One of the most widely used regression models in survival analysis is the Cox proportional hazards model (Cox 1972, 1975).

Let $X_1, \cdots, X_n$; and $C_1, \cdots, C_n$ be independent random variables. Think of C_i as the censoring time associated with the survival time X_i. Due to censoring, we can only observe $(T_1, \delta_1), \cdots, (T_n, \delta_n)$ where

$$T_i = \min(X_i, C_i) \quad \text{and} \quad \delta_i = \begin{cases} 1 & \text{if } X_i \le C_i \\ 0 & \text{if } X_i > C_i \end{cases} . \tag{1}$$

Also available are $z_1, \cdots, z_n$, which are covariates associated with the responses $X_i, \cdots, X_n$ and we assume z_i do not change with time here.

According to Cox's proportional hazards model, the cumulative hazard function of X_i is related to the covariate z_i.

$$\Lambda_{X_i}(t) = \Lambda_i(t) = \Lambda(t|z_i) = \Lambda_0(t)\exp(\beta_0 z_i) \tag{2}$$

where β_0 is the unknown regression coefficient and $\Lambda_0(t)$ is the so called baseline cumulative hazard function. Another way to think of $\Lambda_0(t)$ is that it is the cumulative hazard for an individual with zero covariate, $z = 0$.

The semiparametric Cox proportional hazards model assumes that the baseline cumulative hazard function Λ_0 is completely unknown and arbitrary.

We study here the inference in the Cox model where we have some information on the baseline hazard. But it remains an infinite dimensional nuisance parameter.

For example, we may know that the baseline hazard has median 45. Or median is between 44 and 46. For stratified Cox model, we may know that one baseline hazard is stochastically smaller than the other baseline, or the two hazards cross at $t = 50$, etc. When comparing a placebo against a new treatment in a two sample case, we often have extra knowledge about the survival experience for the placebo group, may be from past experiences, other studies, etc.

Empirical Likelihood approach is used to obtain inference about β_0 in the presence of this new information. We show that the modified estimator also has asymptotic normal distribution and the empirical likelihood ratio also follows a Wilks theorem under null hypothesis.

The modified estimator of β is more accurate and the test have better power compared to the regular Cox partial likelihood estimator/test.

We made use of extra information. The information is on the baseline. It helps improve estimate β.

We show how the extra information can be incorporated into the Cox model via empirical likelihood, and provides a unified analysis.

Some known results about regular Cox estimates:

For simplicity we gave detailed formula for the case dim$(z_i) = 1$. For the case where dim$(z_i) = k$, parallel results to those obtained here can be obtained similarly.

Let $\Re_i = \{j : T_j \geq T_i\}$, the risk set at time T_i. Define

$$\ell(\beta) = \sum_{i=1}^{n} \delta_i z_i - \sum_{i=1}^{n} \delta_i \frac{\sum_{j \in \Re_i} z_j \text{exp}(\beta z_j)}{\sum_{j \in \Re_i} \text{exp}(\beta z_j)}, \tag{3}$$

and

$$I(\beta_0) = \sum_{i=1}^{n} \delta_i \left(\frac{\sum\limits_{j\in\Re_i} z_j^2 \exp(\beta_0 z_j)}{\sum\limits_{j\in\Re_i} \exp(\beta_0 z_j)} - \left[\frac{\sum\limits_{j\in\Re_i} z_j \exp(\beta_0 z_j)}{\sum\limits_{j\in\Re_i} \exp(\beta_0 z_j)} \right]^2 \right) = -\ell'(\beta_0). \quad (4)$$

If $\widehat{\beta}_c$ is the solution of (3), i.e. $\ell(\widehat{\beta}_c) = 0$, then $\widehat{\beta}_c$ is called the Cox partial likelihood estimate of regression coefficient β_0.

Theorem 1 (Andersen and Gill 1982) *Under regularity conditions we have following results:*
(1). If $\widehat{\beta}_c$ *is the solution of* (3), *then, as* $n \to \infty$, $\widehat{\beta}_c \xrightarrow{\text{P}} \beta_0$.
(2). If $\widehat{\beta}_c$ *is the solution of* (3), *then, as* $n \to \infty$,

$$\sqrt{n}(\widehat{\beta}_c - \beta_0) \xrightarrow{\mathcal{D}} N(0, \Sigma^{-1}), \quad (5)$$

where $\Sigma = Plim_{n\to\infty}\dfrac{1}{n}I(\beta_0)$.

(3). If $\beta_n^\star \xrightarrow{\text{Pr}} \beta_0$, *then* $\frac{1}{n}I(\beta_n^\star) \xrightarrow{\text{P}} \Sigma$.

Define the asymptotic (Poisson) empirical likelihood function of T_i, δ_i as

$$(\Delta\Lambda_i(T_i))^{\delta_i} \exp\{-\Lambda_i(T_i)\}.$$

Under Cox's proportional hazards model,

$$\Delta\Lambda_i(T_i) = \Delta\Lambda_0(T_i) \exp(\beta z_i), \quad \text{and} \quad \Lambda_i(T_i) = \Lambda_0(T_i) \exp(\beta z_i) \ .$$

If we use $\mathcal{A}L^c(\beta, \Lambda_0)$ to denote the (asymptotic) empirical likelihood function under the Cox's model for all observations, then we have

$$\mathcal{A}L^c(\beta, \Lambda_0) = \prod_{i=1}^{n} (\Delta\Lambda_0(T_i) e^{\beta z_i})^{\delta_i} \exp\{-e^{\beta z_i} \sum_{j: T_j \le T_i} \Delta\Lambda_0(T_j)\}, \qquad (6)$$

where we shall require $\Lambda_0 \ll \hat{\Lambda}_{NA}$, the Nelson-Aalen estimator. This restriction is similar to the restriction for CDFs to have same support as the empirical distribution in Owen (1988).

Theorem 2 (Pan 1997) *Under the same conditions as in Theorem 1, we have the following empirical likelihood ratio result:*

$$-2log \frac{\max\limits_{\{\Lambda_0 \ll \widehat{\Lambda}_{NA}\}} \mathcal{A}L^c(\beta_0, \Lambda_0)}{\max\limits_{\{\beta, \Lambda_0 \ll \widehat{\Lambda}_{NA}\}} \mathcal{A}L^c(\beta, \Lambda_0)} = I(\xi)(\beta_0 - \widehat{\beta}_c)^2 \xrightarrow{\mathcal{D}} \chi^2_{(1)} ,$$

where ξ *is between* β_0 *and* $\widehat{\beta}_c$.

Bartlett adjustment: (Gu and Zheng)

Denote the left-hand-side above by $-2LLR$. For one dimensional β, we have

$$P(-2LLR \le u) = \chi^2_1(u) + \left\{ \frac{C_1}{n}\chi^2_1(u) + \frac{C_2}{n}\chi^2_3(u) + \frac{C_3}{n}\chi^2_5(u) \right\} + O(n^{-3/2})$$

where C_i are constants.

The following lemma will be used in the next section.

Lemma 1 (Joint CLT): *Under the same conditions as in Theorem 1, assume also that* $g(\cdot)$ *is a square integrable function with respect to* $\Lambda_0(\cdot)$, *we have*

$$\left[\frac{\ell(\beta_0)}{\sqrt{n}}, \sqrt{n} \cdot m(\beta_0, 0)\right] \xrightarrow{\mathcal{D}} N(0, V),$$

where m *is defined as*

$$m(\beta, \lambda) = \sum_{i=1}^{n} \frac{\delta_i g(T_i)}{\sum_{j \in \Re_i} e^{\beta z_j} + n\lambda g(T_i)} - \int g(t) d\Lambda_0(t) \ .$$

The variance-covariance matrix V is diagonal (or covariate between ℓ

and m is zero): $V = diag(\Sigma, V_{22})$, where Σ as given in Theorem 1, and

$$V_{22} = \lim_{n\to\infty} \int \frac{g^2(s)d\Lambda_0(s)}{\frac{1}{n}\sum_j \exp(\beta_0 z_j) I_{[T_j \ge s]}} = \lim \sum_{i=1}^{n} \frac{n g^2(T_i)\delta_i}{\left[\sum_{j\in\Re_i} \exp(\beta_0 z_j)\right]^2} .$$

Assume the observations are ordered according to T_i values from now on.

Another way to pose the question:

What is there in between $I(\beta_0)$, the information(Cox model) and the information(exponential regression)?

How to get them?

2 Empirical Likelihood Ratio Statistic for β_0 with Information on Baseline

The simplest form of the extra information on the baseline is given in terms of the following equation:

$$\int g(s) d\Lambda_0(s) = \sum g(T_i) \Delta\Lambda_0(T_i) = \theta \qquad (7)$$

where θ is a given constant, and $g(\cdot)$ is a given function. The second expression above assumes a discrete hazard that only have possible jumps at the observed survival times, T_i's (like the Nelson-Aalen estimator). This type of constraint include many situations. For example, if $g(s) = I_{[s \leq 45]}$ and $\theta = -\log 0.5$, then the extra information can be interpreted as "median equal to 45".

The modified estimator of β is defined via the empirical likelihood. For the Wilks theorem, the log of the empirical likelihood ratio becomes the difference of two terms. We compute each term separately:

We will first compute the maximum of the log empirical likelihood (6) when β is fixed at β_0, and with the extra information (7).

Let $w_i^0 = \Delta\Lambda_0(T_i)$ for $i = 1, 2, \cdots, n$. We write the logarithm of $\mathcal{A}L^c(\beta_0)$ in terms of w_i^0's as follows

$$\log \mathcal{A}L^c(\beta_0) \;=\; \sum_{i=1}^{n} \delta_i \log w_i^0 + \sum_{i=1}^{n} \delta_i \beta_0 z_i - \sum_{i=1}^{n} \sum_{j=1}^{i} w_j^0 \exp(\beta_0 z_i)$$

After some calculation, it follows that

$$w_i^0 = \frac{\delta_i}{\sum_{j=i}^n \exp(\beta_0 z_j) + n\gamma g(T_i)\delta_i}$$

for $i = 1, 2, \cdots, n$. The value of the γ in the above can be obtained as the solution of the equation

$$0 = m(\beta_0, \gamma) = \sum_{i=1}^n \frac{g(T_i)\delta_i}{\sum_{j=i}^n \exp(\beta_0 z_j) + n\gamma g(T_i)\delta_i} - \theta \ . \tag{8}$$

The derivative of $m(\beta_0, \gamma)$ with respect to γ is always negative, so there is a unique γ solution, for the feasible values of θ.

Lemma 2 *The solution, $\hat{\gamma}$, of (8) with $\theta = \int g(s) d\Lambda_0(s)$ has the following representation*

$$\hat{\gamma} = m(\beta_0, 0) \times A^{-1} + o_p(1/\sqrt{n})$$

where

$$A = \sum_{i=1}^{n} \frac{n g^2(T_i)\delta_i}{\left[\sum_{j\in\Re_i} \exp(\beta_0 z_j)\right]^2} \ .$$

Proof: Use Taylor expansion on (8). We point out that $\lim A = V_{22}$. QED

Thus the log likelihood maximized over baselines that satisfy the extra constraint, and with β fixed at β_0, is

$$\log \mathcal{A}L^c(\beta_0, \widehat{\Lambda}(\beta_0)) =$$

$$\sum_{i=1}^{n} \delta_i \beta_0 z_i - \sum_{i=1}^{n} \delta_i \log \left(\sum_{j=i}^{n} e^{\beta_0 z_j} + n\widehat{\gamma} g(T_i)\delta_i \right) - \sum_{i=1}^{n} \frac{\delta_i \sum_{j=i}^{n} \exp(\beta_0 z_j)}{\sum_{j=i}^{n} e^{\beta_0 z_j} + n\widehat{\gamma} g(T_i)\delta_i}.$$

where $\widehat{\gamma}$ is the solution of the equation (8).

Step II: We now compute the maximum of the log empirical likelihood without fixing the β, (i.e. maximize over β) but still have extra information on the baseline hazard, (7). Here we shall also define the modified estimator of the regression coefficient.

Again let $w_i^0 = \Delta\Lambda(T_i)$ for $i = 1, 2, \cdots, n$. We rewrite the log likelihood $\log \mathcal{A}L^c(\beta, \Lambda_0)$ in terms of w_i^0 similar as in step I:

$$\log \mathcal{A}L^c(\beta, \Lambda_0) = \sum_{i=1}^{n} \delta_i \log w_i^0 + \sum_{i=1}^{n} \delta_i \beta z_i - \sum_{i=1}^{n} w_i^0 \sum_{j=i}^{n} \exp(\beta z_j).$$

After some calculation we have

$$w_i^0 = \frac{\delta_i}{\sum_{j=i}^n \exp(\beta z_j) + n\lambda g(T_i)\delta_i} \quad i = 1, 2, \cdots, n. \tag{9}$$

The λ in the above equation is the solution of

$$m(\beta, \lambda) = \sum_{i=1}^n \frac{\delta_i g(T_i)}{\sum_{j=i}^n \exp(\beta z_j) + n\lambda g(T_i)\delta_i} - \theta = 0. \tag{10}$$

Substituting (9) into

$$\frac{\partial G}{\partial \beta} = \sum_{i=1}^n \delta_i z_i - \sum_{i=1}^n w_i^0 \sum_{j=i}^n z_j \exp(\beta z_j) = 0 \tag{11}$$

we get the equation

$$\ell^*(\beta, \lambda) = \sum_{i=1}^{n} \delta_i z_i - \sum_{i=1}^{n} \delta_i \frac{\sum\limits_{j \in \Re_i} z_j \mathrm{exp}(\beta z_j)}{\sum\limits_{j \in \Re_i} \mathrm{exp}(\beta z_j) + n\lambda g(T_i)\delta_i} = 0. \tag{12}$$

Solving (12) and (10) for β and λ simultaneously requires iterative methods. We shall discuss the computation in more detail in the next section. Let us use $\widehat{\beta}$, $\widehat{\lambda}$ to denote the solution of (12) and (10). The solution $\widehat{\beta}$ is also our modified estimator of the regression coefficient.

Lemma 3 *The simultaneous solution of (12) and (10) has the following representation:*

$$\sqrt{n}[\widehat{\beta} - \beta_0, \widehat{\lambda}] = -\left[\frac{\ell(\beta_0)}{\sqrt{n}}, \sqrt{n} m(\beta_0, 0)\right] D^{-1}\sqrt{n} + o(1)$$

where D is the matrix

$$D = \begin{pmatrix} -I(\beta_0)/\sqrt{n} & -\sqrt{n}B \\ \sqrt{n}B & -\sqrt{n}A \end{pmatrix} \ .$$

The quantity A is defined in Lemma 2. The quantity B is defined as

$$B = \sum_{i=1}^{n} \frac{\delta_i g(T_i) \sum_{j \in \Re_i} z_j \exp(\beta_0 z_j)}{\left[\sum_{j \in \Re_i} \exp(\beta_0 z_j)\right]^2} \ .$$

Proof: The $\hat{\beta}$ and $\hat{\lambda}$ is the solution of $(0,0) = [\frac{\ell^*(\beta,\lambda)}{\sqrt{n}}, m(\beta,\lambda)\sqrt{n}]$. By Taylor expansion

$$[\frac{\ell^*(\beta,\lambda)}{\sqrt{n}}, m(\beta,\lambda)\sqrt{n}] = [\frac{\ell^*(\beta_0,0)}{\sqrt{n}}, m(\beta_0,0)\sqrt{n}] + (\beta-\beta_0,\lambda)\cdot D + o(|\beta-\beta_0|+|\lambda|)$$

where D is the matrix of the first derivatives of the vector. We let $\beta = \widehat{\beta}$ and $\lambda = \widehat{\lambda}$ in the above to get

$$(0,0) = [\frac{\ell^*(\beta_0, 0)}{\sqrt{n}}, m(\beta_0, 0)\sqrt{n}] + (\widehat{\beta} - \beta_0, \widehat{\lambda}) \cdot D + o(|\beta - \beta_0| + |\lambda|) \ .$$

Notice $\ell^*(\beta_0, 0) = \ell(\beta_0)$, which gives

$$[\widehat{\beta} - \beta_0, \widehat{\lambda}] = -[\frac{\ell(\beta_0)}{\sqrt{n}}, \sqrt{n} m(\beta_0, 0)] \times D^{-1} + o_p(1/\sqrt{n}) \ .$$

QED

Theorem 3 *As* $n \to \infty$ *the regression estimator with extra information,* $\widehat{\beta}$*, has the following limiting distribution*

$$\sqrt{n}(\widehat{\beta} - \beta_0) \xrightarrow{\mathcal{D}} N(0, (\Sigma^*)^{-1})$$

where $\Sigma^* = \Sigma + B^2 A^{-1}$ *and thus the variance is smaller then that of the regular Cox estimator.*

Proof: From Lemma 3 we can compute

$$\sqrt{n}(\widehat{\beta} - \beta_0) = \frac{\frac{\ell(\beta_0)}{\sqrt{n}} A + \sqrt{n} m(\beta_0, 0) B}{A I(\beta_0)/n + B^2} + o_p(1)$$

The asymptotic normality is immediate from Lemma 1. We compute the asymptotic variance. Since ℓ and m are asymptotically independence (Lemma 1), we compute

$$Var(\sqrt{n}(\widehat{\beta} - \beta_0)) = \frac{A^2 \Sigma + B^2 V_{22}}{(\Sigma A + B^2)^2} + o(1) \ .$$

Since $\lim V_{22} = \lim A$, we have

$$\lim Var(\sqrt{n}(\widehat{\beta} - \beta_0)) = \frac{A}{\Sigma A + B^2} = \frac{1}{\Sigma + B^2 A^{-1}} = \frac{1}{\Sigma^*} \ .$$

We know $A > 0$ therefore we have

$$\lim Var(\sqrt{n}(\widehat{\beta} - \beta_0)) = (\Sigma^*)^{-1} \leq \Sigma^{-1} = \lim Var\sqrt{n}(\widehat{\beta}_c - \beta_0) \ .$$

QED $\diamondsuit$

It is interesting to note that the variance of $\widehat{\beta}$ above is smaller than that of a regular Cox estimator, but if $B = 0$ then there is no improvement.

Substituting β in (9) by $\widehat{\beta}$, λ by $\widehat{\lambda}$, we get the expression of w_i^0:

$$w_i^0 = \frac{\delta_i}{\sum\limits_{j \in \Re_i} \exp(\widehat{\beta} z_j) + n\widehat{\lambda} g(T_i)\delta_i}, \qquad i = 1, 2, \cdots, n. \tag{13}$$

Therefore we can write the max of log likelihood as:

$$\log \max_{\{\beta, \Lambda \ll \widehat{\Lambda}_{NA}, \text{satisfy (7)}\}} \mathcal{A}L^c(\beta, \Lambda) =$$

$$\sum_{i=1}^{n} \delta_i \widehat{\beta} z_i - \sum_{i=1}^{n} \delta_i \log\left(\sum_{j=i}^{n} e^{\widehat{\beta} z_j} + n\widehat{\lambda} g(T_i)\delta_i\right) - \sum_{i=1}^{n} \frac{\delta_i \sum_{j=i}^{n} e^{\widehat{\beta} z_j}}{\sum_{j=i}^{n} e^{\widehat{\beta} z_j} + n\widehat{\lambda} g(T_i)\delta_i}.$$

If we let $C(\beta, \lambda) =$

$$\sum_{i=1}^{n} \delta_i \beta z_i - \sum_{i=1}^{n} \delta_i \log \left(\sum_{j=i}^{n} e^{\beta z_j} + n\lambda g(T_i)\delta_i \right) - \sum_{i=1}^{n} \frac{\delta_i \sum_{j=i}^{n} e^{\beta z_j}}{\sum_{j=i}^{n} e^{\beta z_j} + n\lambda g(T_i)\delta_i},$$

and combine step I and II, we have that the Wilks statistic

$$-2\log\mathcal{ALR}^c(\beta_0) = -2\log \frac{\max\limits_{\{\Lambda \ll \hat{\Lambda}_{NA},\ \Lambda \text{satisfy}(7)\}} \mathcal{A}L^c(\beta_0, \Lambda)}{\max\limits_{\{\beta, \Lambda \ll \hat{\Lambda}_{NA},\ \Lambda \text{satisfy}(7)\}} \mathcal{A}L^c(\beta, \Lambda)}$$

$$= 2(C(\hat{\beta}, \hat{\lambda}) - C(\beta_0, \hat{\gamma})) = T_1 - T_2. \qquad (say)$$

We can verify easily that for any β value we have

$$\frac{\partial C(\beta,\lambda)}{\partial \lambda}|_{\lambda=0} = -\sum \frac{n\delta_i g(T_i)}{\sum e^{\beta z_j} + n\lambda g(T_i)}|_{\lambda=0} + \sum \frac{n\delta_i g(T_i)\sum e^{\beta z_j}}{(\sum e^{\beta z_j} + n\lambda g(T_i))^2}|_{\lambda=0} = 0 \tag{14}$$

and

$$\frac{\partial^2 C(\beta,\lambda)}{\partial \lambda^2}|_{\lambda=0,\beta=\beta_0} = \sum \frac{\delta_i n^2 g^2(T_i)}{(\sum)^2} - 2\sum \frac{\delta_i n^2 g^2(T_i)}{(\sum)^2} = -n^2 \sum \frac{\delta_i g^2(T_i)}{[\sum e^{\beta_0 z_j}]^2} =$$

We have the following Taylor expansion

$$T_2 = 2\{C(\beta_0,0) + \hat{\gamma} C'(\beta_0,0) + 1/2 C''(\beta_0,0)\hat{\gamma}^2 + o(1)\} = 2C(\beta_0,0) - nA\hat{\gamma}^2 + o($$

On the other hand

$$T_1 = 2\{C(\beta_0,0) + (\hat{\beta}-\beta_0,\hat{\lambda})(C'_\beta(\beta_0,0), C'_\lambda(\beta_0,0))^T + 1/2(\hat{\beta}-\beta_0,\hat{\lambda})C''(\beta_0,0)($$

$$= 2C(\beta_0, 0) + 2(\hat{\beta} - \beta_0)C'_\beta(\beta_0, 0) + (\hat{\beta} - \beta_0, \hat{\lambda})Q(\hat{\beta} - \beta_0, \hat{\lambda})^T + o(1)\}$$

where Q is the second derivative matrix of $C(\beta, \lambda)$ at $\beta = \beta_0, \lambda = 0$.

Now we compute Q. Notice also that

$$\frac{\partial C(\beta, \lambda)}{\partial \beta}|_{\lambda=0} = \sum \delta_i z_i - \sum \delta_i \frac{\sum z_j e^{\beta z_j}}{\sum e^{\beta z_j}} = \ell(\beta) \ .$$

and

$$\frac{\partial^2 C(\beta, \lambda)}{\partial \beta^2}|_{\lambda=0} = -I(\beta).$$

Also

$$\frac{\partial^2 C(\beta, \lambda)}{\partial \lambda \partial \beta}|_{\lambda=0} = \frac{\partial^2 C(\beta, \lambda)}{\partial \beta \partial \lambda}|_{\lambda=0} = 0.$$

Thus we have a diagonal matrix Q:

$$Q = \mathrm{diag}[-I(\beta_0), -nA].$$

Put these all together we have

$$\begin{aligned}
& -2\log\mathcal{ALR}^c(\beta_0) \\
= \ & \{nA\hat{\gamma}^2 + 2(\hat{\beta} - \beta_0)C'_\beta(\beta_0, 0) + (\hat{\beta} - \beta_0, \hat{\lambda})Q(\hat{\beta} - \beta_0, \hat{\lambda})^T + o(1)\} \\
= \ & (\sqrt{n}\frac{m(\beta_0, 0)}{\sqrt{A}} + o_p(1))^2 + 2((\hat{\beta} - \beta_0)C'_\beta(\beta_0, 0)) \\
& \quad + \ [\frac{\ell(\beta_0)}{\sqrt{n}}, \sqrt{n}m(\beta_0, 0)]D^{-1}\sqrt{n}\frac{Q}{n}(\sqrt{n}D^{-1})^T[\frac{\ell(\beta_0)}{\sqrt{n}}, \sqrt{n}m(\beta_0, 0)]^T \\
= \ & [\frac{\ell(\beta_0)}{\sqrt{n}}, \sqrt{n}m(\beta_0, 0)] \cdot (I + II + III) \cdot [\frac{\ell(\beta_0)}{\sqrt{n}}, \sqrt{n}m(\beta_0, 0)]^T + o_p(1) \\
= \ & [\frac{\ell(\beta_0)}{\sqrt{n}}, \sqrt{n}m(\beta_0, 0)] \cdot M \cdot [\frac{\ell(\beta_0)}{\sqrt{n}}, \sqrt{n}m(\beta_0, 0)]^T + o_p(1) \qquad (15)
\end{aligned}$$

where

$$I = \begin{pmatrix} 0 & 0 \\ 0 & A^{-1} \end{pmatrix} , \quad II = \frac{1}{A\Sigma + B^2} \begin{pmatrix} 2A & B \\ B & 0 \end{pmatrix} \quad \text{and} \quad III = \frac{1}{A\Sigma + B^2} \begin{pmatrix} -A \\ 0 \end{pmatrix}$$

and

$$M = \frac{1}{A\Sigma + B^2} \begin{pmatrix} A & B \\ B & B^2 A^{-1} \end{pmatrix} .$$

Now we are ready to present the Wilks theorem.

Theorem 4 *Assume all the conditions of Theorem 1. In addition we assume $g(\cdot)$ is square integrable wrt Λ_0. Finally assume the true baseline hazard satisfy (7). Then we have, as $n \to \infty$,*

$$-2 \log \mathcal{ALR}^c(\beta_0) \xrightarrow{\mathcal{D}} \chi^2_{(1)} \ .$$

Proof: We compute

$$M \cdot V = \frac{1}{A\Sigma + B^2} \begin{pmatrix} A\Sigma & AB \\ B\Sigma & B^2 \end{pmatrix} \ .$$

It is easy to verify that the above matrix is idempotent and has rank 1. In view of (15) and by Lemma 1 and 4 we have the desired result. QED

Lemma 4 (Graybill 1976) *Suppose* $Y \xrightarrow{\mathcal{D}} N(0, V)$ *and* M *is a symmetric matrix. Then* $YMY^T \xrightarrow{\mathcal{D}} \chi^2_p$ *if and only if* MV *is idempotent and* $rank(MV) = p$*.*

Remark If the regression coefficient β is a vector, then the same proof still holds with the limiting distribution becomes a χ^2_p where the integer $p = \text{dim}(\beta)$.

Remark We also get an improved estimator of the baseline hazard function, $\Lambda_0(t)$, which satisfy (7).

3 Computation of the Improved Estimator

We have modified the programs for the regular Cox model in R language (Gentleman and Ihaka 1996) survival package (Therneau) to do the computation for the new estimator here (it is open source). The package is called `coxEL`. The relevant function is `coxphEL()`. This function is similar to the function `coxph()` in the `survival` package for regular Cox model. But you need to supply additional input: a value `lam` and a function $g(\cdot)$ when calling `coxphEL()`.

It will solve (by iterative method) for β the equation (12) for the given λ value and $g(\cdot)$ function. It does not solve (10).

The program will output, among other things, $\widehat{\beta}$, the modified regression coefficient estimator, and the value

$$\sum_{i=1}^{n} \frac{\delta_i g(T_i)}{\sum_{j \in \Re_i} e^{\widehat{\beta} z_j^*} + n\lambda g(T_i)\delta_i} = \sum \delta_i g(T_i) \widehat{w}_i^* = \int g(t) d\widehat{\Lambda}_0^*(t) \qquad (16)$$

where $z_j^* = z_j - \bar{z}$ ...... since R/Splus (also SAS) re-centers the covariates, z, automatically.

If you happen to pick θ equal to this value, (16), then this λ also solves (12). If not, you need to adjust λ until you get (16) to equal to your θ. Notice (16) is monotone in λ so this is not too hard.

Therefore the baseline hazard is actually the hazard for a subject with $z = \bar{z}$ instead of $z = 0$. If you would rather recover the constraint value for the hazard at $z = 0$, we need to multiply the value obtained in (16) by $\exp(-\widehat{\beta}\bar{z})$.

For `lam` $= 0$, you get the regular Cox estimator, $\widehat{\beta}_c$, and the constraint value is the NPMLE of the integral $\int g d\Lambda_0$.

3.1 Some Preliminary Simulation Results

We use a two sample situation and both samples are exponentially distributed, and have same sample size. Sample 1 $\sim$ exp(0.2). Sample 2 $\sim$ exp(0.3). We use a binary covariate, z, to indicate the samples: if $z_i = 0$ then y_i is from sample 1; if $z_i = 1$ then y_i is from sample 2.

The risk ratio or hazard ratio is 0.3/0.2. In Cox model, this imply the true coefficient, β_0, should be $\log(0.3/0.2) = 0.4054651$, since $\exp(coef) * 0.2 = 0.3$.

We did not impose censoring in this simulation.

The extra information we suppose we have is that the integration

$$\int \exp(-t) d\Lambda_0(t) = \theta = rate1 \ .$$

When both sample have 200 observations, i.e. (y_i, z_i) $i = 1, \cdots, 400$, we obtained the following results:

We generate 400 such samples (each of size 400) and for each sample we compute the two Cox estimators of the regression coefficient. The mean and variance of the regular Cox estimator is (sample mean and sample variance based on 400 trials):

```
> mean(result1[2,1:400])
[1] 0.4160447
> var(result1[2,1:400])
[1] 0.009736113
```

The mean and variance of the improved Cox estimator:

```
> mean(result1[1,1:400])
[1] 0.4129862
> var(result1[1,1:400])
[1] 0.008310867
```

But for smaller sample sizes, the iteration computation sometimes has problem to converge for larger λ values. One reason is that the sample is too far away from the true value of the extra info required. Similar to "the true mean is zero, but the observations in the sample happens to be all positive" then the empirical likelihood computation is impossible. Those needs to be redefined as having infinite likelihood ratio.

Remark: The above example actually demonstrated a better log-rank test in the two sample case. The estimator we compared can be think of as the Hodges-Lehmann estimator derived from tests.

Information on the baseline is that the integration of baseline hazard is in an interval. We took the interval to be = true value ± 0.1.

Sample size = 90 + 90.

```
> mean( result1[1,1:500])
[1] 0.4187548
> mean( result1[2,1:500])
[1] 0.4194698
> var( result1[1,1:500])
[1] 0.02708653
> var( result1[2,1:500])
[1] 0.02715997
```

result1 is better!! and is closer to the true value: $\log(1.5) = 0.4054651$.

If the integral is inside the interval, no adjustment. If the integration is outside, adjust to the boundary.

Multiple regression. (no problem).

More information on baseline.

For extra information in the form of many equations like (7), with many $g()$ functions, $m(\cdot)$ in (8) becomes a vector, and

$$\sqrt{n}(\widehat{\beta}-\beta_0) = \frac{\ell(\beta_0)}{\sqrt{n}}(I/n+BA^{-1}B)^{-1}+\sqrt{n}m(\beta_0,0)[A+B(I/n)^{-1}B]^{-1}B(I/n)$$

with the obvious definition of matrix A and vector B. This leads to an estimator $\widehat{\beta}$ that is asymptotically normal with asymptotic variance given by

$$[\Sigma^*]^{-1} = [\Sigma + B^T A^{-1} B]^{-1} \ .$$

Let us call the Fisher information of β in the restricted baseline Cox model as

$$\Sigma^* = [\Sigma + B^T A^{-1} B] \ .$$

The quantity $B^T A^{-1} B$ is the increment of the Fisher information due to the restriction on baseline. It is a special matrix, by a Lemma of Kim and Zhou (2002) this can be written as a summation that approximates an integration.

When $g_i(t)$ are indicator functions: $g_i(t) = I_{[t \leq u_i]}$ for several constants u_i, the increment in the Fisher information, $B^T A^{-1} B$ takes a particular simple form:

$$B^T A^{-1} B = \sum \frac{[h(u_i) - h(u_{i-1})]^2}{V(u_i) - V(u_{i-1})} \ ,$$

where $h(u_i) = B_i$ and $A_{ij} = V(\min(u_i, u_j))$.

$$h(t) = \sum_{T_i \le t} \frac{\delta_i \sum_{j \in \Re_i} z_j e^{\beta_0 z_j}}{\left[\sum_{j \in \Re_i} e^{\beta_0 z_j}\right]^2};$$

$$V(t) = \sum_{T_i \le t} \frac{\delta_i n}{\left[\sum_{j \in \Re_i} e^{\beta_0 z_j}\right]^2} \ .$$

It will approach from below the integral (when u_i become dense)

$$\int \frac{[h'(t)]^2}{V'(t)} dt$$

In the limit, this integration is also equal to

$$\lim \int \frac{[\sum z_j e^{\beta z_j} / \sum e^{\beta z_j}]^2}{n / \sum e^{\beta z_j}} d\Lambda_0(t) = \lim \sum_{i=1}^{n} \left(\frac{\sum z_j e^{\beta z_j}}{\sum e^{\beta z_j}} \right)^2 \frac{\delta_i}{n}.$$

In view of the expression of $I(\beta_0)$ in (4), we see that the Fisher information of the restricted baseline hazard Cox model can approach but never exceed the upper bound

$$\Sigma^* = [\Sigma + B^T A^{-1} B] \leq \Sigma^{**}$$

with

$$\Sigma^{**} = \lim \frac{1}{n} \sum_{i=1}^{n} \delta_i \frac{\sum\limits_{j \in \Re_i} z_j^2 \exp(\beta_0 z_j)}{\sum\limits_{j \in \Re_i} \exp(\beta_0 z_j)} = \lim \frac{I^{**}(\beta_0)}{n}$$

The relation between $I(\beta)$ and $I^{**}(\beta)$ is like that of a variance and a second moment.

We have the following equality in expected information.

Theorem The fully parametric proportional hazards model (where the baseline is completely specified) has the expected information for β easily calculated (when no censoring) as

$$I_{para}(\beta) = \sum_{i=1}^{n} z_i^2 \ ,$$

where we used the fact that $EH_i(Y_i) = 1$ since $H_i(Y_i) \sim \exp(1)$. With censoring, the information is

$$E \sum z_i^2 H_i(\min(Y_i, C_i)) = \sum z_i^2 E H_i(\min(Y_i, C_i)).$$

We have the following

Theorem At least without censoring, the expected information

$$EI^{**} = I_{para}$$

where the expectation is over all the possible ordering of the obsverations (when no censoring). This can be proved by induction.

Summerizing the results above paints the following picture: more restrictions on the baseline hazard increases the Fisher information of β. They form a continuous spectrum from the completely unspecified baseline model (i.e. Cox model with information I) to completely specified baseline model (parametric model) with information $I_{para} = EI^{**}$.

The maximum empirical likelihood estimator (MELE) have variances given by the inverse of those informations in the spectrum.

This also show that empirical likelihood is a continuous extension of parametric likelihood when nuisance parameter is of infinite dimesional, in the sense that, in terms of information, it reduces to parametric likelihood with added restrictions on the infinite dimensional nuisance parameter.

This “Information Spectrum” phenomena also showing up in the Envelope Empirical Likelihood as described by Zhou (2000) where on one end is the Fisher information of a location estimator (like median) with arbitrary distribution, on the other end is the Fisher information of location parameter with a symmetric (but unknown) distribution.

Both ends are semiparametric models.

References

Andersen, P.K., Borgan, O., Gill, R. and Keiding, N. (1993), *Statistical Models Based on Counting Processes.* Springer, New York.

Cox, D. R. (1972). Regression Models and Life Tables (with discussion) *J. Roy. Statist. Soc B.*, **34**, 187-220.

Cox, D. R. (1975). Partial Likelihood. *Biometrika*, **62**, 269-276.

Li, G. (1995). On nonparametric likelihood ratio estimation of survival probabilities for censored data. *Statist. & Prob. Letters*, **25** , 95-104.

Owen, A. (1988). *Empirical Likelihood Ratio Confidence Intervals for a Single Functional.* Biometrika, 75 237-249.

Owen, A. (2001). *Empirical likelihood.* Chapman Hall, London.

Pan, X.R. (1997). *Empirical Likelihood Ratio Method for Censored Data. Ph.D. Thesis, Univ. of Kentucky, Dept. of Statist.*

Pan, X.R. and Zhou, M. (1999). Empirical likelihood in terms of cumulative hazard function for censored data. *Univ. of Kentucky, Dept. of Statist. Tech Report # 361* J. Multi vari analysis.

Gentleman, R. and Ihaka, R. (1996). R: A Language for data analysis and graphics. *J. of Computational and Graphical Statistics*, **5**, 299-314.

Thomas, D. R. and Grunkemeier, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. *J. Amer. Statist. Assoc.* **70**, 865-871.

The Regularity Conditions A - D

A.(Finite interval).

B.(Asymptotic stability).

C.(Lindeberg condition).

D.(Asymptotic regularity conditions).

Proof of The joint CLT of martingales (Lemma 1): It is now a standard result that we have the following martingale representation:

$$m(\beta_0, 0) = \sum_{i=1}^{n} \frac{\delta_i g(T_i)}{\sum_{T_j \geq T_i} e^{\beta_0 z_j}} - \int g(s) d\Lambda_0(s) = \sum_{i=1}^{n} \int \frac{g(s)}{\sum_{j=1}^{n} e^{\beta_0 z_j} I_{[T_j \geq s]}} dM_i(s)$$

and

$$\ell(\beta_0) = \sum_{i=1}^{n} \int \left(z_i - \frac{\sum_{j=1}^{n} z_j e^{\beta_0 z_j} I_{[T_j \geq s]}}{\sum_{j=1}^{n} e^{\beta_0 z_j} I_{[T_j \geq s]}} \right) dM_i(s)$$

where

$$M_i(t) = I_{[T_i < t]} - \int^t I_{[T_i > s]} \exp(\beta_0 z_i) d\Lambda_0(s)$$

with

$$< M_i(t) >= \int_0^t I_{[T_i > s]} \exp(\beta_0 z_i) d\Lambda_0(s)$$

Using the standard computation of the predictable quadratic process for the martingales we get

$$< \sqrt{n} m(\beta_0, 0), \ell(\beta_0)/\sqrt{n} >= 0 \qquad \text{and} \qquad < \sqrt{n} m(\beta_0, 0) >= V_{22}$$

etc.

Similar to the proof of Theorem 1, we may use the CLT for the martingales (Rebolledo theorem) to get the convergence in distribution result.

$$\Sigma = < \ell(\beta_0)/\sqrt{n} >$$