KM: 8.3pn2

3. Infinite Estimates
   a. If values of single covariate are in same order as event times, then estimator of associated β is $\pm\infty$
      i. Each term of ℓ' for covariate j is $z_{ij} - \frac{\sum_{k\in\mathcal{R}(i)} z_{kj}\exp(z_k\beta))}{\sum_{k\in\mathcal{R}(i)}\exp(z_k\beta))}$
      ii. Hence MLE $\hat{\boldsymbol{\beta}}$ satisfies $z_{ij} - \frac{\sum_{k\in\mathcal{R}(i)} z_{kj}\exp(z_k\hat{\beta}))}{\sum_{k\in\mathcal{R}(i)}\exp(z_k\hat{\beta}))} = 0\,.$
      iii. "same order" condition means $z_{kj} \geq z_{ij}$ for all $k \in \mathcal{R}(i)$.
      iv. Second part is weighted average of z_{kj}
      v. In order to make weighted average $= z_{ij}$ all weight must be on z_{ij}
      vi. If $z_{kj} \leq z_{ij}$ for $k \in \mathcal{R}(i)$ then $\beta_k = \infty$ works
   b. Algorithm can't converge in standard sense.
   c. Diagnose from convergence behavior.
      i. Warning message says algorithm hasn't converged.
      ii. Or Very large parameter estimates.
   d. Also can happen with linear combination of $\boldsymbol{\beta}$
      i. Harder to see by looking at data R Code SAS Code
4. Factors that make infinite estimators more likely
   a. Covariates are dichotomous
      i. Since a single observation with time and covariate in the opposite direction makes the estimate finite.
   b. The sample size is small.
   c. The model has a large number of covariates.

   d. At least one covariate has a large effect.
5. Bayesian Inference and Regularization
   a. Recall: Regression model has unknown baseline hazard, unknown parameters
   b. Baysian 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 $\varpi(\boldsymbol{\beta})$ be prior density on parameter space.
6. The posterior as a regularization method.
   a. Posterior is $\propto L(\boldsymbol{\beta})\varpi(\boldsymbol{\beta})$
      i. Parameter estimate maximizes posterior.
      ii. If $\lim_{\|\beta\|\to\infty}\varpi(\boldsymbol{\beta}) = 0$, 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 $\ell(\boldsymbol{\beta}) - \varsigma(\boldsymbol{\beta}) + C$ for $\varsigma(\boldsymbol{\beta}) = -\log(\varpi(\boldsymbol{\beta}))$
      i. Equivalent to frequentist technique of regularization
      ii. $\varsigma(\boldsymbol{\beta}) = \lambda\sum_j|\beta_j|^2$ if $\boldsymbol{\beta}$ independent $N(0, 1/\lambda)$
      iii. $\varsigma(\boldsymbol{\beta}) = \lambda\sum_j|\beta_j|$ if $\boldsymbol{\beta}$ independent Laplace with scale $1/\sqrt{\lambda}$
      iv. We investigated both of these regularizations for multiple regession
         - $\ell(\boldsymbol{\beta}) \propto \sum_j(Y_j - \boldsymbol{\beta}\boldsymbol{z}_j)^2$

      v. Most common procedure is to use Jeffreys prior. R Code SAS Code

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.
      v. Move to model with numerical criteria improved.
      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. Same warnings and rules for including variables
      i. No interactions without main effects
      ii. Watch for multiple comparisons R Code SAS Code

KM: 8.8

F. Baseline survival estimation.
1. Introduction
   a. Before was treated as nuisance parameter
   b. Now might be of interest
      i. Corresponds to person with $\boldsymbol{z} = \mathbf{o}$
      ii. With suitable redefining, can refer to any fixed $\boldsymbol{z}$
   c. We will consider only no-ties case
2. Estimate via Cumulative Hazard
   a. Order event times

b. Estimator of baseline hazard at event time k is

$$\frac{D_k}{\sum_{j \in \mathcal{R}(k)} \exp(\boldsymbol{z}_j \boldsymbol{\beta})}$$

c. $\hat{H}(t) = \sum_{k | t_k \leq t} \frac{D_k}{\sum_{j \in \mathcal{R}(k)} \exp(\boldsymbol{z}_j \boldsymbol{\beta})}$.

d. $\hat{S}(t_i) = \exp\left(-\sum_{k=1}^{i} \frac{d_k}{\sum_{j \in \mathcal{R}(i)} \exp(\boldsymbol{z}_j \boldsymbol{\beta})}\right)$

   i. With no covariates this corresponds to exponentiated Nelson–Aalen estimator

e. Can estimate S at arbitrary $\boldsymbol{z}$ by $\hat{S}(t)^{\exp(\boldsymbol{z}\boldsymbol{\beta})}$

f. If $\boldsymbol{\beta}$ known, can calculate SE just as for Kaplan–Meier

g. Must be increased for having to estimate $\boldsymbol{\beta}$

3. Alternate estimator:

   a. First estimate survival function

   $\hat{S}(t_i) = \prod_{k=1}^{i} \left(1 - \frac{d_k}{\sum_{j \in \mathcal{R}(i)} \exp(\boldsymbol{z}_j \boldsymbol{\beta})}\right)$

      i. Made by substituting $\exp(-x) \approx 1 - x$

      ii. Weighted Kaplan–Meier curve

      iii. Adjusted for ties if necessary

   b. Then estimate cumulative hazard function

   $$\hat{H}(t_i) = \sum_{k=1}^{i} \log\left(1 - \frac{d_k}{\sum_{j \in \mathcal{R}(i)} \exp(\boldsymbol{z}_j \boldsymbol{\beta})}\right).$$

   c. Both estimators have expressions for standard error

   SAS Code R Code

KM: 9.4

G. Late Entry

1. Subject not observed at beginning of study

   a. As with Kaplan–Meier, get data set representing survival conditional on having entered

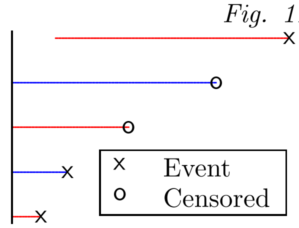
   b. As before, adjust risk set to only contain those who have already entered

   c. Requires entry time to be independent of life time.

   d. treatment is different from adding entry time as covariate

      i. See Fig. 11.

Fig. 11: Delayed Entry



      ii. Partial likelihood :

$$L(\beta) = \frac{\exp(\beta)}{\exp(\beta) + \exp(0) + \exp(\beta) + \exp(0)} \times \frac{\exp(0)}{\exp(0) + \exp(\beta) + \exp(0) + \exp(\beta)} \times \frac{\exp(\beta)}{\exp(\beta)}$$

SAS Code R Code

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 struture 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. R Code

This page intentionally left blank.