Semiparametric Regression

Patrick Breheny

October 19

Introduction

- Over the past few weeks, we've introduced a variety of regression models under the proportional hazards and accelerated failure time frameworks
- I say "variety" in the sense that under each modeling framework, we can assume various parametric distributions for the failure time and arrive at different models
- In this course, we only considered a few examples, concentrating mainly on exponential and Weibull regression, but Chapter 2 of Kalbfleish and Prentice introduce a number of additional possibilities (of ever-increasing complexity)

Parametric vs. nonparametric

- One approach to modeling, then, is to try out and compare various parametric models in an attempt to determine which parametric form best fits the observed distribution
- An alternative approach, however, is to avoid parametric assumptions concerning the distribution of survival times altogether in an attempt to address the problem in a nonparametric manner
- Broadly speaking, these nonparametric approaches are something of a last resort in the industrial testing and reliability field, but the models of choice in medical research

Semiparametric modeling

- The models we will consider, however, are not entirely nonparametric, in that we would still like parameters to describe the way in which the covariates affect survival
- Thus, we would like our models to be *semiparametric*:
  - Nonparametric portion: The underlying survival distribution
  - Parametric portion: The way in which covariates affect that underlying distribution
- The development of semiparametric models can be pursued under both the PH and AFT frameworks; our goal for today is to introduce the main idea behind both approaches

Introduction

- First, let us consider the AFT model:

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + W_i,$$

  where $Y_i = \log T_i$
- Here, the parametric aspect of the model is the $\mathbf{x}_i^T \boldsymbol{\beta}$ portion, while the nonparametric aspect involves assuming that $W_i \overset{\perp\!\!\!\perp}{\sim} F$, where F is some generic, unspecified distribution (note that for the sake of identifiability, our model cannot contain an intercept unless we introduce a restriction on the "location" of W)
- In other words, we want to carry out inference concerning $\boldsymbol{\beta}$ without deciding on a specific distribution F

Rank regression

- The regular linear regression version of this problem is a well-studied problem in nonparametric statistics
- A widely used method, *rank regression*, generalizes the basic idea of the Wilcoxon rank sum test to the regression setting
- Consider the case of a single covariate, and let $x_{(i)}$ denote the covariate value associated with the ith largest response
- A nonparametric rank-based test for the association between x and the outcome can then be based on the test statistic

$$\sum_i (i - \bar{i})(x_{(i)} - \bar{x}),$$

where $\bar{i} = (n+1)/2$

Extension to censored data

- To extend this idea to censored data, consider modifying the test statistic to

$$U = \sum_j (x_{(j)} - \bar{x}_{(j)}),$$

where j indexes the unique failure times, $x_{(j)}$ denote the covariate associated with the jth failure time, and $\bar{x}_{(j)}$ denotes the average of the covariate values of all subjects at risk at time t_j

- Furthermore, under the null, the variance of the above sum is

$$V = \sum_j (x_{(j)} - \bar{x}_{(j)})^2$$

Rank-based estimation

- Thus, we can base our test of $H_0 : \beta = 0$ upon

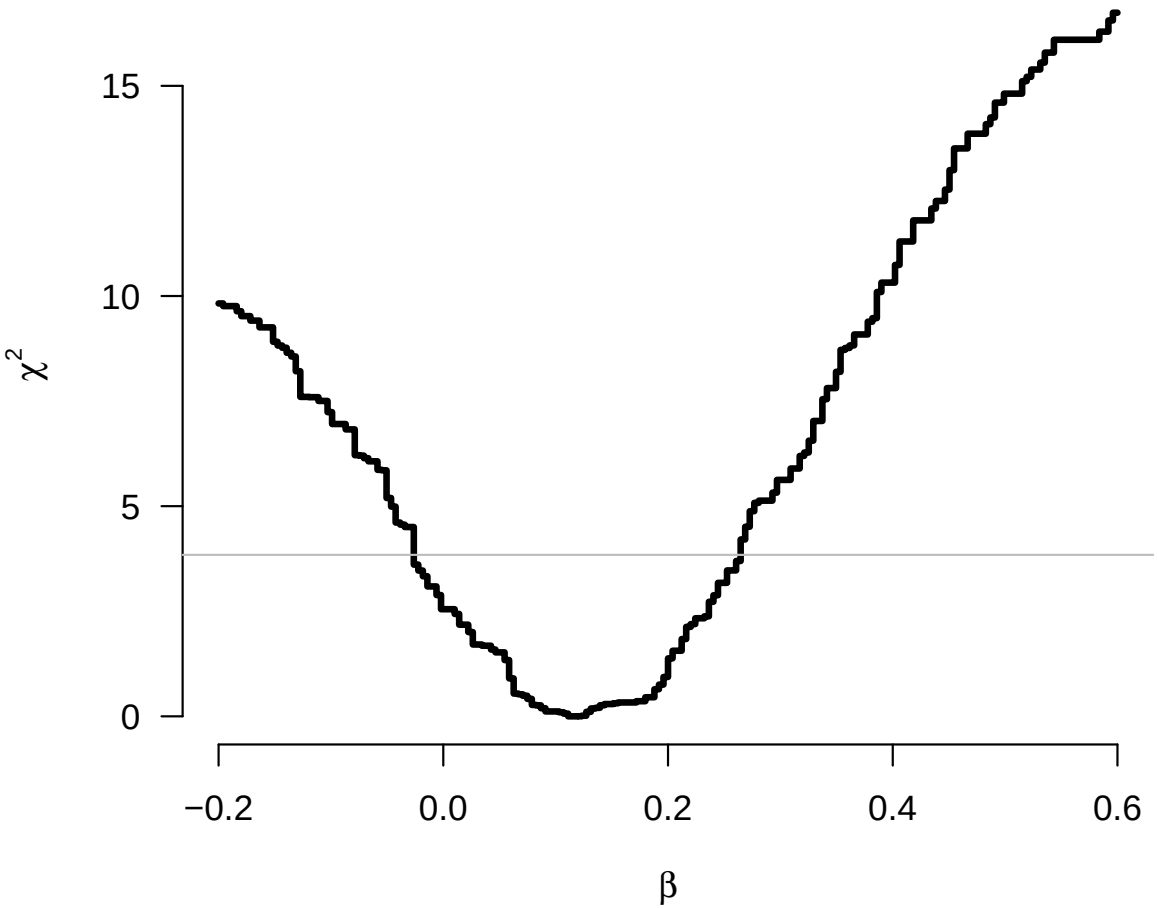
$$\frac{U^2}{V} \dot{\sim} \chi^2_1$$

- In order to use this idea for estimation and confidence interval construction, however, we need to be able to test the general null hypothesis $H_0 : \beta = \beta_0$
- Consider, therefore, testing whether the residuals of the AFT model are associated with the covariate: i.e., we apply the same procedure as before, but instead of ranking the outcomes, we rank $W_i = Y_i - \mathbf{x}_i^T \boldsymbol{\beta}_0$

Pike rat: Setup

- To get a sense of how this works, let's apply the idea to the Pike rat data and use it to estimate the effect of pretreatment regimen (`Group`)
- Over a grid of values for β_0, we will compute $W_i = Y_i - x_i\beta_0$, then calculate U and V along with the test statistic U^2/V
- We will retain inside our confidence interval all values of β_0 such that $U^2/V \leq \chi^2_{1,1-\alpha}$
- Furthermore, we will obtain a point estimate by finding the value $\widehat{\beta}$ such that $U(\widehat{\beta}) = 0$

Pike rat: Results



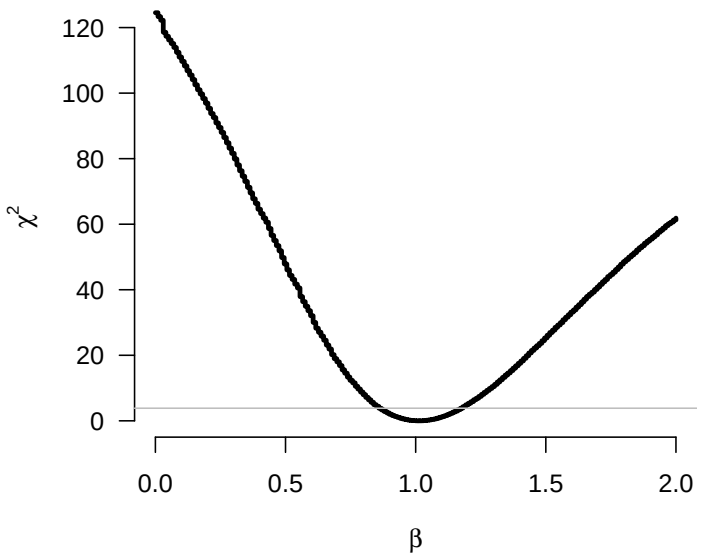
Pike rat: Comparison

		95% CI	
	$\widehat{\beta}$	Lower	Upper
Rank-based	0.11	-0.03	0.26
Weibull	0.13	0.01	0.25
Exponential	0.09	-0.56	0.75
Lognormal	0.09	-0.04	0.23

We get wildly different results depending on the assumed distribution; it is reassuring, however, the the rank-based results resemble the Weibull, which previous diagnostics suggested was a reasonable choice

Simulated example

True distribution: Weibull with $\gamma = 1.5, \beta = 1$



	$\widehat{\beta}$	95% CI Lower	95% CI Upper
Rank-based	1.01	0.86	1.17
Weibull	1.00	0.85	1.15
Exponential	1.25	1.02	1.48
Lognormal	1.09	0.93	1.26

Rank-based AFT modeling: Advantages

- The obvious advantage of the rank-based approach is that often, all one wants is to estimate the effects of various covariates – specifying the distribution of W is a nuisance that one only cares about to the extent that it affects the estimation of β
- In those cases, the rank-based AFT is a very robust method: regardless of the underlying distribution, it produces reasonable estimates (assuming that the AFT assumptions hold, of course)
- Furthermore, the loss of efficiency (compared to choosing the correct parametric form, if one exists) is typically rather mild

Rank-based AFT modeling: Disadvantages

- The overwhelming disadvantage, however, is that the test statistic is not a continuous function of β
- Typically, changing β by a small amount will not change the ranking of the residuals at all (thereby leading the test statistic unchanged as well)
- Only for a countable number of values does changing β produce changes in the test statistic, and when it does, those changes occur in discrete jumps as observations get reordered

Rank-based AFT modeling: Disadvantages (cont'd)

- This lack of differentiability has some important consequences for using these models
- First, solving for the MLE is computationally intensive in multiple dimensions (no Newton-Raphson algorithm)
- Second, inference is also problematic as we can't apply the usual Wald-type approach (no Information matrix) in order to get confidence intervals
- For these reasons, rank-based estimation in the AFT model is not widely used in practice

Semiparametric PH modeling

- It turns out, however, that the proportional hazards model is much more amenable to semiparametric modeling
- Recall the PH model:

$$\lambda_i(t) = \lambda_0(t) \exp(\mathbf{x}_i^T \boldsymbol{\beta})$$

- The parametric part of the model is the $\exp(\mathbf{x}_i^T \boldsymbol{\beta})$ portion, while the nonparametric aspect is the lack of any assumptions concerning the baseline hazard $\lambda_0(t)$
- In this approach, we will assume that our model does not contain an intercept (if we introduced an intercept, we would have to introduce some restrictions on λ_0 in order to maintain identifiability)

Special case: Two observations

- To understand how estimation works in this model, let's start by considering the special case of two observations, T_1 and T_2, with hazard functions $\lambda_1(t)$ and $\lambda_2(t)$
- Suppose that the first failure occurs at time t; how likely is it that the subject who failed was subject 1?
- **Proposition:**

$$\mathbb{P}(T_1 = t | T_{(1)} = t) = \frac{\lambda_1(t)}{\lambda_1(t) + \lambda_2(t)}$$

Elimination of the baseline hazard

- Under the proportional hazards assumption, therefore, we have

$$\mathbb{P}(T_1 < T_2) = \frac{\exp(\mathbf{x}_1^T \boldsymbol{\beta})}{\exp(\mathbf{x}_1^T \boldsymbol{\beta}) + \exp(\mathbf{x}_2^T \boldsymbol{\beta})}$$

- The remarkable result here is that the baseline hazard, $\lambda_0(t)$, has canceled out of the expression
- Extending this logic to the somewhat more general case in which we have multiple subjects, but no censoring, we have

$$\mathbb{P}(T_1 < T_2 < \cdots < T_J) = \prod_{j=1}^{J} \frac{\exp(\mathbf{x}_j^T \boldsymbol{\beta})}{\sum_{k=j}^{J} \exp(\mathbf{x}_k^T \boldsymbol{\beta})},$$

where $\mathbf{x}_j$ denotes the vector of covariates for the subject with the jth failure time

Rank-based likelihood

- We therefore have an expression for the the likelihood of $\boldsymbol{\beta}$ given the ranks of the failure times, and have found that this likelihood is free of $\lambda_0(t)$
- One would expect the ranks to contain most of the information about $\boldsymbol{\beta}$ and that, say, the duration of time between t_j and t_{j+1} probably wouldn't add a great deal of information to our knowledge of $\boldsymbol{\beta}$
- At any rate, drawing further conclusions about $\boldsymbol{\beta}$ based on the gaps between failure times is going to be highly dependent on making distributional assumptions concerning λ_0
- Focusing on the ranks, therefore, is likely to be both efficient and robust

Censoring

- Extending these results to account for censoring is mathematically straightforward, but perhaps a bit challenging philosophically
- The probability that subject j fails at time t given that one of the subjects from the risk set $R(t)$ failed at time t is

$$\frac{\exp(\mathbf{x}_j^T \boldsymbol{\beta})}{\sum_{k \in R(t)} \exp(\mathbf{x}_k^T \boldsymbol{\beta})}$$

- Thus, the full likelihood is

$$L(\boldsymbol{\beta}) = \prod_j \frac{\exp(\mathbf{x}_j^T \boldsymbol{\beta})}{\sum_{k \in R(t_j)} \exp(\mathbf{x}_k^T \boldsymbol{\beta})},$$

where the product is taken over the observed failure times

A likelihood?

- But is this a likelihood?
- Not exactly – it does not specify the probability of observing $\mathbf{t}$ and $\mathbf{d}$ given $\mathbf{X}$ and $\boldsymbol{\beta}$
- It doesn't even specify the probability of observing a given ranking of failure times, like the likelihood in the uncensored case
- Thus, it is not entirely appropriate to call this product of conditional probability statements a "likelihood", in that it does not specify the probability of observing a given set of data

Partial likelihood

- To get around this difficulty, Sir David Cox proposed the name *partial likelihood* in 1975 for the expression on slide 20, providing a general definition for what constitutes a partial likelihood and a justification for why it can be treated like a regular likelihood
- This material is summarized in Section 4.2.1 of Kalbfleish & Prentice
- The take-home message is that the partial likelihood still yields a score with mean zero and a variance given by the negative Hessian matrix; thus, we can use all the likelihood-based techniques we have developed thus far in the course to study it

Final remarks

- Although both the AFT and PH frameworks can be extended to allow nonparametric specification of the underlying survival distribution, the PH framework is much more convenient in that we end up with a regular, differentiable (partial) likelihood
- For that reason, regression based on the partial likelihood from earlier, originally proposed by Cox in 1972, is far more widely used than rank-based AFT regression models, and is indeed by far the most common regression modeling approach in survival data analysis
- Most of the remainder of the course will be spent covering Cox regression in greater depth