Lecture 5

THE PROPORTIONAL HAZARDS REGRESSION MODEL

Now we will explore the relationship between survival and explanatory variables by mostly semiparametric regression modeling. We will first consider a major class of semiparametric regression models (Cox 1972, 1975):

Proportional Hazards (PH) models

$$\lambda(t|\mathbf{Z}) = \lambda_0(t)e^{\boldsymbol{\beta}'\mathbf{Z}}$$

Here $\mathbf{Z}$ 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)

Later on we will talk about time-varying covariate $\mathbf{Z}(t)$, but we will first focus on time-independent covariates.

Under the PH model, we will try to 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 relate to the logrank test or the Wilcoxon test?

The Cox (1972) Proportional Hazards model

$$\lambda(t|\mathbf{Z}) = \lambda_0(t)\exp(\boldsymbol{\beta}'\mathbf{Z})$$

is the most commonly used regression model for survival data.

Why?

- suitable for survival type data
- flexible choice of covariates
- fairly easy to fit
- standard software exists

Note: some books or papers use $h(t;\mathbf{X})$ as their standard notation for the hazard instead of $\lambda(t;\mathbf{Z})$, and $H(t)$ for the cumulative hazard instead of $\Lambda(t)$.

Why do we call it proportional hazards?

Think of an earlier example on Leukemia data, 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, e^β, which does NOT depend on time, t. In other words, the hazards of the two groups remain proportional over time.

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

- e^β is referred to as the **hazard ratio** (HR) or **relative risk** (RR)
- β is the **log hazard ratio** or **log relative risk**.

This applied to any types of Z, as they are the (log) HR for one unit increase in the value of Z.

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:

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

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

$$\lambda_i(t) = \lambda(t|\mathbf{Z}_i) = \lambda_0(t) \exp(\beta_1 Z_{1i} + \cdots + \beta_p Z_{pi})$$

Q: Should the model have an intercept in the linear combination?

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

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

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

One of the main advantages of the framework of the Cox PH model is that we can estimate the parameters $\boldsymbol{\beta}$ without having to estimate $\lambda_0(t)$.

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

This second part is what makes the model *semi-parametric*.

Questions:

Why don't we just model the hazard ratio, $\phi = \lambda_i(t)/\lambda_0(t)$, directly as a linear function of the covariates Z?

Likelihood Estimation for the PH Model

Cox (1972, 1975) proposed a **partial likelihood** for $\boldsymbol{\beta}$ without involving $\lambda_0(t)$.

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

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

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

For now, assume there are no tied death times.

The idea is similar to the log-rank test, we look at (i.e. condition on) each observed failure time.

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

The partial likelihood is a product over the observed failure times of conditional probabilities, of seeing the observed failure, given the risk set at that time and that one failure is to happen.

In other words, these are the conditional probabilities of the observed individual, being chosen from the risk set to fail.

At each failure time X_j, the contribution to the likelihood is:

$$\begin{aligned} L_j(\boldsymbol{\beta}) &= P(\text{individual j fails}|\text{one failure from } \mathcal{R}(X_j)) \\ &= \frac{P(\text{individual } j \text{ fails}| \text{ at risk at } X_j)}{\Sigma_{\ell \in \mathcal{R}(X_j)} P(\text{individual } \ell \text{ fails}| \text{ at risk at } X_j)} \\ &= \frac{\lambda(X_j|\mathbf{Z}_j)}{\Sigma_{\ell \in \mathcal{R}(X_j)} \lambda(X_j|\mathbf{Z}_\ell)} \end{aligned}$$

Under the PH assumption, $\lambda(t|\mathbf{Z}) = \lambda_0(t)e^{\boldsymbol{\beta}'\mathbf{Z}}$, so the partial likelihood is:

$$\begin{aligned} L(\boldsymbol{\beta}) &= \prod_{j=1}^{K} \frac{\lambda_0(X_j)e^{\boldsymbol{\beta}'\mathbf{Z}_j}}{\Sigma_{\ell \in \mathcal{R}(X_j)} \lambda_0(X_j)e^{\boldsymbol{\beta}'\mathbf{Z}_\ell}} \\ &= \prod_{j=1}^{K} \frac{e^{\boldsymbol{\beta}'\mathbf{Z}_j}}{\Sigma_{\ell \in \mathcal{R}(X_j)} e^{\boldsymbol{\beta}'\mathbf{Z}_\ell}} \end{aligned}$$

Partial likelihood as a rank likelihood

Notice that the partial likelihood only uses the ranks of the failure times. In the absence of censoring, Kalbfleisch and Prentice derived the same likelihood as the marginal likelihood of the ranks of the observed failure times.

In fact, suppose that T follows a PH model:

$$\lambda(t|\mathbf{Z}) = \lambda_0(t)e^{\boldsymbol{\beta}'\mathbf{Z}}$$

Now consider $T^* = g(T)$, where g is a strictly increasing transformation. We can show that T^* also follows a PH model, with the same relative risk, $e^{\boldsymbol{\beta}'\mathbf{Z}}$ [**HW 2**].

The above has to do with the fact that $\lambda_0(t)$ is unspecified. It is one reason when we consider likelihood methods for estimating $\boldsymbol{\beta}$, we can use just the ranks of the survival times.

A censored-data likelihood derivation:

Recall that in general, the likelihood contributions for right-censored data fall into two categories:

- **Individual is censored at X_i:**

$$L_i(\boldsymbol{\beta}) = S_i(X_i)$$

- **Individual fails at X_i:**

$$L_i(\boldsymbol{\beta}) = f_i(X_i) = S_i(X_i)\lambda_i(X_i)$$

So the full likelihood is:

$$\begin{aligned} L(\boldsymbol{\beta}) &= \prod_{i=1}^{n} \lambda_i(X_i)^{\delta_i} \, S_i(X_i) \\ &= \prod_{i=1}^{n} \left[\frac{\lambda_i(X_i)}{\Sigma_{j \in \mathcal{R}(X_i)} \, \lambda_j(X_i)} \right]^{\delta_i} \left[\sum_{j \in \mathcal{R}(X_i)} \lambda_j(X_i) \right]^{\delta_i} S_i(X_i) \end{aligned}$$

in the above we have multiplied and divided by the term $\left[\Sigma_{j \in \mathcal{R}(X_i)} \, \lambda_j(X_i)\right]^{\delta_i}$.

Cox (1972) argued that the first term in this product contained almost all of the information about $\boldsymbol{\beta}$, while the last two terms contained the information about $\lambda_0(t)$, the baseline hazard.

If we keep only the first term, then under the PH assumption:

$$\begin{aligned} L(\boldsymbol{\beta}) &= \prod_{i=1}^{n} \left[\frac{\lambda_i(X_i)}{\Sigma_{j \in \mathcal{R}(X_i)} \lambda_j(X_i)} \right]^{\delta_i} \\ &= \prod_{i=1}^{n} \left[\frac{\lambda_0(X_i) \exp(\boldsymbol{\beta}' \mathbf{Z}_i)}{\Sigma_{j \in \mathcal{R}(X_i)} \lambda_0(X_i) \exp(\boldsymbol{\beta}' \mathbf{Z}_j)} \right]^{\delta_i} \\ &= \prod_{i=1}^{n} \left[\frac{\exp(\boldsymbol{\beta}' \mathbf{Z}_i)}{\Sigma_{j \in \mathcal{R}(X_i)} \exp(\boldsymbol{\beta}' \mathbf{Z}_j)} \right]^{\delta_i} \end{aligned}$$

This is the partial likelihood defined by Cox. Note that it does not depend on the underlying hazard function $\lambda_0(\cdot)$.

A simple example:

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:

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

The partial likelihood would be the product of these four terms.

Partial likelihood inference

Cox recommended to treat the partial likelihood as a regular likelihood for making inferences about β, in the presence of the nuisance parameter $\lambda_0(\cdot)$. This turned out to be valid (Tsiatis 1981, Andersen and Gill 1982, Murphy and van der Vaart 2000).

The **log-partial likelihood** is:

$$\begin{aligned}
\ell(\boldsymbol{\beta}) &= \log \left[\prod_{i=1}^{n} \frac{e^{\boldsymbol{\beta}' \mathbf{Z}_i}}{\sum_{\ell \in \mathcal{R}(X_i)} e^{\boldsymbol{\beta}' \mathbf{Z}_\ell}} \right]^{\delta_i} \\
&= \sum_{i=1}^{n} \delta_i \left[\boldsymbol{\beta}' \mathbf{Z}_i - \log \left\{ \sum_{\ell \in \mathcal{R}(X_i)} e^{\boldsymbol{\beta}' \mathbf{Z}_\ell} \right\} \right] \\
&= \sum_{i=1}^{n} l_i(\boldsymbol{\beta})
\end{aligned}$$

where l_i is the log-partial likelihood contribution from individual i.

The **partial likelihood score function** is:

$$U(\beta) = \frac{\partial}{\partial \beta}\ell(\beta) = \sum_{i=1}^{n} \delta_i \left[Z_i - \frac{\Sigma_{\ell \in \mathcal{R}(X_i)} Z_\ell e^{\beta' Z_\ell}}{\Sigma_{\ell \in \mathcal{R}(X_i)} e^{\beta' Z_\ell}} \right].$$

Denote

$$\pi_j(\beta; t) = \frac{Y_j(t) e^{\beta' Z_j}}{\Sigma_{l=1}^{n} Y_l(t) e^{\beta' Z_l}},$$

where $Y_j(t) = I(X_j \geq t)$. These are the conditional probabilities that contribute to the partial likelihood.

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

$$U(\beta) = \frac{\partial}{\partial \beta}\ell(\beta) = \sum_{i=1}^{n} \delta_i \{ Z_i - E(Z; X_i) \},$$

where $E(Z; X_i)$ is the expectation of the covariate Z w.r.t. the (discrete) probability distribution $\{\pi_j(\beta; X_i)\}_{j=1}^{n}$.

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

Analogous to standard likelihood theory, it can be shown that asymptotically

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

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

$$\widehat{\text{Var}}(\hat{\beta}) = \left[-\frac{\partial^2}{\partial \beta^2} \ell(\hat{\beta}) \right]^{-1}.$$

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

$$-\frac{\partial^2}{\partial \beta^2} \ell(\beta) = \sum_{i=1}^{n} \delta_i \left[\frac{\Sigma_{\ell \in \mathcal{R}(X_i)} \, Z_\ell^{\otimes 2} e^{\beta' Z_\ell}}{\Sigma_{\ell \in \mathcal{R}(X_i)} \, e^{\beta' Z_\ell}} - E(Z; X_i)^{\otimes 2} \right],$$

where $a^{\otimes 2} = aa'$ for a vector a. Notice that in [ ] is the variance of Z with respect to the probability distribution $\{\pi_j(\beta; X_i)\}_{j=1}^{n}$.

Eg. Leukemia Data

SAS PROC PHREG Output:

The PHREG Procedure

Data Set: WORK.LEUKEM
Dependent Variable: FAILTIME Time to Relapse
Censoring Variable: FAIL
Censoring Value(s): 0
Ties Handling: BRESLOW

Summary of the Number of
Event and Censored Values

Total	Event	Censored	Percent Censored
42	30	12	28.57

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	187.970	172.759	15.211 with 1 DF (p=0.0001)
Score	.	.	15.931 with 1 DF (p=0.0001)
Wald	.	.	13.578 with 1 DF (p=0.0002)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
TRTMT	1	-1.509191	0.40956	13.57826	0.0002	0.221

Compare this with the logrank test

```
The LIFETEST Procedure

Rank Tests for the Association of FAILTIME with Covariates
Pooled over Strata

Univariate Chi-Squares for the LOG RANK Test

                Test       Standard                    Pr >
Variable    Statistic    Deviation    Chi-Square    Chi-Square

TRTMT         10.2505       2.5682       15.9305        0.0001
```

Notes:

- The **logrank test = score test** under the PH model (more later).
- In general, the score test would be for *all* of the variables in the model, but in this case, we have only "`trtmt`".
- In R you can use coxph() to fit the PH model.

More Notes:

- 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, $\hat{\phi} = e^{\hat{\beta}} = 0.221$, 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.

- As you see, the software does not immediately give estimates of the survival function $\hat{S}(t)$ for each treatment group. **Why?**

Constructing Confidence intervals for the Hazard Ratio:

Many software packages provide estimates of β, but the hazard ratio, or relative risk, RR= $\exp(\beta)$ is often the parameter of interest.

Confidence intervals for $\exp(\beta)$

Form a confidence interval for $\hat{\beta}$, and then exponentiate the endpoints:

$$[L, U] \;=\; [e^{\hat{\beta}-1.96se(\hat{\beta})}, e^{\hat{\beta}+1.96se(\hat{\beta})}]$$

Should we try to use the delta method for $e^{\hat{\beta}}$?

Hypothesis Tests:

For each covariate of interest, the null hypothesis is

$$H_0 : RR_j = 1 \Leftrightarrow \beta_j = 0$$

A Wald test of the above hypothesis is constructed as:

$$Z = \frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)} \qquad \text{or} \qquad \chi^2 = \left(\frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)}\right)^2$$

Note: if we have a factor A with a levels, then we would need to construct a χ^2 test with $(a-1)$ df, using a test statistic based on a quadratic form:

$$\chi^2_{(a-1)} \;=\; \widehat{\boldsymbol{\beta}}'_A \text{Var}(\widehat{\boldsymbol{\beta}}_A)^{-1} \widehat{\boldsymbol{\beta}}_A$$

where $\boldsymbol{\beta}_A = (\beta_1, ..., \beta_{a-1})'$ are the $(a-1)$ coefficients corresponding to the binary variables $Z_1, ..., Z_{a-1}$.

Likelihood Ratio Test:

Suppose there are $(p+q)$ explanatory variables measured:

$$Z_1, \ldots, Z_p, Z_{p+1}, \ldots, Z_{p+q}.$$

Consider the following models:

- **Model 1:** (contains only the first p covariates)

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

- **Model 2:** (contains all $(p+q)$ covariates)

$$\lambda_i(t|\mathbf{Z}) = \lambda_0(t) \exp(\beta_1 Z_1 + \cdots + \beta_{p+q} Z_{p+q})$$

We can construct a **likelihood ratio** test for testing

$$H_0 : \beta_{p+1} = \cdots = \beta_{p+q} = 0$$

as:

$$\chi^2_{LR} \;=\; -2 \left\{ \log L(M1) - \log L(M2) \right\},$$

where $L(M\cdot)$ is the maximized partial likelihood under each model. Under H_0, this test statistic is approximately distributed as χ^2 with q df.

An example:

MAC Disease Trial

ACTG 196 was a randomized clinical trial to study the effects of combination regimens on prevention of MAC (mycobacterium avium complex) disease, which is one of the most common opportunistic infections in AIDS patients and is associated with high mortality and morbidity.

The **treatment regimens** were:

- clarithromycin (new)
- rifabutin (standard)
- clarithromycin plus rifabutin

The study also recorded the patients' performance status (Karnofski score) and CD4 counts.

Model 1:

```
No. of subjects =         1151              Number of obs   =      1151
No. of failures =          121
Time at risk    =       489509
                                            LR chi2(3)      =     32.01
Log likelihood  =   -754.52813              Prob > chi2     =    0.0000

------------------------------------------------------------------------------
      _t |
      _d |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
  karnof |  -.0448295   .0106355    -4.215  0.000    -.0656747   -.0239843
     rif |   .8723819   .2369497     3.682  0.000     .4079691    1.336795
   clari |   .2760775   .2580215     1.070  0.285    -.2296354    .7817903
------------------------------------------------------------------------------
```

Model 2:

```
No. of subjects =         1151              Number of obs   =      1151
No. of failures =          121
Time at risk    =       489509
                                            LR chi2(4)      =     63.74
Log likelihood  =   -738.66225              Prob > chi2     =    0.0000

------------------------------------------------------------------------------
      _t |
      _d |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
  karnof |  -.0368538   .0106652    -3.456  0.001    -.0577572   -.0159503
     rif |    .880338   .2371111     3.713  0.000     .4156089    1.345067
   clari |   .2530205   .2583478     0.979  0.327     -.253332    .7593729
     cd4 |  -.0183553   .0036839    -4.983  0.000    -.0255757   -.0111349
------------------------------------------------------------------------------
```

Notes:

- We can compute the hazard ratio of CD4, for example, by exponentiating the coefficients:

$$RR_{cd4} = \exp(-0.01835) = 0.98$$

What is the interpretation of this RR?

Although this RR is highly significant, it is very close to 1, why?

- The likelihood ratio test for the effect of CD4 is twice the difference in minus log-likelihoods between the two models:

$$\chi^2_{LR} \;=\; 2 * (754.533 - 738.66) = 31.74$$

How does this test statistic compare to the Wald χ^2 test?

- In the MAC study, there were three treatment arms (rif, clari, and the rif+clari combination). Because we have only included the `rif` and `clari` effects in the model, the combination therapy is the "reference" group.

- Stata allows a Wald test comparing the treatment effect across three groups:

```
. test rif clari

 ( 1)  rif = 0.0
 ( 2)  clari = 0.0

           chi2(  2) =   17.01
         Prob > chi2 =    0.0002
```

  This tests whether both treatment coefficients are equal to 0.

[Reading] A special case: the two-sample problem

Previously, we derived the logrank test for
$(X_{01}, \delta_{01}) \ldots (X_{0n_0}, \delta_{0n_0})$ from group 0,
and $(X_{11}, \delta_{11}), \ldots, (X_{1n_1}, \delta_{1n_1})$ from group 1.

Just as the chi-squared test for 2x2 tables 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 regression model.

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

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

In other words, we have n_0 rows of data $(X_i, \delta_i, 0)$ for the group 0 subjects, then n_1 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:

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

Taking the derivative with respect to β, we get:

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

where

$$\bar{Z}_j = \frac{\Sigma_{\ell\in\mathcal{R}(X_j)}\, Z_\ell e^{\beta Z_\ell}}{\Sigma_{\ell\in\mathcal{R}(X_j)}\, e^{\beta Z_\ell}} \overset{\beta=0}{=} \frac{\Sigma_{\ell\in\mathcal{R}(X_j)}\, Z_\ell}{|\mathcal{R}(X_j)|} = \frac{r_{1j}}{r_j}.$$

$U(\beta)$ is the **"score"**.

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 under H_0 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 X_j

$\delta_j \bar{Z}_j$ **expected** number of deaths in group 1 at X_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 X_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 X_j, it is from group 1.

When there are no ties,

$$U(0) = \sum \delta_j \left\{ d_{1j} - \frac{r_{1j}}{r_j} \right\},$$

also

$$I(0) = -U'(0) = \sum \delta_j \left\{ \frac{r_{1j}}{r_j} - \left(\frac{r_{1j}}{r_j} \right)^2 \right\}.$$

Therefore the score test under the Cox model for two-group comparison, which has $U(0)/\sqrt{I(0)} \overset{H_0}{\sim} N(0,1)$, is the log-rank test.

[Reading] Adjusting for ties

The proportional hazards model assumes a continuous hazard – ties should not be 'heavy'. However, when they do happen, there are a few proposed modifications to the partial 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** (stata)

Some notation:

$\tau_1, ...\tau_K$	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 right before the j-th death or failure time
$i_{j1}, ...i_{jd_j}$	the identities of the d_j individuals who fail at τ_j

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

Cox's method assumes that if there are tied failure times, they truly happened at the same time.

The **partial likelihood** is then:

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

where

- $s(j, d_j)$ is the set of all possible sets of d_j individuals that can possibly be drawn from the risk set at time X_j
- S_j is the sum of the Z's for all the d_j individuals who fail at X_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 $s(j, d_j)$

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

Simple Example (with 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	**Likelihood 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	$e^0/2 = 1/2$

The tie occurs at $t = 6$, when $\mathcal{R}(X_j) = \{Z = 0 : (6, 8^+, 9, 10^+), Z = 1 : (6, 8^+)\}$. 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 large numbers of ties, the denominator can have many many terms and be difficult to calculate.

(2) Breslow method:

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

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

Justification:

Suppose individuals 1 and 2 fail from $\{1, 2, 3, 4\}$ at time τ_j. Let $\phi(i) = \exp(\boldsymbol{\beta}\mathbf{Z}_i)$ be the relative risk for individual i.

$$\begin{aligned} & P\{\text{ 1 and 2 fail} \,|\, \text{two failures from } R(\tau_j)\,\} \\ = \;& \frac{\phi(1)}{\phi(1)+\phi(2)+\phi(3)+\phi(4)} \times \frac{\phi(2)}{\phi(2)+\phi(3)+\phi(4)} \\ & \text{or } \frac{\phi(2)}{\phi(1)+\phi(2)+\phi(3)+\phi(4)} \times \frac{\phi(1)}{\phi(1)+\phi(3)+\phi(4)} \\ \approx \;& \frac{\phi(1)\phi(2)}{[\phi(1)+\phi(2)+\phi(3)+\phi(4)]^2} \end{aligned}$$

This approximation will break down when the number of ties are large relative to the size of the risk sets, and then tends to yield estimates of β which are biased toward 0.

This is the default for most software programs, because it is computationally simple.

(3) Efron's (1977) method:

Efron suggested an even closer approximation to the discrete likelihood:

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

Like the Breslow approximation, Efron's method also assumes that the failures occur one at a time, and will yield estimates of β which are biased toward 0 when there are many ties.

However, the Efron approximation is much faster than the exact methods and tends to yield much closer estimates than the Breslow approach.

This is the default in R coxph().

(4) Exact method (Kalbfleisch and Prentice):

What we discussed in (1) is an exact method assuming that tied events truly are tied.

This second exact method is based on 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 probabilities of each are summed.

Example with 2 tied events (1,2) from riskset (1,2,3,4):

$$\frac{e^{\beta Z_1}}{e^{\beta Z_1}+e^{\beta Z_2}+e^{\beta Z_3}+e^{\beta Z_4}} \times \frac{e^{\beta Z_2}}{e^{\beta Z_2}+e^{\beta Z_3}+e^{\beta Z_4}}$$
$$+ \frac{e^{\beta Z_2}}{e^{\beta Z_1}+e^{\beta Z_2}+e^{\beta Z_3}+e^{\beta Z_4}} \times \frac{e^{\beta Z_1}}{e^{\beta Z_1}+e^{\beta Z_3}+e^{\beta Z_4}}$$

Implications of Ties

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

(2) **When there are only a few ties**, it won't make much difference which method is used.

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

(4) **The choice of which exact method to use** could 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.

Example: The fecundability study

Women who had recently given birth (or had tried to get pregnant for at least a year) were asked to recall how long it took them to become pregnant, and whether or not they smoked during that time. The outcome of interest is time to pregnancy (measured in menstrual cycles).

```
data fecund;
  input   smoke      cycle     status    count;
  cards;
0          1          1        198
0          2          1        107
0          3          1         55
0          4          1         38
0          5          1         18
0          6          1         22
.....................................

1         10          1          1
1         11          1          1
1         12          1          3
1         12          0          7
;

proc phreg;
  model cycle*status(0) = smoke /ties=breslow;   /* default in SAS */
  freq count;

proc phreg;
  model cycle*status(0) = smoke /ties=discrete;
  freq count;

proc phreg;
  model cycle*status(0) = smoke /ties=exact;
  freq count;

proc phreg;
  model cycle*status(0) = smoke /ties=efron;
  freq count;
```

SAS Output for Fecundability study: Accounting for Ties

```
*************************************************************************
Ties Handling: BRESLOW

                  Parameter   Standard       Wald         Pr >      Risk
Variable    DF    Estimate     Error     Chi-Square    Chi-Square   Ratio
 SMOKE      1    -0.329054    0.11412       8.31390      0.0039     0.720


*************************************************************************
Ties Handling: DISCRETE

                  Parameter   Standard       Wald         Pr >      Risk
Variable    DF    Estimate     Error     Chi-Square    Chi-Square   Ratio
 SMOKE      1    -0.461246    0.13248      12.12116      0.0005     0.630


*************************************************************************
Ties Handling: EXACT

                  Parameter   Standard       Wald         Pr >      Risk
Variable    DF    Estimate     Error     Chi-Square    Chi-Square   Ratio
 SMOKE      1    -0.391548    0.11450      11.69359      0.0006     0.676


*************************************************************************
Ties Handling: EFRON

                  Parameter   Standard       Wald         Pr >      Risk
Variable    DF    Estimate     Error     Chi-Square    Chi-Square   Ratio
 SMOKE      1    -0.387793    0.11402      11.56743      0.0007     0.679

*************************************************************************
```

For this particular dataset, does it seem like it would be important to consider the effect of tied failure times? Which method would be best?

When there are ties and comparing two or more groups, the score test under the PH model can correspond to different versions of the log-rank test.

Typically (depending on software):

discrete/exactp	$\rightarrow$ Mantel-Haenszel logrank test
breslow	$\rightarrow$ linear rank version of the logrank test