Chapter 3
The Cox Proportional Hazards Model

The *proportional hazards* (PH) or *Cox model* holds on E, if the hazard rate has the form

$$\lambda_{x(\cdot)}(t) = r\{x(t)\}\,\lambda_0(t), \quad x(\cdot) \in E, \tag{3.1}$$

where $\lambda_0(\cdot)$ is an unspecified *baseline hazard rate* function, and $r(\cdot)$ is a positive function on E. The function $r(\cdot)$ explains the *summing effect* of all $x_i(\cdot)$ on the distribution of T. If r is unknown it is a *nonparametric model*. In most applications the function r is parameterized in the form

$$r(x) = \exp\{\beta^T x\},$$

where $\beta = (\beta_1, \dots, \beta_m)^T$ is the vector of *regression parameters*. Under this parameterization we obtain the classical *semiparametric* Cox model on E with time-dependent covariables:

$$\lambda_{x(\cdot)}(t) = e^{\beta^T x(t)}\lambda_0(t), \quad x(\cdot) \in E. \tag{3.2}$$

If $\lambda_0(\cdot)$ is further taken from some parametric family of hazard rates (the Weibull family, for example), then we have a parametric model. The following example illustrates that the *parametric* Weibull family fulfills the setting of Cox model with covariates.

Example: Parametric example of Cox's model
Consider a class of Weibull distributions indexed by a set of covariates $\mathbf{X} = (X_1, \dots, X_p)^T$: let $T_{\mathbf{X}}$ be distributed as Weibull with survival function

$$S_{T_{\mathbf{X}}}(t) = \mathbf{P}\{T_{\mathbf{X}} \geq t\} = \exp(-a_* t^b).$$

M. Nikulin and H.-D.I. Wu, *The Cox Model and Its Applications*,
SpringerBriefs in Statistics, DOI 10.1007/978-3-662-49332-8_3

Here the scale parameter a_* is modeled as $\exp(\beta_0 + \beta^T \mathbf{X})$ with

$$\beta^T \mathbf{X} = \beta_1 X_1 + \cdots + \beta_p X_p.$$

Then the cumulative hazard

$$\Lambda_{T_{\mathbf{X}}}(t) = \exp(\beta^T \mathbf{X})(at^b)$$

for $a = \exp(\beta_0)$. By this expression, the Weibull family has the structure of Cox PH model: $\lambda(t; \mathbf{X}) = \lambda_0(t) \exp(\beta^T \mathbf{X})$ if the baseline hazard $\lambda_0(t)$ equals to abt^{b-1} (or the baseline cumulative hazard $\Lambda_0(t) = at^b$). An example for the analysis of Stanford Heart Transplant (SHT) data using Weibull regression is presented in Sect. 3.5 to compare with the estimation of Cox model.

Note that in this example the shape parameter b is a fixed and unknown positive constant. If it is further modeled as $b = \exp(\gamma^T \mathbf{X})$ for a set of parameters $\gamma = (\gamma_1, \ldots, \gamma_p)^T$, then we have the so-called *heteroscedastic hazards regression* (HHR) model proposed by Hsieh (2001). See Sect. 5.4 for further details. ♣

The PH model is mostly applied for analysis of survival data but the statisticians working in reliability are very cautious to use it, especially for analysis of accelerated life testing data. It can be explained by the fact that the PH model has one unnatural property: the conditional probability to fail in a time interval $(t, t+s)$ given that a unit is functioning at the moment t depends only on the values of the stress (or covariable) $x(\cdot)$ in that interval but does not depend on the values of the stress until the moment t:

$$\mathbf{P}(T \leq t+s \mid T > t) = 1 - \exp\left\{ -\int_t^{t+s} e^{\beta^T x(u)} d\Lambda_0(u) \right\},$$

here T is failure time, λ_0 is the baseline hazard function which does not depend on stress. For this reason we can say that PH model has the property of *absence of memory*.

The common sense says that if a unit functioned in high stress conditions, it used a large amount of resource and is "*older*" than a unit which functioned in *mild* stress conditions at the same time t. So the conditional probabilities of failure after the moment t should be different for those two units. More exactly, the proportional hazards model suppose that the ratio of resource using at the moment t depends only on values of covariates at this moment and does not depend on resources used until this moment.

Let x_1 and x_2 be two *constant* stresses, $x_1, x_2 \in E_1$, and $x(\cdot) \in E_2$ is a *simple step stress* of the form

$$x(t) = \begin{cases} x_1, & \text{if } t < t_1 \\ x_2, & \text{if } t \geq t_1, \end{cases} \tag{3.3}$$

then for all $t \geq t_1$

$$\lambda_{x(\cdot)}(t) = \lambda_{x_2}(t).$$

If x_1 is some *accelerated stress* with respect to the "*normal*" stress x_2, the resource $\Lambda_{x(\cdot)}(t_1)$ used until the moment t_1 under the stress $x(\cdot)$ is larger than the resource used till the moment t_1 under the stress x_2. Nevertheless, the proportional hazards model states that the rate of resource using is the same after the moment t_1. If, for example, individuals are *aging*, it is not very natural. The hazard rate of individuals who used more of the resource would be higher after the moment t_1. So we need a generalization of the model which includes dependence of the rate of resource using on the *used resource*.

Nevertheless, in survival analysis the PH model usually works quite well, because the values of covariates under which estimation of survival is needed are in the range of covariate values used in experiments. In epidemiologic studies, *cumulative effect* of some variable(s) up to moment t can be naturally created in case the *data history* is known.

So, using a simple model (which could be *not very exact*) often is preferable than using a more complicated model. The case is similar to the classical linear regression models: the mean of the dependent variable is rarely a linear function of the independent variables but the linear approximation works reasonably well in some range of independent variable values.

3.1 Some Properties of the Cox Model on E_1

From (3.1), $\lambda_x(t) = r(x)\ \lambda_0(t), \quad x \in E_1$, it follows that for any constant in time stress $x \in E_1$ the corresponding survival function S_x has the form :

$$S_x(t) = S_0^{r(x)}(t) = \exp\{-r(x)\Lambda_0(t)\}, \quad x \in E_1,$$

where

$$S_0(t) = e^{-\Lambda_0(t)}$$

and

$$\Lambda_0(t) = \int_0^t \lambda_0(s)ds$$

are the baseline survival and cumulative hazards functions.

It is evident that

$$\Lambda_0(t) = -\ln S_0(t),$$

and for any x

$$\Lambda_x(t) = -\ln S_x(t), \quad x \in E_1.$$

Note that for any $x_0 \in E_1$ the Cox model implies

$$\lambda_x(t) = \rho(x_0, x)\lambda_{x_0}(t), \quad \Lambda_x(t) = \rho(x_0, x)\Lambda_{x_0}(t)$$

and

$$S_x(t) = S_{x_0}^{\rho(x_0,x)}(t),$$

where

$$\rho(x_0, x) = \frac{r(x)}{r(x_0)}.$$

Under the PH model on E_1, the hazard ratio $HR(t, x_0, x)$ between different covariates x and x_0 is constant over time:

$$HR(t, x_0, x) = \rho(x_0, x).$$

From the definition of the PH model on E, we have

$$\Lambda_{x(\cdot)}(t) = \int_0^t r(x(u))d\Lambda_0(u)$$

and

$$S_{x(\cdot)}(t) = \exp\left\{-\int_0^t r(x(u))d\Lambda_0(u)\right\}.$$

3.1.1 Tampered Failure Time Model

Now we are able to give several important properties of PH model for simple step-stresses $x(\cdot) \in E_2$.

Let $x(\cdot)$ has the form given by (3.3). Then it is easy to show that under PH model for any $x(\cdot) \in E_2$

$$\lambda_{x(\cdot)}(t) = \begin{cases} \lambda_{x_1}(t), & 0 \le t < t_1, \\ \lambda_{x_2}(t), & t \ge t_1, \end{cases}$$
$$= \begin{cases} r(x_1)\lambda_0(t), & 0 \le t < t_1, \\ r(x_2)\lambda_0(t), & t \ge t_1. \end{cases}$$

From this, for any $x_0 \in E_1$, we have

$$\lambda_{x(\cdot)}(t) = \begin{cases} \rho(x_0, x_1)\lambda_{x_0}(t), & 0 \le t < t_1 \\ \rho(x_0, x_2)\lambda_{x_0}(t), & t \ge t_1. \end{cases}$$

Taking $x_0 = x_1$ we also have

$$\lambda_{x(\cdot)}(t) = \begin{cases} \lambda_{x_1}(t), & 0 \le t < t_1, \\ \rho(x_1, x_2)\lambda_{x_1}(t), & t \ge t_1. \end{cases}$$

By the same manner,

$$\begin{aligned} S_{x(\cdot)}(t) &= \begin{cases} S_{x_1}(t), & 0 \le t < t_1, \\ S_{x_1}(t_1)\frac{S_{x_2}(t)}{S_{x_2}(t_1)}, & t \ge t_1. \end{cases} \\ &= \begin{cases} S_0^{r(x_1)}(t), & 0 \le t < t_1, \\ S_0^{r(x_1)}(t_1)\left(\frac{S_0(t)}{S_0(t_1)}\right)^{r(x_2)}, & t \ge t_1. \end{cases} \end{aligned}$$

And, for any $x_0 \in E_1$,

$$S_{x(\cdot)}(t) = \begin{cases} S_{x_0}^{\rho(x_0,x_1)}(t), & 0 \le t < t_1, \\ S_{x_0}^{\rho(x_0,x_1)}(t_1)\left(\frac{S_{x_0}(t)}{S_{x_0}(t_1)}\right)^{\rho(x_0,x_2)}, & t \ge t_1. \end{cases}$$

Taking $x_0 = x_1$ we then obtain

$$S_{x(\cdot)}(t) = \begin{cases} S_{x_1}(t), & 0 \le t < t_1, \\ S_{x_1}(t_1)\left(\frac{S_{x_1}(t)}{S_{x_1}(t_1)}\right)^{\rho(x_1,x_2)}, & t \ge t_1. \end{cases}$$

The Cox model on E_2 for simple step-stresses of the form (3.3) is known as the *tampered failure rate* (TFR) model (Bhattacharyya and Stoejoeti 1989); see also Bagdonavicius et al. (2002).

3.1.2 Model GM

Let us consider the so-called model GM (*Generalized Multiplicative*) proposed by Bagdonavicius and Nikulin (1995), which generalizes a little the PH model on E. We suppose that there exist a positive function r on E and a survival function S_0 such that for all $x(\cdot) \in E$

$$\frac{\partial f_{x(\cdot)}^G(t)}{\partial t} = r[x(t)]\frac{\partial f_0^G(t)}{\partial t}$$

with the initial conditions

$$f_{x(\cdot)}^G(0) = f_0^G(0) = 0, \quad \text{where} \quad f_0^G(t) = H(S_0(t)), \quad x(\cdot) \in E.$$

We call S_0 the *baseline* survival function. The GM model means that the rate of resource using at the moment t is proportional to some *baseline rate*. The proportionality constant is a function of the stress applied at t.

Taking any two stresses $x(\cdot)$ and $y(\cdot)$ from E we obtain from the definition of the Model GM

$$\frac{\partial f_{x(\cdot)}^G(t)}{\partial t} / \frac{\partial f_{y(\cdot)}^G(t)}{\partial t} = r[x(t)]/r[y(t)],$$

in which one can see that the ratio of resources using at the moment t depends only on the stresses $x(t)$ and $y(t)$. The definition of the model GM implies

$$S_{x(\cdot)}(t) = G\left(\int_0^t r[x(\tau)]dH(S_0(\tau))\right), \quad x(\cdot) \in E.$$

where $H = G^{-1}$. Note that if $x(t) \equiv x = const$, the GM model implies

$$S_{x(\cdot)}(t) = G(r(x)H(S_0(t))), \quad x \in E_1.$$

Consider some *submodels* of GM with G specified. If the distribution of the resource R^G is exponential,

$$G(t) = e^{-t}, \quad t \geq 0,$$

then under the exponential resource we have

$$H = G^{-1}, \quad H(p) = -ln(p), \quad 0 < p < 1,$$

and

$$H(S_0(t)) = -lnS_0(t) = \Lambda_0(t),$$

$$f_{x(\cdot)}^G(t) = H(S_{x(\cdot)}(t)) = -lnS_{x(\cdot)}(t) = \Lambda_{x(\cdot)}(t) = \int_0^t r(x(s))d\Lambda_0(s), \quad x(\cdot) \in E.$$

In this case we obtain the Cox model since

$$\frac{\partial f_{x(\cdot)}^G(t)}{\partial t} = \lambda_{x(\cdot)}(t), \quad x(\cdot) \in E,$$

i.e., the rate of resource using is the hazards rate, and in this case the GM model can be presented by the next way:

$$\lambda_{x(\cdot)}(t) = r(x(t))\lambda_0(t), \quad x(\cdot) \in E.$$

It is the PH model of Cox on E, given by (3.1).

If the distribution of the resource is *log-logistic*,

$$G(t) = \frac{1}{1+t}, \quad t \geq 0,$$

then the GM model can be formulated as

$$\frac{\lambda_{x(\cdot)}(t)}{S_{x(\cdot)}(t)} = r(x(t))\frac{\lambda_0(t)}{S_0(t)}, \quad x(\cdot) \in E.$$

If the stresses are constant in time then we obtain a simple expression on E_1:

$$\frac{1}{S_x(t)} - 1 = r(x)\left(\frac{1}{S_0(t)} - 1\right), \quad x \in E_1.$$

It is the analogue of *logistic regression* model used for the analysis of *dichotomous outcomes* when the probability of *success* depends on some factors. The model given here is close to the Cox model when t is small. It could be useful in practice when the constructed Cox model is not in accordance with the data for small t.

We consider the case when the resource is *lognormal*,

$$G(t) = \Phi(\ln t), \quad t \geq 0,$$

where Φ is the distribution function of the *standard normal law*. If covariates are constant in time then in terms of survival functions the model GM can be written as

$$\Phi^{-1}(S_x(t)) = \ln r(x) + \Phi^{-1}(S_0(t)), \quad x \in E_1.$$

It is the famous *generalized probit model*, see Dabrowska and Doksum (1988).

3.2 Some Simple Examples of Alternatives for the PH Models

Although the PH model works well in many studies, there exist situations when the proportional hazards assumption is not feasible; in particular, when the hazards ratio under different fixed covariates is not constant in time (say, monotone), or when the survivor functions under different stresses intersect, etc. Here are some simple models that could be considered as applicable alternatives to the Cox PH model.

Example 1. Additive Hazard models
The *additive hazard* (AH) model holds on E if the hazard rate under a covariate $x(\cdot)$ is given by

$$\lambda_{x(\cdot)}(t) = \lambda_0(t) + a\left(x(t)\right), \quad x(\cdot) \in E.$$

This model is *nonparametric* if $\lambda_0(\cdot)$ and $a(\cdot)$ are unknown functions. The AH model also has the *absence of memory property* as the Cox model, since the value $\lambda_{x(\cdot)}(t)$ of the hazard rate function $\lambda_{x(\cdot)}(\cdot)$ at the moment t does not depend on the values $x(s)$ for $0 < s < t$. That is, the value of the hazard rate function $\lambda_{x(\cdot)}(\cdot)$ does not depend on the history.

The AH model looks very simple on E_1:

$$\lambda_x(t) = \lambda_0(t) + a(x), \quad x \in E_1.$$

The differences of hazard rates for different covariates $x, y \in E_1$ *does not depend* on the baseline function $\lambda_0(\cdot)$ and t:

$$\lambda_x(t) - \lambda_y(t) = a(x) - a(y), \quad x, y \in E_1.$$

The function $a(\cdot)$ is often parameterized as

$$a(x) = \beta^T x, \quad \beta = (\beta_1, \ldots, \beta_m)^T.$$

In this case we have the classical *semiparametric* AH regression model on E:

$$\lambda_{x(\cdot)}(t) = \lambda_0(t) + \beta^T x(t), \quad x(\cdot) \in E,$$

which is also known as the *McKeague–Sasieni model* (1994). Application of this model is still restrictive because it does not take into account the history represented by the covariate $x(\cdot)$. For semiparametric analysis of McKeague–Sasieni model, one can refer to Lin and Ying (1994) and Martinussen and Scheike (2006).

Example 2. Model of Lin and Ying
The PH and AH models are special cases of the so-called *additive–multiplicative hazard* (AMH) model on E (Lin and Ying 1996):

$$\lambda_{x(\cdot)}(t) = \beta^T x(t)\lambda_0(t) + \gamma^T x(t), \quad x(\cdot) \in E,$$

where $\gamma = (\gamma_1, \ldots, \gamma_m)^T$. Hereafter, it is called the Lin–Ying model. From this formula one can see that Lin–Ying model also has the property of absence of memory.

Example 3. Aalen's additive risk (AAR) model
A different version of the AH model on E_1 was proposed earlier by Aalen (1980). According to the *Aalen's model* on E_1

$$\lambda_x(t) = x^T \lambda_0(t), \quad x \in E,$$

where $\lambda_0(t) = (\lambda_{01}(t), \ldots, \lambda_{0m}(t))^T$ is an unknown vector baseline hazard function, which permits the effects (contribution) of each covariate x_i to be functions of time. Combining the two models (Lin–Ying and AAR) give the well-known *partly parametric additive risk* (PPAR) *model of McKeague and Sasieni* (1994):

$$\lambda_x(t) = y^T \lambda_0(t) + \beta^T z, \quad x(\cdot) \in E_1,$$

where

$$\lambda_0(t) = (\lambda_{01}(t), \ldots, \lambda_{0p}(t))^T, \quad \beta = (\beta_1, \ldots, \beta_q)^T,$$

and

$$y = (x_{1_1}, \ldots x_{1_p})^T \quad \text{and} \quad z = (x_{2_1}, \ldots, x_{2_q})^T$$

are p and q dimensional subvectors ($p \leq m$, $q \leq m$) of the covariate $x = (x_1, \ldots, x_m)^T$.

In general all models considered here are semiparametric, but one can easily parameterize these models as what could be done on the PH model.

3.3 Partial Likelihood Estimation

Consider the set of right-censored data:

$$T_i = \min(T_i^*, C_i), \text{ and } \delta_i = 1_{\{T_i^* \leq C_i\}}$$

where T_i^* and C_i are survival and censoring times, respectively, associated with the ith individual, and $X_i(t)$ is the time-dependent covariate vector. Let t_i be the realization of T_i. Without loss of generality, we assume $t_1 < \cdots < t_n$ when there are no ties. The partial likelihood proposed by Cox (1972, 1975) is

$$L_p = \prod_i \left\{ \frac{e^{\beta^T X_i(t_i)}}{\sum_j Y_j(t_i) e^{\beta^T X_j(t_i)}} \right\}^{\delta_i}. \tag{3.4}$$

Taking partial derivative of $\log\{L_p\}$ with respect to β, we have the following score function:

$$U(\beta) = \sum_i \left\{ X_i(t_i) - \frac{\sum_j Y_j(t_i) X_j(t_i) e^{\beta^T X_j(t_i)}}{\sum_j Y_j(t_i) e^{\beta^T X_j(t_i)}} \right\}^{\delta_i}. \tag{3.5}$$

Note that, in (3.5), β, $U(\beta)$, and $X(\cdot)$ are all p-dimensional vectors. Setting $U(\beta) = 0$ and using iterative scheme for numerical calculation, we can solve $\widehat{\beta}$ such that $U(\widehat{\beta}) = 0$. For a column vector u, $u^{\otimes 2} = u \cdot u^T$. Further define

$$S^{(0)}(\beta, t) = \frac{1}{n} \sum_i Y_i(t) \exp(\beta^T X_i(t)) \tag{3.6}$$

$$S^{(1)}(\beta, t) = \frac{1}{n} \sum_i Y_i(t) X_i(t) \exp(\beta^T X_i(t)) \tag{3.7}$$

$$S^{(2)}(\beta, t) = \frac{1}{n} \sum_i Y_i(t) X_i(t)^{\otimes 2} \exp(\beta^T X_i(t)) \text{ and} \tag{3.8}$$

$$V(\beta, t) = \frac{S^{(2)}(\beta, t)}{S^{(0)}(\beta, t)} - \left\{ \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \right\}^{\otimes 2}. \tag{3.9}$$

Under suitable regularity conditions (Tsiatis 1981; Andersen et al. 1993), large sample properties of $\widehat{\beta}$ can be obtained. In brief, one can show that $\widehat{\beta}$ is consistent and asymptotic normal:

$$\widehat{\beta} \rightarrow_p \beta, \quad \sqrt{n}(\widehat{\beta} - \beta) \rightarrow_d N\left(0, \Sigma_\beta^{-1}\right). \tag{3.10}$$

The inverse of the asymptotic variance

$$\Sigma = \int_0^\tau v(\beta, t) s^{(0)}(\beta, t) \lambda_0(t) dt,$$

where the integration is taken over the entire observational time interval $(0, \tau)$, and $v(\beta, t)$, $s^{(0)}(\beta, t)$, $s^{(1)}(\beta, t)$, and $s^{(2)}(\beta, t)$ are the asymptotic limits of

$$V(\beta, t), S^{(0)}(\beta, t), S^{(1)}(\beta, t) \quad \text{and} \quad S^{(2)}(\beta, t),$$

respectively.

Based on the partial likelihood estimation, the PH analysis has the merits of easy implementation, semiparametric efficient, nice interpretation in hazard ratio, and readily available packages in various statistical softwares (such as SAS, S-plus, R, SPSS, etc.)

3.3.1 Breslow Estimator for the Baseline Cumulative Hazard Function

Once $\widehat{\beta}$ has been obtained, the baseline cumulative hazard can be estimated by

$$\widehat{\Lambda}_0(t) = \sum_{i=1}^n \int_0^t \frac{dN_i(u)}{\sum_j Y_j(t_i) \exp\{\widehat{\beta}^T Z_j(t_i)\}}.$$

It is called the Breslow estimator of $\Lambda_0(t)$ (Breslow 1974). It has been argued that, profiled on $\widehat{\beta}$, the estimator $\widehat{\Lambda}_0(t)$ is the MLE for the full likelihood $L(\beta, \Lambda_0)$; moreover, $L_p(\beta) = \max_{\Lambda_0(\cdot)} L(\beta, \Lambda_0)$ (Johansen 1983; Andersen et al. 1993).

3.3.2 *The Stanford Heart Transplant Data as an Example*

For the Stanford Heart Transplant (SHT) data introduced in Example 1 of Chap. 1, see Wu and Hsieh (2009) for more explorations. A complete data with 154 observations are used here to implement the conventional PH analysis and compared with the parametric Weibull regression mentioned previously. One can see from the following table that relevant parameter estimates are quite close for the two analysis.

	Cox model	Weibull regression
Variable	Parameter ($\times 10^{-2}$) (p-value)	Parameter ($\times 10^{-2}$) (p-value)
Age	$-12.51(0.0227)$	$-13.34(0.0154)$
Age^2	0.20(0.0050)	0.22(0.0028)
Mismatch score	13.71(0.4548)	14.79(0.4200)

The above Weibull regression has the baseline estimate $\widehat{a}t^{\widehat{b}}$, $\widehat{a} = 0.06119$, and $\widehat{b} = 0.5976$. Comparison for the baseline cumulative hazards for the two models (Cox and Weibull) are shown in Fig. 3.1. It exhibits a lower (smaller) estimate for the baseline cumulative hazard using the Cox regression than the parametric Weibull regression.

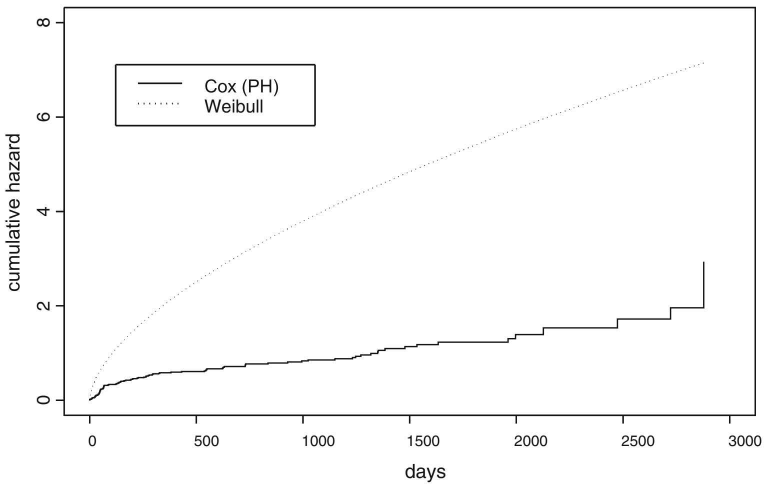

Fig. 3.1 Comparison of baselines: Cox (PH) versus Weibull

3.4 Log-Rank Test and Robust Tests for Treatment Effect

3.4.1 Log-Rank Test and Weighted Log-Rank Test

For the ordered failured times $t_{(1)} < \cdots < t_{(k)}$, let $Z = 1$ codes for the treatment group and $Z = 0$ for the control group. Assume that there are d_i failures observed at $t_{(i)}$, then a 2×2 contingency table is formed as

	Died	Survived	Total
$Z = 1$	d_{1i}	$Y_{1i} - d_{1i}$	Y_{1i}
$Z = 0$	d_{0i}	$Y_{0i} - d_{0i}$	Y_{0i}
total	d_i	$Y_i - d_i$	Y_i

Here Y_{1i} and Y_{0i} are the at-risk number of the group $Z = i$ $(i = 1, 0)$ with $Y_i = Y_{1i} + Y_{0i}$, and d_{1i} and d_{0i} are, respectively, the number of failures observed at $t_{(i)}$. Conditional on the marginal totals, the random variable d_{1i} is distributed as a *hypergeometric distribution* with mean and variance:

$$E(d_{1i}) = d_i \frac{Y_{1i}}{Y_i},$$
$$Var(d_{1i}) = d_i \frac{Y_{1i} Y_{0i}(Y_i - d_i)}{Y_i^2 (Y_i - 1)}.$$

By this setting, a series of "correlated" 2×2 tables are formed. However, if the data history satisfies the requirement of *increasing σ-algebras* (a filtration; Sect. 2.3), then the *martingale central limit theorem* can be applied so that

$$\mathscr{T}_{LR} = \frac{\left\{\sum_{i=1}^{k} (d_{1i} - E(d_{1i}))\right\}^2}{\sum_{i=1}^{k} Var(d_{1i})} \tag{3.11}$$

has asymptotically a χ_1^2 distribution. If there are **no ties**, then $d_i = 1$. The log-rank statistic (Mantel 1966; Peto 1972) can be simplified as

$$\mathscr{T}_{LR} = \frac{\left\{\sum_{i=1}^{k} (d_{1i} - \bar{d}_{1i})\right\}^2}{\sum_{i=1}^{k} \bar{d}_{1i}(1 - \bar{d}_{1i})}, \tag{3.12}$$

where $\bar{d}_{1i} = Y_{1i}/Y_i$ is the sample mean of the indicators of treatment group. A class of *predictable* weight processes $\mathscr{K}(t)$ can be imposed to produce *weighted log-rank* tests (Klein and Moeschberger 1997):

$$\mathscr{T}_{LR} = \frac{\left\{\sum_{i=1}^{k} \mathscr{K}(t_{(i)})(d_{1i} - E(d_{1i}))\right\}^2}{\sum_{i=1}^{k} \mathscr{K}^2(t_{(i)})Var(d_{1i})}. \tag{3.13}$$

See Sect. 7.3 for more details for the selection of $\mathscr{K}(t)$. It is easy practice to derive the log-rank test as a score test under the proportional hazards model and using the partial likelihood if there are no covariates and the only explanatory variable is the treatment indicator.

3.4.2 *Robust Inference: Preliminary*

The Cox model (Cox 1972) is the most popular model used in survival analysis for epidemiological data, clinical studies, and many other fields. When there could be model misspecification, robust inferential procedures should be considered. Model misspecification involves several aspects: First, some observations do not actually respond to the covariate in proportional hazards setting. Second, there could be omitted covariates. And, third, variables may have measurement errors. With the general concern of model misspecification, Lin and Wei (1989) derived the asymptotic distribution of the maximum partial likelihood estimator (MPLE) with a *sandwich* variance estimator, and proposed robust tests for treatment effect possibly with covariates adjustment. The procedures suggested in Lin and Wei (1989) are as follows.

Consider a set of random failure times $T_1^*, \ldots, T_n^*$ subject to random right censoring times $C_1, \ldots, C_n$. We observe $T_i = \min(T_i^*, C_i)$ and $\delta_i = 1(T_i^* \leq C_i)$; and $Y_i(t) = 1_{\{T_i \geq t\}}$ is the "at-risk" process. Let the hazard function associated with a set of treatment/covariates be

$$\lambda(t; Z, X) = \lambda_0(t)\exp\{\phi Z + \beta_1 X_1(t) + \cdots + \beta_p X_p(t)\},$$

where Z is the indicator of treatment. Denote $\theta = (\phi, \beta_1, \ldots, \beta_p)^T = (\phi, \beta^T)^T$, $\tilde{X}(t) = (X_1(t), \ldots, X_p(t))^T$, and $X(t) = (Z, \tilde{X}(t)^T)^T$. For individual i, $\lambda_i(t) = \lambda(t; X_i(t))$, $X_i(t) = (Z_i, \tilde{X}_i(t)^T)^T$. Assume that the treatment is assigned by suitable random allocation so that Z is reasonably independent of $\tilde{X}(t)$. Let $l(\theta)$ be the partial likelihood for the "specified" model and further denote

$$H_\theta = -(1/n)\partial^2 l(\theta)/\partial\theta^2 = \begin{pmatrix} H_{\phi\phi} & H_{\phi\beta} \\ H_{\beta\phi} & H_{\beta\beta} \end{pmatrix},$$

$$J_\theta = (1/n)\sum(U_i(\theta)U_i(\theta)^T) = \begin{pmatrix} J_{\phi\phi} & J_{\phi\beta} \\ J_{\beta\phi} & J_{\beta\beta} \end{pmatrix},$$

where $U_i(\theta) = \partial l_i(\theta)/\partial\theta$ with l_i being the contribution from the ith observation to $l(\theta)$; $\sum U_i(\theta) = U_\theta = (U_\phi, U_\beta^T)^T$, $U_\phi = \sum U_{i\phi}$, $U_\beta = \sum U_{i\beta}$.

As was discussed by Lin and Wei (1989) for "fully parametric" problems, a possible *robust* Wald test for $H_0 : \phi = 0$ (means "no treatment effect") is, on one hand,

$$T_{\mathscr{W}1} = n\widehat{\phi}^2\{\widehat{V}_{\phi\phi}\}^{-1},$$

with $\widehat{V}_{\phi\phi}$ being the (1, 1)th component of $H_\theta^{-1} J_\theta H_\theta^{-1}$ evaluated at $\widehat{\theta} = (0, \widehat{\beta}_0)$, where $\widehat{\beta}_0$ is the restricted MPLE under $H_0 : \phi = 0$. On the other hand, a *robust* score test can be constructed as

$$T_{\mathscr{S}1} = \frac{U_\phi^2(0, \widehat{\beta}_0)}{\sum\{U_{i\phi}(0, \widehat{\beta}_0) - H_{\phi\beta}(0, \widehat{\beta}_0) H_{\beta\beta}^{-1}(0, \widehat{\beta}_0) U_{i\beta}(0, \widehat{\beta}_0)\}^2}.$$

3.4.3 Robust Test with Covariates Adjustment

Lin and Wei's test based on Cox model

We need the following notations:

$$S^{(r)}(t) = n^{-1}\sum_i Y_i(t)\lambda_i(t)X_i(t)^{\otimes r}, s^{(r)}(t) = ES^{(r)}(t),$$
$$S^{(r)}(\theta, t) = n^{-1}\sum_i Y_i(t)\exp\{\theta^T X_i(t)\}X_i(t)^{\otimes r}, s^{(r)}(\theta, t) = ES^{(r)}(\theta, t),$$

where $r = 0, 1, 2$ and $q^{\otimes 2}$ equals qq^T for a column vector q. With possible model misspecification, assume that there exists a θ^* satisfying

$$\int s^{(1)}(t)dt - \int \frac{s^{(1)}(\theta, t)}{s^{(0)}(\theta, t)} s^{(0)}(t)dt = 0. \tag{3.14}$$

It was proved by Lin and Wei that the MPLE $\widehat{\theta}$ is consistent and asymptotic normal:

$$\sqrt{n}(\widehat{\theta} - \theta^*) \to N(0, H_\theta^{-1} J_\theta^* H_\theta^{-1}),$$

where J_θ^* can be consistently estimated by $\widehat{J_\theta^*}$ calculated at $\widehat{\theta} = (0, \widehat{\beta}_0)$:

$$\begin{aligned}\widehat{J_\theta^*} &= \frac{1}{n}\sum_i \delta_i \left\{\left\{X_i(t) - \frac{S^{(1)}(\theta, t_i)}{S^{(0)}(\theta, t_i)}\right\} - \sum_j \frac{\delta_j Y_i(t_j) e^{\theta^T X_i(t_j)}}{nS^{(0)}(\theta, t_j)}\left\{X_i(t_j) - \frac{S^{(1)}(\theta, t_j)}{S^{(0)}(\theta, t_j)}\right\}\right\}^{\otimes 2} \\ &\equiv \frac{1}{n}\sum_i \delta_i \mathscr{W}_i^{\otimes 2},\end{aligned} \tag{3.15}$$

The term in the big brackets, $\mathscr{W}_i$, can be decomposed as $(\mathscr{W}_{i\phi}, \mathscr{W}_{i\beta}^T)^T$. Based on this property, Lin and Wei proposed another robust score test:

$$T_{LW} = \frac{U_\phi^2(0, \widehat{\beta}_0)}{\sum\{\mathscr{W}_{i\phi}(0, \widehat{\beta}_0) - H_{\phi\beta}(0, \widehat{\beta}_0) H_{\beta\beta}^{-1}(0, \widehat{\beta}_0) \mathscr{W}_{i\beta}(0, \widehat{\beta}_0)\}^2}. \tag{3.16}$$

The condition (3.14) can be easily interpreted as follows. For individual i, let the hazard of the *true model* be $\lambda_i(t)$ and

$$\lambda(t; Z_i, X_i(t)) = \lambda_0(t) \exp\{\phi Z_i + \beta_1 X_{1i}(t) + \cdots + \beta_p X_{pi}(t)\},$$

be the hazard of the *working model*. For ease of further explorations, assume that under the true model the estimating equation corresponding to the treatment Z can be expressed as

$$\sum \delta_i \left\{ Z_i - \frac{\sum_j Y_j(t_i)\lambda_j(t_i) Z_j}{\sum_j Y_j(t_i)\lambda_j(t_i)} \right\} = 0. \tag{3.17}$$

For the working model, the estimating equation is

$$\sum \delta_i \left\{ Z_i - \frac{\sum_j Y_j(t_i) \exp(\theta^T X_j(t_i)) Z_j}{\sum_j Y_j(t_i) \exp(\theta^T X_j(t_i))} \right\} = 0. \tag{3.18}$$

Combining (3.17) and (3.18), we have

$$\frac{1}{n} \sum \delta_i \frac{\frac{1}{n}\sum_j Y_j(t_i)\lambda_j(t_i) Z_j}{\frac{1}{n}\sum_j Y_j(t_i)\lambda_j(t_i)} = \frac{1}{n} \sum \delta_i \frac{\frac{1}{n}\sum_j Y_j(t_i) \exp(\theta^T X_j(t_i)) Z_j}{\frac{1}{n}\sum_j Y_j(t_i) \exp(\theta^T X_j(t_i))}. \tag{3.19}$$

According to the arguments of Tsiatis (1981), the left-hand side of (3.19) converges in probability to

$$\int E\left(\frac{1}{n}\sum_j Y_j(t)\lambda_j(t) Z_j\right) dt;$$

and the right-hand side converges to

$$\int \frac{E\left[\frac{1}{n}\sum_j Y_j(t) Z_j \exp\{\theta^T X_j(t)\}\right]}{E\left[\frac{1}{n}\sum_j Y_j(t) \exp\{\theta^T X_j(t)\}\right]} E\left(\frac{1}{n}\sum_j Y_j(t)\lambda_j(t)\right) dt.$$

Or, using the notations in Lin and Wei's paper (1989, formula (2.1)), (3.19) implies

$$\int s^{(1)}(t)dt = \int \frac{s^{(1)}(\theta, t)}{s^{(0)}(\theta, t)} s^{(0)}(t)dt,$$

which is exactly (3.14).

An Improvement

Evidently, Lin and Wei's robust variance estimate $\widehat{H}_\theta^{-1} \widehat{J}_\theta^* \widehat{H}_\theta^{-1}$ can be reexpressed as

$$\frac{1}{n} \sum \widehat{IF}_i \widehat{IF}_i^T,$$

where $IF_i = H_\theta^{-1} \mathscr{W}_i$ is the *influence function* and $\widehat{IF}_i$ is the counterpart when the involved θ is replaced by its consistent estimate (Reid and Crépeau 1985; Heritier et al. 2009.) For the purpose of robust estimation, Bednarski (1993) and Minder and Bednarski (1996) proposed an alternative robust procedure based on the following weighted partial score:

$$U_W(\beta) = \sum_i W(t_i, X_i(t_i)) \left\{ X_i(t_i) - \frac{\sum_j Y_j(t_i) W(t_i, X_j(t_i)) X_j(t_i) e^{\beta^T X_j(t_i)}}{\sum_j Y_j(t_i) W(t_i, X_j(t_i)) e^{\beta^T X_j(t_i)}} \right\}^{\delta_i}.$$

The weight functions appeared at two places have different purposes: for that at the outer sum, it downweights observations with large $t \exp(\beta^T X)$; for the two weights at the numerator and denominator in the curly brackets, they downweight observations with large $\beta^T X$ (Minder and Bednarski 1996). Three basic weights can be used: linear, exponential, and quadratic. We denote the resultant robust estimate as $\widehat{\beta}_{RE}$. Let $\beta = (\phi, \gamma)$ and if the effects of a subset of the covariates are tested, say, $H_0 : \phi = 0$. Then, using the approach of Bednarski (1993) and Minder and Bednarski (1996), robust Wald test and score test can be constructed from replacing the partial score involved in the Lin and Wei's procedure by its counterpart $U_W(\beta)$ and the consistent estimate $\widehat{\beta}_{RE}$. For more details, see Sect. 7.3 of Heritier et al. (2009).

Kong and Slud's Test

Assuming independence between Z and the covariates $\tilde{X}(t)$, Kong and Slud (1997) showed that $U_\phi(0, \widehat{\beta}_0)$ is asymptotically distributed as a zero-mean Gaussian distribution with covariance Σ which can be consistently estimated by

$$\widehat{\Sigma} = \frac{1}{n} \sum_i \delta_i (\mathscr{W}_i^0 - \bar{\mathscr{W}}^0)^2.$$

In the above expression

$$\mathscr{W}_i^0 = \left\{ Z_i - \frac{\sum_k Z_k Y_k(t_i)}{\sum_k Y_k(t_i)} \right\} - \sum_j \frac{\delta_j Y_i(t_j) e^{\theta^T X_i(t_j)}}{n S^{(0)}(\theta, t_j)} \left\{ Z_i - \frac{\sum_k Z_k Y_k(t_j)}{\sum_k Y_k(t_j))} \right\}$$

calculated at $\widehat{\theta} = (0, \widehat{\beta}_0)$, and $\bar{\mathscr{W}}^0$ is the sample mean of $\mathscr{W}_i^0 (i = 1, \ldots, n)$. A χ^2-statistic can then be constructed as

$$T_{KS} = \frac{U_\phi^2(0, \widehat{\beta}_0)}{\frac{1}{n} \sum_i \delta_i \left(\mathscr{W}_i^0 - \bar{\mathscr{W}}^0 \right)^2}. \tag{3.20}$$