Applied survival analysis

Constantin T Yiannoutsos, Ph.D.

January 17, 2016

Introduction

Now we will explore the relationship between survival and explanatory variables by modeling. In this class, we consider two broad classes of regression models:

$\boxed{\textbf{Proportional Hazards (PH) models}}$

$$\lambda(t; \mathbf{Z}) = \lambda_\mathbf{0}(\mathbf{t})\mathbf{\Psi}(\mathbf{Z})$$

Most commonly, we write the second term as: $\Psi(\mathbf{Z}) = \mathbf{e}^{\beta \mathbf{Z}}$

Suppose $Z = 1$ for treated subjects and $Z = 0$ for untreated subjects. Then this model says that the hazard is increased by a factor of e^β for treated subjects versus untreated subjects (c^β might be < 1).

This is an example of a **semi-parametric** model.

Accelerated Failure Time (AFT) models

These types of models are as follows:

$$\log(T) = \mu + \beta \mathbf{Z} + \sigma \mathbf{w}$$

where w is an "error distribution". Typically, we place a **parametric** assumption on w:

- exponential, Weibull, Gamma
- lognormal

Covariates

In general, $\mathbf{Z}$ is a *vector* of covariates of interest.
$\mathbf{Z}$ may include:

- continuous factors (eg, age, blood pressure)
- discrete factors (gender, marital status)
- possible interactions (age by sex interaction)

Just as in standard linear regression, if we have a discrete covariate A with a levels, then we will need to include $(a-1)$ dummy variables $(U_1, U_2, \ldots, U_a)$ such that $U_j = 1$ if $A = j$. Then

$$\lambda_i(t) = \lambda_0(t) \exp(\beta_2 U_2 + \beta_3 U_3 + \cdots + \beta_a U_a)$$

(In the above model, the subgroup with $A = 1$ or $U_1 = 1$ is the reference group.)

Interactions

Two factors, A and B, interact if the hazard of death depends on the combination of levels of A and B.

We follow the principle of hierarchical models, and only include interactions if all of the associated main effects are also included.

The example I just gave was based on a proportional hazards model, but the description of the types of covariates we might want to include in our model applies to both the AFT and PH model.

Introduction

We'll start out by focusing on the Cox PH model, and address some of the following questions:

- What does the term $\lambda_0(t)$ mean?
- What's "proportional" about the PH model?
- How do we estimate the parameters in the model?
- How do we interpret the estimated values?
- How can we construct tests of whether the covariates have a significant effect on the distribution of survival times?
- How do these tests compare to the logrank test or the Wilcoxon test?

The Cox Proportional Hazards model

$$\lambda(t; \mathbf{Z}) = \lambda_{\mathbf{0}}(\mathbf{t}) \exp(\beta \mathbf{Z})$$

This is the most common model used for survival data. Why?

- flexible choice of covariates
- fairly easy to fit
- standard software exists

References

References: Collett, Chapter 3
Allison, Chapter 5
Cox and Oakes, Chapter 7
Kleinbaum, Chapter 3
Klein and Moeschberger, Chapters 8 & 9
Kalbfleisch and Prentice
Lee

Some books (like Collett) use $h(t; \mathbf{X})$ as their standard notation instead of $\lambda(t; \mathbf{Z})$.

Why do we call it proportional hazards?

Think of the first example, where $Z = 1$ for treated and $Z = 0$ for control. Then if we think of $\lambda_1(t)$ as the hazard rate for the treated group, and $\lambda_0(t)$ as the hazard for control, then we can write:

$$\begin{aligned} \lambda_1(t) &= \lambda(t; Z = 1) = \lambda_0(t)\exp(\beta Z) \\ &= \lambda_0(t)\exp(\beta) \end{aligned}$$

This implies that the ratio of the two hazards is a constant, ϕ, which does NOT depend on time, t. In other words, the hazards of the two groups remain proportional over time.

$$\phi = \frac{\lambda_1(t)}{\lambda_0(t)} = e^{\beta}$$

ϕ is referred to as the **hazard ratio**. **What is the interpretation of β here?**

The Baseline Hazard Function

In the example of comparing two treatment groups, $\lambda_0(t)$ is the hazard rate for the control group.

In general, $\lambda_0(t)$ is called the **baseline hazard function**, and reflects the underlying hazard for subjects with all covariates $Z_1, ..., Z_p$ equal to 0 (i.e., the "reference group").

The general form is:

$$\lambda(t; \mathbf{Z}) = \lambda_0(t) \exp(\beta_1 Z_1 + \beta_2 Z_2 + \cdots + \beta_p Z_p)$$

So when we substitute all of the Z_j's equal to 0, we get:

$$\begin{aligned} \lambda(t, \mathbf{Z} = \mathbf{0}) &= \lambda_0(t) \exp(\beta_1 * 0 + \beta_2 * 0 + \cdots + \beta_p * 0) \\ &= \lambda_0(t) \end{aligned}$$

The baseline hazard function (cont'd)

In the general case, we think of the *i*-th individual having a set of covariates $\mathbf{Z_i} = (\mathbf{Z_{1i}}, \mathbf{Z_{2i}}, ..., \mathbf{Z_{pi}})$, and we model their hazard rate as some multiple of the baseline hazard rate:

$$\lambda_i(t, \mathbf{Z_i}) \quad = \quad \lambda_0(t) \exp(\beta_1 Z_{1i} + \cdots + \beta_p Z_{pi})$$

This means we can write the log of the hazard ratio for the *i*-th individual to the reference group as:

$$\log\left(\frac{\lambda_i(t)}{\lambda_0(t)}\right) \quad = \quad \beta_1 Z_{1i} + \beta_2 Z_{2i} + \cdots + \beta_p Z_{pi}$$

The Cox Proportional Hazards model is a linear model
for the log of the hazard ratio

Advantages of the Cox PH model

One of the biggest advantages of the framework of the Cox PH model is that we can estimate the parameters β which reflect the effects of treatment and other covariates without having to make any assumptions about the form of $\lambda_0(t)$.

In other words, we don't have to assume that $\lambda_0(t)$ follows an exponential model, or a Weibull model, or any other particular parametric model.

That's what makes the model *semi-parametric*.

Questions

We will answer the following questions:

1. Why don't we just model the hazard ratio, $\phi = \lambda_i(t)/\lambda_0(t)$ directly as a linear function of the covariates $\mathbf{Z}$?
2. Why doesn't the model have an intercept?

Estimation of the model parameters

The basic idea is that under PH, information about β can be obtained from the relative orderings (i.e., ranks) of the survival times, rather than the actual values. Why?

Suppose T follows a PH model:

$$\lambda(t; \mathbf{Z}) = \lambda_{\mathbf{0}}(\mathbf{t})\mathbf{e}^{\beta \mathbf{Z}}$$

Now consider $T^* = g(T)$, where g is a monotonic increasing function. We can show that T^* also follows the PH model, with the same multiplier, $e^{\beta \mathbf{Z}}$. Therefore, when we consider likelihood methods for estimating the model parameters, we only have to worry about the ranks of the survival times.

Likelihood Estimation for the PH Model

Kalbfleisch and Prentice derive a likelihood involving only β and $\mathbf{Z}$ (not $\lambda_0(t)$) based on the marginal distribution of the ranks of the observed failure times (in the absence of censoring).

Cox (1972) derived the same likelihood, and generalized it for censoring, using the idea of a **partial likelihood**.

Suppose we observe $(X_i, \delta_i, \mathbf{Z}_i)$ for individual i, where

- X_i is a censored failure time random variable
- δ_i is the failure/censoring indicator (1=fail,0=censor)
- $\mathbf{Z}_i$ represents a set of covariates

The covariates may be continuous, discrete, or time-varying.

Suppose there are K distinct failure (or death) times, and let $\tau_1, \ldots . \tau_K$ represent the K ordered, distinct death times.

For now, assume there are no tied death times.

Let $\mathcal{R}(t) = \{i : x_i \geq t\}$ denote the set of individuals who are "at risk" for failure at time t.

More about risk sets:

- I will refer to $\mathcal{R}(\tau_j)$ as the risk set at the jth failure time
- I will refer to $\mathcal{R}(X_i)$ as the risk set at the failure time of individual i
- There will still be r_j individuals in $\mathcal{R}(\tau_j)$.
- r_j is a number, while $\mathcal{R}(\tau_j)$ identifies the actual subjects at risk

What is the partial likelihood?

Intuitively, it is a product over the set of observed death times of the conditional probabilities of seeing the observed deaths, given the set of individuals at risk at those times.

At each death time τ_j, the contribution to the likelihood is:

$$\begin{aligned}
L_j(\beta) &= Pr(\text{individual j fails}|\text{1 failure from } \mathcal{R}(\tau_j)) \\
&= \frac{Pr(\text{individual } j \text{ fails}|\text{ at risk at } \tau_j)}{\sum_{\ell \in \mathcal{R}(\tau_j)} Pr(\text{individual } \ell \text{ fails}|\text{ at risk at } \tau_j)} \\
&= \frac{\lambda(\tau_j; \mathbf{Z}_j)}{\sum_{\ell \in \mathcal{R}(\tau_j)} \lambda(\tau_j; \mathbf{Z}_\ell)}
\end{aligned}$$

Under the PH assumption, $\lambda(t; \mathbf{Z}) = \lambda_0(t) e^{\beta \mathbf{Z}}$, so we get:

$$L^{partial}(\beta) = \prod_{j=1}^{K} \frac{\lambda_0(\tau_j) e^{\beta \mathbf{Z}_\mathbf{j}}}{\sum_{\ell \in \mathcal{R}(\tau_j)} \lambda_0(\tau_j) e^{\beta \mathbf{Z}_\ell}}$$

Another derivation

In general, the likelihood contributions for censored data fall into two categories:

- **Individual is censored at X_i:**

$$L_i(\beta) = \mathbf{S}(\mathbf{X_i}) = \exp[-\int_{\mathbf{0}}^{\mathbf{X_i}} \lambda_{\mathbf{i}}(\mathbf{u})\mathbf{du}]$$

- **Individual fails at X_i:**

$$L_i(\beta) = \mathbf{S}(\mathbf{X_i})\lambda_{\mathbf{i}}(\mathbf{X_i}) = \lambda_{\mathbf{i}}(\mathbf{X_i})\exp[-\int_{\mathbf{0}}^{\mathbf{X_i}} \lambda_{\mathbf{i}}(\mathbf{u})\mathbf{du}]$$

Thus, everyone contributes $S(X_i)$ to the likelihood, and only those who fail contribute $\lambda_i(X_i)$.

This means we get a total likelihood of:

$$L(\beta) = \prod_{i=1}^{n} \lambda_i(X_i)^{\delta_i} \exp[-\int_0^{X_i} \lambda_i(u)du]$$

The above likelihood holds for all censored survival data, with general hazard function $\lambda(t)$. In other words, we haven't used the Cox PH assumption at all yet.

Now, let's multiply and divide by the term $\left[\sum_{j\in\mathcal{R}(X_i)}\lambda_i(X_i)\right]^{\delta_i}$:

$$L(\beta)=\prod_{\mathbf{i=1}}^{\mathbf{n}}\left[\frac{\lambda_{\mathbf{i}}(\mathbf{X_i})}{\sum_{\mathbf{j}\in\mathcal{R}(\mathbf{X_i})}\lambda_{\mathbf{i}}(\mathbf{X_i})}\right]^{\delta_{\mathbf{i}}}\left[\sum_{\mathbf{j}\in\mathcal{R}(\mathbf{X_i})}\lambda_{\mathbf{i}}(\mathbf{X_i})\right]^{\delta_{\mathbf{i}}}\exp[-\int_{\mathbf{0}}^{\mathbf{X_i}}\lambda_{\mathbf{i}}(\mathbf{u})\mathbf{du}]$$

Cox (1972) argued that the first term in this product contained almost all of the information about β, while the second two terms contained the information about $\lambda_0(t)$, i.e., the baseline hazard.

If we just focus on the first term, then under the Cox PH assumption:

$$
\begin{aligned}
L(\beta) &= \prod_{i=1}^{n} \left[\frac{\lambda_i(X_i)}{\sum_{j \in \mathcal{R}(X_i)} \lambda_i(X_i)} \right]^{\delta_i} \\
&= \prod_{i=1}^{n} \left[\frac{\lambda_0(X_i)\exp(\beta \mathbf{z_i})}{\sum_{j \in \mathcal{R}(X_i)} \lambda_0(X_i)\exp(\beta \mathbf{z_j})} \right]^{\delta_i} \\
&= \prod_{i=1}^{n} \left[\frac{\exp(\beta \mathbf{z_i})}{\sum_{j \in \mathcal{R}(X_i)} \exp(\beta \mathbf{z_j})} \right]^{\delta_i}
\end{aligned}
$$

This is the partial likelihood defined by Cox. Note that it does not depend on the underlying hazard function $\lambda_0(\cdot)$. Cox recommends treating this as an ordinary likelihood for making inferences about β in the presence of the nuisance parameter $\lambda_0(\cdot)$.

A simple example

Consider the following small data set:

individual	X_i	δ_i	Z_i
1	9	1	4
2	8	0	5
3	6	1	7
4	10	1	3

Now let's compile the pieces that go into the partial likelihood contributions at each failure time:

j	ordered failure time X_i	$\mathcal{R}(X_i)$	i_j	Likelihood contribution $\left[e^{\beta Z_i} / \sum_{j \in \mathcal{R}(X_i)} e^{\beta Z_j}\right]^{\delta_i}$
1	6	{1,2,3,4}	3	$e^{7\beta}/[e^{4\beta}+e^{5\beta}+e^{7\beta}+e^{3\beta}]$
2	8	{1,2,4}	2	1
3	9	{1,4}	1	$e^{4\beta}/[e^{4\beta}+e^{3\beta}]$
4	10	{4}	4	$e^{3\beta}/e^{3\beta}=1$

Notes on the partial likelihood

$$
\begin{aligned}
L(\beta) &= \prod_{j=1}^{n} \left[\frac{e^{\beta \mathbf{z_j}}}{\sum_{\ell \in \mathcal{R}(X_j)} e^{\beta \mathbf{z}_\ell}} \right]^{\delta_j} \\
&= \prod_{j=1}^{K} \frac{e^{\beta \mathbf{Z_j}}}{\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta \mathbf{Z}_\ell}}
\end{aligned}
$$

where the product is over the K death (or failure) times.

- contributions only at the death times
- the partial likelihood is NOT a product of independent terms, but of conditional probabilities
- There are other choices besides $\Psi(\mathbf{z}) = \mathbf{e}^{\beta \mathbf{z}}$, but this is the most common and the one for which software is generally available.

Partial Likelihood inference

Inference can be conducted by treating the partial likelihood as though it satisfied all the regular likelihood properties (take the more advanced failure time course to see why!!) The **log-partial likelihood** is

$$\begin{aligned}
\ell(\beta) &= \log\left[\prod_{j=1}^{n} \frac{e^{\beta \mathbf{z_j}}}{\sum_{\ell\in\mathcal{R}(X_j)} e^{\beta \mathbf{z}_\ell}}\right]^{\delta_j} \\
&= \log\left[\prod_{j=1}^{K} \frac{e^{\beta \mathbf{z_j}}}{\sum_{\ell\in\mathcal{R}(\tau_j)} e^{\beta \mathbf{z}_\ell}}\right] \\
&= \sum_{j=1}^{K}\left[\beta \mathbf{z_j} - \log[\sum_{\ell\in\mathcal{R}(\tau_\mathbf{j})} \mathbf{e}^{\beta \mathbf{z}_\ell}]\right] = \sum_{j=1}^{K} l_j(\beta)
\end{aligned}$$

where l_j is the log-partial likelihood contribution at the j-th ordered death time.

Partial Likelihood inference (cont'd)

Suppose there is only one covariate (β is one-dimensional).

The **partial likelihood score equations** are:

$$U(\beta) = \frac{\partial}{\partial\beta}\ell(\beta) = \sum_{j=1}^{n} \delta_j \left[Z_j - \frac{\sum_{\ell \in \mathcal{R}(X_j)} Z_\ell e^{\beta Z_\ell}}{\sum_{\ell \in \mathcal{R}(X_j)} e^{\beta Z_\ell}} \right]$$

We can express $U(\beta)$ intuitively as a sum of "observed" minus "expected" values:

$$U(\beta) = \frac{\partial}{\partial\beta}\ell(\beta) = \sum_{j=1}^{n} \delta_j (Z_j - \bar{Z}_j)$$

where $\bar{Z}_j$ is the "weighted average" of the covariate Z over all the individuals in the risk set at time τ_j. Note that β is involved through the term $\bar{Z}_j$.

The maximum partial likelihood estimators can be found by solving $U(\beta) = 0$.

Inference from the partial likelihood (cont'd)

Like standard likelihood theory, it can be shown (not easily) that

$$\frac{(\widehat{\beta} - \beta)}{se(\hat{\beta})} \sim N(0,1)$$

The variance of $\hat{\beta}$ can be obtained by inverting the second derivative of the partial likelihood,

$$var(\hat{\beta}) \sim \left[-\frac{\partial^2}{\partial \beta^2} \ell(\beta) \right]^{-1}$$

From the above expression for $U(\beta)$, we have:

$$\frac{\partial^2}{\partial \beta^2} \ell(\beta) = \sum_{j=1}^{n} \delta_j \left[- \frac{\sum_{\ell \in \mathcal{R}(\tau_j)} (Z_j - \bar{Z}_j)^2 e^{\beta Z_\ell}}{\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta Z_\ell}} \right]$$

Note: The true variance of $\hat{\beta}$ is a function of the unknown β. We calculate the "observed" information by substituting the partial likelihood estimate of β into the above variance formula.

Simple Example for 2-group comparison: (no ties)

Group 0: $4^+, 7, 8^+, 9, 10^+ \implies Z_i = 0$
Group 1: $3, 5, 5^+, 6, 8^+ \implies Z_i = 1$

	ordered failure	Number at risk		Likelihood contribution
j	time X_i	Group 0	Group 1	$\left[e^{\beta Z_i} / \sum_{j \in \mathcal{R}(X_i)} e^{\beta Z_j}\right]^{\delta_i}$
1	3	5	5	$e^{\beta}/[5 + 5e^{\beta}]$
2	5	4	4	$e^{\beta}/[4 + 4e^{\beta}]$
3	6	4	2	$e^{\beta}/[4 + 2e^{\beta}]$
4	7	4	1	$e^{0}/[4 + 1e^{\beta}] = 1/[4 + e^{\beta}]$
5	9	2	0	$e^{0}/[2 + 0] = 1/2$

Again, we take the product over the likelihood contributions, then maximize to get the partial MLE for β.

What does β represent in this case?

Notes

- The "observed" information matrix is generally used because in practice, people find it has better properties. Also, the "expected" is very hard to calculate.
- There is a nice analogy with the score and information matrices from more standard regression problems, except that here we are summing over observed death times, rather than individuals.
- Newton Raphson is used by many of the computer packages to solve the partial likelihood equations.

Fitting Cox PH model with R

R uses the “coxph” command.

```
coxph(formula, data=, weights, subset,
      na.action, init, control,
      ties=c("efron","breslow","exact"),
      singular.ok=TRUE, robust=FALSE,
      model=FALSE, x=FALSE, y=TRUE, tt, method, ...)
```

Example Leukemia Data

```
Call:
coxph(formula = Surv(weeks, remiss) ~ trt, data = leukemia, ties = "breslow")

  n= 42, number of events= 30

       coef exp(coef) se(coef)      z Pr(>|z|)
trt -1.5092    0.2211   0.4096 -3.685 0.000229 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

    exp(coef) exp(-coef) lower .95 upper .95
trt    0.2211      4.523   0.09907    0.4934

Concordance= 0.69  (se = 0.053 )
Rsquare= 0.304   (max possible= 0.989 )
Likelihood ratio test= 15.21  on 1 df,   p=9.615e-05
Wald test            = 13.58  on 1 df,   p=0.0002288
Score (logrank) test = 15.93  on 1 df,   p=6.571e-05
```

More Notes:

Here are some comments:

- The Cox Proportional hazards model has the advantage over a simple logrank test of giving us an estimate of the "risk ratio" (i.e., $\phi = \lambda_1(t)/\lambda_0(t)$). This is more informative than just a test statistic, and we can also form confidence intervals for the risk ratio.
- In this case, $\hat{\phi} = 0.221$, which can be interpreted to mean that the hazard for relapse among patients treated with 6-MP is less than 25% of that for placebo patients.
- From the STS LIST command in Stata or PROC LIFETEST in SAS, we were able to get estimates of the entire survival distribution $\hat{S}(t)$ for each treatment group; we can't immediately get this from our Cox model without further assumptions. **Why not?**

Adjustments for ties

The proportional hazards model assumes a continuous hazard – ties are not possible. There are four proposed modifications to the likelihood to adjust for ties.

(1) Cox's (1972) modification: "discrete" method

(2) Peto-Breslow method

(3) Efron's (1977) method

(4) Exact method (Kalbfleisch and Prentice)

(5) Exact marginal method

Some notation

$\tau_1, \ldots \tau_K$	the K ordered, distinct death times
d_j	the number of failures at τ_j
H_j	the "history" of the entire data set, up to the j-th death or failure time, including the <u>time</u> of the failure, but not the identities of the d_j who fail there.
$i_{j1}, \ldots i_{jd_j}$	the identities of the d_j individuals who fail at τ_j

Cox's (1972) modification: The "discrete" method

Cox's method assumes that if there are tied failure times, they truly happened at the same time. It is based on a discrete likelihood.

The **partial likelihood** is:

$$
\begin{aligned}
L(\beta) &= \prod_{j=1}^{K} Pr(i_{j1},\ldots i_{jd_j} \text{ fail} \mid d_j \text{ fail at } \tau_j, \text{ from } \mathcal{R}) \\
&= \prod_{j=1}^{K} \frac{Pr(i_{j1},\ldots i_{jd_j} \text{ fail} \mid \text{ in } \mathcal{R}(\tau_j))}{\sum_{\ell \in s(j,d_j)} Pr(\ell_1,\ldots.\ell_{d_j} \text{ fail} \mid \text{ in } \mathcal{R}(\tau_j))} \\
&= \prod_{j=1}^{K} \frac{\exp(\beta \mathbf{z}_{\mathbf{i_{j1}}})\cdots\exp(\beta \mathbf{z}_{\mathbf{i_{jd_j}}})}{\sum_{\ell \in s(j,d_j)} \exp(\beta \mathbf{z}_{\ell_1})\cdots\exp(\beta \mathbf{z}_{\ell_{\mathbf{d_j}}})} \\
&= \prod_{j=1}^{K} \frac{\exp(\beta \mathbf{S_j})}{\sum_{\ell \in s(j,d_j)} \exp(\beta \mathbf{S_{j\ell}})}
\end{aligned}
$$

In the previous formula

- $s(j, d_j)$ is the set of all possible sets of d_j individuals that can possibly be drawn from the risk set at time τ_j
- S_j is the sum of the Z's for all the d_j individuals who fail at τ_j
- $S_{j\ell}$ is the sum of the Z's for all the d_j individuals in the ℓ-th set drawn out of $s(j, d_j)$

Simple Example (with ties)

What does this all mean??!!

Let's modify our previous simple example to include ties.

Group 0: $4^+, 6, 8^+, 9, 10^+ \implies Z_i = 0$

Group 1: $3, 5, 5^+, 6, 8^+ \implies Z_i = 1$

j	Ordered failure time X_i	Number at risk Group 0	Number at risk Group 1	Lik. Contribution $e^{\beta S_j} / \sum_{\ell \in s(j,d_j)} e^{\beta S_{j\ell}}$
1	3	5	5	$e^{\beta}/[5+5e^{\beta}]$
2	5	4	4	$e^{\beta}/[4+4e^{\beta}]$
3	6	4	2	$e^{\beta}/[6+8e^{\beta}+e^{2\beta}]$
4	9	2	0	$e^{0}/2 = 1/2$

Comments

The tie occurs at $t = 6$, when $\mathcal{R}(\tau_j) = \{Z = 0 : (6, 8^+, 9, 10^+), Z = 1 : (6, 8^+)\}$. Of the $\binom{6}{2} = 15$ possible pairs of subjects at risk at t=6, there are 6 pairs formed where both are from group 0 ($S_j = 0$), 8 pairs formed with one in each group ($S_j = 1$), and 1 pairs formed with both in group 1 ($S_j = 2$).

Problem: With numbers of ties, the denominator can have many many terms and be difficult to calculate.

The Breslow method: (default)

Breslow and Peto suggested replacing the term $\sum_{\ell \in s(j,d_j)} e^{\beta S_{j\ell}}$ in the denominator by the term $\left(\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta Z_\ell}\right)^{d_j}$, so that the following modified partial likelihood would be used:

$$L(\beta) = \prod_{j=1}^{K} \frac{e^{\beta S_j}}{\sum_{\ell \in s(j,d_j)} e^{\beta S_{j\ell}}} \approx \prod_{j=1}^{K} \frac{e^{\beta S_j}}{\left(\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta Z_\ell}\right)^{d_j}}$$

Justification

Suppose individuals 1 and 2 fail from $\{1, 2, 3, 4\}$ at time τ_j. Let $\phi(i)$ be the hazard ratio for individual i (compared to baseline).

$$\begin{aligned}
\frac{e^{\beta S_j}}{\sum_{\ell \in s(j,d_j)} e^{\beta S_{j\ell}}} &= \frac{\phi(1)}{\phi(1)+\phi(2)+\phi(3)+\phi(4)} \times \frac{\phi(2)}{\phi(2)+\phi(3)+\phi(4)} \\
&\quad + \frac{\phi(2)}{\phi(1)+\phi(2)+\phi(3)+\phi(4)} \times \frac{\phi(1)}{\phi(1)+\phi(3)+\phi(4)} \\
&\approx \frac{2\phi(1)\phi(2)}{[\phi(1)+\phi(2)+\phi(3)+\phi(4)]^2}
\end{aligned}$$

The Peto (Breslow) approximation will break down when the number of ties are relative to the size of the risk sets, and then tends to yield estimates of β which are biased toward 0.

Efron's (1977) method

Efron suggested an even closer approximation to the discrete likelihood:

$$L(\beta) = \prod_{j=1}^{K} \frac{e^{\beta S_j}}{\left(\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta Z_\ell} + \frac{j-1}{d_j} \sum_{\ell \in \mathcal{D}(\tau_j)} e^{\beta Z_\ell}\right)^{d_j}}$$

Like the Breslow approximation, Efron's method will yield estimates of β which are biased toward 0 when there are many ties.

However, (1995) Allison recommends the Efron approximation since it is much faster than the exact methods and tends to yield much closer estimates than the default Breslow approach.

Exact method (Kalbfleisch and Prentice)

The “discrete” option that we discussed in (1) is an exact method based on a discrete likelihood (assuming that tied events truly ARE tied).

This second exact method is based on the continuous likelihood, under the assumption that if there are tied events, that is due to the imprecise nature of our measurement, and that there must be some *true* ordering.

All possible orderings of the tied events are calculated, and the probabilites of each are summed.

Example

Here is an example with 2 tied events (1,2) from risk set (1,2,3,4):

$$
\begin{aligned}
\frac{e^{\beta S_j}}{\sum_{\ell \in s(j,d_j)} e^{\beta S_{j\ell}}} \quad = \quad & \frac{e^{\beta S_1}}{e^{\beta S_1} + e^{\beta S_2} + e^{\beta S_3} + e^{\beta S_4}} \times \frac{e^{\beta S_2}}{e^{\beta S_2} + e^{\beta S_3} + e^{\beta S_4}} \\
& + \frac{e^{\beta S_2}}{e^{\beta S_1} + e^{\beta S_2} + e^{\beta S_3} + e^{\beta S_4}} \times \frac{e^{\beta S_1}}{e^{\beta S_1} + e^{\beta S_3} + e^{\beta S_4}}
\end{aligned}
$$

Bottom Line

Implications of Ties (See Allison (1995), p.127-137):

(1) **When there are no ties**, all four options give *exactly* the same results.

(2) **When there are only a few ties**, it won't make much difference which method is used. However, since the exact methods won't take much extra computing time, you might as well use one of them.

(3) **When there are many ties** (relative to the number at risk), the Breslow option (default) performs poorly (Farewell & Prentice, 1980; Hsieh, 1995). Both of the approximate methods, Breslow and Efron, yield coefficients that are attenuated (biased toward 0).

Implication of ties (cont'd)

(4) **The choice of which exact method to use** should be based on substantive grounds - are the tied event times truly tied? ...or are they the result of imprecise measurement?

(5) **Computing time of exact methods** is much longer than that of the approximate methods. However, in most cases it will still be less than 30 seconds even for the exact methods.

(6) **Best approximate method** - the Efron approximation nearly always works better than the Breslow method, with no increase in computing time, so use this option if exact methods are too computer-intensive.

R Commands for PH Model with Ties

R offers four options for adjustments with tied data:

- `breslow` (default)
- `efron`
- `exactp` (same as the "discrete" option in SAS)
- `exactm` - an exact marginal likelihood calculation (different than the "exact" option in SAS)

Fecundability data example

```
Call:
coxph(formula = Surv(cycle, censor) ~ smoker, data = fecund)

  n= 586, number of events= 567

          coef exp(coef) se(coef)      z Pr(>|z|)
smoker -0.3878    0.6786   0.1140 -3.401 0.000671 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

       exp(coef) exp(-coef) lower .95 upper .95
smoker    0.6786      1.474    0.5427    0.8485

Concordance= 0.537  (se = 0.014 )
Rsquare= 0.021   (max possible= 1 )
Likelihood ratio test= 12.57  on 1 df,   p=0.000392
Wald test            = 11.57  on 1 df,   p=0.0006712
Score (logrank) test = 11.71  on 1 df,   p=0.0006218
```

A special case: the two-sample problem

Previously, we derived the logrank test from an intuitive perspective, assuming that we have $(X_{01}, \delta_{01}) \ldots (X_{0n_0}, \delta_{0n_0})$ from group 0 and $(X_{11}, \delta_{11}), \ldots, (X_{1n_1}, \delta_{1n_1})$ from group 1.

Just as a χ^2 test for binary data can be derived from a logistic model, we will see here that the logrank test can be derived as a special case of the Cox Proportional Hazards model.

First, let's re-define our notation in terms of (X_i, δ_i, Z_i):

$$\begin{array}{lcl} (X_{01}, \delta_{01}), \ldots, (X_{0n_0}, \delta_{0n_0}) & \Longrightarrow & (X_1, \delta_1, 0), \ldots, (X_{n0}, \delta_{n0}, 0) \\ (X_{11}, \delta_{11}), \ldots, (X_{1n_1}, \delta_{1n_1}) & \Longrightarrow & (X_{n0+1}, \delta_{n0+1}, 1), \ldots, (X_{n0+n1}, \delta_{n0+n1}, 1) \end{array}$$

In other words, we have $n0$ rows of data $(X_i, \delta_i, 0)$ for the group 0 subjects, then $n1$ rows of data $(X_i, \delta_i, 1)$ for the group 1 subjects.

Using the proportional hazards formulation, we have

$$\lambda(t; Z) = \lambda_0(t)\, e^{\beta Z}$$

Group 0 hazard: $\lambda_0(t)$

Group 1 hazard: $\lambda_0(t)\, e^{\beta}$

The log-partial likelihood is:

$$\begin{aligned} logL(\beta) &= \log\left[\prod_{j=1}^{K} \frac{e^{\beta Z_j}}{\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta Z_\ell}}\right] \\ &= \sum_{j=1}^{K}\left[\beta Z_j - \log[\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\beta Z_\ell}]\right] \end{aligned}$$

Taking the derivative with respect to β, we get:

$$\begin{aligned} U(\beta) &= \frac{\partial}{\partial\beta}\ell(\beta) \\ &= \sum_{j=1}^{n} \delta_j \left[Z_j - \frac{\sum_{\ell\in\mathcal{R}(\tau_j)} Z_\ell e^{\beta Z_\ell}}{\sum_{\ell\in\mathcal{R}(\tau_j)} e^{\beta Z_\ell}} \right] \\ &= \sum_{j=1}^{n} \delta_j (Z_j - \bar{Z}_j) \end{aligned}$$

$$\text{where} \quad \bar{Z}_j = \frac{\sum_{\ell\in\mathcal{R}(\tau_j)} Z_\ell e^{\beta Z_\ell}}{\sum_{\ell\in\mathcal{R}(\tau_j)} e^{\beta Z_\ell}}$$

$U(\beta)$ is called the **“score”**.

The Score test

As we discussed earlier in the class, one useful form of a likelihood-based test is the **score test**. This is obtained by using the score $U(\beta)$ evaluated at H_o as a test statistic.

Let's look more closely at the form of the score:

$\delta_j Z_j$	**observed** number of deaths in group 1 at τ_j
$\delta_j \bar{Z}_j$	**expected** number of deaths in group 1 at τ_j

Why?

Under $H_0 : \beta = 0$, $\bar{Z}_j$ is simply the number of individuals from group 1 in the risk set at time τ_j (call this r_{1j}), divided by the total number in the risk set at that time (call this r_j). Thus, $\bar{Z}_j$ approximates the probability that given there is a death at τ_j, it is from group 1.

Thus, the score statistic is of the form:

$$\sum_{j=1}^{n}(O_j - E_j)$$

When there are ties, the likelihood has to be replaced by one that allows for ties.

Implementation in R

R produces the score test as follows (for the fecundability example above):

```
Likelihood ratio test= 12.68  on 1 df,   p=0.0003695
Wald test            = 12.12  on 1 df,   p=0.0004985
Score (logrank) test = 12.25  on 1 df,   p=0.0004642
```

... which is the same as what would have been produced by the logrank test:

```
Call:
survdiff(formula = Surv(cycle, censor) ~ smoker, data = fecund)

           N Observed Expected (O-E)^2/E (O-E)^2/V
smoker=0 486      474      446      1.76      12.3
smoker=1 100       93      121      6.50      12.3

 Chisq= 12.3  on 1 degrees of freedom, p= 0.000464
```