Guan, Z. and Peng, C. 2010. Two-sample Semiparametric Proportional (Reverse) Hazard Model. In JSM Proceedings, Section on Nonparametric Statistics. Alexandria, VA: American Statistical Association. 2067-2080.

Two-sample Semiparametric Proportional (Reverse) Hazard Model

Zhong Guan*   Cheng Peng†

Abstract
The proportional hazards and proportional reverse hazards model is shown to be equivalent in some sense. The rank-based semiparametric full likelihood is constructed and used to find maximum likelihood estimate of the proportionality parameter and the underlying nonparametric distribution simultaneously. The proposed estimates are based on rank statistics and thus distribution-free. Simulation methods for constructing confidence interval and for approximating p-values of the hypothesis tests are described.

Key Words: Distribution-free; Rank-based likelihood; Order statistics; Proportional (reverse) hazards model; Ranks; Semiparametric model; Reliability; Survival data

1. Introduction

Although the two-sample proportional hazards model is less important than the more general proportional hazards regression model of Cox (1972, 1975) in survival analysis of life-tables in the area of biostatistics, there are still important applications in the area of reliability. Another important model in the area of reliability is the so-called two-sample proportional reverse hazard rate model of Lehmann (1953) which fits the reverse transformed data of two-sample proportional hazards model. In both survival analysis and reliability, it is also important to estimate the survival function. However, Cox's partial likelihood method focus only on the estimation of the parameters. The baseline survival function is estimated parametrically in a very computationally efficient way but with the true value of the regression coefficient being replaced by Cox's partial likelihood estimate. The commonly used method of estimating survival functions is that of Kalbfleisch and Prentice (1973) who suggest an alternative discrete proportional hazards model that is differently from the one in Cox (1972).

1.1 Two-sample Proportional Hazards Model:

Let $F(x)$ be a baseline (or control group) distribution of random variable X with survival function $\bar{F}(x) = 1 - F(x)$. Let $G(y)$ be the treatment group distribution of random variable Y with survival function $\bar{G}(y) = 1 - G(y)$. The hazard rate is defined as $h_F(t) = d \log \bar{F}(t)/dt$. Then the proportional hazards model is defined by:

$$\frac{h_G(t)}{h_F(t)} = \frac{g(t)/\bar{G}(t)}{f(t)/\bar{F}(t)} = \theta \tag{1}$$

where $g(t) = G'(t)$, $f(t) = F'(t)$, and $\theta > 0$. Model (1) is equivalent to $\log \bar{G}(t) = \theta \log \bar{F}(t)$ and

$$\bar{G}(t) = \{\bar{F}(t)\}^{\theta}. \tag{2}$$

Clearly $G(t)$ and $F(t)$ share the same support $S \in R$.

*Department of Mathematical Sciences, Indiana University South Bend
†Department of Mathematics and Statistics, University of Southern Maine

1.2 Two-sample Proportional Reverse Hazard Rate Model:

Lehmann (1953) proposed alternative distribution G of F as an exponentiated function of $F(t)$, i.e.,

$$G(t) = \{F(t)\}^{\theta}, \tag{3}$$

for $\theta > 0$ and t in the common support of F and G. Clearly $\log G(t)/\log F(t) = \theta$ and

$$dG(t) = \theta\{F(t)\}^{\theta-1} dF(t). \tag{4}$$

The inverse of G is

$$t = G^{-1}(u) = F^{-1}(u^{1/\theta}). \tag{5}$$

If F is uniform$(0, 1)$, $F(t) = t$, $0 < t < 1$, then $G(t) = t^{\theta}$, $G^{-1}(u) = u^{1/\theta}$.

The reverse hazard rate associated with $F(t)$ is defined as

$$r_F(t) = \frac{d \log F(t)}{dt}.$$

Because $r_G(t)/r_F(t) = \theta$, (3) is called proportional reversed hazard rate model (PRHRM) (Gupta and Gupta, 2007). In reliability modeling and survival analysis, the proportional model has applications in modeling left censored data. For the semi-parametric proportional reverse hazard model, except the exponential structure, the form (4) is simpler than the ones in the proportional hazards model and the proportional odds model.

1.3 Relationship Between the Two Models:

Because $G^*(s) = \bar{G}(-s)$ and $F^*(s) = \bar{F}(-s)$ are the cdf's of the reversed random variables $X^* = -X$ and $Y^* = -Y$ respectively and have the same support $S^* = -S = \{-s \mid s \in S\}$. Hence (2) is equivalent to the following so-called proportional reverse hazard rate model (3) for F^* and G^*: $G^*(s) = \{F^*(s)\}^{\theta}$ and Cox's partial likelihood method applies.

Because the proportional reverse hazard rate model (3) is a little simpler than the celebrated proportional hazard rate model (2) and is convenient in notations, we will investigate the statistical inference for parameter θ and the underlying distribution F in the semiparametric model (3) using a rank-based likelihood method. In case where the proportional hazard rate model (2) is concerned, one can reverse the sample values. Similar to the proportional odds rate model (Guan and Peng, 2009), if one switches F and G, (3), the ratio of reverse hazard rates becomes $r_F(t)/r_G(t) = 1/\theta$. So a good estimator of θ must be *reciprocal symmetric* in the sense that $\hat{\theta}(x, y) = 1/\hat{\theta}(y, x)$.

It is also worth to note that (3) is a special case with $c = 0$ of the following more general one.

$$F^{c+1}(t) dG(t) = \theta G^{c+1}(t) dF(t), \quad c \geqslant 0. \tag{6}$$

If $c = 1$, (6) is the proportional odds ratio model (see Dabrowska and Doksum, 1988a,b; Guan and Peng, 2009; Wu, 1995, for example)

2. The Full Likelihood

Assume that the two independent samples $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$ are from F and G respectively. Let the combined sample be $\{z_1, \ldots, z_N\} = \{x_1, \ldots, x_m; y_1, \ldots, y_n\}$. Denote the order statistics of $z_1, \ldots, z_N$ by $z_{(1)}, \ldots, z_{(N)}$. Let $p_k = dF\{z_{(k)}\}$, $k = 1, \ldots, N$, be the nonnegative jumps of F with total mass unity, so that

$$F(x) = \sum_{k=1}^{N} p_k I_{\{z_{(k)} \leqslant x\}},$$

where I_A is the indicator of A.

Let r_j be the rank of y_j in the pooled sample $z_1, \ldots, z_N$(in case of tie, maximum rank is used). So $F(y_j) = \sum_{k=1}^{r_j} p_k$. Denote $\boldsymbol{p} = (p_1, \ldots, p_N)^{\mathrm{T}}$. We can define the semiparametric likelihood function

$$\mathscr{L}(\theta, \boldsymbol{p}) = \prod_{i=1}^{m} dF(x_i) \prod_{j=1}^{n} dG(y_j) = \theta^n \prod_{k=1}^{N} p_k \prod_{j=1}^{n} \Big(\sum_{u=1}^{r_j} p_u \Big)^{\theta - 1} \tag{7}$$

and semiparametric log-likelihood function

$$\ell(\theta, \boldsymbol{p}) = n \log \theta + \sum_{k=1}^{N} \log p_k + (\theta - 1) \sum_{j=1}^{n} \log \Big(\sum_{u=1}^{r_j} p_u \Big), \tag{8}$$

where

$$\sum_{k=1}^{N} p_k = 1, \quad \theta > 0, \quad p_k \geqslant 0, \quad k = 1, \ldots, N. \tag{9}$$

Define Lagrangian function

$$\Lambda(\theta, \boldsymbol{p}) = \ell(\theta, \boldsymbol{p}) + \tau \left(1 - \sum_{k=1}^{N} p_k \right),$$

where τ is a Lagrange multiplier. Differentiating $\Lambda(\theta, \boldsymbol{p})$ with respect to θ, we have score

$$h_0(\theta, \boldsymbol{p}) \equiv \frac{\partial \Lambda(\theta, \boldsymbol{p})}{\partial \theta} = \frac{n}{\theta} + \sum_{j=1}^{n} \log \Big(\sum_{u=1}^{r_j} p_u \Big) \tag{10}$$

with derivatives

$$\frac{\partial h_0(\theta, \boldsymbol{p})}{\partial \theta} = -\frac{n}{\theta^2}, \tag{11}$$

$$\frac{\partial h_0(\theta, \boldsymbol{p})}{\partial p_k} = \sum_{j=1}^{n} \frac{I_{\{k \leqslant r_j\}}}{\sum_{u=1}^{r_j} p_u}. \tag{12}$$

Differentiating $\Lambda(\theta, \boldsymbol{p})$ with respect to p_k, we have

$$\frac{\partial \Lambda(\theta, \boldsymbol{p})}{\partial p_k} = \frac{1}{p_k} + \sum_{j=1}^{n} \frac{(\theta - 1) I_{\{k \leqslant r_j\}}}{\sum_{u=1}^{r_j} p_u} - \tau = 0. \tag{13}$$

It follows from (13) that

$$0 = \sum_{k=1}^{N} p_k \frac{\partial \Lambda(\theta, \boldsymbol{p})}{\partial p_k} = N + \sum_{j=1}^{n} \frac{(\theta - 1) \sum_{u=1}^{r_j} p_u}{\sum_{u=1}^{r_j} p_u} - \tau = N + n(\theta - 1) - \tau \tag{14}$$

and thus

$$\tau = N + n(\theta - 1).$$

Then equation (13) can be rewritten as

$$h_k(\theta, \boldsymbol{p}) \equiv \frac{1}{p_k} + (\theta - 1) \sum_{j=1}^{n} \frac{I_{\{k \leqslant r_j\}}}{\sum_{u=1}^{r_j} p_u} - N - n(\theta - 1) = 0. \tag{15}$$

The derivatives of the score function $h_k(\theta, \boldsymbol{p})$ are

$$\frac{\partial h_k(\theta, \boldsymbol{p})}{\partial \theta} = \sum_{j=1}^{n} \frac{I_{\{k \leqslant r_j\}} - \sum_{u=1}^{r_j} p_u}{\sum_{u=1}^{r_j} p_u} = \sum_{j=1}^{n} \frac{I_{\{k \leqslant r_j\}}}{\sum_{u=1}^{r_j} p_u} - n,$$

and, for $k, r = 1, \ldots, N$,

$$\frac{\partial h_k(\theta, \boldsymbol{p})}{\partial p_r} = -\frac{\delta_{k,r}}{p_k^2} - (\theta - 1) \sum_{j=1}^{n} \frac{I_{\{k \vee r \leqslant r_j\}}}{\left(\sum_{u=1}^{r_j} p_u\right)^2},$$

where $\delta_{k,r} = I_{\{k=r\}}$ is the Kronecker delta and $a \vee b = \max\{a, b\}$.

Let $\boldsymbol{p}(\theta)$ be the solution of the equation (15) for any given θ. Clearly, $\tilde{\boldsymbol{p}} \equiv \boldsymbol{p}(1) = (1/N, \ldots, 1/N)^{\mathrm{T}}$. The profile log-likelihood of θ is

$$\ell(\theta) = n \log \theta + \sum_{k=1}^{N} \log p_k(\theta) + (\theta - 1) \sum_{j=1}^{n} \log \Big\{ \sum_{u=1}^{r_j} p_u(\theta) \Big\}. \tag{16}$$

By (15), and that for any θ, $\sum_{k=1}^{N} p_k(\theta) = 1$,

$$\ell'(\theta) = \frac{n}{\theta} + \sum_{j=1}^{n} \log \Big\{ \sum_{u=1}^{r_j} p_u(\theta) \Big\}. \tag{17}$$

Therefore

$$\ell''(\theta) = -\frac{n}{\theta^2} + \sum_{j=1}^{n} \frac{\sum_{u=1}^{r_j} p'_u(\theta)}{\sum_{u=1}^{r_j} p_u(\theta)}. \tag{18}$$

The Fisher information can be defined as

$$I(\theta) \equiv -\lim_{m,n \to \infty} \frac{\ell''(\theta)}{N}. \tag{19}$$

The maximum likelihood estimate $\hat{\theta}$ is the maximizer of $\ell(\theta)$. Thus the maximum likelihood estimate $\boldsymbol{p}$ is $\boldsymbol{p}(\hat{\theta})$. Clearly, $\hat{\theta}$ is reciprocal symmetric and distribution-free. The log-likelihood-ratio statistic for hypotheses $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ is

$$\mathcal{LR}(\theta_0) = 2\{\ell(\hat{\theta}, \hat{\boldsymbol{p}}) - \ell(\theta_0)\} = 2\{\ell(\hat{\theta}) - \ell(\theta_0)\}. \tag{20}$$

When $\theta_0 = 1$, the above hypotheses are equivalent to the hypotheses $H_0 : F = G$ against the alternative (3) with $\theta \neq 1$. Except for $\theta = 1$, numerical method has to be used to find $\hat{p}_k$ and $\hat{\theta}$ which maximize $\ell(\theta, \boldsymbol{p})$.

3. Numerical Procedures

3.1 Newton Iterations

Denote $H(\theta, \boldsymbol{p}) = \{h_1(\theta, \boldsymbol{p}), \ldots, h_N(\theta, \boldsymbol{p})\}^{\mathrm{T}}$, $q_j = \sum_{u=1}^{r_j} p_u$, $\boldsymbol{q} = (q_1, \ldots, q_n)^{\mathrm{T}}$, $E_n = (1, \ldots, 1)^{\mathrm{T}}_{n\times 1}$, $\Pi = \mathrm{diag}\{p_1, \ldots, p_N\}$, and

$$D = \begin{pmatrix} \frac{I_{\{1\leqslant r_1\}}}{q_1} & \cdots & \frac{I_{\{1\leqslant r_n\}}}{q_n} \\ \vdots & \ddots & \vdots \\ \frac{I_{\{N\leqslant r_1\}}}{q_1} & \cdots & \frac{I_{\{N\leqslant r_n\}}}{q_n} \end{pmatrix}.$$

It is easy to see that

$$D^{\mathrm{T}}\Pi E_N = E_n,$$

Then score functions and Jacobian matrix can be written as

$$\begin{aligned} h_0(\theta, \boldsymbol{p}) &= \frac{n}{\theta} + E_n^{\mathrm{T}} \log \boldsymbol{q}, \\ H(\theta, \boldsymbol{p}) &= [\Pi^{-1} - \{N + (\theta - 1)n\}I_N]E_N + (\theta - 1)DE_n, \\ A(\theta, \boldsymbol{p}) &\equiv \begin{pmatrix} \frac{\partial h_0(\theta,\boldsymbol{p})}{\partial\theta} & \frac{\partial h_0(\theta,\boldsymbol{p})}{\partial \boldsymbol{p}^{\mathrm{T}}} \\ \frac{\partial H(\theta,\boldsymbol{p})}{\partial\theta} & \frac{\partial H(\theta,\boldsymbol{p})}{\partial \boldsymbol{p}^{\mathrm{T}}} \end{pmatrix} \equiv \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \\ &= \begin{pmatrix} -\frac{n}{\theta^2} & E_n^{\mathrm{T}} D^{\mathrm{T}} \\ DE_n - nE_N & B(\theta, \boldsymbol{p}) \end{pmatrix} \end{aligned}$$

where

$$B(\theta, \boldsymbol{p}) \equiv \frac{\partial}{\partial \boldsymbol{p}^{\mathrm{T}}} H(\theta, \boldsymbol{p}) = -\{\Pi^{-2} + (\theta - 1)DD^{\mathrm{T}}\}.$$

The inverse of $A(\theta, \boldsymbol{p})$ as partitioned matrix is

$$[A(\theta, \boldsymbol{p})]^{-1} = \begin{pmatrix} A_{11\cdot 2}^{-1} & -A_{11\cdot 2}^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}A_{11\cdot 2}^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}A_{11\cdot 2}^{-1}A_{12}A_{22}^{-1} \end{pmatrix}$$

where $A_{11\cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$ and $A_{22}^{-1} = [B(\theta, \boldsymbol{p})]^{-1}$.

$$A_{22}^{-1} = [B(\theta, \boldsymbol{p})]^{-1} = -[\Pi^2 - (\theta - 1)\Pi^2 D\{I_n + (\theta - 1)D^{\mathrm{T}}\Pi^2 D\}^{-1}D^{\mathrm{T}}\Pi^2].$$

These formulas are useful in implementing the following Newton algorithms in computer languages.

Newton iteration for finding $\hat{\theta}$ and $\hat{p}$ simultaneously:

$$\begin{pmatrix} \theta^{(s+1)} \\ \boldsymbol{p}^{(s+1)} \end{pmatrix} = \begin{pmatrix} \theta^{(s)} \\ \boldsymbol{p}^{(s)} \end{pmatrix} - \{A(\theta^{(s)}, \boldsymbol{p}^{(s)})\}^{-1} \begin{pmatrix} h_0(\theta^{(s)}, \boldsymbol{p}^{(s)}) \\ H(\theta^{(s)}, \boldsymbol{p}^{(s)}) \end{pmatrix}, \quad s = 0, 1, 2, \ldots.$$

start with initial values $p_k^{(0)} = \tilde{p}_k(\theta^{(0)})$, $k = 1, \ldots, N$, and $\theta^{(0)}$. The initial value $\theta^{(0)}$ of θ can be found using Newton iteration for uniform model with y_j being replaced by $\hat{y}_j = \hat{F}(y_j)$. Another good initial guess of θ is $\hat{\theta}_0 = \hat{\vartheta}_0(y, x)$.

Newton iteration for finding $\boldsymbol{p}(\theta)$ with fixed θ: We use the following Newton iteration to calculate $\mathcal{LR}(\theta)$ for each given value of θ.

$$\boldsymbol{p}^{(s+1)}(\theta) = \boldsymbol{p}^{(s)}(\theta) - \{B(\theta, \boldsymbol{p}^{(s)})\}^{-1} H\{\theta^{(s)}, \boldsymbol{p}^{(s)}(\theta)\}, \quad s = 0, 1, 2, \ldots.$$

For each θ, the initial values for p_k are $p_k^{(0)} = \tilde{p}_k(\theta)$, $k = 1, \ldots, N$.

It is easy to show that for a fixed θ,

$$\boldsymbol{p}'(\theta) = [B(\theta, \boldsymbol{p})]^{-1}(DE_n - nE_N)$$

and therefore

$$\ell''(\theta) = -\frac{n}{\theta^2} + (DE_n)^{\mathrm{T}}\boldsymbol{p}'(\theta) = -\frac{n}{\theta^2} + (DE_n)^{\mathrm{T}}[B(\theta, \boldsymbol{p})]^{-1}(DE_n - nE_N)$$

3.2 Initial value of θ for Newton method

It is easy to see that (cf. Gupta and Gupta, 2007) if F is known, then the parametric maximum likelihood estimate of θ is

$$\hat{\theta}_F = \frac{-n}{\sum_{j=1}^n \log F(y_j)}.$$

Let $\hat{F}$ and $\hat{G}$ be the empirical distribution functions of samples $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$ respectively. That is,

$$\hat{F}(x) = \frac{1}{m}\sum_{i=1}^m I_{\{x_i \leqslant x\}}, \quad \hat{G}(y) = \frac{1}{n}\sum_{j=1}^n I_{\{y_j \leqslant y\}}.$$

An initial value for θ can be chosen as

$$\tilde{\theta}_0 = \frac{-n}{\sum_{j=1}^n \log \hat{F}(y_j)}.$$

Since model (3) is equivalent to $\log G(t)/\log F(t) = \theta$, one can also choose

$$\theta_0^* = \frac{1}{n_0}\sum_{i=1}^{n_0} \frac{\log \hat{G}(t_i)}{\log \hat{F}(t_i)},$$

where $\max\{\min_i x_i, \min_j y_j\} < t_i < \min\{\max_i x_i, \max_j y_j\}$.

A preferred initial can be obtained using Cox's partial likelihood method. We then use this initial value of θ to obtain the initial estimates of p's using the method described in the next subsection.

3.3 Approximation of $\hat{\boldsymbol{p}}(\theta)$ for a given θ

To simplify notation, we define $\lambda_N = m/N$. Assume that $\lambda_N \to \lambda \in (0, 1)$, as $m, n \to \infty$. The empirical distribution of the combined sample is

$$\hat{K}_N(t) = \lambda_N \hat{F}(t) + (1 - \lambda_N)\hat{G}(t).$$

The mixture distribution function is defined as

$$K_N(t) = \lambda_N F(t) + (1 - \lambda_N)G(t) = \lambda_N F(t) + (1 - \lambda_N)\{F(t)\}^\theta.$$

If $\theta = 1$, $F(t) = K_N(t)$.

$$K_N'(t) = F'(t)[\lambda_N + \theta(1 - \lambda_N)\{F(t)\}^{\theta - 1}].$$

For any given $\theta > 0$ and $v \in [0, 1]$, let $u = u_\theta(v)$ be the solution of equation

$$\varphi_N(u) \equiv \lambda_N u + (1 - \lambda_N)u^\theta = v.$$

It is clear that, for all $\theta > 0$, $u_\theta(0) = 0$, $u_\theta(1) = 1$, and $0 \leqslant u_\theta(v) \leqslant 1$. Define $\tilde{F}_\theta(z) = u_\theta\{\hat{K}_N(z)\}$, for $z \in S_F$. It follows from Glivenko-Cantelli theorem that if model (3) is true then one can approximate $\hat{p}_k$ by $\tilde{p}_k(\theta) \equiv \Delta\tilde{F}_\theta(z_{(k)}) = \tilde{F}_\theta(z_{(k)}) - \tilde{F}_\theta(z_{(k-1)})$, $k = 1, \ldots, N$, with $z_{(0)} = -\infty$. Clearly, if $\theta = 1$, $u_1(v) = v$ and $\tilde{p}_k(1) = 1/N$, $k = 1, \ldots, N$. Also for all $\theta > 0$,

$$\sum_{k=1}^{N} \tilde{p}_k(\theta) = \tilde{F}_\theta(z_{(N)}) - \tilde{F}_\theta(z_{(0)}) = 1.$$

If $v \in (0, 1]$, then for all $\theta > 0$,

$$u'_\theta(v) = \frac{du_\theta(v)}{dv} = \frac{1}{\lambda_N + \theta(1-\lambda_N)u_\theta^{\theta-1}(v)} > 0.$$

and that $u_\theta(v)$ is strictly increasing function of v. Therefore $\tilde{p}_k(\theta) > 0$, $k = 1, \ldots, N$. For each given θ and $k = 1, \ldots, N-1$, $\tilde{F}_\theta(z_{(k)}) = u_\theta\{\hat{K}_N(z_{(k)})\} = u_\theta(k/N)$ can be found using Newton iteration

$$u_{i+1} = u_i - \frac{\lambda_N u_i + (1-\lambda_N)u_i^\theta - k/N}{\lambda_N + \theta(1-\lambda_N)u_i^{\theta-1}}, \quad i = 0, 1, \ldots,$$

with initial

$$u_0 = \frac{1}{m}\sum_{i=1}^{m} I_{\left\{\frac{i-1}{m-1} \leqslant v_{(k)}\right\}},$$

where $v_{(k)}$ is the kth smallest value of

$$\left\{\frac{i-1}{m-1}, \quad i = 1, \ldots, m; \quad \left(\frac{j}{n}\right)^{1/\theta}, \quad j = 1, \ldots, n\right\}.$$

It is also clear that $\tilde{p}_k(\theta)$ depends on $(\theta, k/N, \lambda_N)$ but independent of the sample values. However, $\tilde{F}_\theta(z)$ depends on both x_i's and y_j's.

Because $0 \leqslant u'_\theta(v) \leqslant \lambda_N^{-1}$,

$$\tilde{p}_k(\theta) = u_\theta\left(\frac{k}{N}\right) - u_\theta\left(\frac{k-1}{N}\right) = \mathcal{O}(\mathcal{N}^{-\infty}).$$

Remark on calculation: Based on our experience, the Newton iteration is more likely to converge, but slower, for $\theta > 1$ than $\theta < 1$. So if the estimated initial $\theta^{(0)} < 1$ and the iteration is not convergent, try to switch the x and y samples and to obtain estimates $\hat{\vartheta}$ and $\hat{\boldsymbol{q}}$ of $\vartheta = 1/\theta$ and the probability masses $\boldsymbol{q} = (q_1, \ldots, q_N)^{\mathrm{T}}$ of G, respectively. Then θ and F can be estimated, respectively, by $\hat{\theta} = 1/\hat{\vartheta}$ and

$$\hat{F}(x) = \left\{\hat{G}(x)\right\}^{\hat{\vartheta}} = \left\{\sum_{k=1}^{N} \hat{q}_k I_{\{z_{(k)} \leqslant x\}}\right\}^{\hat{\vartheta}}.$$

If one wants to fit the data using the proportional hazards model $1 - G(t) = \{1 - F(t)\}^\theta$, then transform the samples to $x_i^* = -x_i$, $i = 1, \ldots, m$, and $y_i^* = -y_j$, $j = 1, \ldots, n$. Let $\hat{\boldsymbol{p}}^* = (p_1^*, \ldots, p_N^*)^{\mathrm{T}}$ be the estimated masses of the baseline distribution $F^*(t) = 1 - F(-t)$. Then one can estimate the probability masses of F by $\hat{\boldsymbol{p}} = (p_N^*, \ldots, p_1^*)^{\mathrm{T}}$ and therefore estimate the distribution function by

$$\tilde{F}(x) = \sum_{k=1}^{N} \hat{p}^*_{N-k+1} I_{\{z_{(k)} \leqslant x\}}.$$

If sample sizes are large, the Newton iteration works very well. For small samples or divergent Newton iteration, the downhill simplex method (Nelder and Mead, 1965) works fine. An R package, rel, will be available on CRAN and from the authors.

4. Estimations and Hypothesis Tests

In this section, we give some procedures for constructing confidence intervals and hypotheses testing for the parameter and the model.

4.1 Confidence Interval for θ

One can construct a $100(1-\alpha)\%$ confidence interval for θ as

$$\mathcal{I} = \{\theta \mid \mathcal{LR}(\theta) \leqslant \mathcal{C}_\alpha\} = [\hat{\theta}_{\mathcal{L}}, \hat{\theta}_{\mathcal{U}}], \tag{21}$$

where C_α is the upper α percentile of the sampling distribution of $\mathcal{LR}(\theta_0)$.

For large sample sizes m and n, since $\mathcal{LR}(\theta)$ is asymptotically χ^2_1, C_α is approximately the upper α percentile $\chi^2_1(\alpha)$ of the chi-squared distribution χ^2_1. For small sample sizes, we can approximate C_α by simulation as follows.

1. Generate $x^*_1, \ldots, x^*_m$ and $u_1, \ldots, u_n$ from uniform$(0,1)$. Calculate $y^*_j = {u_j}^{1/\hat{\theta}}$, $j = 1, \ldots, n$.
2. Calculate likelihood ratio $\mathcal{LR}^*(\hat{\theta})$ based on simulated samples x^*_i and y^*_j.
3. Repeat the above simulation B times to get $\mathcal{LR}^*_\lfloor(\hat{\theta})$, $b = 1, \ldots, B$.
4. Approximate C_α by the sample's upper α percentile of $\mathcal{LR}^*_\lfloor(\hat{\theta})$, $b = 1, \ldots, B$.

4.2 Hypothesis Tests for $H_0 : \theta = \theta_0$

The p-values of the test can be approximated using simulation method.

1. Calculate $\hat{\theta}$ based on the sample data.
2. Generate $x^*_1, \ldots, x^*_m$ and $u_1, \ldots, u_n$ from uniform$(0,1)$. Set $y^*_j = {u_j}^{1/\theta_0}$, $j = 1, \ldots, n$.
3. Calculate the estimated value $\hat{\theta}^*$ of θ based on samples x^*_i and y^*_j.
4. Repeat the above simulation B times to get $\hat{\theta}^*_b$, $b = 1, \ldots, B$.
5. The p-value of the test for H_0 can be approximated as follows.
   (i) For alternative $H_1 : \theta \neq \theta_0$: $p_{val} = \frac{1}{B}\sum_{b=1}^{B} I\{|\hat{\theta}^*_b - \theta_0| > |\hat{\theta} - \theta_0|\}$.
   (ii) For alternative $H_1 : \theta > \theta_0$: $p_{val} = \frac{1}{B}\sum_{b=1}^{B} I\{\hat{\theta}^*_b > \hat{\theta}\}$.
   (iii) For alternative $H_1 : \theta < \theta_0$: $p_{val} = \frac{1}{B}\sum_{b=1}^{B} I\{\hat{\theta}^*_b < \hat{\theta}\}$.

The above procedure works for any sample sizes. If sample size is large, The p-value of the two-sided test can be approximated using χ^2_1 distribution: $p_{val} = \Pr\{\chi^2_1 > \mathcal{LR}(\theta_\prime)\}$. The p-value of the one-sided test can be approximated using standard normal distribution based on the signed root $z_0 = \mathrm{sgn}(\hat{\theta} - \theta_0)\sqrt{\mathcal{LR}(\theta_\prime)}$.

4.3 Estimation of F and G and Goodness-of-Fit Test

An empirical reverse hazards ratio plot: The scatter plot of the empirical odds ratio

$$\frac{\log \hat{G}(t)}{\log \hat{F}(t)}, \quad x_{(1)} \vee y_{(1)} < t < x_{(m)} \wedge y_{(n)},$$

against t would be convenient for a preliminary test for the goodness of fit of the model. A horizontal pattern shown by the scatter plot would indicate a good fit of the model.
Kolmogorov-Smirnov test for model (3): The semiparametric estimates of F and G are respectively

$$\tilde{F}(x) = \sum_{k=1}^{N} p_k(\hat{\theta}) I(z_{(k)} \leqslant x) = \sum_{k=1}^{N} \hat{p}_k I(z_{(k)} \leqslant x), \quad \tilde{G}(y) = \{\tilde{F}(y)\}^{\hat{\theta}}$$

These estimates are better than the empirical distribution in the sense that they are smoother and use the estimated probability masses on all data values in the pooled sample.

We propose to use the following Kolmogorov-Smirnov type statistic to test the goodness of fit of model (3)

$$ks = m^{1/2} \sup_x |\tilde{F}(x) - \hat{F}(x)| + n^{1/2} \sup_y |\tilde{G}(y) - \hat{G}(y)|. \tag{22}$$

It is easy to see that if F is continuous, then under the null hypothesis H_0 that (3) is true, the distribution of the test statistic ks is independent of F. Simulation method can be used to approximate the null distribution of ks and thus the p-values of the goodness-of-fit test.

By adopting a more powerful GOF test (Owen, 1995; Zhang, 2002), we propose the following test statistic

$$T_N = m^{1/2} \sup_x K\{\hat{F}(x), \tilde{F}(x)\} + n^{1/2} \sup_y K\{\hat{G}(y), \tilde{G}(y)\}, \tag{23}$$

where

$$K(x, y) = x \log(x/y) + (1 - x) \log\{(1 - x)/(1 - y)\}, \quad 0 \leqslant x, y \leqslant 1,$$

and $0 \log 0 = 0$. It is easy to obtain that

$$\sup_x K\{\hat{F}(x), \tilde{F}(x)\} = \max_{x_{(1)} \leqslant z_{(k)} < x_{(m)}} K\left\{\hat{F}(z_{(k)}), \tilde{F}(z_{(k)})\right\},$$
$$\sup_y K\{\hat{G}(y), \tilde{G}(y)\} = \max_{y_{(1)} \leqslant z_{(k)} < y_{(n)}} K\left\{\hat{G}(z_{(k)}), \tilde{G}(z_{(k)})\right\}.$$

5. Simulation and Example

5.1 Simulation Results

Because the Cox's partial likelihood method (Cox, 1972) is the most popular method so far in applying the proportional hazards model, in this section we focus only on the comparison between the methods of Cox and the present paper. Let $\hat{\theta}_{cox}$ be the maximum partial likelihood estimator of θ and $\hat{F}_{cox}$ the the estimate of the baseline distribution F using function `survfit.coxph` in R package `survival`. For a cumulative distribution F, we define the "bias" of its estimator $\hat{F}$ as

$$Bias(\hat{F}) = \int_{-\infty}^{\infty} \{\hat{F}(x) - F(x)\} dF(x) = \int_{-\infty}^{\infty} \hat{F}(x) dF(x) - \frac{1}{2}.$$

In the following simulation studies, we compare the biases and the mean squared errors of $\hat{\theta}_{cox}$ and $\hat{\theta}$. We also compare the biases and the Cramér-von-Mises criteria of $\hat{F}_{cox}$ and $\tilde{F}$. Since the proposed method is distribution-free, we consider the case where $F =$ uniform(0,1). For each of 1,000 runs, we generate $x_1, \ldots, x_m$ and $u_1, \ldots, u_n$ from uniform(0,1), and set $y_j = 1 - u_j^{1/\theta}$, $j = 1, \ldots, n$. The sample sizes are $m = n = 10, 20, 30, 40, 100,$ and 200 and parameter values are $\theta = 1, 1.5, \ldots, 4$.

If F has support (0,1), then

$$Bias(\hat{F}) = \int_{-\infty}^{\infty} \hat{F}(x)dF(x) - \frac{1}{2} = \sum_{k=1}^{N} \hat{F}(z_{(k)})\Delta F(z_k) - \frac{1}{2},$$

where $\Delta F(z_k) = F(z_{(k+1)}) - F(z_{(k)})$, $k = 1, \ldots, N$, $z_{(N+1)} = 1$. The Cramér-von-Mises criterion is

$$\begin{aligned} C(\hat{F}) &= \int_{-\infty}^{\infty} \{\hat{F}(x) - F(x)\}^2 dF(x) \\ &= \sum_{k=1}^{N} \hat{F}^2(z_{(k)})\Delta F(z_k) - \sum_{k=1}^{N} \hat{F}(z_{(k)})\Delta F^2(z_k) + \frac{1}{3}, \end{aligned}$$

where $\Delta F^2(z_k) = F^2(z_{(k+1)}) - F^2(z_{(k)})$, $k = 1, \ldots, N$.

In Table 1, $B(\hat{F})$ and $C(\hat{F})$ represent the bias and the Cramér-von-Mises criterion of an estimator $\hat{F}$. From the simulation results (see Table 1), we see that our proposed baseline distribution estimate $\tilde{F}$ performs very similarly as $\hat{F}_{cox}$ in the sense that both estimators have similar absolute biases and similar Cramér-von-Mises criteria in most cases. It should be noted that both estimators seem to underestimate F. This is caused by the overestimation of θ by both methods and $\theta \geqslant 1$ in this simulation. In Table 2, $B(\hat{\theta})$ and $M(\hat{\theta})$ represent the bias and the mean squared error of an estimator $\hat{\theta}$. We see in Table 2 that both Cox's partial likelihood and the proposed full likelihood methods underestimate θ but the maximum full likelihood estimate seems to have a bigger bias for small sample sizes. Both estimates seem to be asymptotically unbiased because the biases get close as the sample sizes increase.

In both the simulation study and the following examples, we use R functions `coxph` and `survfit.coxph` with default `efron` method in R package `survival` by Therneau and original R port by Thomas Lumley (2009) to fit the Cox proportional hazards model and to estimate survival function as competitors of the proposed ones.

5.2 Example: Failure times of electrical cable insulations

Analyzed by Lawless (2003), the data given by Table 3 contain the voltage levels (in kilovolts per millimeter) in increasing order at which failures occurred in two types of electrical cable insulation when specimens were subjected to an increasing voltage stress in a laboratory test. Twenty specimens of each type were involved.

We take Type 2 as baseline x and Type 1 as y. If we take Type 1 as baseline, then the estimated initial $\theta_0 < 1$ and the Newton iteration does not converge (See Remark on calculation in Section 3.) Using proportional hazards model to fit the data, we transform x's and y's to $-x$'s and $-y$'s respectively. The maximum (full) likelihood estimate of θ is $\hat{\theta} = 8.044413$. The goodness-of-fit test statistic $T_n = 0.2845$ with simulated p-value 0.547 based on 1,000 runs. Therefore, we conclude that the data is fitted very well by the proportional hazards model. This can also be convinced by Figure 1 in which we see that the semiparametric estimates of F and G are very close to the empirical distributions. The likelihood ratio for $\theta_0 = 1$ is $\mathcal{LR}(\theta_0) = 25.3668$. The p-value of the test for $H_0 : \theta = 1$

Table 1: Simulated biases and Cramér-von-Mises criteria of $\hat{F}_{cox}$ and $\tilde{F}$

	$(m,n)=(10,10)$				$(m,n)=(20,20)$			
θ	$B(\hat{F}_{cox})$	$B(\tilde{F})$	$C(\hat{F}_{cox})$	$C(\tilde{F})$	$B(\hat{F}_{cox})$	$B(\tilde{F})$	$C(\hat{F}_{cox})$	$C(\tilde{F})$
1.0	-0.0138	-0.0519	0.0125	0.0144	-0.0066	-0.0263	0.0061	0.0062
1.5	-0.0136	-0.0323	0.0129	0.0135	-0.0086	-0.0170	0.0063	0.0061
2.0	-0.0167	-0.0255	0.0133	0.0140	-0.0085	-0.0150	0.0064	0.0062
2.5	-0.0157	-0.0222	0.0132	0.0145	-0.0082	-0.0107	0.0066	0.0068
3.0	-0.0180	-0.0211	0.0137	0.0152	-0.0084	-0.0096	0.0066	0.0075
3.5	-0.0193	-0.0139	0.0140	0.0142	-0.0101	-0.0065	0.0071	0.0072
4.0	-0.0203	-0.0147	0.0138	0.0147	-0.0100	-0.0089	0.0069	0.0071
	$(m,n)=(30,30)$				$(m,n)=(40,40)$			
1.0	-0.0037	-0.0181	0.0041	0.0040	-0.0033	-0.0125	0.0030	0.0032
1.5	-0.0052	-0.0110	0.0042	0.0039	-0.0034	-0.0079	0.0031	0.0032
2.0	-0.0063	-0.0088	0.0041	0.0044	-0.0051	-0.0040	0.0032	0.0031
2.5	-0.0054	-0.0100	0.0044	0.0045	-0.0039	-0.0075	0.0032	0.0033
3.0	-0.0067	-0.0073	0.0045	0.0047	-0.0052	-0.0042	0.0034	0.0033
3.5	-0.0052	-0.0036	0.0045	0.0046	-0.0040	-0.0068	0.0033	0.0036
4.0	-0.0061	-0.0043	0.0046	0.0049	-0.0041	-0.0036	0.0034	0.0038
	$(m,n)=(100,100)$				$(m,n)=(200,200)$			
1.0	-0.0013	-0.0055	0.0012	0.0011	-0.0003	-0.0022	0.0006	0.0006
1.5	-0.0003	-0.0034	0.0012	0.0013	-0.0010	-0.0014	0.0006	0.0006
2.0	-0.0014	-0.0024	0.0013	0.0013	-0.0009	-0.0010	0.0006	0.0006
2.5	-0.0014	-0.0012	0.0013	0.0013	-0.0011	-0.0004	0.0006	0.0007
3.0	-0.0024	-0.0024	0.0013	0.0013	-0.0007	-0.0022	0.0007	0.0007
3.5	-0.0018	-0.0014	0.0014	0.0013	-0.0014	-0.0017	0.0007	0.0007
4.0	-0.0017	-0.0025	0.0014	0.0014	-0.0012	-0.0009	0.0007	0.0007

Table 2: Simulated biases and mean square errors of $\hat{\theta}_{cox}$ and $\hat{\theta}$

	$(m,n)=(10,10)$				$(m,n)=(20,20)$			
θ	$B(\hat{\theta}_{cox})$	$B(\hat{\theta})$	$M(\hat{\theta}_{cox})$	$M(\hat{\theta})$	$B(\hat{\theta}_{cox})$	$B(\hat{\theta})$	$M(\hat{\theta}_{cox})$	$M(\hat{\theta})$
1.0	0.3194	0.7557	0.6446	1.7246	0.1182	0.2921	0.1475	0.2496
1.5	0.3541	1.0457	1.6187	6.6385	0.1378	0.3746	0.3480	0.5765
2.0	0.4714	1.5017	2.7430	11.4695	0.2055	0.5384	0.8284	1.4537
2.5	0.7185	2.3298	5.1062	26.4694	0.2684	0.7196	1.4806	2.8268
3.0	0.7794	2.8906	6.5186	37.7709	0.4338	1.0585	2.6587	5.4862
3.5	0.8486	3.5468	8.3601	54.8353	0.4238	1.2003	3.8557	8.7002
4.0	1.0708	4.6212	11.2006	81.5548	0.6140	1.6698	6.5516	21.2780
	$(m,n)=(30,30)$				$(m,n)=(40,40)$			
1.0	0.0644	0.1796	0.0742	0.1103	0.0506	0.1376	0.0574	0.0777
1.5	0.0734	0.2217	0.2083	0.2901	0.0680	0.1779	0.1505	0.1950
2.0	0.1186	0.3190	0.4187	0.6019	0.0649	0.2075	0.2599	0.3402
2.5	0.1894	0.4584	0.7563	1.1448	0.1532	0.3450	0.4982	0.6917
3.0	0.2553	0.6036	1.2306	1.9419	0.1588	0.4007	0.7903	1.1010
3.5	0.2525	0.6788	1.6622	2.7741	0.2333	0.5386	1.1476	1.6616
4.0	0.3555	0.8887	2.4248	4.0891	0.2767	0.6502	1.7491	2.6058
	$(m,n)=(100,100)$				$(m,n)=(200,200)$			
1.0	0.0174	0.0547	0.0211	0.0240	0.0051	0.0250	0.0103	0.0109
1.5	0.0224	0.0654	0.0501	0.0558	0.0039	0.0258	0.0234	0.0244
2.0	0.0201	0.0745	0.0953	0.1056	0.0136	0.0407	0.0490	0.0517
2.5	0.0198	0.0890	0.1683	0.1871	0.0210	0.0551	0.0806	0.0858
3.0	0.0631	0.1510	0.2594	0.2980	0.0566	0.0998	0.1324	0.1438
3.5	0.0654	0.1730	0.3525	0.4080	0.0358	0.0880	0.1754	0.1889
4.0	0.0997	0.2303	0.5085	0.5987	0.0420	0.1047	0.2592	0.2797

Table 3: Failure times of two types of Electrical Cable Insulations

Type 1:	32.0	35.4	36.2	39.8	41.2	43.3	45.5	46.0	46.2	46.4
	46.5	46.8	47.3	47.3	47.6	49.2	50.4	50.9	52.4	56.3
Type 2:	39.4	45.3	49.2	49.4	51.3	52.0	53.2	53.2	54.9	55.5
	57.1	57.2	57.5	59.2	61.0	62.4	63.8	64.3	67.3	67.7

versus $H_1 : \theta > 1$ is 0 based on a simulation of 10,000 runs, and $\Pr\{N(0,1) > (\hat{\theta} - \theta_0)\sqrt{\mathcal{LR}(\theta_0)}\} = 0$ by the normal approximation. So we do reject $H_0 : \theta = 1$ in favor of $H_1 : \theta > 1$. Confidence intervals of level 95% for θ are $(\tilde{\theta}_L, \tilde{\theta}_U) = (3.523, 19.695)$ based on chi-square approximation, and $(\hat{\theta}_L, \hat{\theta}_U) = (3.009, 23.7994)$ based on a simulation using $C_{0.05} = 4.991$. The former is a subinterval of the latter but both exclude $\theta = 1$.

Using Cox's method with Efron's approximation which is the default of R function `coxph` of the `survival` package, one can obtain an estimate of θ: $\hat{\theta} = 6.55$ with p-value of 8.1×10^{-6} and likelihood ratio test statistic of 21.5 with p-value of 3.48×10^{-6}. In Figure 1, the estimated baseline distribution function $\hat{F}_{Cox}$ is plotted with solid line and is located between the empirical distribution functions $\hat{F}$ and $\hat{G}$. From this we see that the Cox's method underestimates the baseline survival function $\bar{F} = 1 - F$. This is also why the Cox's method gives a smaller estimate of θ.

6. Concluding Remarks

The major advantage of the proposed method is distribution-free property. So the "exact" p-values and critical values for the hypothesis tests and the confidence intervals/bands can be obtained using simulation. Although the performance of the estimate $\hat{\theta}$ is not as good as that of $\hat{\theta}_{cox}$, the performance of the baseline distribution function estimate $\hat{F}$ is very similar to the existing one. The round-off errors of the numerical procedure which involves inverting large matrices may be blamed for the worse performance of $\hat{\theta}$.

References

Cox, D. R. (1972), "Regression models and life-tables," *J. Roy. Statist. Soc. Ser. B*, 34, 187–220.

— (1975), "Partial likelihood," *Biometrika*, 62, 269–276.

Dabrowska, D. M. and Doksum, K. A. (1988a), "Estimation and testing in a two-sample generalized odds-rate model," *J. Amer. Statist. Assoc.*, 83, 744–749.

— (1988b), "Partial likelihood in transformation models with censored data," *Scand. J. Statist.*, 15, 1–23.

Guan, Z. and Peng, C. (2009), "A rank-based empirical likelihood approach to two-sample proportional odds model and its goodness-of-fit," Submitted Manuscript.

Gupta, R. C. and Gupta, R. D. (2007), "Proportional reversed hazard rate model and its applications," *J. Statist. Plann. Inference*, 137, 3525–3536.

Kalbfleisch, J. D. and Prentice, R. L. (1973), "Marginal likelihoods based on Cox's regression and life model," *Biometrika*, 60, 267–278.

Lawless, J. F. (2003), *Statistical Models and Methods for Lifetime Data*, Hoboken: John Wiley and Sons, 2nd ed.

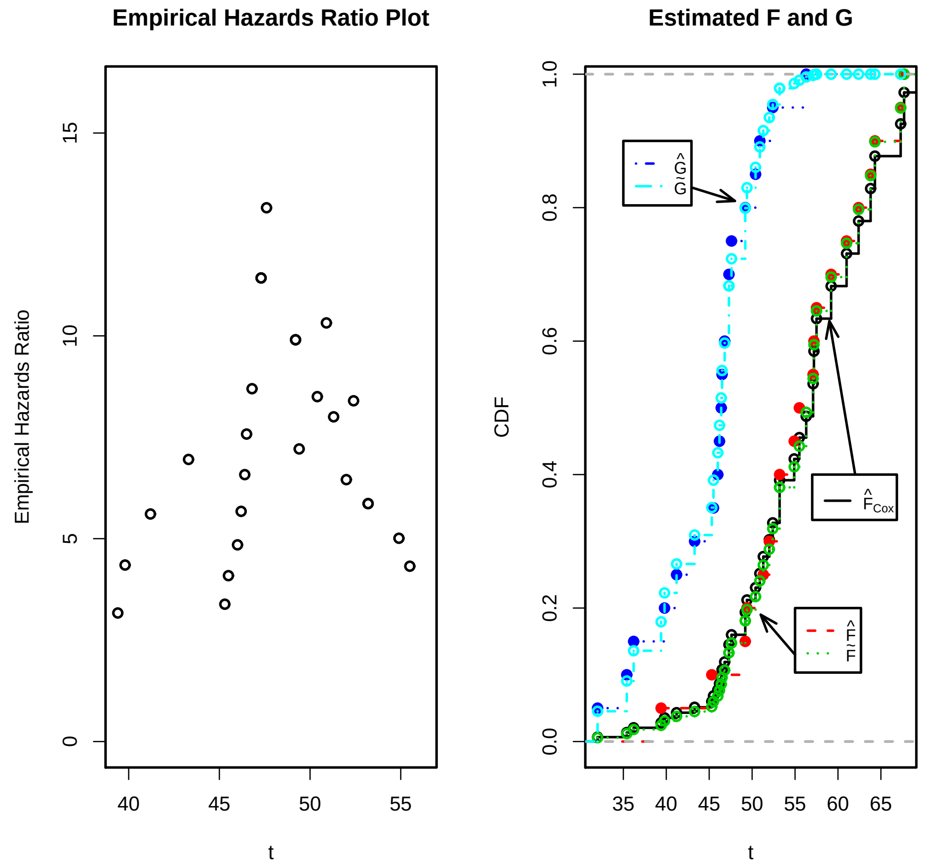

Figure 1: Plots of hazards ratio and Estimated CDF's

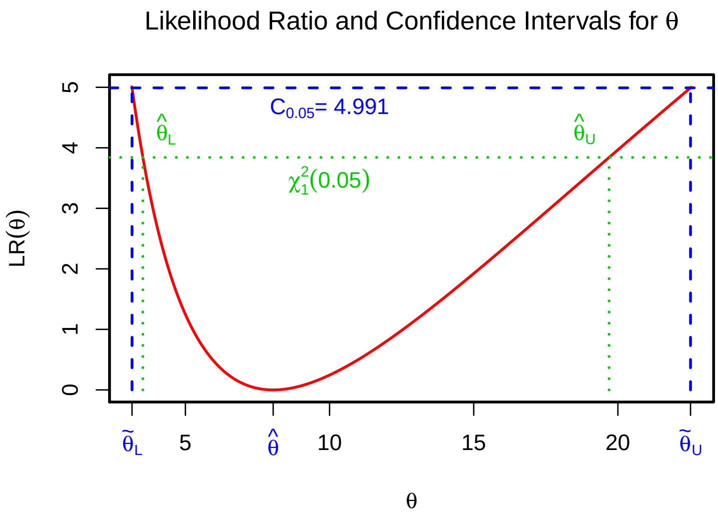

Figure 2: Likelihood ratio and confidence intervals of θ

Lehmann, E. L. (1953), "The power of rank tests," *Ann. Math. Statistics*, 24, 23–43.

Nelder, J. A. and Mead, R. (1965), "A Simplex Method for Function Minimization," *The Computer Journal*, 7, 308–313.

Owen, A. B. (1995), "Nonparametric likelihood confidence bands for a distribution function," *J. Amer. Statist. Assoc.*, 90, 516–521.

Therneau, T. and original R port by Thomas Lumley (2009), *survival: Survival analysis, including penalised likelihood.*, R package version 2.35-8.

Wu, C. O. (1995), "Estimating the real parameter in a two-sample proportional odds model," *Ann. Statist.*, 23, 376–395.

Zhang, J. (2002), "Powerful Goodness-of-fit Tests Based on the Likelihood Ratio," *Journal of the Royal Statistical Society, Series B, Methodological*, 64, 281–294.