I. Introduction

A. Examples
   a. Medicine
      i. Deaths from certain kinds of disease
      ii. Recovery from disease
      iii. Time until disease progresses to a certain point
          • sometimes a surrogate endpoint
   b. Equipment failures
   c. Mortgage prepayments
   d. etc., etc., ... SAS Code R Code

B. Describing life distributions
   KM: 2.1
   1. Most elemental description of distributions is CDF
      a. Let \( F(x) \) be the CDF.

Lecture 1

1. Analysis of data that are event times
   1. Examples
      a. Medicine
         i. Deaths from certain kinds of disease
         ii. Recovery from disease
         iii. Time until disease progresses to a certain point
          • sometimes a surrogate endpoint
      b. Equipment failures
      c. Mortgage prepayments
      d. etc., etc., ...

2. Life data analysis differs from other forms of data analysis
   a. Additional complications arising from missing data
      i. Missingness is in the form of censoring
      ii. To be defined later this lecture.
   b. Different probabilistic models are appropriate
      i. For ex., wouldn’t be interested in modeling times as normal
      ii. Asymmetric (skewed right)
   c. Different effects of interventions are usually expected
      i. For ex., horizontal shifts are usually not of interest
   d. We are probably interested in estimating not just a mean but all quantiles of a distribution
      KM: 2.2

B. Describing life distributions

1. Most elemental description of distributions is CDF
   a. Let \( F(x) \) be the CDF.

Lecture 1

1. Analysis of data that are event times
   1. Examples
      a. Medicine
         i. Deaths from certain kinds of disease
         ii. Recovery from disease
         iii. Time until disease progresses to a certain point
          • sometimes a surrogate endpoint
      b. Equipment failures
      c. Mortgage prepayments
      d. etc., etc., ...

2. Life data analysis differs from other forms of data analysis
   a. Additional complications arising from missing data
      i. Missingness is in the form of censoring
      ii. To be defined later this lecture.
   b. Different probabilistic models are appropriate
      i. For ex., wouldn’t be interested in modeling times as normal
      ii. Asymmetric (skewed right)
   c. Different effects of interventions are usually expected
      i. For ex., horizontal shifts are usually not of interest
   d. We are probably interested in estimating not just a mean but all quantiles of a distribution

B. Describing life distributions

1. Most elemental description of distributions is CDF
   a. Let \( F(x) \) be the CDF.

Lecture 1

1. Analysis of data that are event times
   1. Examples
      a. Medicine
         i. Deaths from certain kinds of disease
         ii. Recovery from disease
         iii. Time until disease progresses to a certain point
          • sometimes a surrogate endpoint
      b. Equipment failures
      c. Mortgage prepayments
      d. etc., etc., ...

2. Life data analysis differs from other forms of data analysis
   a. Additional complications arising from missing data
      i. Missingness is in the form of censoring
      ii. To be defined later this lecture.
   b. Different probabilistic models are appropriate
      i. For ex., wouldn’t be interested in modeling times as normal
      ii. Asymmetric (skewed right)
   c. Different effects of interventions are usually expected
      i. For ex., horizontal shifts are usually not of interest
   d. We are probably interested in estimating not just a mean but all quantiles of a distribution

B. Describing life distributions

1. Most elemental description of distributions is CDF
   a. Let \( F(x) \) be the CDF.

Lecture 1

1. Analysis of data that are event times
   1. Examples
      a. Medicine
         i. Deaths from certain kinds of disease
         ii. Recovery from disease
         iii. Time until disease progresses to a certain point
          • sometimes a surrogate endpoint
      b. Equipment failures
      c. Mortgage prepayments
      d. etc., etc., ...

2. Life data analysis differs from other forms of data analysis
   a. Additional complications arising from missing data
      i. Missingness is in the form of censoring
      ii. To be defined later this lecture.
   b. Different probabilistic models are appropriate
      i. For ex., wouldn’t be interested in modeling times as normal
      ii. Asymmetric (skewed right)
   c. Different effects of interventions are usually expected
      i. For ex., horizontal shifts are usually not of interest
   d. We are probably interested in estimating not just a mean but all quantiles of a distribution

B. Describing life distributions

1. Most elemental description of distributions is CDF
   a. Let \( F(x) \) be the CDF.

Lecture 1

1. Analysis of data that are event times
   1. Examples
      a. Medicine
         i. Deaths from certain kinds of disease
         ii. Recovery from disease
         iii. Time until disease progresses to a certain point
          • sometimes a surrogate endpoint
      b. Equipment failures
      c. Mortgage prepayments
      d. etc., etc., ...

2. Life data analysis differs from other forms of data analysis
   a. Additional complications arising from missing data
      i. Missingness is in the form of censoring
      ii. To be defined later this lecture.
   b. Different probabilistic models are appropriate
      i. For ex., wouldn’t be interested in modeling times as normal
      ii. Asymmetric (skewed right)
   c. Different effects of interventions are usually expected
      i. For ex., horizontal shifts are usually not of interest
   d. We are probably interested in estimating not just a mean but all quantiles of a distribution

B. Describing life distributions

1. Most elemental description of distributions is CDF
   a. Let \( F(x) \) be the CDF.

Lecture 1

1. Analysis of data that are event times
   1. Examples
      a. Medicine
         i. Deaths from certain kinds of disease
         ii. Recovery from disease
         iii. Time until disease progresses to a certain point
          • sometimes a surrogate endpoint
      b. Equipment failures
      c. Mortgage prepayments
      d. etc., etc., ...

2. Life data analysis differs from other forms of data analysis
   a. Additional complications arising from missing data
      i. Missingness is in the form of censoring
      ii. To be defined later this lecture.
   b. Different probabilistic models are appropriate
      i. For ex., wouldn’t be interested in modeling times as normal
      ii. Asymmetric (skewed right)
   c. Different effects of interventions are usually expected
      i. For ex., horizontal shifts are usually not of interest
   d. We are probably interested in estimating not just a mean but all quantiles of a distribution

B. Describing life distributions

1. Most elemental description of distributions is CDF
   a. Let \( F(x) \) be the CDF.

Lecture 1

1. Analysis of data that are event times
   1. Examples
      a. Medicine
         i. Deaths from certain kinds of disease
         ii. Recovery from disease
         iii. Time until disease progresses to a certain point
          • sometimes a surrogate endpoint
      b. Equipment failures
      c. Mortgage prepayments
      d. etc., etc., ...

2. Life data analysis differs from other forms of data analysis
   a. Additional complications arising from missing data
      i. Missingness is in the form of censoring
      ii. To be defined later this lecture.
   b. Different probabilistic models are appropriate
      i. For ex., wouldn’t be interested in modeling times as normal
      ii. Asymmetric (skewed right)
   c. Different effects of interventions are usually expected
      i. For ex., horizontal shifts are usually not of interest
   d. We are probably interested in estimating not just a mean but all quantiles of a distribution

B. Describing life distributions

1. Most elemental description of distributions is CDF
   a. Let \( F(x) \) be the CDF.
4. Generalization of Exponential
a. Sum of \( k \) independent exponentials with same (hazard) rate \( \lambda \), 
b. Density is \( \lambda^k x^{k-1} \exp(-\lambda x)/\Gamma(k) \).
c. Survival function given by incomplete gamma function 
d. \( h(x) = \lambda + \frac{1}{x} \).
i. See Fig. 1.

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

e. \( f(x) = \phi((\log(x) - \mu)/\sigma)/(\sigma x) \)
f. \( \phi(x) = \exp(-x^2/2)/\sqrt{2\pi} \) the standard normal density, 
   \( \Phi(x) = \int_{-\infty}^{x} \exp(-y^2/2) \, dy/\sqrt{2\pi} \) the standard normal CDF.
g. MRL is pretty complicated.
h. Hazard is no simpler. See Fig. 3.

6. Generalization of Gamma:
a. \( f(x) = \alpha \lambda^k x^{\alpha k-1} \exp(-\lambda x)/\Gamma(k) \)
b. Integral \( S(x) \) involves incomplete gamma function 
c. Hazard is complicated.
d. Contains Weibull (\( k = 1 \)) and log normal as special cases.

e. \( f(x) = \phi((\log(x) - \mu)/\sigma)/(\sigma x) \)
f. \( \phi(x) = \exp(-x^2/2)/\sqrt{2\pi} \) the standard normal density, 
   \( \Phi(x) = \int_{-\infty}^{x} \exp(-y^2/2) \, dy/\sqrt{2\pi} \) the standard normal CDF.
g. MRL is pretty complicated.
h. Hazard is no simpler. See Fig. 3.

7. Pareto distribution:
a. \( h(x) = \theta/x \) for \( x > \lambda \): decreasing.
b. \( S(x) = \exp \left( - \int_{x}^{\infty} \frac{\theta}{s} \, ds \right) = \exp(\theta \log(x) - \theta \log(\lambda)) = \frac{\lambda^\theta}{x^\theta} \)
c. \( f(x) = -\frac{d}{dx} S(x) = \theta x^{-\theta-1} \lambda^\theta \)
   KM: 3.1-2

II. Censoring and Truncation
A. Types of censoring
1. Right censoring
Lecture 1

Fig. 4: Relationships between distributions

<table>
<thead>
<tr>
<th>Gamma; Parameters: shape $\gamma$, scale $\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma = 1$</td>
</tr>
<tr>
<td>$\alpha = 1$</td>
</tr>
</tbody>
</table>

Exponential; Parameters: scale $\lambda$

Generalized Gamma; Parameters: shape $\gamma$, power $\alpha$, scale $\lambda$

<table>
<thead>
<tr>
<th>Weibull; Parameters: scale $\lambda$, power $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 1$</td>
</tr>
<tr>
<td>$\gamma = 1$</td>
</tr>
<tr>
<td>$\lambda \to \infty$; $\alpha$, $\lambda$ move accordingly</td>
</tr>
</tbody>
</table>

Log Normal: scale, shape parameter

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

Fig. 5: Calendar Time Diagram

Calendar Time

- 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.

d. Taxonomy
   i. 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 a enrollment and the fixed end of a study. See Fig. 5.
      - Makes time on study more relevant than calendar time. See Fig. 6.
   ii. Type II censoring: Study proceeds until $r < n$ events.
      - $C_i = X_{(r)}$
   iii. random censoring:
      - For each $X_i$ associate a censoring time $C_i$

Lecture 2

3. More exotic censoring mechanisms
   a. double censoring if either may happen
   b. interval censoring if your information is an interval.
      i. Ex., if a person is periodically screened for a disease.

KM: 3.4

4. Truncation:
   a. Certain subjects omitted from data set.
   b. left truncation:
      i. result of delayed entry
      ii. Those who have event before start are not recorded
   c. right truncation: Those who haven't had event are not recorded
      i. Ex.: data from death records.

KM: 3.3

Fig. 6: Time on Study Diagram

Time on Study

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

End of Study

Fig. 7: Lexis Diagram

Calendar Time

- easiest is when censoring is ⊥ mechanism under study
- May have a mixture of these mechanisms.
- Display on a Lexis diagram: Time on study by calendar time. See Fig. 7.

2. Left censoring:
   a. Knowledge that failure time is less than some value