4. Bayesian Inference and Regularization
   a. Recall: Regression model has unknown baseline hazard, unknown parameters
   b. Bayesian Paradigm:
      i. Treat unknown quantities as random
      ii. Put prior distribution on these
      iii. Calculate distribution conditional on data: posterior
   c. Partial likelihood approach: remove baseline hazard via profiling.
      i. We continue this here.
      ii. Alternatively, can put a prior on function space.
      iii. Let \( \pi(\beta) \) be prior density on parameter space.
   d. Bayesian Paradigm:
      i. Corresponds to person with \( z \in \mathbb{R} \)
      ii. Test and training set.
      iii. Adjusted for ties if necessary
   e. Same warnings and rules for including variables
      i. No interactions without main effects
      ii. Watch for multiple comparisons
      iii. Try nonlinear terms, interactions, etc
      iv. Dichotomize continuous variables.
   f. Model parameters measure effect of explanatory variable in light of all other variables in model.

5. The posterior as a regularization method.
   a. Posterior is \( \propto L(\beta)\pi(\beta) \)
      i. Parameter estimate maximizes posterior.
      ii. If \( \lim_{|\beta| \to \infty} \pi(\beta) = 0 \), then the posterior does not have the monotonicity problem that we saw could arise in frequentist approach.
   b. On log scale, log (partial) posterior is \( \ell(\beta) - \omega(\beta) + C \) for \( \omega(\beta) = -\log(\pi(\beta)) \)
      i. Equivalent to frequentist technique of regularization
      ii. \( \omega(\beta) = \lambda \sum_j |\beta_j|^2 \) if \( \beta \) independent \( \mathcal{N}(0, 1/\lambda) \)
   c. Inference after selection has multiple-comparisons issue.
      i. The posterior as a regularization method.
      ii. Test and training set.
      iii. Adjust for ties if necessary
      iv. With no covariates this corresponds to exponentiated Nelson–Aalen estimator.
   d. Can estimate \( S \) at arbitrary \( z \) by \( \hat{S}(t)^{\exp(z \beta)} \)
   e. If \( \beta \) known, can calculate SE just as for Kaplan–Meier
   f. Must be increased for having to estimate \( \beta \)

2. Interpretation after selection:
   a. Model parameters measure effect of explanatory variable in light of all other variables in model.
   b. Estimator of baseline hazard at event time \( k \) is
      \[
      D_k/\sum_{j \in R(k)} \exp(z_j \beta)
      \]
   c. \( \hat{S}(t_i) = \exp \left( -\sum_{k=1}^{i-1} \frac{d_k}{\sum_{j \in R(i)} \exp(z_j \beta)} \right) \)
      i. With no covariates this corresponds to exponentiated Nelson–Aalen estimator.
   d. Can estimate \( S \) at arbitrary \( z \) by \( \hat{S}(t)^{\exp(z \beta)} \)
   e. If \( \beta \) known, can calculate SE just as for Kaplan–Meier
   f. Must be increased for having to estimate \( \beta \)

3. Alternate estimator:
   a. First estimate survival function
      \[
      \hat{S}(t_i) = \prod_{i=1}^{k-1} \left( 1 - \frac{d_k}{\sum_{j \in R(i)} \exp(z_j \beta)} \right)
      \]
      i. Made by substituting \( \exp(-x) \approx 1 - x \)
      ii. Weighted Kaplan–Meier curve
   b. Most common procedure is to use Jeffreys prior.
   c. KM: 8.7

E. Model Building
   1. Same regression techniques, constraints.
      a. Measure quality: AIC, \( p \)-value.
         i. using Akaike’s Information criterion: \(-2\ell + 2p\)
         • for \( p \) the number of parameters
         • Lower is better
         ii. Using test \( p \)-value
            • Typically set significance higher: 0.15?
            • To ensure stability, put level for removal higher than that of inclusion.
   b. Search through models using stepwise:
      i. Start with an initial model.
      ii. Consider models with separate (groups of) parameters added or removed, one at a time.
      iii. Try nonlinear terms, interactions, etc
      iv. Dichotomize continuous variables.
      vi. Gives a local, rather than guaranteed global, optimum.
   vii. At each step, one can add or remove variables.
      • Only considering additions: Forward stepwise.
      • Only considering deletions: Backwards stepwise.

2. Interpretation after selection:
   a. Model parameters measure effect of explanatory variable in light of all other variables in model.
   b. Hence interpretation of parameter changes as other variables move in and out of the model.
   c. Inference after selection has multiple-comparisons issue.
      i. Effect of variables in a best-fitting model will be exaggerated relative to a model selected a priori.
      ii. One must adjust for this exaggerated effect.
      iii. Solutions:
         • test and training set.
         • build model without explanatory variable for primary hypothesis.
   d. Model selection is impacted by coordinate system for variables.
      i. Ex., a model containing baseline value and a change from baseline will be treated differently from a model containing baseline and later value.
   e. Similar warnings and rules for including variables
      i. No interactions without main effects
      ii. Watch for multiple comparisons
   f. KM: 8.8

F. Estimating nuisance baseline survival function
   1. Introduction
      a. Before was treated as nuisance parameter
      b. Now might be of interest
         i. Corresponds to person with \( z = 0 \)
         ii. With suitable redefining, can refer to any fixed \( z \)
      c. We will consider only no-ties case
   2. Estimate via Cumulative Hazard
      a. Order event times
      b. Nelson–Aalen estimator
      c. KM: 9.4
   d. Treatment is different from adding entry time as covariate
Fig. 11: Delayed Entry

i. See Fig. 11.
ii. Partial likelihood:
\[ L(\beta) = \frac{\exp(\beta)}{\exp(\beta) + \exp(0) + \exp(\beta) + \exp(0)} \times \frac{\exp(0)}{\exp(0) + \exp(\beta) + \exp(0) + \exp(\beta)} \times \frac{\exp(\beta)}{\exp(\beta)} \]

2. Subject removed and returned leaves partial likelihood unchanged
   a. Subject now has two lines
      i. First entry censored
      ii. Second entry late.
   b. Each risk set under initial structure containing subject now has either first copy or second copy.
      i. Hence partial likelihood unchanged.
   c. Can be repeated to give as many records for a subject as desired.