Kamakura, Toshinari

Working Paper

Computational Methods in Survival Analysis

Papers / Humboldt-Universität Berlin, Center for Applied Statistics and Economics (CASE), No. 2004,29

Provided in cooperation with:
Humboldt-Universität Berlin

Computational Methods in Survival Analysis

Toshinari Kamakura[1]

Chuo University, Kasuga 1-13-27, Bunkyo-ku, Tokyo 112-8551, Japan
kamakura@indsys.chuo-u.ac.jp

Survival analysis is widely used in the fields of medical science, pharmaceutics, reliability and financial engineering, and many others to analyze positive random phenomena defined by event occurrences of particular interest. In the reliability field, we are concerned with the time to failure of some physical component such as an electronic device or a machine part. This article briefly describes statistical survival techniques developed recently from the standpoint of statistical computational methods focussing on obtaining the good estimates of distribution parameters by simple calculations based on the first moment and conditional likelihood for eliminating nuisance parameters and approximation of the likelihoods. The method of partial likelihood (Cox, 1972, 1975) was originally proposed from the view point of conditional likelihood for avoiding estimating the nuisance parameters of the baseline hazards for obtaining simple and good estimates of the structure parameters. However, in case of heavy ties of failure times calculating the partial likelihood does not succeed. Then the approximations of the partial likelihood have been studied, which will be described in the later section and a good approximation method will be explained. We believe that the better approximation method and the better statistical model should play an important role in lessening the computational burdens greatly.

1 Introduction

Let T be a positive random variable with density function $f(t)$ and distribution function $F(t)$. The survival function $S(t)$ is then defined as

$$S(t) = 1 - F(t) = \Pr\{T > t\},$$

and the hazard function or hazard rate as

$$\lambda(t) = \lim_{h \to 0} \frac{\Pr\{t < T \le t + h | T > t\}}{h}.$$

The hazard function can also be expressed as

$$\lambda(t) = \frac{f(t)}{S(t)}. \tag{1}$$

The right-hand side (RHS) of Eq.(1) becomes

$$\frac{f(t)}{S(t)} = -\frac{d}{dt} \log S(t),$$

and inversely

$$S(t) = \exp\left\{-\int_0^t \lambda(u)du\right\}. \tag{2}$$

1.1 Nonparametric Model

We assume that the observed data set consists of failure or death times t_i and censoring indicators δ_i, $i = 1, \cdots, n$. The indicator δ is unity for the case of failure and zero for censoring. The censoring scheme is an important concept in survival analysis in that one can observe partial information associated with the survival random variable. This is due to some limitations such as loss to follow-up, drop-out, termination of the study, and others.

The Kaplan-Meier method (Kaplan and Meier, 1958) is currently the standard for estimating the nonparametric survival function. For the case of a sample without any censoring observations, the estimate exactly corresponds to the derivation from the empirical distribution. The dataset can be arranged in table form, i.e.,

Table 1. Failure time data

failure times	t_1	t_2	$\cdots$	t_i	$\cdots$	t_k
the number of failures	d_1	d_2	$\cdots$	d_i	$\cdots$	d_k
the number of individuals of risk set	n_1	n_2	$\cdots$	n_i	$\cdots$	n_k

where, t_i is the i-th order statistic when they are arranged in ascending order for distinct failure times, d_i is the number of failures at the time of t_i, and n_i is the number of survivors at time $t_i - 0$. Under this notation the Kaplan-Meier estimate becomes

$$\hat{S}(t) = \prod_{j:t_j<t} \left(1 - \frac{d_j}{n_j}\right). \tag{3}$$

The standard error of the Kaplan-Meier estimate is

$$SE\left\{\hat{S}(t)\right\} = \left[\hat{S}(t)\right] \left\{\sum_{j:t_j<t} \frac{d_j}{n_j(n_j - d_j)}\right\}^{1/2}. \tag{4}$$

The above formula is called "Greenwood's formula" described by Greenwood (1926).

1.2 Parametric Models

The most important and widely-used models in survival analysis are exponential, Weibull, log-normal, log-logistic, and gamma distributions. The first two models will be introduced for later consideration. The exponential distribution is simplistic and easy to handle, being similar to a standard distribution in some respects, while the Weibull distribution is a generalization of the exponential distribution and allows inclusion of many types of shapes. Their density functions are

$$f(t;\lambda) = \lambda e^{-\lambda t} \quad (\lambda, t > 0) \tag{5}$$

$$f(t;m,\eta) = \frac{m}{\eta}\left(\frac{t}{\eta}\right)^{m-1} \exp\left\{-\left(\frac{t}{\eta}\right)^{m}\right\} \quad (m,\eta,t>0), \tag{6}$$

where the parameter λ is sometimes called the failure rate in reliability engineering. Two models may include additional threshold parameters, or guarantee times. Let γ be this threshold parameter. The Weibull density function then becomes

$$f(t;m,\eta,\gamma) = \frac{m}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{m-1} \exp\left\{-\left(\frac{t-\gamma}{\eta}\right)^{m}\right\} \quad (m,\eta,\gamma,t>0). \tag{7}$$

Here, note that in the case of $m = 1$, the Weibull probability density function is exactly the exponential density function placing $\lambda = 1/\eta$, and that we cannot observe any failure times before threshold time ($t < \gamma$) or an individual cannot die before this time.

As the Weibull distribution completely includes the exponential distribution, only the Weibull model will be discussed further. The Weibull distribution is widely used in reliability and biomedical engineering because of goodness of fit to data and ease of handling. The main objective in lifetime analysis sometimes involves (i) estimation of a few parameters which define the Weibull distribution, and (ii) evaluation of the effects of some environmental factors on lifetime distribution using regression techniques. Inference on the quantiles of the distribution has been previously studied in detail (Johnson et al., 1994).

The maximum likelihood estimate (MLE) is well known, yet it is not expressed explicitly in closed form. Accordingly, some iterative computational methods are used. Menon (Menon (1963)) provided a simple estimator of $1/m$, being a consistent estimate of $1/m$, with a bias that tends to vanish as the sample size increases. Later, Cohen (Cohen, 1965; Cohen and Whitten, 1988) presented a practically useful chart for obtaining a good first approximation to the shape parameter m using the property that the coefficient of variation of the Weibull distribution is a function of the shape parameter m, i.e., it does not depend on η. This is described as follows.

Let T be a random variable with probability density function (6), the r-th moment around the origin is then calculated as

$$E[T^r] = \eta^r \Gamma\left(1 + \frac{r}{m}\right).$$

Here $\Gamma(\cdot)$ is the complete gamma function. ¿From this, the first two moments obtained are the mean life and variance, i.e.,

$$E[T] = \eta\Gamma\left(1 + \frac{1}{m}\right),$$

$$Var[T] = \eta^2\left\{\Gamma\left(1 + \frac{2}{m}\right) - \Gamma^2\left(1 + \frac{1}{m}\right)\right\}.$$

Considering that the coefficient of variation

$$CV = \sqrt{(Var[T])}/E[T]$$

does not depend on the parameter η allows obtaining simple and robust moment estimates, which may be the initial values of the maximum likelihood calculations. Dubey (1967) studied the behavior of the Weibull distribution in detail based on these moments, concluding that the Weibull distribution with shape parameter $m = 3.6$ is relatively similar to the normal distribution.

Regarding the three-parameter Weibull described by (7), Cohen and Whitten (1988) suggested using the method of moments equations, noting that

$$\begin{aligned} E[T] &= \gamma + \eta\Gamma_1(m), \\ Var[T] &= \eta^2\left\{\Gamma_2(m) - \Gamma_1^2(m)\right\}, \\ E[X_{(1)}] &= \gamma + \frac{\eta}{n^{1/m}}\Gamma_1(m), \end{aligned}$$

and equating them to corresponding samples, where $\Gamma_r(m) = \Gamma(1 + r/m)$.

As for obtaining an inference on the parameter of the mean parameter $\mu = E(T)$, this has not yet been investigated and will now be discussed. When one would like to estimate μ, use of either the MLE or the standard sample mean is best for considering the case of an unknown shape parameter. This is true because the asymptotic relative efficiency of the sample mean to the MLE is calculated as

$$\begin{aligned} ARE(\bar{T}) &= \frac{nAvar(\tilde{\mu})}{nAvar(\bar{T})} \\ &= \frac{6}{m^2\pi^2}\cdot\frac{1}{CV^2}\left[\frac{\pi^2}{6} + \{c - 1 + \psi(1 + 1/m)\}^2\right], \end{aligned} \tag{8}$$

where c is Euler's constant, $\psi(\cdot)$ a digamma function, $\tilde{\mu}$ the MLE, and $\bar{T}$ the sample mean.

Table 2. ARE of the sample mean to the MLE

m	eff	m	eff	m	eff
.1	0.0018	1.1	.9997	2.1	.9980
.2	0.1993	1.2	.9993	2.2	.9981
.3	0.5771	1.3	.9988	2.3	.9982
.4	0.8119	1.4	.9984	2.4	.9983
.5	0.9216	1.5	.9981	2.5	.9984
.6	0.9691	1.6	.9980	2.6	.9984
.7	0.9890	1.7	.9979	2.7	.9985
.8	0.9968	1.8	.9979	2.8	.9985
.9	0.9995	1.9	.9979	2.9	.9985
1.0	1.0000	2.0	.9980	3.0	.9986

Table 2 gives the ARE with respect to various values of m. Note the remarkably high efficiency of the sample mean, especially for $m \geq 0.5$, where more than 90% efficiency is indicated. The behavior of $ARE(\bar{T})$ form $m > 1$ is that $ARE(\bar{X})$ has a local minimum 0.9979 at $m = 1.7884$ and a local maximum 0.9986 at $m = 3.1298$, and that for the larger m, $ARE(\bar{T})$ monotonically decreases in m and the infimum of $ARE(\bar{T})$ is given in $m \to \infty$;

$$\lim_{m\to\infty} ARE(\bar{T}) = \frac{6(\pi^2 + 6)}{\pi^4} \cong 0.9775. \tag{9}$$

When m is known and tends to infinity, the behavior of $ARE(\bar{T})$ is as follows:

$$\lim_{m\to\infty} \frac{1}{(mCV)^2} = \frac{6}{\pi^2} \cong 0.6079. \tag{10}$$

A higher relative efficiency of the sample mean for unknown m is shown compared to known m. From a practical standpoint, the sample mean is easily calculated for a point estimation of the Weibull mean if no censored data are included. These results support the benefits of using the sample mean for the complete sample.

2 Estimation of Shape or Power Parameter

Let us now consider the class of the lifetime distributions, whose distribution functions are expressed by

$$F(t; \alpha, \gamma, \sigma) = G\left(\left(\frac{t-\gamma}{\sigma}\right)^{\alpha}\right), \tag{11}$$

where $G(\cdot)$ is also a distribution function. For the Weibull model, $G(t) = 1 - \exp(-t)$ is an exponential distribution. Nagatsuka and Kamakura (Nagatsuka and Kamakura, 2003, 2004) proposed a new method using the location-scale-free transformation of data set to estimate the power parameter in the

Castillo-Hadi model (Castillo and Hadi, 1995). That is, let $T_1, \dots, T_n$ be independently distributed according to the distribution function (11). Consider the W-transformation to be defined as

$$W_i = \frac{T_i - T_{(1)}}{T_{(n)} - T_{(1)}}, \quad (i = 2, \dots, n-1), \tag{12}$$

where $T_{(k)}$ is the k-th order statistic of T_i's. The new random variables W_i's derived by this W-transformation are then free from location and scale parameter. The arithmetic mean of W_i's gives the approximation to the original distribution of T. Let $V_i, i = 1, \dots, n$ be i.i.d. distributed with common distribution function $F_V(v)$, and let the i-th order statistic $V_{(i)}$ have the marginal distribution function $F_{V_{(i)}}(v)$. Then

$$F_v(v) = \frac{1}{n}\sum_{i=1}^{n} F_{V_{(i)}}(v). \tag{13}$$

This equation indicates that the arithmetic mean of the marginal distributions of n order statistics is exactly the original distribution. In the case of the Cstillo-Hadi Model, Nagatsuka and Kamakura (2004) provided a theorem regarding this approximation, i.e.,

Theorem 1. *(Nagatsuka and Kamakura, 2004)*
The mixture of the marginal distributions of $W_{(i)}$*,* $i = 2, \dots, n-1$*:*

$$F^{(n)}(w) = \frac{1}{n-2}\sum_{i=2}^{n-1} F_{W_{(i)}}(w) \tag{14}$$

is the approximate distribution of W_i*'s and the limiting distribution (14) is the power function distribution with parameter* $1/\alpha$*. That is*

$$\lim_{n\to\infty} \frac{1}{n-2}\sum_{i=2}^{n-1} F_{W_{(i)}}(w) = w^{\frac{1}{\alpha}}, \quad 0 < w < 1.$$

In the case of the Weibull distribution, the marginal distribution of $W_{(i)}$ is calculated as

$$
\begin{aligned}
F_{W_{(i)}}(w) &= \Pr\left(W_{(i)} \le w\right) \\
&= \Pr\left(\frac{T_{(i)} - T_{(1)}}{T_{(n)} - T_{(1)}} \le w\right) \\
&= \int_0^\infty \int_u^\infty n(n-1) f(u) f(v) \Bigg[\sum_{k=i-1}^{n-2} \binom{n-2}{k} \\
&\quad \times \left\{F((1-w)u + wv) - F(u)\right\}^k \\
&\quad \times \left\{F(v) - F((1-w)u + wv)\right\}^{n-k-2} \Bigg] dv du \\
&= \int_0^1 \int_u^1 n(n-1) \sum_{k=i-1}^{n-2} \binom{n-2}{k} \\
&\quad \times \left[1 - \exp\left\{-\alpha(w,m,u,v)\right\} - u\right]^k \\
&\quad \times \left[v - \left(1 - \exp\left\{-\alpha(w,m,u,v)\right\}\right)\right]^{n-k-2}, \qquad (15)
\end{aligned}
$$

where

$$
\alpha(w,m,u,v) = \left[(1-w)\left\{-\log(1-u)\right\}^{\frac{1}{m}} + w\left\{-\log(1-v)\right\}^{\frac{1}{m}}\right]^m .
$$

Calculations show that $F^{(n)}(w)$ has a first moment of

$$
\begin{aligned}
\mu_n(m) &= \int_0^\infty \left\{1 - F^{(n)}(w)\right\} dw \\
&= -\frac{1}{n-2} + \frac{n(n-1)}{m} \int_0^1 \int_u^1 (v-u)^{n-3} \\
&\quad \times \frac{\Gamma(\frac{1}{m}, -\log(1-u), -\log(1-v))}{\left\{-\log(1-v)\right\}^{\frac{1}{m}} - \left\{-\log(1-u)\right\}^{\frac{1}{m}}} dv du. \qquad (16)
\end{aligned}
$$

where $\Gamma(\cdot,\cdot,\cdot)$ is the incomplete generalized gamma function defined by

$$
\Gamma(a, z_0, z_1) = \int_{z_0}^{z_1} t^{a-1} e^{-t} dt.
$$

Now, an estimating of the shape parameter m is obtained by equating the theoretical population mean with sample mean of W-transformed W's. Nagatsuka and Kamakura (2003) provided a table for obtaining estimates and concluded based on simulation studies that the robust estimate of m is possible without using any existing threshold parameter.

3 Regression Models

Survival analysis is now a standard statistical method for lifetime data. Fundamental and classical parametric distributions are also very important, but regression methods are very powerful to analyze the effects of some covariates on life lengths. Cox (1972) introduced a model for the hazard function $\lambda(t;x)$ with survival time T for an individual with possibly time-dependent covariate x, i.e.,

$$\lambda(t;x) = \lambda_0(t)\exp(\beta^\top x), \tag{17}$$

where λ_0(t) is an arbitrary and unspecified base-line hazard function and $x^\top = (x_1, \ldots, x_p)$ and $\beta^\top = (\beta_1, \ldots, \beta_p)$. Cox generalized (17) this to a discrete logistic model expressing y as

$$\frac{\lambda(t;x)}{1-\lambda(t;x)} = \frac{\lambda_0(t)}{1-\lambda_0(t)}\exp(\beta^\top x). \tag{18}$$

Kamakura and Yanagimoto (1983) compared the estimators of regression parameters in the proportional hazards model (17) or (18) when we take the following methods; the Breslow-Peto (Breslow, 1974; Peto, 1972) method, the partial likelihood (Cox, 1972, 1975) method and the generalized maximum likelihood method (Kalbfleish and Prentice, 1980; Miller, 1981).

3.1 The Score Test

In many applications it is necessary to test the significance of the estimated value, using for example the score test or the likelihood ratio test based on asymptotic results of large sample theory. First we express the three likelihood factors defined at each failure time as L_{BP}, L_{PL}, L_{GML} corresponding to the Breslow-Peto, the partial likelihood and the generalized maximum likelihood methods, respectively;

$$L_{BP}(\beta) = \frac{\prod_{i=1}^r \exp(\beta^\top x_i)}{\{\sum_{i=1}^n \exp(\beta^\top x_i)\}^r}, \tag{19}$$

$$L_{PL}(\beta) = \frac{\prod_{i=1}^r \exp(\beta^\top x_i)}{\sum_\Psi \prod_{i=1}^r \exp(\beta^\top x_{\psi_i})}, \tag{20}$$

$$L_{GML}(\beta) = \frac{\prod_{i=1}^r \lambda\exp(\beta^\top x_i)}{\prod_{i=1}^n \{1+\lambda\exp(\beta^\top x_i)\}}, \tag{21}$$

where $x_1, \ldots, x_n$ denote covariate vectors for n individuals at risk at a failure time and $x_1, \ldots, x_r$ correspond to the failures, and Ψ denotes the set of all subsets $\{\psi_1, \ldots, \psi_r\}$ of size r from $\{1, \ldots, n\}$. The overall likelihood obtained by each method is the product of these cases of many failure times. It can be shown that the first derivatives of the three log likelihoods with respect β have the same values, i.e.,

$$\sum_{i=1}^r x_{ji} - \frac{r}{n}\sum_{i=1}^n x_{ji} \quad (j = 1, \ldots, p)$$

at $\beta = 0$.

The Hessian matrices of the log likelihoods evaluated at $\beta = 0$ are respectively,

$$-\left(\frac{r}{n}\right) S,$$
$$-\left\{\frac{r(n-r)}{n(n-1)}\right\} S,$$
$$-\left\{\frac{r(n-r)}{n^2}\right\} S,$$

where S is a matrix whose elements $s_j k$ are defined by

$$s_{jk} = \sum_{i=1}^{n} (x_{ji} - \bar{x}_{j.})(x_{ki} - \bar{x}_{k.}).$$

The first two results were derived by Farewell and Prentice (1980). Maximizing out λ from L_{GML} gives the last one, which is obtained in an unpublished manuscript. Since

$$\frac{r}{n} \geq \frac{r(n-r)}{n(n-1)} > \frac{r(n-r)}{n^2},$$

we conclude that the Breslow-Peto approach is the most conservative one.

3.2 Evaluation of Estimators in the Cox Model

Farewell and Prentice (1980) pointed out in their simulation study that when the discrete logistic model is true the Breslow-Peto method causes downward bias compared to the partial likelihood method. This was proven in Kamakura and Yanagimoto (1983) for any sample when β is scalar-valued, i.e.,

Theorem 2. *(Kamakura and Yanagimoto, 1983)*
Let $\hat{\beta}_{BP}$ be the maximum likelihood estimator of $L_{BP}(\beta)$ and $\hat{\beta}_{PL}$ be that of $L_{BP}(\beta)$. Suppose that all x_i's are not identical. Then both $\hat{\beta}_{BP}$ and $\hat{\beta}_{PL}$ are unique, if they exist, and $sgn(\hat{\beta}_{BP}) = sgn(\hat{\beta}_{PL})$ and

$$\left|\hat{\beta}_{BP}\right| \leq \left|\hat{\beta}_{PL}\right|. \tag{22}$$

The equality in (22) holds when $\hat{\beta}_{PL}$ is equal to zero or the number of ties r is equal to one.

Corollary 1. *(Kamakura and Yanagimoto, 1983)*
The likelihood ratio test for $\beta = 0$ against $\beta \neq 0$ is also conservative if we use the Preslow-Peto method. The statement is also valid in the multivariate case.

This theorem and corollary confirm the conservatism of the Breslow-Peto approximation in relation to Cox's discrete model (Oaks, 2001).

3.3 Approximation of Partial Likelihood

Yanagimoto and Kamakura (1984) proposed an approximation method using full likelihood for the case of Cox's discrete model. Analytically the same problems appear in various fields of statistics. Prentice and Breslow (1978) and Farewell (1979) remarked that the inference procedure using the logistic model contains the same problems in case-control studies where data are summarized in multiple 2×2 or $k \times 2$ tables. The proportional hazards model provides a type of logistic model for the contingency table with ordered categories (Pregibon, 1982). As an extension of the proportional hazards model, the proportional intensity model in the point process is employed to describe an asthma attack in relation to environmental factors (Korn and Whittemoore, 1979; Yanagimoto and Kamakura, 1984). For convenience, although in some cases partial likelihood becomes conditional likelihood, we will use the term of partial likelihood.

It is worthwhile to explore the behavior of the maximum full likelihood estimator even when the maximum partial likelihood estimator is applicable. Both estimators obviously behave similarly in a rough sense, yet they are different in details. Identifying differences between the two estimators should be helpful in choosing one of the two.

We use the notation described in the previous section for expressing the two likelihoods. Differentiating $\log L_{PL}$ gives

$$LP(\beta) = \sum_{i=1}^{r} x_i - \frac{\sum_{\Psi} \sum_{\psi} x_j \exp\left(\beta^\top \sum_{\psi} x_j\right)}{\sum_{\Psi} \exp\left(\beta^\top \sum_{\psi} x_j\right)} = 0.$$

Differentiating $\log L_{GML}$ with respect to β and λ allows obtaining the maximum full likelihood estimator, i.e.,

$$\sum_{i=1}^{r} x_i - \sum_{i=1}^{n} \lambda x_i \frac{\exp(\beta^\top x_i)}{1 + \lambda \exp(\beta^\top x_i)} = 0$$

and

$$\frac{r}{\lambda} - \sum_{i=1}^{n} \frac{\exp(\beta^\top x_i)}{1 + \lambda \exp(\beta^\top x_i)}.$$

¿From the latter equation $\lambda(\beta)$ is uniquely determined for any fixed β. Using $\lambda(\beta)$, we define

$$LF(\beta) = \sum_{i=1}^{r} x_i - \sum_{i=1}^{n} \lambda(\beta) x_i \frac{\exp(\beta^\top x_i)}{1 + \lambda \exp(\beta^\top x_i)}.$$

The maximum full likelihood estimator, $\hat{\beta}_{GML}$, is a root of the equation $LF(\beta) = 0$. We denote $\lambda(\beta)$ by λ for simplicity.

Note that the entire likelihoods are the products over all distinct failure times T. Thus the likelihood equations in a strict sense are $\sum LP_t(\beta) = 0$ and $\sum LF_t(\beta) = 0$, where the summations extend over t in T. As far as we are concerned, the results in a single failure time can be straightforwardly extended to those with multiple failure times. Let us now focus on likelihood equations of a single failure time and suppress the suffix t.

Proposition 1. *(Yanagimoto and Kamakura, 1984)*
Let $K(\beta)$ be either of $LF(\beta)$ or $LP(\beta)$. Denote $\sum_{i=1}^n x_i/n$ by $\bar{x}$, and $x_{(1)} + \cdots + x_{(r)}$ and $x_{(n-r+1)} + \cdots + x_{(n)}$ by $L(x;r)$ and $U(x;r)$ respectively, where $x_{(1)}, \ldots, x_{(n)}$ are ordered covariates in ascending order. $K(\beta)$ accordingly has the following four properties:

(i) $K(0) = x_1 + \cdots + x_r - r\bar{x}$.
(ii) $K'(\beta)$ *is negative for any* β, *that is,* $K(\beta)$ *is strictly decreasing.*
(iii) $\lim_{\beta \to -\infty} K(\beta) = U(x;r)$.
(iv) $\lim_{\beta \to \infty} K(\beta) = L(x;r)$.

Extension to the case of vector parameter β is straightforward. ¿From Proposition 1 it follows that if either of the two estimators exists, then the other also exists and they are uniquely determined. Furthermore, both the estimators have a common sign.

Theorem 3. *(Yanagimoto and Kamakura, 1984)*
Suppose that $\sum(x_i - \bar{x})^2 \neq 0$. The functions $LP(\beta)$ and $LF(\beta)$ then have a unique intersection at $\beta = 0$. It also holds that $LP(\beta) < LF(\beta)$ for $\beta > 0$.The reverse inequality is valid for $\beta < 0$.

The above theorem proves that $\hat{\beta}_{GML} > \hat{\beta}_{PL}$ for the case of $LP(0) = LF(0) > 0$.

To quantitatively compare the behaviors of $LF(\beta)$ and $LP(\beta)$, their their power expansions are presented near the origin. Since both functions behave similarly, it is expected that the quantitative difference near the origin is critical over a wide range of β. Behavior near the origin is of practical importance for studying the estimator and test procedure.

Proposition 2. *(Yanagimoto and Kamakura, 1984)*
The power expansions of $LF(\beta)$ and $LP(\beta)$ near the origin up to the third order are as follows: for $n \geq 4$,

(i)

$$LF(\beta) \approx \sum_{i=1}^{r} x_i - \left[r\bar{x} + \frac{r(n-r)}{n^2} s_2 \beta + \frac{1}{2} \frac{r(n-r)(n-2r)}{n^3} s_3 \beta^2 + \frac{1}{6} \frac{r(n-r)}{n^5} \left\{ n(n^2 - 6rn + 6r^2) s_4 - 3(n-2r)^2 s_2^2 \right\} \beta^3 \right],$$

(ii) (Cox, 1970)

$$LP(\beta) \approx \sum_{i=1}^{r} x_i - \left[r\bar{x} + \frac{r(n-r)}{n(n-1)} s_2 \beta + \frac{1}{2} \frac{r(n-r)(n-2r)}{n(n-1)(n-2)} s_3 \beta^2 + \frac{1}{6} \frac{r(n-r)}{n^2(n-1)(n-2)(n-3)} \left\{ n(n^2 - 6rn + 6r^2 + n) s_4 + 3(r-1)n(n-r-1)s_2^2 \right\} \beta^3 \right],$$

where $s_k = \sum (x_i - \bar{x})^k$, $k = 2, 3$ *and* 4.

The function $LF(\beta)$ has a steeper slope near the origin than $LP(\beta)$. The relative ratio is $n/(n-1)$, which indicates that $LF(n\beta/(n-1))$ is close to $LP(\beta)$ near the origin. The power expansion of $LA(\beta) = LF(n\beta/(n-1))$ is expressed by

$$LA(\beta) \approx \sum_{i=1}^{r} x_i - \left\{ r\bar{x} + \frac{r(n-r)}{n(n-1)} s_2 \beta + \left(\frac{n}{n-1} \right)^2 c_3 \beta^2 + \left(\frac{n}{n-1} \right)^3 c_4 \beta^3 \right\}, \tag{23}$$

where c_3 and c_4 are coefficients of order 2 and 3 of $LF(\beta)$. Although $LA(\beta)$ is defined to adjust the coefficient of $LF(\beta)$ of order 1 to that of $LP(\beta)$, the coefficient of order 2 of $LA(\beta)$ becomes closer to that of $LP(\beta)$ than that of $LF(\beta)$. The following approximations are finally obtained.

$$LP(\beta) \approx LA(\beta), \tag{24}$$

$$\hat{\beta}_{PL} \approx \frac{(n-1)\hat{\beta}_{GML}}{n}. \tag{25}$$

The proposed approximated estimator and test statistic are quite helpful in cases of multiple 2×2 table when the value of both n and r are large (Yanagimoto and Kamakura, 1984).

4 Multiple Failures and Counting Processes

The standard methods of survival analysis can be generalized to include multiple failures simply defined as a series of well-defined event occurrences. For example, in software reliability, engineers are often interested in detecting software bugs. Inference from a single counting process has been studied in detail (Cox and Lewis, 1966; Musa et al., 1987), with multiple independent processes being considered as a means to estimate a common cumulative mean function from a nonparametric or semi-parametric viewpoint (Lawless and Nadeau, 1993; Nelson, 1992). Kamakura (1996) discussed problems associated with parametric conditional inference in models with a common trend parameter or possibly different base-line intensity parameters.

4.1 Intensity function

For multiple failures, intensity functions correspond to hazard functions in that the intensity function is defined as discussed next.

In time interval $[t_0, t]$ we define the number of occurrences of events or failures as $N(t)$. The Poisson counting process $\{N(t) : t \geq t_0\}$ is given such that it satisfies the following three conditions for $t \geq t_0$.

1. $\Pr\{N(t_0) = 0\} = 1$
2. The increment $N_{s,t} = N(t) - N(s)$ $(t_0 \geq s < t)$has a Poisson distribution with the mean parameter $\Lambda_t - \Lambda_s$, for some positive and increasing function in t.
3. $\{N_t : t \geq t_0\}$ is a process of independent increments. That is, for any $(t_0 <)t_1 < t_2 < \cdots < t_n$, n increments,$N(t_1) - N(t_0), \ldots, N(t_n) - N(t_{n-1})$ are mutually independent.

For this counting process $\{N(t) : t \geq t_0\}$ we can define the intensity function as

$$\lambda(t) = \lim_{\Delta \to 0} \frac{1}{\Delta} \Pr\{N(t + \Delta t) - N(t) = 1 | H(t)\}, \tag{26}$$

where $H(t)$is the history of the process up to t:

$$H(t) = \{N(u) : \; t_0 \leq u \leq t\}.$$

Note that

$$\Lambda(t) = \int_{t_0}^{t} \lambda(t) dt.$$

Expectation of $E[N_{s,t}]$becomes

$$E[N_{s,t}] = \sum_{n=0}^{\infty} n \Pr\{N_{n,s} = n\} = \Lambda_t - \Lambda_s, \tag{27}$$

and

$$\lambda(t) = \frac{d}{dt} \Lambda_t = \frac{d}{dt} E[N(t)]. \tag{28}$$

The nonparametric estimate of the intensity function is easy to determine and is quite useful for observing the trend of a series of events. If a data set of failure times $\{t_1, t_2, \ldots, t_n\}$ is available, assuming constant intensity in $(t_{k_1}, t_k]$, then

$$\lambda(t) = \lambda_k \quad (t_{k-1} < t \leq t_k),$$

and the nonparametric ML estimates becomes

$$\lambda_k = \frac{1}{t_k - t_{k-1}} \quad (k = 1, \ldots, n), \tag{29}$$

where $t_0 = 0$.

4.2 Multiple Counting Processes

We assume several independent counting processes $\{N_k(t_k), i.e., 0 < t_k \leq \tau_k, k = 1, \ldots, K\}$. The cumulative mean function for $N_k(t)$ is expressed by

$$M_k(t) = E\left\{N_k(t)\right\}. \tag{30}$$

Nelson (1992) described a method for estimating the cumulative mean function of an identically distributed process without assuming any Poisson process structure, while Lawless and Nadeau (1993) developed robust variance estimates based on the Poisson process. All these methods are basically concerned with nonparametric estimation. Here, parametric models for effectively acquiring information on the trend of an event occurrence are dealt with. Kamakura (1996) considered generalized versions of two primal parametric models to multiple independent counting processes under the framework of a nonhomogeneous Poisson process.

Cox and Lewis (Cox and Lewis, 1966) considered a log-linear model for trend testing a singe counting process, i.e.,

$$\lambda(t) = \exp(\alpha + \beta t), \tag{31}$$

where $\lambda(t)$ is the intensity function corresponding to the derivative of the mean function in the continuous case. Note that for a single case the subscript k is omitted. They assumed the above nonhomogeneous Poisson process and gave a simple test statistic for $H_0 : \beta = 0$ against $H_A : \beta \neq 0, i.e.,$

$$U = \frac{\sum_{i=1}^{n} t_i - \frac{1}{2}\tau_0}{\tau_0\sqrt{\frac{n}{12}}}. \tag{32}$$

The distribution of this statistic steeply converges to the standard normal distribution when $n \to \infty$. This statistic is sometimes called the U statistic and is frequently applied to trend testing in reliability engineering.

Kamakura (1996) generalized this log-linear model to the multiple case, with the log-linear model for k-th individual being

$$\lambda_k(t) = \exp\left(\alpha_k + \beta t\right). \tag{33}$$

In this modeling we assume the common trend parameter β and are mainly interested in estimating and testing this parameter. The full likelihood for the model becomes

$$\begin{aligned} L(\beta, \alpha_1, \alpha_2, \ldots, \alpha_K) &= \prod_{k=1}^{K}\left[\left\{\prod_{i=1}^{n_k}\lambda_k(t_{ki})\right\}\exp\left\{-\int_0^{\tau_k}\lambda_k(u)du\right\}\right] \\ &= \exp\left\{\sum_{k=1}^{K} n_k\alpha_k + \beta\sum_{k=1}^{K}\sum_{i=1}^{n_k} t_{ki} - \frac{1}{\beta}\sum_{k=1}^{K} e^{\alpha_k}\left(e^{\beta\tau_k} - 1\right)\right\}. \end{aligned} \tag{34}$$

If K is large, it is difficult to compute all parameter estimates based on such full likelihood.

Given $N_k(\tau_k) = n_k, k = 1, 2, \ldots, K$, conditional likelihood is considered as

$$CL(\beta | N_k(\tau_k) = n_k, i = 1, \ldots, K) = \frac{\prod_{k=1}^{K} (n_k!) \beta^{\sum n_k} e^{\beta \sum \sum t_{ki}}}{\prod_{k=1}^{K} \left(e^{\beta \tau_k} - 1\right)^{n_k}}. \tag{35}$$

Note that the nuisance parameter α_k's do not appear. Fisher information is calculated as

$$\begin{aligned} I(\beta) &= E\left[-\frac{\partial^2 \log CL}{\partial \beta^2}\right] \\ &= \begin{cases} \sum_{k=1}^{K} n_k \left\{ \frac{1}{\beta^2} - \frac{\tau_k^2 e^{-\beta\tau_k}}{\left(1-e^{-\beta\tau_k}\right)^2} \right\} & (\beta \neq 0) \\ \frac{1}{12} \sum_{k=1}^{K} n_k \tau_k^2 & (\beta = 0) \end{cases}. \end{aligned} \tag{36}$$

The test statistic obtained from the above calculations becomes

$$\begin{aligned} U_k &= \frac{\log CL|_{\beta=0}}{\sqrt{I(0)}} \\ &= \frac{\sum_{k=1}^{K} \sum_{i=1}^{n_k} t_{ki} - \frac{1}{2} \sum_{k=1}^{K} n_k \tau_k}{\sqrt{\frac{1}{12} \sum_{k=1}^{K} n_k \tau_k^2}}. \end{aligned} \tag{37}$$

To obtain the conditional estimate, numerical calculations are required such as Newton-Raphson method. However, the log conditional likelihood and its derivatives are not computable at the origin of the parameter β. In such a case, Taylor series expansions of the log conditional likelihood are used around the origin (Kamakura, 1996).

4.3 Power Law Model

Crow (1982) considered the power law model, sometimes called the Weibull process model. This model was generalized to the multiple case using the following intensity for the k-th individual (Kamakura, 1996):

$$\lambda_k(t) = \theta_k m t^{m-1}. \tag{38}$$

In this case it is easy to calculate the MLE. Direct calculation of the likelihood gives rise to the MLE $\hat{m}$ and $\hat{\theta_k}$ i.e.,

$$\hat{m} = \frac{\sum_{k=1}^{K} n_k}{\sum_{k=1}^{K} \sum_{i=1}^{n_k} \log\left(\frac{\tau_k}{t_{ki}}\right)}, \tag{39}$$

$$\hat{\theta_k} = \frac{n_k}{\tau_k^{\hat{m}}}. \tag{40}$$

Putting

$$Z = \frac{2m \sum_{k=1}^{K} n_k}{\hat{m}}, \tag{41}$$

the distribution of Z becomes a chi-square with $2 \sum_{k=1}^{K} n_k$ degrees of freedom. Based on this result we can make an inference of the common parameter m.

4.4 Models Suitable for Conditional Estimation

Estimation based on conditional likelihood allows effectively eliminating the nuisance parameter and obtaining information on the structure parameter. Let us now consider the class of nonhomogeneous Poisson process models which are specified by the intensity parameterized by two parameters. The first parameter α is concerned with the base line occurrences for the individual, while the second parameter β is concerned with the trend of intensity. For simplicity, the property of the intensity for $K = 1$ is examined. Using conditional likelihood is convenient because the nuisance parameter α need not be known. This is of great importance in multiple intensity modeling, i.e.,

Theorem 4. *(Kamakura, 1996) Conditional likelihood does not include the nuisance parameter α iff the intensity is factorized as two factors, a function of α and a function of β and the time t, in the class of nonhomogeneous Poisson process models. That is, the intensity is expressed as*

$$\lambda(t; \alpha, \beta) = h(\alpha) g(\beta; t). \quad a.s. \tag{42}$$

Several intensity models for software reliability are described in Musa et al. (1987): the log-linear model, geometric model, inverse linear model, inverse polynomial model, and power law model, all of which are included in this class satisfying the condition of the theorem.

Acknowledgements

This work has been partially supported financially by Chuo University as one of the 2003 Research Projects for Promotion of Advanced Research at Graduate School.

References

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics*, 30:89–99.

Castillo, E. and Hadi, A. S. (1995). Modeling lifetime data with application to fatigue models. *Journal of American Statistical Association*, 90:1041–1054.

Cohen, A. C. (1965). Maximum likelihood estimation in the Weibull distribution based on complete and censored samples. *Technometrics*, 5:579–588.

Cohen, A. C. and Whitten, B. J. (1988). *Parameter estimation in reliability and life span models*. Marcel Dekker, New York.

Cox, D. R. (1970). *The analysis of binary data*. Methuen, London.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of Royal Statistical Society:B*, 34:187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62:269–276.

Cox, D. R. and Lewis, P. A. W. (1966). *The statistical analysis of series of events*. Methuen, London.

Crow, L. H. (1982). Confidence interval procedures for the Weibull process with applications to reliability growth. *Technometrics*, 24:67–72.

Dubey, S. D. (1967). Normal and Weibull distributions. *Naval Research Logistics Quarterly*, 14:69–79.

Farewell, V. T. (1979). Some results on the estimation of logistic models based on retrospective data. *Biometrika*, 66:27–32.

Farewell, V. T. and Prentice, R. L. (1980). The approximation of partial likelihood with emphasis on case-control studies. *Biometrika*, 67:273–278.

Greenwood, M. (1926). *A report on the natural duration of cancer: Appendix I The 'Errors of Sampling' of the survivorshi tables: Reports on Public Health and Medical Subjects*. His Majesty's Stationery Office, London.

Johnson, N. L., Kotz, S., and Balakrishna, N. (1994). *Continuous Univariate Distributions*. John Wiley, New York.

Kalbfleish, J. D. and Prentice, R. L. (1980). *The statistical analysis of failure time data*. John Wiley, New York.

Kamakura, T. (1996). Trend analysis of multiple counting processes. In Jewell, N. P., Kimber, A. C., Lee, M.-L. T., and Whitmore, G. A., editors, *Lifetime Data: Models in Reliability and Survival Analysis*, pages 149–156. Kluwer Academic Publishers, Dordrecht.

Kamakura, T. and Yanagimoto, T. (1983). Evaluation of the regression parameter estimators in the proportional hazard model. *Biometrika*, 70:530–533.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of American Statistical Association*, 53:457–481.

Korn, E. L. and Whittemoore, A. (1979). Methods for analysing panel studies of acute health effects of air pollution. *Biometrics*, 35:795–802.

Lawless, J. F. and Nadeau, J. C. (1993). Some simple robust methods for the analysis of recurrent events. University of Waterloo IIQP Research Report.

Menon, M. (1963). Estimation of the shape and scale parameters of the Weibull distributions. *Technometrics*, 5:175–182.

Miller, R. G. (1981). *Survival analysis*. John Wiley, New York.

Musa, J. D., Iannino, A., and Okumoto, K. (1987). *Software Reliablility*. McGraw-Hill, New York.

Nagatsuka, H. and Kamakura, T. (2003). A new method of inference for Weibull shape parameter (in Japanese). *Journal of Reliability Engineering Association of Japan*, 25:583–593.

Nagatsuka, H. and Kamakura, T. (2004). Parameter estimation of the shape parameter of the Castillo-Hadi model. *Communications in Statistics: Theory and Methods*, 33:15–27.

Nelson, W. B. (1992). Confidence limits for recurrence data - applied to cost or number of product repairs and of disease episodes. *Technometrics*, 22:1023–1031.

Oaks, D. (2001). Biometrika centenary: Survival analysis. *Biometrika*, 88:99–142.

Peto, R. (1972). Discussion of paper by D. R. Cox. *Journal of Royal Statistical Society:B*, 34:205–207.

Pregibon, D. (1982). Resistant fits for some commonly used logistic models with medical applications. *Biometrics*, 38:485–498.

Prentice, R. L. and Breslow, N. E. (1978). Restrospective studies and failure time models. *Biometrika*, 65:153–158.

Yanagimoto, T. and Kamakura, T. (1984). The maximum full and partial likelihood estimators in the proportional hazard model. *Annals of the Institute of Statistical Mathematics*, 36:363–373.

Index

2×2 table, 12

approximations, 12
asymptotic relative efficiency, 4

base-line intensity, 12
Breslow-Peto method, 8, 9

Castillo-Hadi model, 6
coefficient of variation, 3, 4
common cumulative mean function, 12
common parameter, 16
common trend parameter, 12
conditional likelihood, 10, 15, 16
counting process, 12, 14
covariate, 8
Cox model, 8, 9
Cox's discrete model, 10
cumulative mean function, 14

digamma function, 4
discrete logistic model, 8

efficiency of the sample mean, 5
Euler's constant, 4
exponential density function, 3
exponential distribution, 3

Fisher information, 15
full likelihood, 10

gamma distribution, 3
generalized maximum likelihood method, 8
Greenwood's formula, 3

hazard function, 1, 13
hazard rate, 1

independent increments, 13
intensity functions, 13

Kaplan-Meier method, 2

location-scale-free transformation, 5
log-linear model, 14
log-logistic distribution, 3
log-normal distribution, 3

marginal distribution function, 6
maximum full likelihood, 10
maximum likelihood estimate, 3
maximum partial likelihood, 10
method of moments, 4
multiple counting processes, 14
multiple failures, 12, 13

Newton-Raphson method, 15
nonhomogeneous Poisson process, 14, 16
nuisance parameter, 15, 16

partial likelihood, 8, 10
Poisson distribution, 13
power expansions, 11
power parameter, 5
proportional hazards model, 10

sample mean, 4

shape parameter, 3
software reliability, 12
structure parameter, 16
survival function, 1

Taylor series expansions, 15
threshold parameters, 3

W-transformation, 6
Weibull density function, 3
Weibull distribution, 3
Weibull process model, 15