Doubly Censored Data

In paired comparison experiments, suppose we have n pairs of observations from two treatments on n subjects.

To estimate the treatment difference, it is customary to focus on the n pairwise differences. If there is no censoring, a paired t-test or some rank based test could be used.

However, when right censoring occurs in either treatments, the pairwise difference can be right censored, or left censored, as the following table shows.

Trt 1	Trt 2	diff= $(trt1 - trt2)$
14+	6	8+
12	7	5
9	5+	4−

$$\cdots \; \cdots$$

A reasonable model for the paired experiment is as follows: for the ith subject (or pair) $(i = 1, 2, \ldots, n)$ we observe Y_{1i} and Y_{2i} where

$$Y_{1i} = \tau_d + S_i + \epsilon_{1i} \ ,$$

$$Y_{2i} = \tau_p + S_i + \epsilon_{2i} \ ,$$

where $\tau_d(\tau_p)$ is the main effect for drug (placebo), S_i is the subject effect, ϵ_{ki} is the random error. The difference of Y_{1i} and Y_{2i} is:

$$D_i = (\tau_d - \tau_p) + (\epsilon_{1i} - \epsilon_{2i}),$$

which is free from S_i. If we assume ϵ_{1i} and ϵ_{2i} are exchangeable, then the median of D_i is $\tau_d - \tau_p$. Thus a test of $H_0 : \tau_d - \tau_p = 0$ can be carried out by testing if the median of D_i is zero.

In the case where ϵ_{ki} are i.i.d. with a distribution of $\exp(\lambda) - 1/\lambda$ (mean zero exponential), D_i has double exponential distribution with location parameter $\tau_d - \tau_p$. Since the sample median is the MLE of the location parameter for a double exponential distribution, we can expect the test to perform well in this case.

```
> myfun <- function(t) { as.numeric(t <= 0) }
>
> el.cen.EM( Dvec, delta, fun= myfun, mu=0.5)
```

Simulation: (based on 5000 runs)

Difference	Size	No Censored	Light Censored	Medium Censored
0.0	$n = 100$	0.053	0.054	0.058
	$n = 25$	0.077	0.048	0.053
0.3	$n = 100$	0.968	0.907	0.819
	$n = 25$	0.487	0.374	0.284
0.2	$n = 100$	0.767	0.635	0.487
	$n = 25$	0.276	0.205	0.146

The Percentage Of Rejecting $H_0 : \tau_d = \tau_p$ *At* $\alpha = 0.05$

When the null hypothesis is true, the percentages of rejecting H_0 are very close to the nominal level 0.05 for small and large samples. When there is a difference between the two treatments (drug and placebo), the rejecting percentages are higher than 0.05 but decreases with the increase of censoring percentages.

Remark The variance of the MLE is hard to estimate, the Wald confidence interval is hard to construct.

(E-step) The EM algorithm used to find the maximum is similar to that of Turnbull (1976) but we modified the M-step to incorporate the constraint on the mean.

(M-step) Find the maximum under mean constraint by using Lagrange multiplier, see Owen (1988).

Cox Proportional Hazards Regression Model

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

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

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.

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

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 on the baseline. It improves estimation of β.

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.

The contribution of T_i, δ_i to the empirical likelihood function is

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

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

$$\mathcal{A}L^c(\beta, \Lambda_0) = \prod_{i=1}^{n} (\Delta\Lambda_0(T_i) e^{\beta z_i})^{\delta_i} \exp\{-e^{\beta z_i} \sum_{j: T_j \le T_i} \Delta\Lambda_0(T_j)\}, \tag{3}$$

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).

2 Empirical Likelihood Ratio Statistic for β_0 with Additional Information on Baseline

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

$$\int g(s) d\Lambda_0(s) = \sum g(T_i) \Delta\Lambda_0(T_i) = \theta \tag{4}$$

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. It is the maximizer of the empirical likelihood subject to the constraint:

$$\max_{\beta, \Lambda \ll \hat{\Lambda}_{NA}, satisfy (4)} \mathcal{A}L^c(\beta, \Lambda)$$

Theorem 3 *As* $n \to \infty$ *the regression estimator with additional information (4),* $\widehat{\beta}$*, 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.*

It is interesting to note that the variance of $\widehat{\beta}$ above is smaller than that of a regular Cox estimator.

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 (4). Then we have, as* $n \to \infty$*,*

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

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 $p = \mathrm{dim}(\beta)$.

Remark We also get an improved estimator of the baseline hazard function, $\Lambda_0(t)$, which satisfy (4).

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 two additional inputs: a value `lam` and a function $g(\cdot)$ when calling `coxphEL()`.

If `lam`$= 0$ then you get the regular Cox estimator, and the NPMLE of $\int g(t) d\Lambda(t)$.

For non-zero `lam` values, you will get an $\widehat{\beta}$ estimator, and the value of the summation on the left of (4).

In Splus/R, 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 $\exp(-\widehat{\beta}\bar{z})$.

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 generated 400 such samples (each of size 400) and for each sample we computed the two Cox estimators of the regression coefficient, β.

Therefore the sample means and sample variances below are based on 400 simulation runs.

	sample mean	sample variance
Regular Cox estimator	0.4160447	0.009736113
Adjusted Cox estimator	0.4129862	0.008310867

Table 1

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.

In the next two simulations, we only adjust the regular Cox estimator when the value of the integration $\int g(t) d\Lambda_0(t)$ is outside the interval $[\theta - \epsilon, \theta + \epsilon]$ where θ is the true value of the integration.

For sample size $n = 400$, $\epsilon = 0.05$ we obtained the following results:

	sample mean	sample variance
Regular Cox estimator	0.4160447	0.009736113
Adjusted Cox estimator	0.4085578	0.009332247

Table 2

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

For sample size $n = 180$ (equal sample size of 90 each), $\epsilon = 0.1$, the results are as follows:

	sample mean	sample variance
Regular Cox estimator	0.4194698	0.02715997
Adjusted Cox estimator	0.4187548	0.02708653

Table 3

The sample mean and sample variance reported above are based on 500 simulation runs.

More information on baseline.

For extra information in the form of many equations like (4), with many $g()$ functions, we have

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

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: $g_i(t) = I_{[t \le u_i]}$ 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

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

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 observations (when no censoring). This can be proved by induction.

Summarizing the results above paints the following picture: more information 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 dimensional, 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.

Efron, B. (1977). Efficiency of Cox's likelihood function for censored data. *JASA* **72**, 557-565.

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

Oakes, D. (1977). The asymptotic information in censored survival data. *Biometrika*, **64**, 441-448.

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.