Tied survival times; estimation of survival probabilities

Patrick Breheny

November 5

Introduction

- Thus far, we have worked with Cox regression under the assumption that no ties are present among the failure times, and thus, that the data can be uniquely sorted with respect to time
- In many data sets, however, ties are present, usually due to the fact that failure times are only reported to the nearest day
- Our first topic for today is how to handle tied survival times in the Cox regression model

Average partial likelihood

- Perhaps the most natural solution would be to consider all possible ways of breaking the ties as equally likely
- In this approach, the Cox partial likelihood would be replaced with the average of the Cox partial likelihoods over all the orderings in which the ties have been broken
- As a simple example, suppose subjects 2 and 3 fail at a given time, and that subject 4 is also in the risk set at that time; the likelihood contribution would be

$$\frac{1}{2}\frac{w_2}{\sum_{2,3,4} w_i}\frac{w_3}{\sum_{3,4} w_i} + \frac{1}{2}\frac{w_3}{\sum_{2,3,4} w_i}\frac{w_2}{\sum_{2,4} w_i}$$

Breslow approximation: Idea

- The averaging method is intuitively reasonable, but from a mathematical and computational standpoint, very messy and time-consuming to work with
- An approximation that greatly simplifies the resulting calculations was proposed by Breslow (1974)
- To continue our simple example from the previous slide, the idea is that

$$\sum_{2,3,4} w_i \sum_{3,4} w_i \approx \sum_{2,3,4} w_i \sum_{2,4} w_i \approx \left(\sum_{2,3,4} w_i \right)^2$$

Breslow approximation: Implementation

- This is a very convenient approximation, as it means that all of the formulas we derived previously still hold
- The only change that needs to be made is a bookkeeping one: when calculating $\pi_{ij} = Y_i(t_j)w_i/W_j$, we simply need to keep track of ties so that the sum W_j includes all subjects who failed at time j

Efron approximation: Motivation

- An alternative approximation was proposed by Efron (1977)
- Clearly, there is an advantage to common denominators in the sum on slide 3, as it allows us to combine the sum into a single term
- It is also clear, however, that the denominator in the Breslow approximation is always larger than it should be

Efron formula

- As a compromise between the two, Efron suggested using the average weight among the failures at time j, $\bar{w} = d_j^{-1} \sum_{k \in D_j} w_k$:

$$\sum_{2,3,4} w_i \sum_{3,4} w_i \approx \sum_{2,3,4} w_i \sum_{2,4} w_i \approx \sum_{2,3,4} w_i \Big(\sum_{2,3,4} w_i - \bar{w} \Big)$$

- In general, then, the likelihood becomes

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{J} \frac{\exp(\mathbf{s}_j^T \boldsymbol{\beta})}{\prod_{r=0}^{d_j - 1} \{ \sum_{k \in R_j} \exp(\mathbf{x}_k^T \boldsymbol{\beta}) - r\bar{w} \}}$$

where $\mathbf{s}_j = \sum_{k \in D_j} \mathbf{x}_j$ and the product is taken over all unique failure times

Remarks

- To summarize:
  - Average partial likelihood: Best accuracy, hardest to work with
  - Efron approximation: Good accuracy, moderately easy to work with
  - Breslow approximation: Least accurate, very easy to work with
- When the number of ties is small, there is typically little difference between the approaches
- Many software programs implement the Breslow approach for its simplicity, but the `survival` package uses the Efron approximation

Discrete Cox model

- However, what if there are a *lot* of ties?
- In that case, it would make sense to treat the failure distribution as discrete, and propose a model in terms of the discrete hazards λ_{ij} (the hazard for the ith subject at time j) as opposed to the hazard density $\lambda_i(t)$
- The semiparametric analog to Cox regression in the discrete case is

$$\frac{\lambda_{ij}}{1-\lambda_{ij}} = \frac{\lambda_{0j}}{1-\lambda_{0j}} \exp(\mathbf{x}_i^T \boldsymbol{\beta})$$

Conditional logistic regression

- By conditioning on the number of failures at each time j, one can construct a likelihood for β that is free of intercept terms, allowing us to avoid specifying λ_{j0}
- This model, which is also widely used in categorical data analysis, is known as *conditional logistic regression*
- In the survival analysis setting, we are fitting a conditional logistic regression model at each time point, then pooling the results (score and information)
- This is often referred to as the "discrete Cox model" and is available in `coxph` through the `ties='exact'` option

Motivation

- As we have remarked several times, the Cox model is a model only for the relative risk comparing subjects versus each other – it makes no predictions as far as absolute risk
- In other words, we can use the Cox model to estimate coefficients, hazard ratios, etc., but we cannot use it to estimate the probability, say, that a subject will survive at least 2 years
- Such quantities are clearly of interest in many applications, however; is there a way we can go back and estimate a baseline hazard?

Basic approach

- It is possible to write down the nonparametric likelihood under proportional hazards and consider the joint maximum likelihood estimation of $\boldsymbol{\beta}$ and the point masses λ_{0j} at each observed failure time
- More simply, however, we can just fix $\boldsymbol{\beta}$ at the Cox regression estimate $\widehat{\boldsymbol{\beta}}$, then maximize the likelihood with respect to the λ_{0j} parameters alone
- Typically, there is little difference between the approaches, as the point masses themselves tend to have little impact on the relative risks

Nonparametric likelihood

- In our lecture on the Kaplan-Meier estimate, we showed that the nonparametric MLE of the survival function $S(t)$ is a discrete function with point masses at the observed failure times (and no mass anywhere else)
- Making the appropriate modifications to allow each observation its own subject-specific hazard, we have:

$$L(\boldsymbol{\lambda}) = \prod_j \Big\{ \prod_{i \in D_j} \lambda_{ij} \prod_{i \in R_j - D_j} (1 - \lambda_{ij}) \Big\},$$

where j indexes failure times, with D_j the set of individuals dying at time j and R_j is the risk set

Proportional hazards on the discrete scale

- The proportional hazards assumption implies the following relationship between subject-specific survival and baseline survival:

$$S_i(t) = S_0(t)^{w_i},$$

  where $w_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$
- Since $S(t) = \prod_{t_j \leq t}(1 - \lambda_j)$ for a discrete survival function, this means that

$$\lambda_{ij} = 1 - (1 - \lambda_{0j})^{w_i}$$

Main result

- Letting $\alpha_j = 1 - \lambda_{0j}$, the nonparametric MLE of S_0 given $\boldsymbol{\beta}$ can be represented with

$$L(\boldsymbol{\alpha}) = \prod_j \Big\{ \prod_{i \in D_j} (1 - \alpha_j^{w_i}) \prod_{i \in R_j - D_j} \alpha_j^{w_i} \Big\},$$

- In the case where only one failure occurs at t_j, we have

$$\hat{\alpha}_j = (1 - \pi_{jj})^{1/w_j};$$

if more than one failure occurs, there is no closed form solution and numerical optimization methods must be used to solve for $\hat{\alpha}_j$

Survival formulas

Once we have obtained $\widehat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$, our estimates of the survival function for a subject with covariate values $\mathbf{x}_i$ is given by

$$\hat{S}_i(t) = \prod_{t_j \leq t} \hat{\alpha}_j^{\exp(\mathbf{x}_i^T \widehat{\boldsymbol{\beta}})},$$

which, like the Kaplan-Meier estimate, will be a step function with discontinuous drops at every observed failure time

survfit.coxph

- The `survfit` function from the `survival` package can be applied to the output of a Cox regression model in order to carry out estimation of the baseline hazard as described on the previous slides:

```
fit <- coxph(S ~ trt + stage + hepato + bili, pbc)
sfit <- survfit(fit)
```

- The methods we described previously, such as `summary(sfit)` and `plot(sfit)`, work exactly as they did before, are compatible with `survminer::ggsurvplot()`, etc.
- The `survival` package also calculates a confidence interval for $\hat{S}_i(t)$, although in the interest of time, we won't go into the details

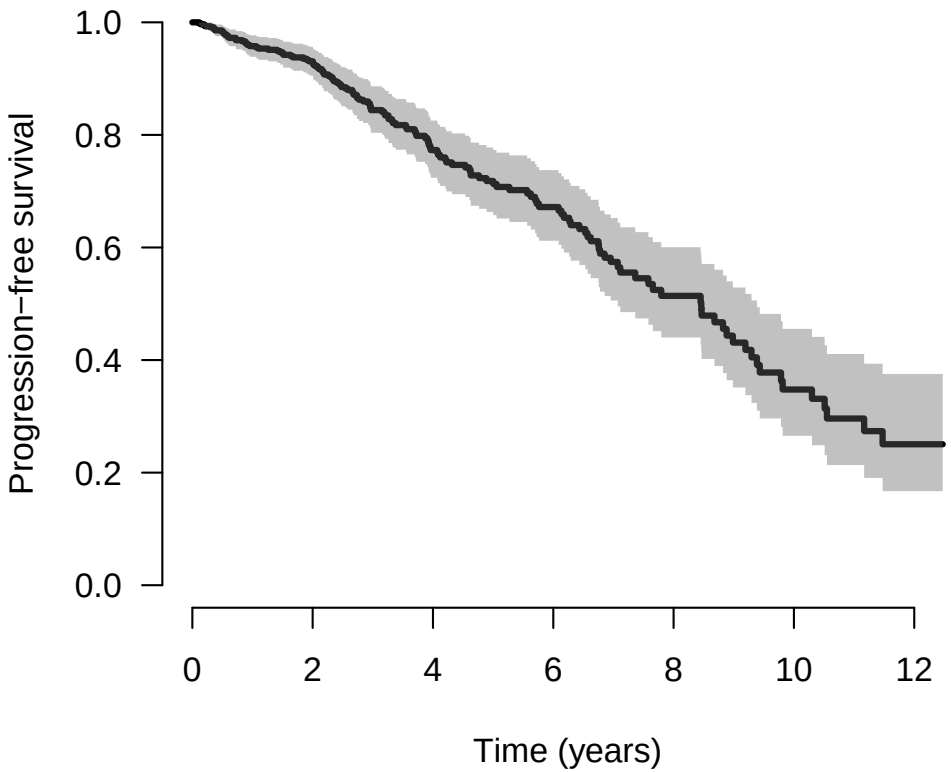
Other estimators

- Several methods for estimating the baseline hazards exist, and differ mainly in how they handle ties; the one we derived in class is usually referred to as the *Kalbfleish-Prentice* estimator
- Others include the *Nelson-Aalen-Breslow* and *Efron* estimators; these are also available as options in the `survival` package
- Actually, the default in the `survival` package is the Nelson-Aalen-Breslow estimator; to obtain the estimator we derived in class, we would need to specify

```
sfit <- survfit(fit, type="kalbfleisch-prentice")
```

although again, unless there are a lot of ties, the methods are all pretty similar

PBC example (baseline)



By default, the curve is drawn at $\mathbf{x} = \mathbf{0}$ for the centered model; i.e., at the mean of each covariate

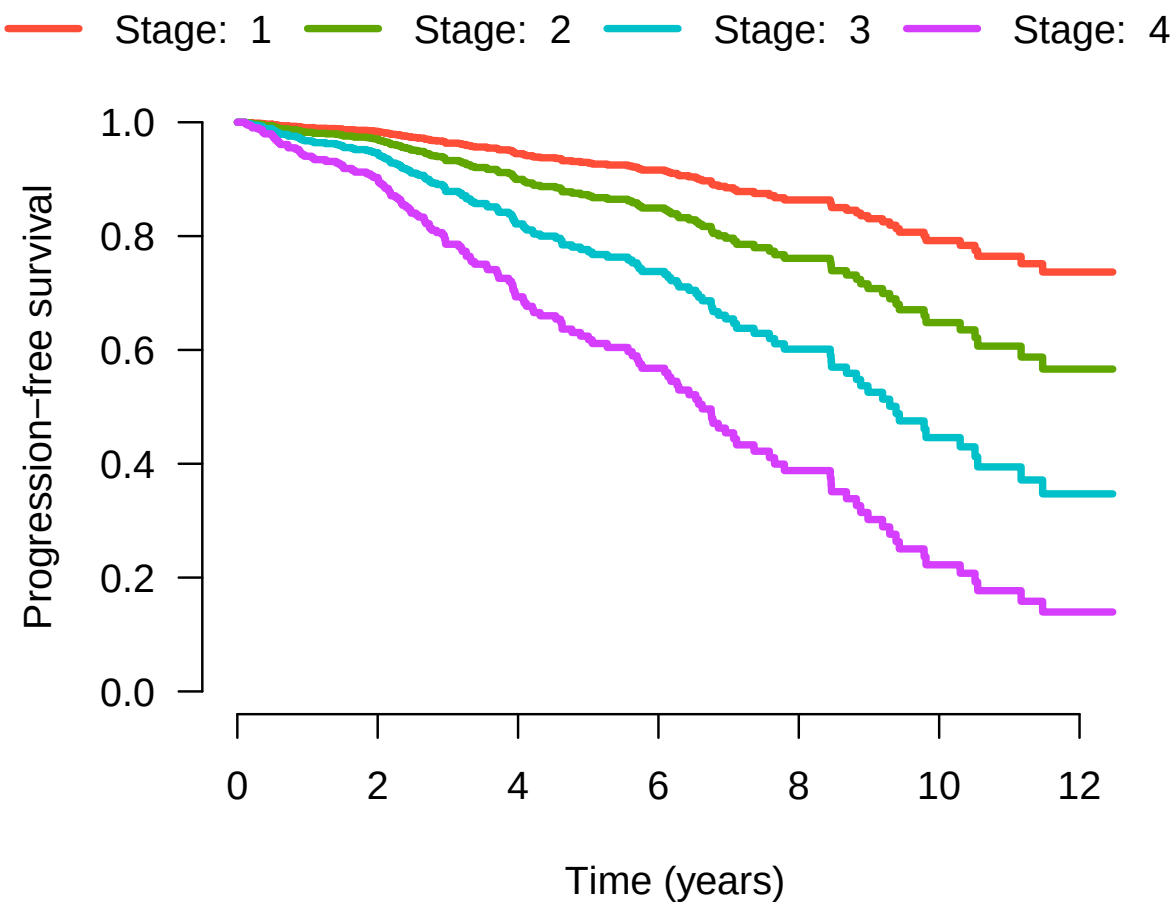
Subject specific curves

- Like `predict`, `survfit.coxph` accepts a `newdata` argument, allowing the calculation of subject-specific survival curves

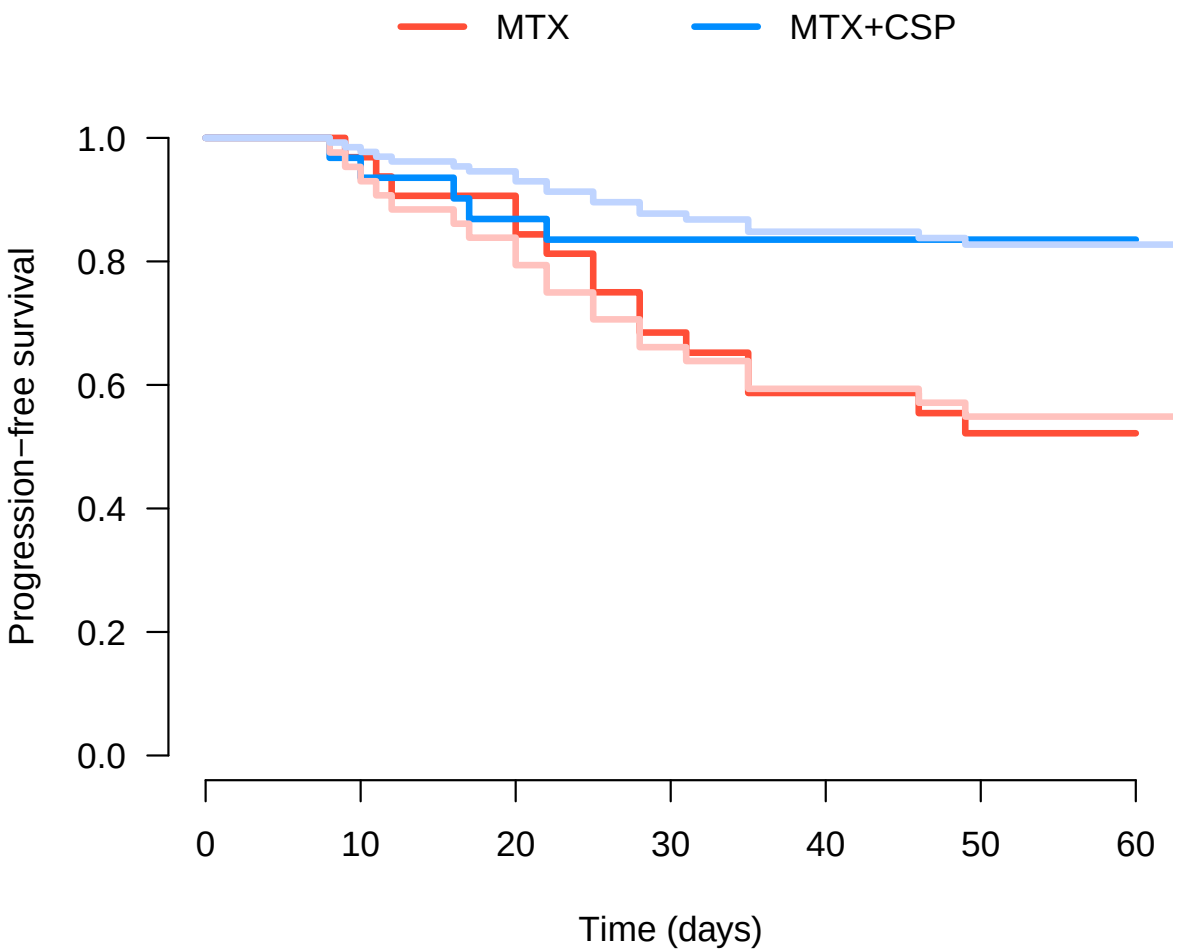
```
ndat <- data.frame(trt=0, stage=1:4, hepato=0, bili=1)
sfit <- survfit(fit, newdata=ndat)
```

- One can then submit `plot(sfit)` to plot the curves or print `sfit`, which will provide estimates and confidence intervals for the median survival time (or use the `Plot()` / `survminer` equivalents)

PBC example: Survival by stage



GVHD example: Cox versus Kaplan-Meier



Remarks

- Although the approaches are similar, it is important to keep in mind that survival function estimation in the Cox model is restricted to obey the proportional hazards assumption, and therefore cannot capture non-proportional aspects of the data
- For example, in the GVHD data, there appears to be little difference between the treatment groups at early times and a more substantial difference at later times; by construction, the Cox estimates show a constant benefit over time
- In later lectures, we will discuss diagnostics and potential remedies for non-proportional hazards