linear parameters, and the baseline survival function.
   - Hence it is infinite-dimensional
   - Prior on baseline is generally \textit{Dirichlet prior}.
For survival analysis regression,

1. **Accelerated failure model (fully parametric):**
   1. $\theta$ are linear parameters, intercept (location parameter for family on the log scale) and scale on the log scale.
   2. Analysis proceeds as usual.

2. **Proportional Hazards model (semi-parametric)**
   1. $\theta$ is linear parameters, and the baseline survival function.
   2. Hence it is infinite-dimensional
   3. Prior on baseline is generally *Dirichlet prior*.
      1. constructed by splitting range into many small bins
For survival analysis regression,

1. **Accelerated failure model (fully parametric):**
   - θ are linear parameters, intercept (location parameter for family on the log scale) and scale on the log scale.
   - Analysis proceeds as usual.

2. **Proportional Hazards model (semi-parametric)**
   - θ is linear parameters, and the baseline survival function.
   - Hence it is infinite-dimensional
   - Prior on baseline is generally *Dirichlet prior.*
     - constructed by splitting range into many small bins
     - Putting a generalization of the beta distribution on each bin
For survival analysis regression,

1. Accelerated failure model (fully parametric):
   1. $\theta$ are linear parameters, intercept (location parameter for family on the log scale) and scale on the log scale.
   2. Analysis proceeds as usual.

2. Proportional Hazards model (semi-parametric)
   1. $\theta$ is linear parameters, and the baseline survival function.
   2. Hence it is infinite-dimensional
   3. Prior on baseline is generally Dirichlet prior.

   1. constructed by splitting range into many small bins
   2. Putting a generalization of the beta distribution on each bin
   3. Or by putting a spline on the curve and putting priors on spline coefficients. R Code
Objectives Lecture 13

1. Bayesian Methods

What are people working on now?: Bayesian methods
Objectives Lecture 13

1. Bayesian Methods
2. Review

Readings:
I'll be available during course time during the reading day (next Tues) to talk about HW, last year's final.

Please fill out course evaluations.
Objectives Lecture 13

1. Bayesian Methods
2. Review
3. Readings:

I'll be available during course time during the reading day (next Tues) to talk about HW, last year's final.

Please fill out course evaluations.
Objectives Lecture 13

1. Bayesian Methods
2. Review
3. Readings:
4. I’ll be available during course time during the reading day (next Tues) to talk about HW, last year’s final.
Objectives Lecture 13

1. Bayesian Methods
2. Review
3. Readings:
4. I’ll be available during course time during the reading day (next Tues) to talk about HW, last year’s final.
5. Please fill out course evaluations.
Section: What are people working on now?

Subsection: Bayesian Statistics
Frequentist statistics was marked by “proof by contradiction”:

1. Most notably with the $p$-value:
Frequentist statistics was marked by “proof by contradiction”:

Most notably with the \( p \)-value:

- Answers question “How often would I see evidence this extreme or more extreme against \( H_0 \), if \( H_0 \) were true?”
Frequentist statistics was marked by “proof by contradiction”:

1. Most notably with the $p$-value:
   1. Answers question “How often would I see evidence this extreme or more extreme against $H_0$, if $H_0$ were true?”
   2. Recall example that likelihood can change very little even as the $p$-value changes dramatically.
Frequentist statistics was marked by “proof by contradiction”:

1. Most notably with the $p$-value:
   1. Answers question “How often would I see evidence this extreme or more extreme against $H_0$, if $H_0$ were true?”
   2. Recall example that likelihood can change very little even as the $p$-value changes dramatically.
2. No information is used about prior beliefs about hypotheses.
Frequentist statistics was marked by “proof by contradiction”:

1. Most notably with the $p$-value:
   1. Answers question “How often would I see evidence this extreme or more extreme against $H_0$, if $H_0$ were true?”
   2. Recall example that likelihood can change very little even as the $p$-value changes dramatically.

2. No information is used about prior beliefs about hypotheses
   1. including information one might have from other studies.
What people intuitively want out of statistical procedures.

Those not indoctrinated into frequentism want to ask “\( P[H_0 \text{ true}] = P[\theta \in \Omega_0] \)”, “\( P[H_A \text{ true}] = P[\theta \in \Omega_A] \)”. 

Probabilities should be conditional on data (denoted by \( X \)).

We really want \( P[H_0 | X = x] \).

In order to do this for every null hypothesis, we need distribution of \( \theta \) conditional on data.

We also have model for \( X \) dependent on \( \theta \).

Now write \( f(x | \theta) \) where as before I wrote \( f(x; \theta) \). Think of it as distribution of data conditional on the parameter.

Math can show that the combination of the distribution in both directions generally define \( P[H_0] \) if the question has a solution.

Hence you cannot do this analysis without specification of \( P[H_0] \). If you want to ask about all possible null hypothesis, you need distribution of \( \theta \). Distribution is called a prior, since it represents ideas about the parameter before seeing data.
What people intuitively want out of statistical procedures.

Those not indoctrinated into frequentism want to ask \( P[H_0\ \text{true}] = P[\theta \in \Omega_0] \), \( P[H_A\ \text{true}] = P[\theta \in \Omega_A] \).

Probabilities should be conditional on data (denoted by \( X \)).
What people intuitively want out of statistical procedures.

1. Those not indoctrinated into frequentism want to ask “\( P[H_0 \text{ true}] = P[\theta \in \Omega_0] \)”, “\( P[H_A \text{ true}] = P[\theta \in \Omega_A] \)”.
2. Probabilities should be conditional on data (denoted by \( X \)).
3. We really want \( P[H_0 | X = x] \).

Math can show that the combination of the distribution in both directions generally define \( P[H_0] \) if the question has a solution. Hence you cannot do this analysis without specification of \( P[H_0] \). If you want to ask about all possible null hypothesis, you need distribution of \( \theta \). Distribution is called a prior, since it represents ideas about the parameter before seeing data.
What people intuitively want out of statistical procedures.

1. Those not indoctrinated into frequentism want to ask “\( P [H_0 \text{ true}] = P [\theta \in \Omega_0] \)”, “\( P [H_A \text{ true}] = P [\theta \in \Omega_A] \)”.

2. Probabilities should be conditional on data (denoted by \( X \)).

3. We really want \( P [H_0 | X = x] \).

4. In order to do this for every null hypothesis, we need distribution of \( \theta \) conditional on data.
What people intuitively want out of statistical procedures.

1. Those not indoctrinated into frequentism want to ask “\( P[H_0 \text{ true}] = P[\theta \in \Omega_0] \)”, “\( P[H_A \text{ true}] = P[\theta \in \Omega_A] \)”.

2. Probabilities should be conditional on data (denoted by \( X \)).

3. We really want \( P[H_0|X = x] \).

4. In order to do this for every null hypothesis, we need distribution of \( \theta \) conditional on data.

5. We also have model for \( X \) dependent on \( \theta \).
What people intuitively want out of statistical procedures.

1. Those not indoctrinated into frequentism want to ask \( P[H_0 \text{ true}] = P[\theta \in \Omega_0] \), “\( P[H_A \text{ true}] = P[\theta \in \Omega_A] \)”.

2. Probabilities should be conditional on data (denoted by \( X \)).

3. We really want \( P[H_0|X = x] \).

4. In order to do this for every null hypothesis, we need distribution of \( \theta \) conditional on data.

5. We also have model for \( X \) dependent on \( \theta \).

6. Now write \( f(x|\theta) \) where as before I wrote \( f(x; \theta) \).
What people intuitively want out of statistical procedures.

1. Those not indoctrinated into frequentism want to ask “$P[H_0 \text{ true}] = P[\theta \in \Omega_0]$, “$P[H_A \text{ true}] = P[\theta \in \Omega_A]$”.

2. Probabilities should be conditional on data (denoted by $X$).

3. We really want $P[H_0|X = x]$.

4. In order to do this for every null hypothesis, we need distribution of $\theta$ conditional on data.

5. We also have model for $X$ dependent on $\theta$.

6. Now write $f(x|\theta)$ where as before I wrote $f(x; \theta)$.

   1. Think of it as distribution of data conditional on the parameter.
What people intuitively want out of statistical procedures.

1. Those not indoctrinated into frequentism want to ask “\( P[H_0 \text{ true}] = P[\theta \in \Omega_0] \)””, “\( P[H_A \text{ true}] = P[\theta \in \Omega_A] \)”.

2. Probabilities should be conditional on data (denoted by \( X \)).

3. We really want \( P[H_0|X=x] \).

4. In order to do this for every null hypothesis, we need distribution of \( \theta \) conditional on data

5. We also have model for \( X \) dependent on \( \theta \).

6. Now write \( f(x|\theta) \) where as before I wrote \( f(x;\theta) \).

   1. Think of it as distribution of data conditional on the parameter.

7. Math can show that the combination of the distribution in both directions generally define \( P[H_0] \)
What people intuitively want out of statistical procedures.

Those not indoctrinated into frequentism want to ask “P[H₀ true] = P[θ ∈ Ω₀]”, “P[Hₐ true] = P[θ ∈ Ωₐ]”.

Probabilities should be conditional on data (denoted by X).

We really want P[H₀|X = x].

In order to do this for every null hypothesis, we need distribution of θ conditional on data.

We also have model for X dependent on θ.

Now write f(x|θ) where as before I wrote f(x;θ).

Think of it as distribution of data conditional on the parameter.

Math can show that the combination of the distribution in both directions generally define P[H₀]

If the question has a solution.
What people intuitively want out of statistical procedures.

1. Those not indoctrinated into frequentism want to ask \( \Pr [H_0 \text{ true}] = \Pr [\theta \in \Omega_0] \), \( \Pr [H_A \text{ true}] = \Pr [\theta \in \Omega_A] \).

2. Probabilities should be conditional on data (denoted by \( X \)).

3. We really want \( \Pr [H_0 | X = x] \).

4. In order to do this for every null hypothesis, we need distribution of \( \theta \) conditional on data.

5. We also have model for \( X \) dependent on \( \theta \).

6. Now write \( f(x|\theta) \) where as before I wrote \( f(x; \theta) \).
   1. Think of it as distribution of data conditional on the parameter.

7. Math can show that the combination of the distribution in both directions generally define \( \Pr [H_0] \)
   1. If the question has a solution.

8. Hence you cannot do this analysis without specification of \( \Pr [H_0] \).
What people intuitively want out of statistical procedures.

1. Those not indoctrinated into frequentism want to ask \( \Pr[H_0 \text{ true}] = \Pr[\theta \in \Omega_0] \), \( \Pr[H_A \text{ true}] = \Pr[\theta \in \Omega_A] \).

2. Probabilities should be conditional on data (denoted by \( X \)).

3. We really want \( \Pr[H_0 | X = x] \).

4. In order to do this for every null hypothesis, we need distribution of \( \theta \) conditional on data.

5. We also have model for \( X \) dependent on \( \theta \).

6. Now write \( f(x | \theta) \) where as before I wrote \( f(x; \theta) \).
   
   Think of it as distribution of data conditional on the parameter.

7. Math can show that the combination of the distribution in both directions generally define \( \Pr[H_0] \).
   
   If the question has a solution.

8. Hence you cannot do this analysis without specification of \( \Pr[H_0] \).
   
   If you want to ask about all possible null hypothesis, you need distribution of \( \theta \).
What people intuitively want out of statistical procedures.

1. Those not indoctrinated into frequentism want to ask \( P[H_0 \text{ true}] = P[\theta \in \Omega_0] \), \( P[H_A \text{ true}] = P[\theta \in \Omega_A] \).

2. Probabilities should be conditional on data (denoted by \( X \)).

3. We really want \( P[H_0|X = x] \).

4. In order to do this for every null hypothesis, we need distribution of \( \theta \) conditional on data.

5. We also have model for \( X \) dependent on \( \theta \).

6. Now write \( f(x|\theta) \) where as before I wrote \( f(x; \theta) \).
   1. Think of it as distribution of data conditional on the parameter.

7. Math can show that the combination of the distribution in both directions generally define \( P[H_0] \)
   1. If the question has a solution.

8. Hence you cannot do this analysis without specification of \( P[H_0] \).
   1. If you want to ask about all possible null hypothesis, you need distribution of \( \theta \).
   2. Distribution is called a prior, since it represents ideas about the parameter before seeing data.
General conditions:

1. $\theta \in \Theta$, for $\Theta$ an interval.
General conditions:

1. \( \theta \in \Theta \), for \( \Theta \) an interval.

2. Distribution is almost always continuous, and can be given by density \( \varpi_\theta(\theta) \).

What are people working on now?: Bayesian Statistics Lecture 13
General conditions:

1. $\theta \in \Theta$, for $\Theta$ an interval.
2. Distribution is almost always continuous, and can be given by density $\omega_\theta(\theta)$.
   1. Sometimes this distribution is determined by parameters.
General conditions:

1. \( \theta \in \Theta \), for \( \Theta \) an interval.

2. Distribution is almost always continuous, and can be given by density \( \varpi_\theta(\theta) \).
   
   - Sometimes this distribution is determined by parameters.
   - These should be known by the analyst.
General conditions:

1. $\theta \in \Theta$, for $\Theta$ an interval.
2. Distribution is almost always continuous, and can be given by density $\varpi_\theta(\theta)$.
   
   1. Sometimes this distribution is determined by parameters.
   2. These should be known by the analyst.
   3. These are termed hyperparameters.
Joint density for $\theta$ and $\mathbf{X}$ is $f_{\theta,\mathbf{X}}(\theta, \mathbf{x}) = \varpi_\theta(\theta) f_{\mathbf{X}|\theta}(\mathbf{x}|\theta)$.

1. $f_{\mathbf{X}|\theta}(\mathbf{x}|\theta)$ is likelihood $L(\theta)$. 
Density for $\theta|X$ is joint density divided by marginal density.

1. Density for $\theta|X$ is

$$f_{\theta|X}(\theta|x) = \frac{f_{\theta,X}(\theta, x)}{f_X(x)}$$

$$= \frac{f_{\theta,X}(\theta, x)}{\int_{\Theta} f_{\theta,X}(\theta, x) \, d\theta}$$

$$= \frac{\varpi_{\theta}(\theta) f_{X|\theta}(x, \theta)}{\int_{\Theta} \varpi_{\theta}(\theta) f_{X|\theta}(x|\theta) \, d\theta},$$

by Bayes theorem.
Density for $\theta|\mathbf{X}$ is joint density divided by marginal density.

1. Density for $\theta|\mathbf{X}$ is

$$f_{\theta|\mathbf{x}}(\theta|x) = \frac{f_{\theta,\mathbf{x}}(\theta|\mathbf{x})}{f_{\mathbf{x}}(\mathbf{x})}$$

$$= f_{\theta,\mathbf{x}}(\theta|\mathbf{x})/\int_{\Theta} f_{\theta,\mathbf{x}}(\theta, \mathbf{x}) \, d\theta$$

$$= \varpi_{\theta}(\theta)f_{\mathbf{x}|\theta}(\mathbf{x}, \theta)/\int_{\Theta} \varpi_{\theta}(\theta)f_{\mathbf{x}|\theta}(\mathbf{x}|\theta) \, d\theta,$$

by Bayes theorem.

2. Same formula for discrete $\mathbf{X}$, this time with mass function.
Density for $\theta|X$ is joint density divided by marginal density.

1. Density for $\theta|X$ is

\[
f_{\theta|X}(\theta|x) = \frac{f_{\theta,X}(\theta,x)}{f_X(x)}
\]

\[
= \frac{f_{\theta,X}(\theta,x)}{\int_{\Theta} f_{\theta,X}(\theta,x) \, d\theta}
\]

\[
= \omega_{\theta}(\theta) f_{X|\theta}(x,\theta|)/\int_{\Theta} \omega_{\theta}(\theta) f_{X|\theta}(x|\theta) \, d\theta,
\]

by Bayes theorem.

2. Same formula for discrete $X$, this time with mass function.

3. Result is called *posterior distribution*. 

What are people working on now?: Bayesian Statistics Lecture 13
Difficulties with Bayesian analyses:

1. Computation of denominator:

   - Families of distributions constructed so that the likelihood times the prior was in the same family as the prior.
   - Integral done by integrating the resulting other member of the family.
   - Not generally useful for survival models.

2. Numerical Integration:
   - Evaluate $L(\theta) \rho(\theta)$ on a grid of points, separated by $\Delta$:
     \[ \zeta_j = L(\theta_0 + (j-1)\Delta) \rho(\theta_0 + (j-1)\Delta) \text{ for } j = \{1, \ldots, m\}. \]
   - Approximation to integral is a linear combination of these evaluations.
   - Trapezoidal rule:
     \[ \int_\Theta L(\theta) \rho_i(\theta) d\theta \approx \Delta (\omega_1 + 2\sum_{j=1}^{m-1} \omega_j + \omega_m) / 2. \]
   - Simpson's rule: if $m$ odd,
     \[ \int_\Theta L(\theta) \rho_i(\theta) d\theta \approx \Delta (\omega_1 + 4\sum_{j=1, j \text{ odd}}^{m-1} \omega_j + 2\sum_{j=1, j \text{ even}}^{m-1} \omega_j + \omega_m) / 3. \]

3. Laplace's method
   - We want
     \[ \int A \exp(\ell(\theta)) \rho(\theta) d\theta \]
   - Let $\hat{\theta}$ be the MLE.
   - Do Taylor series approximation for log likelihood and prior separately.
   - Extends to higher-order approximations.

Integration scales poorly as dimension of $\theta$ increases

Deterministic integration is replaced by simulation.
Difficulties with Bayesian analyses:

1. Computation of denominator:
   - The two examples above illustrate *conjugate priors*:

2. Numerical Integration:
   - Evaluate $L(\theta) \varpi_{\theta}(\theta)$ on a grid of points, separated by $\Delta$:
     
     $\zeta_j = L(\theta_0 + (j - 1)\Delta) \varpi_{\theta}(\theta_0 + (j - 1)\Delta)$ for $j = \{1, \ldots, m\}$.

   - Approximation to integral is a linear combination of these evaluations.

3. Trapezoidal rule:
   
   $\int_{\Theta} L(\theta) \varpi_i(\theta) d\theta \approx \Delta (\omega_1 + 2 \sum_{j=1}^{m-1} \omega_j + \omega_m) / 2$.

4. Simpson's rule:
   - If $m$ odd,
     
     $\int_{\Theta} L(\theta) \varpi_i(\theta) d\theta \approx \Delta (\omega_1 + 4 \sum_{j=1, j \text{ odd}}^{m-1} \omega_j + 2 \sum_{j=1, j \text{ even}}^{m-1} \omega_j + \omega_m) / 3$.

5. Laplace's method:
   - We want $\int_{A} \exp(\ell(\theta)) \varpi(\theta) d\theta$.
   - Let $\hat{\theta}$ be the MLE.
   - Do Taylor series approximation for log likelihood and prior separately.
   - Extends to higher-order approximations.

6. Integration scales poorly as dimension of $\theta$ increases.

Determination integration is replaced by simulation.
Difficulties with Bayesian analyses:

1. Computation of denominator:
   1. The two examples above illustrate *conjugate priors*:
      1. Families of distributions constructed so that the likelihood times the prior was in the same family as the prior.
Difficulties with Bayesian analyses:

1. Computation of denominator:
   1. The two examples above illustrate *conjugate priors*:
      1. Families of distributions constructed so that the likelihood times the prior was in the same family as the prior.
      2. Integral done by integrating the resulting other member of the family.
Difficulties with Bayesian analyses:

1. Computation of denominator:
   1. The two examples above illustrate *conjugate priors*:
      1. Families of distributions constructed so that the likelihood times the prior was in the same family as the prior.
      2. Integral done by integrating the resulting other member of the family.
      3. Not generally useful for survival models.

2. **Numerical Integration:**
   1. Evaluate $L(\theta) \varpi_\theta(\theta)$ on a grid of points, separated by $\Delta$:
      $$\zeta_j = L(\theta_0 + (j - 1)\Delta) \varpi_\theta(\theta_0 + (j - 1)\Delta) \quad \text{for} \quad j = \{1, \ldots, m\}.$$
   2. Approximation to integral is a linear combination of these evaluations.

3. **Trapezoidal rule:**
   $$\int_\Theta L(\theta) \varpi_i,\theta(\theta) \, d\theta \approx \Delta(\omega_1 + 2\sum_{j=1}^{m-1} \omega_j + \omega_m) / 2.$$

4. **Simpson's rule:** If $m$ odd,
   $$\int_\Theta L(\theta) \varpi_i,\theta(\theta) \, d\theta \approx \Delta(\omega_1 + 4\sum_{j=1, j \text{ odd}}^{m-1} \omega_j + 2\sum_{j=1, j \text{ even}}^{m-1} \omega_j + \omega_m) / 3.$$

5. **Laplace's method**
   1. We want
      $$\int_A \exp(\ell(\theta)) \varpi(\theta) \, d\theta.$$
   2. Let $\hat{\theta}$ be the MLE.
   3. Do Taylor series approximation for log likelihood and prior separately.
   4. Extends to higher-order approximations.

Integration scales poorly as dimension of $\theta$ increases.

Deterministic integration is replaced by simulation.
Difficulties with Bayesian analyses:

1. Computation of denominator:
   - The two examples above illustrate *conjugate priors*:
     1. Families of distributions constructed so that the likelihood times the prior was in the same family as the prior.
     2. Integral done by integrating the resulting other member of the family.
     3. Not generally useful for survival models.

2. Numerical Integration:
   - Evaluate \( L(\theta) \varpi(\theta) \) on a grid of points, separated by \( \Delta \):
     \[
     \zeta_j = L(\theta_0 + (j-1)\Delta) \varpi(\theta_0 + (j-1)\Delta)
     \]
     for \( j = \{1, \ldots, m\} \).
   - Approximation to integral is a linear combination of these evaluations.
   - Trapezoidal rule:
     \[
     \int_\Theta L(\theta) \varpi(\theta) d\theta \approx \Delta \left( \omega_1 + 2 \sum_{j=1}^{m-1} \omega_j + \omega_m \right) / 2.
     \]
   - Simpson's rule: if \( m \) odd,
     \[
     \int_\Theta L(\theta) \varpi(\theta) d\theta \approx \Delta \left( \omega_1 + 4 \sum_{j=1, j\text{odd}}^{m-1} \omega_j + 2 \sum_{j=1, j\text{even}}^{m-1} \omega_j + \omega_m \right) / 3.
     \]
   - Laplace's method
     - We want \( \int_A \exp(\ell(\theta)) \varpi(\theta) d\theta \)
     - Let \( \hat{\theta} \) be the MLE.
     - Do Taylor series approximation for log likelihood and prior separately.
     - Extends to higher-order approximations.

3. Integration scales poorly as dimension of \( \theta \) increases
   - Deterministic integration is replaced by simulation.
Difficulties with Bayesian analyses:

1. Computation of denominator:
   - The two examples above illustrate *conjugate priors*:
     - Families of distributions constructed so that the likelihood times the prior was in the same family as the prior.
     - Integral done by integrating the resulting other member of the family.
     - Not generally useful for survival models.

2. Numerical Integration:
   - Evaluate $L(\theta)\varpi_\theta(\theta)$ on a grid of points, separated by $\Delta$: $\zeta_j = L(\theta_0 + (j - 1)\Delta)\varpi_\theta(\theta_0 + (j - 1)\Delta)$ for $j = \{1, \ldots, m\}$.
Difficulties with Bayesian analyses:

1. Computation of denominator:
   - The two examples above illustrate *conjugate priors*:
     - Families of distributions constructed so that the likelihood times the prior was in the same family as the prior.
     - Integral done by integrating the resulting other member of the family.
     - Not generally useful for survival models.

2. Numerical Integration:
   - Evaluate $L(\theta)\varpi_\theta(\theta)$ on a grid of points, separated by $\Delta$: $\zeta_j = L(\theta_0 + (j - 1)\Delta)\varpi_\theta(\theta_0 + (j - 1)\Delta)$ for $j = \{1, \ldots, m\}$.
   - Approximation to integral is a linear combination of these evaluations
Difficulties with Bayesian analyses:

1. Computation of denominator:
   1. The two examples above illustrate conjugate priors:
      1. Families of distributions constructed so that the likelihood times the prior was in the same family as the prior.
      2. Integral done by integrating the resulting other member of the family.
      3. Not generally useful for survival models.

2. Numerical Integration:
   1. Evaluate $L(\theta) \varpi_{\theta}(\theta)$ on a grid of points, separated by $\Delta$: $\zeta_j = L(\theta_0 + (j - 1)\Delta) \varpi_{\theta}(\theta_0 + (j - 1)\Delta)$ for $j = \{1, \ldots, m\}$.
   2. Approximation to integral is a linear combination of these evaluations.
   3. Trapezoidal rule: $\int_{\Theta} L(\theta) \varpi_{i,\theta}(\theta) \, d\theta \approx \Delta(\omega_1 + 2 \sum_{j=1}^{m-1} \omega_j + \omega_m)/2$. 

What are people working on now?: Bayesian Statistics Lecture 13 257 / 260
Difficulties with Bayesian analyses:

1. Computation of denominator:
   1. The two examples above illustrate *conjugate priors*:
      - Families of distributions constructed so that the likelihood times the prior was in the same family as the prior.
      - Integral done by integrating the resulting other member of the family.
      - Not generally useful for survival models.

2. Numerical Integration:
   1. Evaluate $L(\theta)\varpi_{\theta}(\theta)$ on a grid of points, separated by $\Delta$: $\zeta_j = L(\theta_0 + (j - 1)\Delta)\varpi_{\theta}(\theta_0 + (j - 1)\Delta)$ for $j = \{1, \ldots, m\}$.
   2. Approximation to integral is a linear combination of these evaluations.
   3. Trapezoidal rule: $\int_{\Theta} L(\theta)\varpi_{i,\theta}(\theta) \, d\theta \approx \Delta(\omega_1 + 2\sum_{j=1}^{m-1} \omega_j + \omega_m)/2$.
   4. Simpson’s rule: if $m$ odd, $\int_{\Theta} L(\theta)\varpi_{i,\theta}(\theta) \, d\theta \approx \Delta(\omega_1 + 4\sum_{j=1,j\text{odd}}^{m-1} \omega_j + 2\sum_{j=1,j\text{even}}^{m-1} \omega_j + \omega_m)/3$. 

---

What are people working on now?:  Bayesian Statistics Lecture 13  257 / 260
Difficulties with Bayesian analyses:

1. Computation of denominator:
   1. The two examples above illustrate *conjugate priors*:
      1. Families of distributions constructed so that the likelihood times the prior was in the same family as the prior.
      2. Integral done by integrating the resulting other member of the family.
      3. Not generally useful for survival models.

2. Numerical Integration:
   1. Evaluate $L(\theta) \varpi_\theta(\theta)$ on a grid of points, separated by $\Delta$: $\zeta_j = L(\theta_0 + (j - 1)\Delta) \varpi_\theta(\theta_0 + (j - 1)\Delta)$ for $j = \{1, \ldots, m\}$.
   2. Approximation to integral is a linear combination of these evaluations
   3. Trapezoidal rule: $\int_\Theta L(\theta) \varpi_i,\theta(\theta) \, d\theta \approx \Delta (\omega_1 + 2 \sum_{j=1}^{m-1} \omega_j + \omega_m)/2$.
   4. Simpson’s rule: if $m$ odd, $\int_\Theta L(\theta) \varpi_i,\theta(\theta) \, d\theta \approx \Delta (\omega_1 + 4 \sum_{j=1, j\text{odd}}^{m-1} \omega_j + 2 \sum_{j=1, j\text{even}}^{m-1} \omega_j + \omega_m)/3$.

3. Laplace’s method
Difficulties with Bayesian analyses:

1. Computation of denominator:
   1. The two examples above illustrate *conjugate priors*:
      1. Families of distributions constructed so that the likelihood times the prior was in the same family as the prior.
      2. Integral done by integrating the resulting other member of the family.
      3. Not generally useful for survival models.

2. Numerical Integration:
   1. Evaluate $L(\theta)\varpi_{\theta}(\theta)$ on a grid of points, separated by $\Delta$: $\zeta_j = L(\theta_0 + (j - 1)\Delta)\varpi_{\theta}(\theta_0 + (j - 1)\Delta)$ for $j = \{1, \ldots, m\}$.
   2. Approximation to integral is a linear combination of these evaluations
   3. Trapezoidal rule: $\int_{\Theta} L(\theta)\varpi_i,\theta(\theta) \, d\theta \approx \Delta (\omega_1 + 2 \sum_{j=1}^{m-1} \omega_j + \omega_m)/2$.
   4. Simpson’s rule: if $m$ odd, $\int_{\Theta} L(\theta)\varpi_i,\theta(\theta) \, d\theta \approx \Delta (\omega_1 + 4 \sum_{j=1, j \text{ odd}}^{m-1} \omega_j + 2 \sum_{j=1, j \text{ even}}^{m-1} \omega_j + \omega_m)/3$.

3. Laplace’s method
   1. We want $\int_{A} \exp(\ell(\theta))\varpi(\theta) \, d\theta$
Difficulties with Bayesian analyses:

1. Computation of denominator:
   - The two examples above illustrate *conjugate priors*:
     1. Families of distributions constructed so that the likelihood times the prior was in the same family as the prior.
     2. Integral done by integrating the resulting other member of the family.
     3. Not generally useful for survival models.

2. Numerical Integration:
   - Evaluate $L(\theta) \varpi_\theta(\theta)$ on a grid of points, separated by $\Delta$: $\zeta_j = L(\theta_0 + (j - 1)\Delta) \varpi_\theta(\theta_0 + (j - 1)\Delta)$ for $j = \{1, \ldots, m\}$.
   - Approximation to integral is a linear combination of these evaluations
   - Trapezoidal rule: $\int_{\Theta} L(\theta) \varpi_{i,\theta}(\theta) \, d\theta \approx \Delta (\omega_1 + 2 \sum_{j=1}^{m-1} \omega_j + \omega_m)/2$.
   - Simpson’s rule: if $m$ odd, $\int_{\Theta} L(\theta) \varpi_{i,\theta}(\theta) \, d\theta \approx \Delta (\omega_1 + 4 \sum_{j=1,j\text{odd}}^{m-1} \omega_j + 2 \sum_{j=1,j\text{even}}^{m-1} \omega_j + \omega_m)/3$.

3. Laplace’s method
   - We want $\int_A \exp(\ell(\theta)) \varpi(\theta) \, d\theta$
   - Let $\hat{\theta}$ be the MLE.
Difficulties with Bayesian analyses:

1. Computation of denominator:
   - The two examples above illustrate *conjugate priors*:
     - Families of distributions constructed so that the likelihood times the prior was in the same family as the prior.
     - Integral done by integrating the resulting other member of the family.
     - Not generally useful for survival models.

2. Numerical Integration:
   - Evaluate $L(\theta) \varpi_\theta(\theta)$ on a grid of points, separated by $\Delta$: $\zeta_j = L(\theta_0 + (j - 1)\Delta) \varpi_\theta(\theta_0 + (j - 1)\Delta)$ for $j = \{1, \ldots, m\}$.
   - Approximation to integral is a linear combination of these evaluations.
   - Trapezoidal rule: $\int_\Theta L(\theta) \varpi_i,\theta(\theta) \, d\theta \approx \Delta (\omega_1 + 2 \sum_{j=1}^{m-1} \omega_j + \omega_m)/2$.
   - Simpson’s rule: if $m$ odd, $\int_\Theta L(\theta) \varpi_i,\theta(\theta) \, d\theta \approx \Delta (\omega_1 + 4 \sum_{j=1,j\text{odd}}^{m-1} \omega_j + 2 \sum_{j=1,j\text{even}}^{m-1} \omega_j + \omega_m)/3$.

3. Laplace’s method
   - We want $\int_A \exp(\ell(\theta)) \varpi(\theta) \, d\theta$
   - Let $\hat{\theta}$ be the MLE.
   - Do Taylor series approximation for log likelihood and prior separately.

4. Integration scales poorly as dimension of $\theta$ increases.
   - Deterministic integration is replaced by simulation.
Difficulties with Bayesian analyses:

1. Computation of denominator:
   - The two examples above illustrate *conjugate priors*:
     - Families of distributions constructed so that the likelihood times the prior was in the same family as the prior.
     - Integral done by integrating the resulting other member of the family.
     - Not generally useful for survival models.

2. Numerical Integration:
   - Evaluate \( L(\theta) \varpi_\theta(\theta) \) on a grid of points, separated by \( \Delta \):
     \[
     \zeta_j = L(\theta_0 + (j - 1)\Delta) \varpi_\theta(\theta_0 + (j - 1)\Delta)
     \]
   - for \( j = \{1, \ldots, m\} \).
   - Approximation to integral is a linear combination of these evaluations.
   - Trapezoidal rule: \[
     \int_\Theta L(\theta) \varpi_{i,\theta}(\theta) \, d\theta \approx \Delta (\omega_1 + 2 \sum_{j=1}^{m-1} \omega_j + \omega_m)/2.
     \]
   - Simpson’s rule: if \( m \) odd, \[
     \int_\Theta L(\theta) \varpi_{i,\theta}(\theta) \, d\theta \approx \Delta (\omega_1 + 4 \sum_{j=1,j\text{odd}}^{m-1} \omega_j + 2 \sum_{j=1,j\text{even}}^{m-1} \omega_j + \omega_m)/3.
     \]

3. Laplace’s method
   - We want \( \int_A \exp(\ell(\theta)) \varpi(\theta) \, d\theta \)
   - Let \( \hat{\theta} \) be the MLE.
   - Do Taylor series approximation for log likelihood and prior separately.
   - Extends to higher-order approximations.
Difficulties with Bayesian analyses:

1. Computation of denominator:
   1. The two examples above illustrate *conjugate priors*:
      1. Families of distributions constructed so that the likelihood times the prior was in the same family as the prior.
      2. Integral done by integrating the resulting other member of the family.
      3. Not generally useful for survival models.

2. Numerical Integration:
   1. Evaluate $L(\theta) \varpi_{\theta}(\theta)$ on a grid of points, separated by $\Delta$: $\zeta_j = L(\theta_0 + (j - 1)\Delta) \varpi_{\theta}(\theta_0 + (j - 1)\Delta)$ for $j = \{1, \ldots, m\}$.
   2. Approximation to integral is a linear combination of these evaluations.
   3. Trapezoidal rule: $\int_{\Theta} L(\theta) \varpi_{i,\theta}(\theta) \, d\theta \approx \Delta (\omega_1 + 2 \sum_{j=1}^{m-1} \omega_j + \omega_m)/2$.
   4. Simpson’s rule: if $m$ odd, $\int_{\Theta} L(\theta) \varpi_{i,\theta}(\theta) \, d\theta \approx \Delta (\omega_1 + 4 \sum_{j=1, j\text{ odd}}^{m-1} \omega_j + 2 \sum_{j=1, j\text{ even}}^{m-1} \omega_j + \omega_m)/3$.

3. Laplace’s method
   1. We want $\int_A \exp(\ell(\theta)) \varpi(\theta) \, d\theta$.
   2. Let $\hat{\theta}$ be the MLE.
   3. Do Taylor series approximation for log likelihood and prior separately.
   4. Extends to higher-order approximations.

2. Integration scales poorly as dimension of $\theta$ increases
Difficulties with Bayesian analyses:

1. Computation of denominator:
   - The two examples above illustrate *conjugate priors*:
     1. Families of distributions constructed so that the likelihood times the prior was in the same family as the prior.
     2. Integral done by integrating the resulting other member of the family.
     3. Not generally useful for survival models.

2. Numerical Integration:
   - Evaluate $L(\theta)\varpi_{\theta}(\theta)$ on a grid of points, separated by $\Delta$: $\zeta_j = L(\theta_0 + (j - 1)\Delta)\varpi_{\theta}(\theta_0 + (j - 1)\Delta)$ for $j = \{1, \ldots, m\}$.
   - Approximation to integral is a linear combination of these evaluations
   - Trapezoidal rule: $\int_\Theta L(\theta)\varpi_i,\theta(\theta)\ d\theta \approx \Delta(\omega_1 + 2\sum_{j=1}^{m-1} \omega_j + \omega_m)/2$.
   - Simpson’s rule: if $m$ odd, $\int_\Theta L(\theta)\varpi_i,\theta(\theta)\ d\theta \approx \Delta(\omega_1 + 4\sum_{j=1,j\ odd}^{m-1} \omega_j + 2\sum_{j=1,j\ even}^{m-1} \omega_j + \omega_m)/3$.

3. Laplace’s method
   - We want $\int_A \exp(\ell(\theta))\varpi(\theta)\ d\theta$
   - Let $\hat{\theta}$ be the MLE.
   - Do Taylor series approximation for log likelihood and prior separately.
   - Extends to higher-order approximations.

Integration scales poorly as dimension of $\theta$ increases

1. Deterministic integration is replaced by simulation.
Bayesian Inference

1. Estimation:

Generally use expectation of posterior distribution. Minimizes expected squared error loss, analogously to material from lecture 1. Can also use posterior median or mode.
Bayesian Inference

1 Estimation:
   1 Generally use expectation of posterior distribution.
Bayesian Inference

1 Estimation:
   1 Generally use expectation of posterior distribution.
   1 Minimizes expected squared error loss, analogously to material from lecture 1.
Bayesian Inference

1. **Estimation:**
   1. Generally use expectation of posterior distribution.
      - Minimizes expected squared error loss, analogously to material from lecture 1.
   2. Can also use posterior median or mode.
Recall that frequentist inference provides intervals of the form $\theta \in (L, U)$, for $L$ and $U$, such that $P_\theta [L \leq \theta \leq U] = 1 - \alpha$.

Probability uses $f(data; \theta)$.
Recall that frequentist inference provides intervals of the form \( \theta \in (L, U) \), for \( L \) and \( U \), such that \( P_\theta [L \leq \theta \leq U] = 1 - \alpha \).

1. Probability uses \( f(data; \theta) \).
2. A similar interval using posterior probability is called a credible interval.
Recall that frequentist inference provides intervals of the form $\theta \in (L, U)$, for $L$ and $U$, such that $P_{\theta} [L \leq \theta \leq U] = 1 - \alpha$.

1. Probability uses $f(data; \theta)$.
2. A similar interval using posterior probability is called a credible interval.
   1. Recall frequentist intervals were generally equal-tailed
Recall that frequentist inference provides intervals of the form $\theta \in (L, U)$, for $L$ and $U$, such that

$$P_{\theta} [L \leq \theta \leq U] = 1 - \alpha.$$ 

1. Probability uses $f(data; \theta)$.

2. A similar interval using posterior probability is called a credible interval.

   1. Recall frequentist intervals were generally equal-tailed
   2. Bayesian intervals are often highest posterior density R Code
Where do priors come from?

Represent subjective opinion about relative possibility of various values of parameter

What if you don’t have such a subjective opinion? Look for non-informative prior.

Example, \( \theta \) is a location parameter: non-informative prior should be uniform on \((-\infty, \infty)\).

Example, \( \sigma \) is standard deviation: non-informative prior should be uniform on \((0, \infty)\).

Example, \( \tau \) is variance: non-informative prior should be uniform on \((0, \infty)\).

Problem: Density that is uniform for \( \sigma \) is not uniform for \( \sigma^2 \), and vice versa.

Solution: Log standard deviation (and hence log variance) uniform on \((-\infty, \infty)\).

Problem: All of these noninformative priors aren’t really distributions since they integrate to infinity. Such priors are called “improper”. Conceptually justified as the limit of proper priors: \( \theta \sim U[-T, T], T \to \infty \).

In extreme cases, can make integral in denominator be infinite.
Where do priors come from?

1. Represent subjective opinion about relative possibility of various values of parameter.
2. What if you don’t have such a subjective opinion? Look for non-informative prior.

Ex., \( \theta \) is a location parameter: non-informative prior should be uniform on \((-\infty, \infty)\).

Ex., \( \sigma \) is standard deviation: non-informative prior should be uniform on \((0, \infty)\).

If \( \tau \) is variance: non-informative prior should be uniform on \((0, \infty)\).

Problem: Density that is uniform for \( \sigma \) is not uniform for \( \sigma^2 \), and vice versa.

Solution: Log standard deviation (and hence log variance) uniform on \((-\infty, \infty)\).

Problem: All of these noninformative priors aren’t really distributions since they integrate to infinity. Such priors are called “improper”.

Conceptually justified as the limit of proper priors: \( \theta \) U\([-T, \infty] \), \( T \to \infty \).

In extreme cases, can make integral in denominator be infinite.
Where do priors come from?

1. Represent subjective opinion about relative possibility of various values of parameter.

2. What if you don’t have such a subjective opinion? Look for non-informative prior.
   
   - Ex., $\theta$ is a location parameter: non-informative prior should be uniform on $(-\infty, \infty)$.

   - Ex., $\sigma$ is standard deviation: non-informative prior should be uniform on $(0, \infty)$.

   - If $\tau$ is variance: non-informative prior should be uniform on $(0, \infty)$.

   - Problem: Density that is uniform for $\sigma$ is not uniform for $\sigma^2$, and vice versa.

   - Solution: Log standard deviation (and hence log variance) uniform on $(-\infty, \infty)$.

3. Problem: All of these noninformative priors aren’t really distributions since they integrate to infinity.

4. Such priors are called “improper”.

   Conceptually justified as the limit of proper priors: $\theta \sim U[-T, T]$, $T \to \infty$.

5. In extreme cases, can make integral in denominator be infinite.
Where do priors come from?

1. Represent subjective opinion about relative possibility of various values of parameter.

2. What if you don't have such a subjective opinion? Look for non-informative prior.
   - Ex., $\theta$ is a location parameter: non-informative prior should be uniform on $(-\infty, \infty)$.
   - Ex., $\sigma$ is standard deviation: non-informative prior should be uniform on $(0, \infty)$.

3. Problem: Density that is uniform for $\sigma$ is not uniform for $\sigma^2$, and vice versa.

4. Solution: Log standard deviation (and hence log variance) uniform on $(-\infty, \infty)$.

5. Problem: All of these noninformative priors aren't really distributions since they integrate to infinity.

6. Such priors are called "improper".

7. Conceptually justified as the limit of proper priors: $\theta \sim \text{U}[-T, T]$, $T \to \infty$.

8. In extreme cases, can make integral in denominator be infinite.
Where do priors come from?

1. Represent subjective opinion about relative possibility of various values of parameter

2. What if you don’t have such a subjective opinion? Look for non-informative prior.

   1. Ex., $\theta$ is a location parameter: non-informative prior should be uniform on $(-\infty, \infty)$.

   2. Ex., $\sigma$ is standard deviation: non-informative prior should be uniform on $(0, \infty)$?

   3. If $\tau$ is variance: non-informative prior should be uniform on $(0, \infty)$?

Problem: Density that is uniform for $\sigma$ is not uniform for $\sigma^2$, and vice versa.

Solution: Log standard deviation (and hence log variance) uniform on $(-\infty, \infty)$.

Problem: All of these noninformative priors aren’t really distributions since they integrate to infinity.

Such priors are called “improper”. Conceptually justified as the limit of proper priors: $\theta \sim U[-T, T]$, $T \to \infty$.

In extreme cases, can make integral in denominator be infinite.
Where do priors come from?

1. Represent subjective opinion about relative possibility of various values of parameter.

2. What if you don’t have such a subjective opinion? Look for non-informative prior.
   - Ex., $\theta$ is a location parameter: non-informative prior should be uniform on $(-\infty, \infty)$.
   - Ex., $\sigma$ is standard deviation: non-informative prior should be uniform on $(0, \infty)$?
     - If $\tau$ is variance: non-informative prior should be uniform on $(0, \infty)$?
     - Problem: Density that is uniform for $\sigma$ is not uniform for $\sigma^2$, and vice versa.
Where do priors come from?

1. Represent subjective opinion about relative possibility of various values of parameter

2. What if you don’t have such a subjective opinion? Look for non-informative prior.

   1. Ex., \( \theta \) is a location parameter: non-informative prior should be uniform on \((-\infty, \infty)\).
   2. Ex., \( \sigma \) is standard deviation: non-informative prior should be uniform on \((0, \infty)\)?

   1. If \( \tau \) is variance: non-informative prior should be uniform on \((0, \infty)\)?
   2. Problem: Density that is uniform for \( \sigma \) is not uniform for \( \sigma^2 \), and vice versa.

   3. Solution: Log standard deviation (and hence log variance) uniform on \((-\infty, \infty)\).
Where do priors come from?

1. Represent subjective opinion about relative possibility of various values of parameter.

2. What if you don’t have such a subjective opinion? Look for non-informative prior.
   
   - Ex., $\theta$ is a location parameter: non-informative prior should be uniform on $(-\infty, \infty)$.
   
   - Ex., $\sigma$ is standard deviation: non-informative prior should be uniform on $(0, \infty)$?
     
     1. If $\tau$ is variance: non-informative prior should be uniform on $(0, \infty)$?
     
     2. Problem: Density that is uniform for $\sigma$ is not uniform for $\sigma^2$, and vice versa.
     
     3. Solution: Log standard deviation (and hence log variance) uniform on $(-\infty, \infty)$.

3. Problem: All of these noninformative priors aren’t really distributions.
Where do priors come from?

1. Represent subjective opinion about relative possibility of various values of parameter.

2. What if you don’t have such a subjective opinion? Look for non-informative prior.
   - Ex., $\theta$ is a location parameter: non-informative prior should be uniform on $(-\infty, \infty)$.
   - Ex., $\sigma$ is standard deviation: non-informative prior should be uniform on $(0, \infty)$?
     - If $\tau$ is variance: non-informative prior should be uniform on $(0, \infty)$?
     - Problem: Density that is uniform for $\sigma$ is not uniform for $\sigma^2$, and vice versa.
     - Solution: Log standard deviation (and hence log variance) uniform on $(-\infty, \infty)$.

3. Problem: All of these noninformative priors aren’t really distributions since they integrate to infinity.
Where do priors come from?

1. Represent subjective opinion about relative possibility of various values of parameter

2. What if you don’t have such a subjective opinion? Look for non-informative prior.
   
   Ex., $\theta$ is a location parameter: non-informative prior should be uniform on $(-\infty, \infty)$.
   
   Ex., $\sigma$ is standard deviation: non-informative prior should be uniform on $(0, \infty)$?
     
     If $\tau$ is variance: non-informative prior should be uniform on $(0, \infty)$?
     
     Problem: Density that is uniform for $\sigma$ is not uniform for $\sigma^2$, and vice versa.
     
     Solution: Log standard deviation (and hence log variance) uniform on $(-\infty, \infty)$.
     
     Problem: All of these noninformative priors aren’t really distributions since they integrate to infinity.
     
     Such priors are called “improper”.
Where do priors come from?

1. Represent subjective opinion about relative possibility of various values of parameter.

2. What if you don’t have such a subjective opinion? Look for non-informative prior.

   1. Ex., $\theta$ is a location parameter: non-informative prior should be uniform on $(-\infty, \infty)$.

   2. Ex., $\sigma$ is standard deviation: non-informative prior should be uniform on $(0, \infty)$?

      1. If $\tau$ is variance: non-informative prior should be uniform on $(0, \infty)$?

      2. Problem: Density that is uniform for $\sigma$ is not uniform for $\sigma^2$, and vice versa.

      3. Solution: Log standard deviation (and hence log variance) uniform on $(-\infty, \infty)$.

   3. Problem: All of these noninformative priors aren’t really distributions since they integrate to infinity.

      1. Such priors are called “improper”.

      2. Conceptually justified as the limit of proper priors: $\theta \sim U[-T, T]$, $T \to \infty$. 

Where do priors come from?

1. Represent subjective opinion about relative possibility of various values of parameter

2. What if you don’t have such a subjective opinion? Look for non-informative prior.

   Ex., $\theta$ is a location parameter: non-informative prior should be uniform on $(-\infty, \infty)$.

   Ex., $\sigma$ is standard deviation: non-informative prior should be uniform on $(0, \infty)$?

      1. If $\tau$ is variance: non-informative prior should be uniform on $(0, \infty)$?
      2. Problem: Density that is uniform for $\sigma$ is not uniform for $\sigma^2$, and vice versa.
      3. Solution: Log standard deviation (and hence log variance) uniform on $(-\infty, \infty)$.

3. Problem: All of these noninformative priors aren’t really distributions since they integrate to infinity.
   Such priors are called “improper”.
   Conceptually justified as the limit of proper priors: $\theta \sim \text{U}[-T, T]$, $T \to \infty$.
   In extreme cases, can make integral in denominator be infinite. R Code
As before, decide between $H_0 : \theta \in \Omega_0$ vs. $H_A : \theta \in \Omega_a$. 

Choose hypothesis with highest posterior probability.

Often report posterior odds $P[\Omega_0 | \text{data}] / P[\Omega_a | \text{data}]$.

Factor $B$ by which prior odds $P[\Omega_0] / P[\Omega_a]$ was changed is called Bayes factor.

$B = \frac{P[\Omega_0 | \text{data}]P[\Omega_a]}{P[\Omega_a | \text{data}]P[\Omega_0]}$.

When hypothesis $\Omega_0$ and $\Omega_a$ are both simple, Bayes factor is the likelihood ratio.

Point hypotheses are only workable if there's positive prior probability on them.
Bayesian hypothesis testing.

As before, decide between $H_0 : \theta \in \Omega_0$ vs. $H_A : \theta \in \Omega_a$.

Here I used notation similar to that of frequentist analysis.
Bayesian hypothesis testing.

1. As before, decide between $H_0 : \theta \in \Omega_0$ vs. $H_A : \theta \in \Omega_a$.

   1. Here I used notation similar to that of frequentist analysis.
   2. At present, no “null” and “alternate” subtext.

   \[ \text{Bayes factor} = \frac{P[\Omega_0 \mid \text{data}]}{P[\Omega_a \mid \text{data}]} \]

   When hypothesis $\Omega_0$ and $\Omega_a$ are both simple, Bayes factor is the likelihood ratio.

   Point hypotheses are only workable if there’s positive prior probability on them.
Bayesian hypothesis testing.

1 As before, decide between $H_0 : \theta \in \Omega_0$ vs. $H_A : \theta \in \Omega_a$.

   1 Here I used notation similar to that of frequentist analysis.
   2 At present, no “null” and “alternate” subtext.

2 Choose hypothesis with highest posterior probability.
Bayesian hypothesis testing.

1. As before, decide between $H_0 : \theta \in \Omega_0$ vs. $H_A : \theta \in \Omega_a$.
   - Here I used notation similar to that of frequentist analysis.
   - At present, no “null” and “alternate” subtext.

2. Choose hypothesis with highest posterior probability.

3. Often report posterior odds $P[\Omega_0|\text{data}] / P[\Omega_a|\text{data}]$.

Factor $B$ by which prior odds $P[\Omega_0] / P[\Omega_a]$ was changed is called Bayes factor.

When hypothesis $\Omega_0$ and $\Omega_a$ are both simple, Bayes factor is the likelihood ratio.

Point hypotheses are only workable if there’s positive prior probability.
Bayesian hypothesis testing.

1. As before, decide between $H_0 : \theta \in \Omega_0$ vs. $H_A : \theta \in \Omega_a$.
   - Here I used notation similar to that of frequentist analysis.
   - At present, no “null” and “alternate” subtext.

2. Choose hypothesis with highest posterior probability.

3. Often report posterior odds $P[\Omega_0|\text{data}] / P[\Omega_a|\text{data}]$

4. Factor $B$ by which prior odds $P[\Omega_0] / P[\Omega_a]$ was changed is called Bayes factor.
Bayesian hypothesis testing.

1. As before, decide between $H_0 : \theta \in \Omega_0$ vs. $H_A : \theta \in \Omega_a$.
   - Here I used notation similar to that of frequentist analysis.
   - At present, no “null” and “alternate” subtext.

2. Choose hypothesis with highest posterior probability.

3. Often report posterior odds $P[\Omega_0|\text{data}] / P[\Omega_a|\text{data}]$

4. Factor $B$ by which prior odds $P[\Omega_0] / P[\Omega_a]$ was changed is called Bayes factor.
   - $B = (P[\Omega_0|\text{data}]P[\Omega_a])/(P[\Omega_a|\text{data}]P[\Omega_0])$
Bayesian hypothesis testing.

1. As before, decide between $H_0 : \theta \in \Omega_0$ vs. $H_A : \theta \in \Omega_a$.
   - Here I used notation similar to that of frequentist analysis.
   - At present, no “null” and “alternate” subtext.

2. Choose hypothesis with highest posterior probability.

3. Often report posterior odds $P[\Omega_0|\text{data}] / P[\Omega_a|\text{data}]$.

4. Factor $B$ by which prior odds $P[\Omega_0] / P[\Omega_a]$ was changed is called the Bayes factor.
   - $B = (P[\Omega_0|\text{data}] P[\Omega_a]) / (P[\Omega_a|\text{data}] P[\Omega_0])$.
   - When hypothesis $\Omega_0$ and $\Omega_a$ are both simple, Bayes factor is the likelihood ratio.
Bayesian hypothesis testing.

1. As before, decide between $H_0 : \theta \in \Omega_0$ vs. $H_A : \theta \in \Omega_a$.
   1. Here I used notation similar to that of frequentist analysis.
   2. At present, no “null” and “alternate” subtext.

2. Choose hypothesis with highest posterior probability.

3. Often report posterior odds $P[\Omega_0|\text{data}] / P[\Omega_a|\text{data}]$

4. Factor $B$ by which prior odds $P[\Omega_0] / P[\Omega_a]$ was changed is called Bayes factor.
   1. $B = (P[\Omega_0|\text{data}] P[\Omega_a]) / (P[\Omega_a|\text{data}] P[\Omega_0])$
   2. When hypothesis $\Omega_0$ and $\Omega_a$ are both simple, Bayes factor is the likelihood ratio.
   3. Point hypotheses are only workable if there’s positive prior probability on them.