LARGE SAMPLE EFFICIENCIES OF THE
PRODUCT LIMIT SURVIVAL ESTIMATOR

Jerome Klotz and James Fu
LARGE SAMPLE EFFICIENCIES
OF THE
PRODUCT LIMIT SURVIVAL ESTIMATOR

Jerome Klotz
James Fu

Biostatistics Center
University of Wisconsin-Madison

Biostatistics Technical Report No. 37
March 1986
LARGE SAMPLE EFFICIENCIES OF THE PRODUCT LIMIT SURVIVAL ESTIMATOR

BY JEROME KLOTZ 1 AND JAMES FU 2

University of Wisconsin at Madison and University of Manitoba at Winnipeg

Large sample efficiencies are calculated for the Kaplan-Meier product limit estimator when compared to the parametric estimator for exponential distributions. Two types of efficiencies are used for comparison. Local efficiency is calculated as the ratio of the asymptotic variances of each estimator. Large deviation efficiency is computed as the ratio of the exponential rates of convergence to zero for the probabilities that the estimators lie outside an interval about the true survival distribution. In both cases poor performance of the Kaplan-Meier estimator is found for uncensored as well as censored data. An interesting discontinuity is noted for the large deviation efficiency with censoring.

1. Introduction and notation. Let \( X_i \) be independent with cumulative distribution function (c.d.f.) \( F(t), 0 \leq t < \infty, \) for \( i = 1, 2, \ldots, n. \) Let \( Y_i \) be independent with c.d.f. \( G(t), 0 \leq t < \infty, \) for \( i = 1, 2, \ldots, n. \) We are interested in estimating the survival distribution \( 1-F(t) \) on the basis of the data \( (Z_i, D_i) i = 1, 2, \ldots, n \) where \( Z_i = \min (X_i, Y_i) \) and \( D_i = 1 \) if \( X_i \leq Y_i, 0 \) otherwise. We call the nuisance distribution \( G(t) \) the censoring distribution.

To estimate the survival distribution, Kaplan and Meier (1958) proposed the product limit estimator

\[
1 - F_n^0(t) = \prod_{i: Z_i < t} \left( \frac{n - R_i}{n - R_i + 1} \right)^{D_i} \tag{1.1}
\]

where we define the empty product to be 1 and \( R_i \) is the rank of the pair \( (Z_i, 1 - D_i) \) in the increasing lexical ordering of \( (Z_1, 1 - D_1), \ldots, (Z_n, 1 - D_n). \) That is we use the \( Z_i \)'s to rank unless there are tied values and then uncensored observations \( (D_i = 1) \) precede censored observations \( (D_i = 0). \) This is a nonparametric estimator and reduces to the usual sample c.d.f. \( 1 - F_n(t) \) when all observations are uncensored \( (D_i = 1). \)

Among parametric estimators, a popular one is based on the exponential distribution

\[
F(t) = 1 - e^{-\lambda t}, \quad \text{for } t \geq 0, \quad 0 \text{ otherwise.} \tag{1.2}
\]

The maximum likelihood estimator (m.l.e.) of \( \lambda \) is derived using the joint density of \( (Z, D) \) with respect to the product of Lesbesgue and counting measure

\[
f_{Z,D}(z, d) = [(1 - G(z))f(z)]^d [(1 - F(z))g(z)]^{1-d} \tag{1.3}
\]

where \( f \) is the density of \( F, \) \( g \) the density of \( G. \) Maximizing the log likelihood

\[
\sum_{i=1}^{n} \ln f_{Z,D}(z_i, d_i) \tag{1.4}
\]

the m.l.e. for \( \lambda \) is

\[
\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} D_i} \tag{1.5}
\]

and the m.l.e. for \( 1 - F(t) \) is

\[
1 - F(t) = e^{-\hat{\lambda} t}. \tag{1.6}
\]

1 Research supported by National Institute of Health Grant R01-CA18332
2 Research supported by Natural Sciences and Engineering Research Council of Canada Grant NSERC A-9216
Key words and phrases. Survival estimation, product limit estimator, Kaplan-Meier estimator, local efficiency, large deviation efficiency, censoring, exponential distributions, maximum likelihood estimator.
2. Efficiencies in the case of no censoring. We consider this simpler case to illustrate the two efficiency approaches. For the product limit estimator, with no censoring, (1.1) reduces to the sample c.d.f.

$$1 - F_n^0(t) = 1 - F_n(t) = \sum_{i=1}^{n} [Z_i > t]/n$$  

(2.1)

and $n (1 - F_n^0(t))$ has a binomial $(n, 1 - F(t))$ distribution. Thus the variance for $\sqrt{n} (1 - F_n(t))$ is

$$F(t)(1 - F(t)).$$  

(2.2)

For the parametric estimator with no censoring, (1.5) reduces to $\bar{Z} = \sum_{i=1}^{n} Z_i/n$ and $Var (\sqrt{n} (\hat{\theta} - \theta)) = \theta^2$.

Using the central limit theorem for $\hat{\theta}$ and then transforming, we obtain

$$\sigma^2 = \lim_{n \to \infty} Var (\sqrt{n} (e^{-t/\theta} - e^{-t/\theta})) = ((t/\theta)e^{-t/\theta})^2.$$  

(2.3)

Taking the ratio of (2.3) and (2.2) gives the local efficiency

$$\epsilon_{PL,mle} = \frac{((t/\theta)e^{-t/\theta})^2}{(1 - e^{-t/\theta})e^{-t/\theta}} = \frac{T^2e^{-T}}{1 - e^{-T}}$$  

(2.4)

where $T = t/\theta$.

<table>
<thead>
<tr>
<th>T</th>
<th>0</th>
<th>0.5</th>
<th>1.0</th>
<th>1.5</th>
<th>2.0</th>
<th>2.5</th>
<th>3.0</th>
<th>3.5</th>
<th>4.0</th>
<th>\infty</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon_{PL,mle}$</td>
<td>0</td>
<td>0.38537</td>
<td>0.58198</td>
<td>0.64624</td>
<td>0.62607</td>
<td>0.55891</td>
<td>0.47156</td>
<td>0.38144</td>
<td>0.29852</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1. Local efficiency of the product limit estimator without censoring.

To define a large deviation efficiency we use the result that for any $\varepsilon > 0$ as $n \to \infty$

$$P_n = P [\{ (1 - F_n(t)) - (1 - F(t)) \} \geq \varepsilon] \to 0$$  

(2.5)

since both estimators are consistent. In fact, the probabilities converge to zero at an exponential rate. For the product limit estimator using $n$ observations, the probability $\sim \exp(-nK_{pl}(1+o(1)))$ where the Kullback-Liebler information constant $K_{pl} > 0$. Similarly, for the m.l.e. using $m$ observations, we have the probability $\sim \exp(-mK_{mle}(1+o(1)))$ where $K_{mle} > 0$. We then define the large deviation efficiency for fixed $t$, $\varepsilon$ by equating the product limit probability based on $n$ observations and the m.l.e probability based on $m$ observations. Solving, we obtain the large deviation efficiency as the limiting sample size ratio

$$\epsilon_{PL,mle} = \lim_{n \to \infty} \frac{m/n} = K_{pl}/K_{mle}.$$  

(2.6)

To calculate $K_{pl} = \lim_{n \to \infty} -n^{-1} \ln(P_n)$, where $P_n$ is given by (2.5) using estimator (1.1), write

$$P_n = P[F_n(t) \geq F(t) + \varepsilon] + P[F_n(t) \leq F(t) - \varepsilon] = P_{n+} + P_{n-}.$$  

(2.7)

From $\max(P_{n+}, P_{n-}) \leq P_n \leq 2\max(P_{n+}, P_{n-})$ we obtain $K_{pl} = \min(K_{+}, K_{-})$ where

$$-n^{-1} \ln(P_{n+}) \to K_+ = K(F(t) + \varepsilon, F(t))$$

$$-n^{-1} \ln(P_{n-}) \to K_- = K(F(t) - \varepsilon, F(t))$$  

(2.8)

The function $K$ is defined by

$$K(p, r) = p \ln(p/r) + (1 - p) \ln((1 - p)/(1 - r))$$

Since

$$P_{n+} = \sum_{k \geq n(F(t) + \varepsilon)} \left( \begin{array}{c} n \\ k \end{array} \right) [F(t)]^k (1 - F(t))^{n-k}$$
is the probability in the upper tail of the binomial, the large deviation probability limit (2.8) for \( K_+ \) can be derived using Stirling’s formula and a geometric series bound. Details are given by Feller (1957) (vol. 1, page 140) and the result is found in Bahadur (1960), (1971). A similar argument for \( K_- \) follows using the lower tail.

Since the function \( K(p,r) \) is only defined for \( 0 \leq p, r \leq 1 \), we have several cases to consider when computing \( K_{pl} \). If \( F(t) + \varepsilon \leq 1 \), then \( K_+ = K(F(t) + \varepsilon, F(t)) \) is well defined. Otherwise, \( P_{pl} = 0 \) and \( K_+ = \infty \). Summarizing similar cases we have

\[
K_{pl} = \begin{cases} 
\min(K(F(t) + \varepsilon, F(t)), K(F(t) - \varepsilon, F(t))), & \text{for } 0 \leq \varepsilon \leq \min(F(t), (1-F(t))) \\
K(F(t) + \varepsilon, F(t)), & \text{for } F(t) < \varepsilon \leq 1 - F(t) \\
K(F(t) - \varepsilon, F(t)), & \text{for } 1 - F(t) < \varepsilon \leq F(t).
\end{cases}
\quad (2.9)
\]

For the exponential distribution (1.2), writing \( T = t/\theta, E = e^{-t/\theta} \), and

\[
K^*(T,E) = (1-e^{-T}(1-E)) \ln((1-e^{-T}(1-E))/(1-e^{-T})) + e^{-T}(1-E) \ln(1-E),
\]

equation (2.9) becomes

\[
K_{pl} = \begin{cases} 
\min(K^*(T,E), K^*(T,-E)), & \text{for } 0 \leq E \leq \min(1, e^{T-1}) \\
K^*(T,E), & \text{for } e^{T-1} < E \leq 1 \\
K^*(T,-E), & \text{for } 1 < E \leq e^{T-1}.
\end{cases}
\quad (2.10)
\]

For calculating \( K_{mle} = \lim_{m \to \infty} -m^{-1} \ln(P_m) \), where \( P_m \) is given by (2.5) now using estimator (1.6) we similarly consider

\[
P_m = P[\hat{F}(t) \geq F(t) + \varepsilon] + P[\hat{F}(t) \leq F(t) - \varepsilon] = P_{m+} + P_{m-}.
\quad (2.11)
\]

We can write

\[
P_{m+} = P[1-e^{-t/\hat{\theta}} \geq 1-e^{-t/\theta} + \varepsilon] = P[\hat{\theta} \leq \gamma(t, \varepsilon)]
\quad (2.12)
\]

where \( \gamma(t, \varepsilon) = (t/\theta)/(t/\theta - \ln(1-e^{-t/\theta})) \). For this no censoring exponential distribution case,

\[
P_{m+} = P[\sum_{i=1}^{m} \gamma_i \leq m \gamma(t, \varepsilon)] = \sum_{k=0}^{\infty} e^{-\lambda} \lambda^k / k!
\quad (2.13)
\]

where \( \lambda = m \gamma(t, \varepsilon) \). Then direct calculations as \( m \to \infty \) give

\[
-m^{-1} \ln(P_{m+}) \to \gamma(t, \varepsilon) - 1 - \ln(\gamma(t, \varepsilon)).
\quad (2.14)
\]

Similarly

\[
-m^{-1} \ln(P_{m-}) \to \gamma(t, -\varepsilon) - 1 - \ln(\gamma(t, -\varepsilon)).
\]

In terms of \( T = t/\theta, E = e^{-t/\theta} \), and

\[
\gamma^*(T,E) = \gamma(t, \varepsilon) = T/(T - \ln(1-E)),
\]

we summarize to obtain

\[
K_{mle} = \begin{cases} 
\min(h(\gamma^*(T,E)), h(\gamma^*(T,-E))), & \text{for } 0 \leq E \leq \min(1, e^{T-1}) \\
h(\gamma^*(T,E)), & \text{for } e^{T-1} < E \leq 1 \\
h(\gamma^*(T,-E)), & \text{for } 1 < E \leq e^{T-1},
\end{cases}
\quad (2.15)
\]

where \( h(\gamma) = \gamma - 1 - \ln(\gamma) \). Substituting (2.10) and (2.15) in (2.6) gives large deviation efficiency values in table 2.

Both results (2.15) for the uncensored case above, and (5.7) for the censored case considered next, can also be obtained by using the general m.l.e. result of Fu (1973).
We note in this case of no censoring, that the large deviation efficiency approaches the local efficiency as \(\varepsilon \to 0\) (compare table 2 for \(E = 0.1, 0.01, 0.001\) with table 1). This can be shown using l'Hospital's rule.

3. An alternative approach using Sanov's theorem. Sanov (1957) showed under certain regularity conditions for a class of distribution functions \(\Omega\), the sample c.d.f. \(F_n\) satisfies

\[
P \left( F_n \in \Omega \right) = \exp \left\{ -n \, K \left( \Omega \right) + o \left( n \right) \right\}
\]

(3.1)

where

\[
K \left( \Omega \right) = \inf_{F^* \in \Omega} \int \ln \left( \frac{dF^*}{dF} \right) dF^*.
\]

For the product limit estimator, to evaluate \(K_{pl}\) from (2.7), we have

\[
\Omega = \left\{ F^* : |F^*(t) - F(t)| \geq \varepsilon \right\} = \Omega_+ \cup \Omega_-
\]

where

\[
\Omega_+ = \left\{ F^* : F^*(t) \geq F(t) + \varepsilon \right\} \quad \text{and} \quad \Omega_- = \left\{ F^* : F^*(t) \leq F(t) - \varepsilon \right\}.
\]

With this decomposition, \(K \left( \Omega \right) = \min \{K \left( \Omega_+ \right), K \left( \Omega_- \right)\}\). To evaluate

\[
K \left( \Omega_+ \right) = \inf_{F^* \in \Omega_+} \int_0^\infty F^* \left( x \right) \ln \left( \frac{F^* \left( x \right)}{f \left( x \right)} \right) dx
\]

where \(f = dF/dx\), etc., we break the integral on \([0, \infty]\) at \(x = t\) and minimize on \([0, t]\) and \([t, \infty]\) subject to \(F^* \left( t \right) = F \left( t \right) + \varepsilon\). Now

\[
\int_0^t f^* \left( x \right) \ln \left( \frac{f^* \left( x \right)}{f \left( x \right)} \right) dx = -\left( F \left( t \right) + \varepsilon \right) \int_0^t \frac{f^* \left( x \right)}{F \left( t \right) + \varepsilon} \ln \left( \frac{f \left( x \right) \left( F \left( t \right) + \varepsilon \right)}{f^* \left( x \right) \left( F \left( t \right) + \varepsilon \right)} \right) dx
\]

\[
\geq -\left( F \left( t \right) + \varepsilon \right) \ln \left( \int_0^t f \left( x \right) \left( F \left( t \right) + \varepsilon \right) dx \right) = (F \left( t \right) + \varepsilon) \ln \left( (F \left( t \right) + \varepsilon) / F \left( t \right) \right)
\]

(3.2)

using Jensen's inequality for the convex negative log function applied to the distribution \(f^* \left( x \right) / (F \left( t \right) + \varepsilon)\) on \([0, t]\). Equality holds in (3.2) for \(f^* \left( x \right) / f \left( x \right)\) constant and the minimum is then attained. Similarly

\[
\inf_{F^* \in \Omega_-} \int_0^t f^* \left( x \right) \ln \left( f^* \left( x \right) / f \left( x \right) \right) dx = \left( 1 - F \left( t \right) - \varepsilon \right) \ln \left( (1 - F \left( t \right) - \varepsilon) / (1 - F \left( t \right)) \right).
\]

(3.3)

Combining (3.2) and (3.3) gives

\[
K \left( \Omega_+ \right) = (F \left( t \right) + \varepsilon) \ln \left( (F \left( t \right) + \varepsilon) / F \left( t \right) \right) + (1 - F \left( t \right) - \varepsilon) \ln \left( (1 - F \left( t \right) - \varepsilon) / (1 - F \left( t \right)) \right) = K_+.
\]

(3.4)

Similarly \(K \left( \Omega_- \right) = K_-\) and \(K_{pl} = \min \{K_+, K_-\}\) as in (2.8).
For the maximum likelihood estimator, we can similarly evaluate $K_{mle}$ from (3.1) where now

$$\Omega_\gamma = \{F^*: \int_0^\infty x \, dF^*(x) \leq \theta \gamma\}$$

and $\gamma = \gamma(t, \theta, \xi)$ is defined following (2.12). To evaluate $K(\Omega_\gamma)$ for the m.l.e., we solve the calculus of variations problem with a side condition by minimizing

$$\int_0^\infty f^*(x) \ln \left( \frac{f^*(x)}{f(x)} \right) \, dx + \lambda \left| \int_0^\infty x \, dF^*(x) - \theta \gamma \right|$$

with respect to $f^*$. Here $f(x) = \theta^{-1} \exp(-x/\theta)$ and the Lagrange multiplier $\lambda$ is chosen so the inequality in $\Omega_\gamma$ is satisfied with equality. Rewriting (3.5) and using Jensen's inequality we have

$$= -\int_0^\infty f^*(x) \ln \left( \frac{f^*(x)}{f(x)} e^{-\lambda x} \right) \, dx - \lambda \theta \gamma \geq -\ln\left( \int_0^\infty f(x) e^{-\lambda x} \, dx \right) - \lambda \theta \gamma$$

with equality if $f(x) \exp(-\lambda x)/f^*(x)$ is constant. The minimizing $f^*(x) = (\theta \gamma)^{-1} \exp(-x/(\theta \gamma))$ and the resulting minimum of the integral gives (2.14). Similar computations for $\Omega_-$ combine to verify (2.15).

4. Local efficiency for the case of censoring. In this case we must specify the censoring distribution $G(t)$ to calculate efficiencies. For simplicity, we again use the exponential distribution

$$G(t) = 1 - e^{-t/\xi}, \quad \text{for } 0 \leq t < \infty, \quad \xi > 0.$$

For the m.l.e. $\hat{\theta} = \bar{Z}/\bar{D}$ we can show

$$\sqrt{m} \left( \hat{\theta} - \theta \right) \overset{D}{\to} N \left( 0, \sigma^2_{mle} \right)$$

as the sample size $m \to \infty$ where $\rho = \theta/\xi$. Consequently

$$\sqrt{m} \left( e^{-t/\bar{D}} - e^{-t/\theta} \right) \overset{D}{\to} N \left( 0, \sigma^2_{mle} \right), \quad \text{where} \quad \sigma^2_{mle} = \frac{\theta^{-1} e^{-\theta t}}{1 + \rho} \left( 1 + \rho \right).$$

For the product limit estimator, using the result of Breslow and Crowley (1974), we have

$$\sqrt{m} \left( (1 - F_n^0(t)) - (1 - F(t)) \right) \overset{D}{\to} N \left( 0, \sigma^2_{pl} \right), \quad \text{where} \quad \sigma^2_{pl} = e^{-t/\bar{D}} \left( 1 - e^{-t/\theta} \right) \left( 1 + \rho \right).$$

The local efficiency with censoring is then

$$e_{pl,mle} = \frac{\sigma^2_{mle}}{\sigma^2_{pl}} = \frac{(t(1 + \rho)/\theta)^2 e^{-t/(1 + \rho) \theta}}{(1 - e^{-t/(1 + \rho) \theta})}$$

as calculated in table 3 with arguments $t = t/\theta$, and $\rho$.

<table>
<thead>
<tr>
<th>$e_{pl,mle}$</th>
<th>$T$</th>
<th>0.5</th>
<th>1.0</th>
<th>1.5</th>
<th>2.0</th>
<th>2.5</th>
<th>3.0</th>
<th>3.5</th>
<th>4.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$ = 0</td>
<td>.3854</td>
<td>.5820</td>
<td>.6462</td>
<td>.6261</td>
<td>.5589</td>
<td>.4716</td>
<td>.3814</td>
<td>.2985</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>.5036</td>
<td>.6462</td>
<td>.5964</td>
<td>.4716</td>
<td>.3387</td>
<td>.2275</td>
<td>.1454</td>
<td>.0895</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.5820</td>
<td>.6261</td>
<td>.4716</td>
<td>.2985</td>
<td>.1696</td>
<td>.0895</td>
<td>.0447</td>
<td>.0215</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>.6274</td>
<td>.5589</td>
<td>.3387</td>
<td>.1696</td>
<td>.0756</td>
<td>.0311</td>
<td>.0121</td>
<td>.0045</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.6462</td>
<td>.4716</td>
<td>.2275</td>
<td>.0895</td>
<td>.0311</td>
<td>.0100</td>
<td>.0030</td>
<td>.0009</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>.6441</td>
<td>.3814</td>
<td>.1454</td>
<td>.0447</td>
<td>.0121</td>
<td>.0030</td>
<td>.0007</td>
<td>.0002</td>
<td></td>
</tr>
<tr>
<td>3.0</td>
<td>.6261</td>
<td>.2985</td>
<td>.0895</td>
<td>.0215</td>
<td>.0045</td>
<td>.0009</td>
<td>.0002</td>
<td>.0000</td>
<td></td>
</tr>
</tbody>
</table>

We note the entries for $\rho = 0$ correspond to the case of no censoring.
5. The large deviation exponent for the m.l.e. with censoring. For the m.l.e., with censoring, we write (2.12) as

\[ P_{m^+} = P [ \bar{Z} - \bar{D} \theta \gamma \leq 0] \]

Applying a bivariate version of Sanov's theorem, see Bahadur and Zabel (1979), we have as \( m \to \infty \)

\[ -m^{-1} \ln (P_{m^+}) \to K(\Omega_+) \]

where

\[
K(\Omega_+) = \inf_{\sum_{d=0}^{\infty} \int_{z \in \Omega_+} f_{z,d}^*(z,d) \ln \left( \frac{f_{z,d}^*(z,d)}{f_{z,d}(z,d)} \right) dz,}
\]

\[
\Omega_+ = \{ f_{z,d}^*(z,d) : \sum_{d=0}^{\infty} \int_{z \in \Omega_+} (z - d \theta \gamma) f_{z,d}(z,d) dz \leq 0 \},
\]

\( f_{z,d}(z,d) \) is given by (1.3) for \( F \) given by (1.2), \( G \) given by (4.1), and \( f_{z,d}^* \) is the density of \( F_{z,d}^* \). Minimizing, subject to a side condition, we form the Lagrange multiplier

\[
\lambda \sum_{d=0}^{\infty} \int_{z \in \Omega_+} f_{z,d}(z,d) \ln \left( \frac{f_{z,d}(z,d)}{f_{z,d}^*(z,d)} \right) dz + \lambda \int_{z \in \Omega_+} (z - d \theta \gamma) f_{z,d}(z,d) dz.
\]

Rewriting and using Jensen's inequality gives

\[ = - \sum_{d=0}^{\infty} \int_{z \in \Omega_+} f_{z,d}(z,d) \ln \left( \frac{f_{z,d}(z,d)}{f_{z,d}^*(z,d)} e^{-\lambda(z-d\theta\gamma)} \right) dz \geq - \ln \left( \sum_{d=0}^{\infty} \int_{z \in \Omega_+} f_{z,d}(z,d) e^{-\lambda(z-d\theta\gamma)} dz \right) \]

with equality if \( (f_{z,d}(z,d)/f_{z,d}^*(z,d)) \exp (-\lambda(z-d\theta\gamma)) \) is constant. Thus the minimizing density

\[ f_{z,d}^*(z,d) = C f_{z,d}(z,d) e^{-\lambda(z-d\theta\gamma)} \]

where the constant \( C \) is determined so the density integrates to 1 and \( \lambda \) is chosen to satisfy the side condition with equality. As a result

\[ C = \frac{(\theta^{-1} + \xi^{-1} + \lambda)}{(\theta^{-1} e^{\lambda \theta \gamma} + \xi^{-1})} \]

and \( \lambda \) is obtained from the side condition equation

\[ \frac{1}{\theta^{-1} + \xi^{-1} + \lambda} - \frac{\gamma e^{\lambda \theta \gamma}}{\theta^{-1} e^{\lambda \theta \gamma} + \xi^{-1}} = 0. \] (5.2)

If we write \( s = -\lambda \theta \gamma \) then equation (5.2) becomes

\[ (1 + \rho)\gamma(t,\varepsilon) = 1 + s + \rho e^s \] (5.3)

Summarizing, the value of the minimized integral (5.1) is \(-\ln(C^{-1})\) or

\[ K(\Omega_+) = s = \ln(\gamma(t,\varepsilon)), \]

where \( s \) is determined by iteratively solving equation (5.3) and \( \gamma \) is defined after (2.12). Similarly we have

\[ K(\Omega_-) = s = \ln(\gamma(t,-\varepsilon)) \] (5.5)

where now

\[ (1 + \rho)\gamma(t,-\varepsilon) = 1 + s + \rho e^s. \] (5.6)

Combining (5.4) and (5.5) we obtain

\[ K_{mle} = \min\{K(\Omega_+), K(\Omega_-)\}. \] (5.7)
6. The large deviation exponent for the product limit estimator with censoring. When the survival and censoring distributions $F$ and $G$ are continuous, as in the exponential case considered here, the random variables $Z_i$ are distinct with probability 1 and we can write (1.1) in the form

$$- \ln (1 - \hat{F}_n^0(t)) = \sum_{i: Z_i > t} \ln \left( \frac{n - R_i}{n - R_i + 1} \right)$$

$$= n \sum_{i: Z_i > t} \left[ 1 + \frac{1}{n(1 - R_i/n)} \right] \frac{D_i}{n} = n \int_0^t \ln \left[ 1 + \frac{1}{n(1 - H_n(x))} \right] d\hat{H}_n(x)$$

where

$$H_n(x) = \frac{1}{n} \sum_{i=1}^n I[Z_i \leq x], \quad \text{and} \quad \widehat{H}_n(x) = \frac{1}{n} \sum_{i=1}^n I[Z_i \leq x, D_i = 1].$$

Using the inequalities $u - u^2/2 \leq \ln(1 + u) \leq u$, for $u \geq 0$, we have

$$U_n - V_n = - \ln (1 - \hat{F}_n^0(t)) \leq U_n$$

where

$$U_n = \int_0^t \frac{d\hat{H}_n(x)}{1 - H_n(x)}, \quad V_n = \frac{1}{2n} \int_0^t \frac{d\hat{H}_n(x)}{(1 - H_n(x))^2}.$$

We first show the term $V_n$ is negligible compared to $U_n$ when calculating large deviation probabilities. Since $H_n(x) \leq H_n(t)$, for $x \leq t$, we have

$$V_n \leq \frac{1}{2n} \int_0^t \frac{d\hat{H}_n(x)}{(1 - H_n(t))^2} \leq (2n(1 - H_n(t))^2)^{-1}.$$

Thus for any $\delta > 0$

$$P[V_n \geq \delta] \leq P[(2n(1 - H_n(t))^2)^{-1} \geq \delta] = P[n(1 - H_n(t)) \leq a(n)]$$

where $a(n) = \lceil (n/(2\delta))^4 \rceil$ is the largest integer not exceeding $(n/(2\delta))^4$. Using the binomial $(n, p)$ distribution

$$= \sum_{k=a(n)}^{n} \binom{n}{k} p^k q^{n-k} \leq a(n) \left[ \frac{n}{a(n)} \right] p^{a(n)} q^{n-a(n)}$$

where $p = (1 - F(t))(1 - G(t))$, $q = 1 - p$. Direct calculations using Stirling's formula then yield the limit

$$\lim_{n \to \infty} -n^{-1} \ln(P[V_n \geq \delta]) \geq -\ln(q)$$

independent of $\delta$. Writing $P_n = P_{n^+} + P_{n-}$ as in (2.7) we have

$$P[U_n - V_n \geq -\ln(1 - F(t) - \epsilon)] \leq P_{n^+} \leq P[U_n \geq -\ln(1 - F(t) - \epsilon)].$$

Also

$$P[U_n - V_n \geq -\ln(1 - F(t) - \epsilon)] \geq P[U_n - \delta \geq -\ln(1 - F(t) - \epsilon), V_n < \delta]$$

$$= P[U_n - \delta \geq -\ln(1 - F(t) - \epsilon)] - P[U_n - \delta \geq -\ln(1 - F(t) - \epsilon), V_n \geq \delta]$$

$$\geq P[U_n - \delta \geq -\ln(1 - F(t) - \epsilon)] - P[V_n \geq \delta].$$

Thus letting $\delta \to 0$, $-n^{-1} \ln(P_{n^+})$ and $-n^{-1} \ln(P[U_n \geq -\ln(1 - F(t) - \epsilon)])$ will have the same limit provided the latter limit is smaller than $-\ln(q)$ from (6.4).

Applying Sanov's theorem in the bivariate case again gives

$$- \frac{1}{n} \ln(P[\int_0^t \frac{d\hat{H}_n(x)}{1 - H_n(x)} \geq -\ln(1 - F(t) - \epsilon)]) \to K(\Omega_n)$$
\[ \Omega_\epsilon = \{ F^*_\epsilon(z,d) : \int_0^t \frac{f^*_\epsilon(z,1)}{1 - F^*_\epsilon(z)} \, dz \geq -\ln (1 - F(t) - \epsilon) \}. \]  

(6.5)

To simplify the minimization problem define
\[ r^*(z) = \frac{f^*_\epsilon(z,1)}{1 - F^*_\epsilon(z)}, \quad s^*(z) = \frac{f^*_\epsilon(z,0)}{1 - F^*_\epsilon(z)}. \]

Since \( r^*(z) \) is non-negative, we can say it is the hazard for some c.d.f. \( F^*(z) \) and similarly \( s^*(z) \) is the hazard for a c.d.f \( G^*(z) \). Adding and integrating from 0 to \( t \) gives
\[
\int_0^t (r^*(z) + s^*(z)) \, dz = -\ln (1 - F^*(t)) - \ln (1 - G^*(t))
\]
\[
= \int_0^t \frac{f^*_\epsilon(z)}{1 - F^*_\epsilon(z)} \, dz = -\ln (1 - F^*_\epsilon(z)).
\]

It follows that \( 1 - F^*_\epsilon(z) = (1 - F^*(z))(1 - G^*(z)) \) and
\[
f^*_\epsilon(z,d) = [f^*(z)(1 - G^*(z))]^{\#} [g^*(z)(1 - F^*(z))]^{1-d}.
\]

(6.6)

We now can compute (5.1) where the densities are given by (1.3), (6.6) by minimizing with respect to \( F^*, G^* \) in the set
\[ \Omega_\epsilon^* = \{ (F^*, G^*): F^*(t) \geq F(t) + \epsilon \}. \]

(6.7)

We break the integral (5.1) over \([0, \infty)\) into two parts over \([0, t]\) and \([t, \infty)\). We minimize each separately subject to the constraint \( F^*(t) = \alpha = F(t) + \epsilon \). We also impose the constraint \( G^*(t) = \beta \) and then later minimize with respect to \( \beta \) for \( 0 \leq \beta \leq 1 \). The first part over \([0, t]\) is
\[
\sum_{d=0}^1 \int_0^t f^*_\epsilon(z,d) \frac{f^*_\epsilon(z,d)}{f^*_\epsilon(z,d)} \, dz = - (1 - (1-\alpha)(1-\beta)) \sum_{d=0}^1 \int_0^t \frac{f^*_\epsilon(z,d)}{1 - (1-\alpha)(1-\beta)} \frac{f^*_\epsilon(z,d)}{f^*_\epsilon(z,d)} \, dz
\]
and using Jensen’s inequality
\[
\geq - (1 - (1-\alpha)(1-\beta)) \ln \left( \sum_{d=0}^1 \int_0^t \frac{f^*_\epsilon(z,d)}{1 - (1-\alpha)(1-\beta)} \, dz \right)
\]
\[
= (1 - (1-\alpha)(1-\beta)) \ln \left( \frac{1 - (1-\alpha)(1-\beta)}{1 - (1-\alpha)(1-\beta)} \right).
\]

(6.8)

The inequality becomes an equality for the minimizing \( f^*_\epsilon(z,d) = C_1 f^*_\epsilon(z,d) \) where \( C_1 \) is a constant for this interval \([0, t]\). Similarly on the interval \([t, \infty)\) an analogous argument gives the minimized part to be
\[
(1-\alpha)(1-\beta) \ln \left( (1-\alpha)(1-\beta)/(1-F(t))(1-G(t)) \right)
\]
where the minimizing \( f^*_\epsilon(z,d) = C_2 f^*_\epsilon(z,d) \) and \( C_2 \) is a constant for this interval. The actual minimizing \( F^*, G^* \) are complicated and could be obtained from a pair of coupled differential equations with boundary conditions. Fortunately, we are only interested in the value of the integral which is the sum of (6.8) and (6.9)
\[
(1-\alpha)(1-\beta) \ln \left( \frac{(1-\alpha)(1-\beta)}{(1-F(t))(1-G(t))} \right) + (1-\alpha)(1-\beta) \ln \left( \frac{1-(1-\alpha)(1-\beta)}{1-(1-F(t))(1-G(t))} \right).
\]

(6.10)

The minimum of (6.10) with respect to \( \beta \) for \( 0 \leq \beta \leq 1 \) occurs for \( 1-\beta = (1-F(t))(1-G(t))/(1-F(t) - \epsilon) \) if this term does not exceed 1 and gives the value zero. If the term does exceed 1, then the minimum occurs at \( \beta = 0 \) and gives the value
\[
(1-F(t)-\epsilon) \ln \left( \frac{1-F(t)-\epsilon}{(1-F(t))(1-G(t))} \right) + (F(t)+\epsilon) \ln \left( \frac{F(t)+\epsilon}{1-(1-F(t))(1-G(t))} \right)
\]
for \((1-F(t))G(t) < \epsilon\).
Repeating similar steps for \( \lim_{n \to \infty} -n \ln(P_{n-1}) \) we obtain the same expression (6.10) except that now \( \alpha = F(t) - \epsilon \). The minimizing value in this case is \( 1 - \beta = (1 - F(t))((1 - G(t)) / (1 - F(t)) + \epsilon) \) which satisfies \( 0 \leq \beta \leq 1 \) and gives the value zero. There is an exception. If \( G(t) = 0 \), then \( \beta = 0 \) is required to make the integral finite and we get \( K(F(t) - \epsilon, F(t)) \) as in the no censoring case.

We now check that \( K(\Omega_a) \) and \( K(\Omega_b) \) do not exceed \(-\ln(q)\) given by (6.4). This is true when the values are zero. In the non zero case (6.11), consider the function

\[
h(u) = u \ln \left( \frac{u}{1 - F(t)} \right) + (1 - u) \ln \left( \frac{1 - u}{1 - F(t)(1 - G(t))} \right).
\]

We have \( h(0) = -\ln(q) \) and \( h((1 - F(t))(1 - G(t))) = 0 \). For \( u = 1 - F(t) - \epsilon \), \( h(u) \) has the value (6.11). Since \( h(u) \) is decreasing on the interval \( 0 \leq u \leq (1 - F(t))(1 - G(t)) \) we have \( h(0) \geq h(u) \) verifying the result.

Summarizing, \( \lim_{n \to \infty} -n \ln(P \{ \mid F_n(t) - F(t) \mid \geq \epsilon \}) = K_{pl} \) where

\[
K_{pl} = \begin{cases} 
0, & \text{for } 0 \leq \epsilon \leq (1 - F(t))G(t) \\
(1 - F(t) - \epsilon) \ln \left( \frac{1 - F(t) - \epsilon}{1 - F(t)(1 - G(t))} \right) + (F(t) + \epsilon) \ln \left( \frac{F(t) + \epsilon}{1 - F(t)(1 - G(t))} \right), & \text{for } (1 - F(t))G(t) < \epsilon < 1 - F(t).
\end{cases}
\]

7. The efficiency of the product limit estimator with exponential survival and censoring. Writing (6.12) with \( F(t), G(t) \) defined by (1.2), (4.1) and substituting \( T = t / \theta, E = exp(t / \theta) \) as in (2.10), the large deviation efficiency with censoring is

\[
e_{pl, mle} = K_{pl} / K_{mle}.
\]

Here \( K_{mle} \) is given by (5.7) and \( \gamma(t, \epsilon) = \gamma^*(T, E) \) as in (2.15). The exponent \( K_{pl} \) given by (6.12) can be written

\[
K_{pl} = \begin{cases} 
0, & \text{for } 0 \leq E \leq 1 - e^{-\rho T} \\
e^{-T}(1 - E) \ln \left( \frac{1 - E}{e^{-\rho T}} \right) + (1 - e^{-T}(1 - E)) \ln \left( \frac{1 - e^{-T}(1 - E)}{1 - e^{-T(1 + \rho)}} \right), & \text{for } 1 - e^{-\rho T} < E < 1.
\end{cases}
\]

Table 4 gives efficiency values for \( \rho = 0.1, 0.5 \).
Table 4. Large deviation efficiency of the product limit estimator with censoring.

<table>
<thead>
<tr>
<th></th>
<th>$\rho = 1.1$</th>
<th>$\rho = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho$</td>
<td>$T$ 0.5</td>
</tr>
<tr>
<td>$e_{pl,mle}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$0.1$</td>
<td>.1241</td>
<td>.0016</td>
</tr>
<tr>
<td>$0.2$</td>
<td>.3015</td>
<td>.1949</td>
</tr>
<tr>
<td>$0.3$</td>
<td>.4071</td>
<td>.3410</td>
</tr>
<tr>
<td>$0.4$</td>
<td>.4870</td>
<td>.4358</td>
</tr>
<tr>
<td>$0.5$</td>
<td>.5528</td>
<td>.5016</td>
</tr>
</tbody>
</table>

8. Discussion. The zero values for the large deviation efficiencies over part of the range is interesting. This surprising result indicates a slower than exponential rate in that region and it would be of interest to find the actual rate of convergence to zero for (2.5).

Future work for non-exponential $F(t)$, $G(t)$ seems of interest in view of the poor performance of the product limit estimator, especially as censoring increases.

Acknowledgements. Thanks go to Tom Kurtz, Olvi Mangasarian, Dennis Cox, and Brian Yandell for very helpful discussions. Computations were performed on the University of Wisconsin Statistical Research Computer.

REFERENCES


