ON THE ESTIMATION OF SURVIVAL FUNCTIONS UNDER A STOCHASTIC ORDER CONSTRAINT

JAVIER ROJO

ABSTRACT. Consider distribution functions F and G and suppose that $F(x) \leq G(x)$ for all x. The problem of estimating F or G, or both, arises quite naturally in applications. For example, in corrosion engineering it is of interest to estimate the pitting times of metals under two different strengths of corrosive environments. The empirical distribution functions F_m and G_n will not necessary satisfy the order constraint imposed by the experimental conditions. Lo (1987) proposed the estimators $\hat{F}_m = \min(F_m, G_n)$ and $\hat{G}_n = \max(F_m, G_n)$, which satisfy the constraint of interest, and showed that these estimator are asymptotically minimax, under suitable conditions, for a large class of loss functions. Although $\hat{F}_m$ and $\hat{G}_n$ are strongly uniformly consistent when both m and n tend to infinity, neither one is when only m or n go to infinity. Here, the estimators F_m^* (G_n^*) are proposed which are strongly uniformly consistent for F (G) when only m $(n) \to \infty$. The case of censored data is also considered. Under suitable conditions, weak convergence of the processes $\{\sqrt{m}(F_m^*(x) - F(x)), \quad 0 < x < \infty, \quad m = 1, \ldots\}$ and $\{\sqrt{n}(G_n^*(x) - G(x)), \quad 0 < x < \infty, \quad n = 1, \ldots\}$ is demonstrated. As a consequence, asymptotic confidence bands are obtained. For testing the hypothesis of identical distributions against a stochastic order alternative, the asymptotic distribution of the estimators under the assumption that $F(x) = G(x)$ for all x is also discussed. The results of a Monte-Carlo study show that the new estimators perform better than $\hat{F}_m$ and $\hat{G}_n$ and the non-parametric maximum likelihood estimators in terms of bias and mean squared error for a large class of examples.

1. INTRODUCTION

In many experimental sciences it is often of interest to estimate lifetime of experimental units when two different treatments are applied. For example, in corrosion engineering, the times until pitting of metals immersed in a corrosive environment are measured under two different solution corrosivities. Shibata and Takeyama (1977), for example, present data which strongly supports the belief that the times until pitting should be shorter in some sense, for the more corrosive environment. In toxicity studies, cells are grown in environments containing different levels of toxic materials (e.g. Arenaz *et al* (1992)). Invariably, the data supports the intuitive notion that the stronger the toxic solution is, the shorter the lifetimes of the organisms. Thus, it is of interest to estimate the lifetime distributions of these cells when it is know that one lifetime distribution is stochastically smaller.

To make this precise, let X and Y represent the random lifetimes under two different environments and let X and Y have F and G as their distribution functions. Lehmann (1955) defines the random variable X to be stochastically larger than the random variable Y if $F(x) \leq G(x)$ for all x. Lehmann and Rojo (1992) provide characterizations of stochastic ordering in terms of the maximal invariant GF^{-1} with respect to the group of monotone transformations, and connections with other partial orderings are provided. The problem of interest is the estimation of F and G based on independent random samples $X_1, \ldots, X_m$, and $Y_1, \ldots, Y_n$, subject to the constraint that $F(x) \leq G(x)$ for all x. In terms of survival functions $P = 1 - F$ and $Q = 1 - G$, one must estimate P and Q subject to the condition that $P \geq Q$.

As a way of illustrating the wide range of possible applications of partial orders of distributions in general, and of stochastic order in particular, consider Figure 1.1 which shows the Kaplan-Meier estimator for the data obtained in a clinical trial which was run to evaluate the efficiency of maintenance chemotherapy for acute myelongeneous leukemia (AML). The trial was conducted at Stanford University (Embury *et al* (1977)). After reaching a state of remission through treatment by chemotherapy, the patients who entered the study were randomized into two groups. The first group received maintenance chemotherapy; the second group did not. One would then expect that in this case, the survival times in the control group would be stochastically smaller than those in the first group. Figure 1 elucidates this fact. The clinical trial data are as follows: ($+$ denotes a censored observation)

Mantained group: 9, 13, 13^+, 18, 23, 28^+, 31, 34, 45^+, 48, 161^+.
Control Group: 5, 5, 8, 8, 12, 16^+, 23, 27, 30, 33, 43, 45.

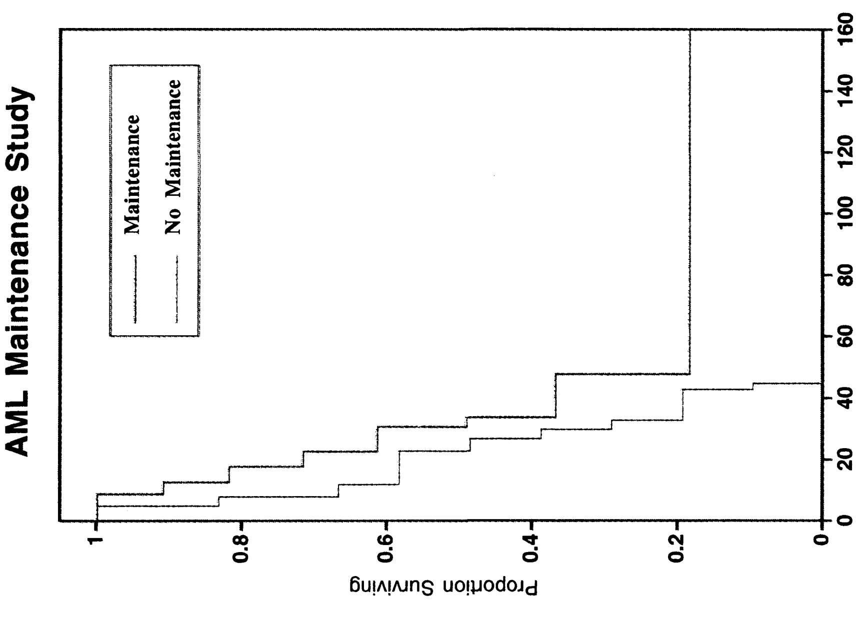


Figure 1.1- Survival Times in Weeks

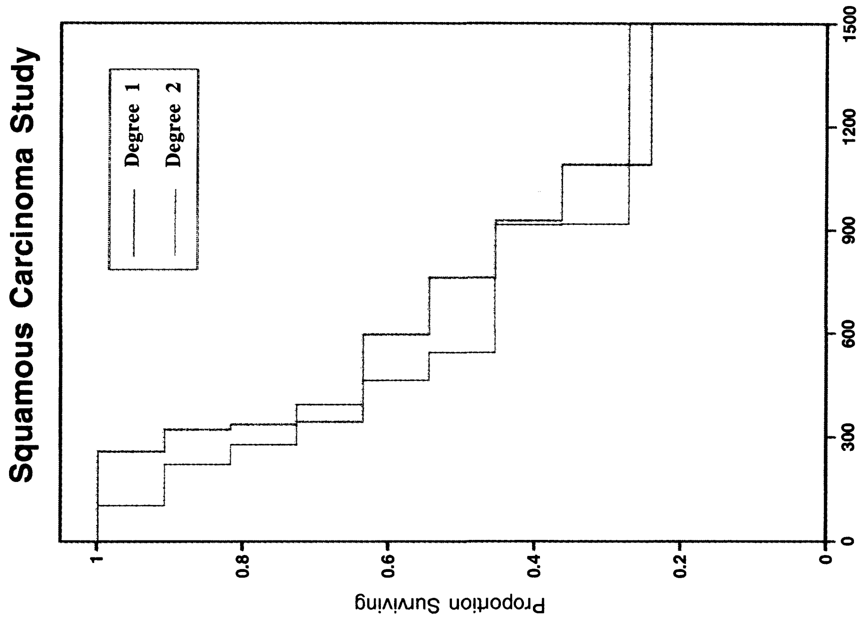


Figure 1.2- Survival Times in Days

Figure 1.2 shows the Kaplan-Meier estimators for the survival functions for two groups of patients with squamous carcinoma in the oropharynx with different degrees of deterioration. The data is taken from Kalbfleisch and Prentice (1980). For patients with degree one and degree two of lymph node deterioration, the survival times (in days) are as follows:

Degree 1: 261, 324, 338, 347, 599, 763, 929, 1086^+, 1092, 1317^+, 1609^+.
Degree 2: 105, 222, 279, 395, 465, 546, 915, 918, 1058^+, 1455^+, 1644^+.

Note that although it is expected that the survival function for patients in the group of degree one will be everywhere larger than that for patients in the group of degree two, due the small samples, the survival functions do not satisfy the constraint. Therefore, alternative estimators must be considered. The nonparametric maximum likelihood approach has been discussed by Dykstra (1982) and Dykstra and Fletz (1989). Dykstra, Kochar, and Robertson (1991), commenting on the availability of statistical inferential procedures when stochastic ordering obtains, conclude that "Unfortunately, statistical inference procedures have not been developed for many problems involving stochastic ordering and the development of the necessary theory for these problems seems to be a difficult task." Since then, however, there has been progress in developing the asymptotic theory needed for the nonparametric maximum likelihood estimator (Praestgaard and Huang (1996)). Nevertheless, the present approach provides estimators which dominate the NPMLE in terms of bias and mean squared error. Ma (1991), Rojo and Ma (1993), and Puri and Singh (1992) considered the one-sample problem and proposed an alternative to the maximum likelihood approach. Lo (1987), proposed estimators in the two sample problem which, under suitable conditions, are asymptotically minimax for a wide class of loss functions. Ma (1991) and Rojo and Ma (1993) considered estimators proposed by Lo (1987) and using Monte-Carlo simulation showed that these estimators behave better than their NPMLE counterparts in terms of bias and mean squared error for a large class of distributions. The case of censored data was also considered by Ma (1991) and Rojo and Ma (1993). Some possible drawbacks of the nonparametric maximum likelihood estimators are their large bias and lack of any distribution theory (finite-sample or asymptotic) that allows for the construction of confidence bands. In the two-sample problem the estimators of Lo (1987), although consistent when both sample sizes go to infinity, fail to be consistent when only one of the sample sizes goes to infinity.

The goal of this paper is to provide new estimators that remedy the problems alluded to earlier, and to discuss their asymptotic distribution theory. Under suitable conditions, the resulting processes converge weakly to the weak limit of the empirical process in the case of complete data, and to the weak limit of the Kaplan-Meier estimator in the case of censored data. As a consequence, standard results for the construction of confidence bands can be applied in the cases considered in this work. For testing purposes, the asymptotic distribution theory of the estimators is also considered when $P(x) = Q(x)$ for all x.

The organization of this paper is as follows: Section 2 defines the estimators and their strong uniform consistency is demonstrated. Section 3 discusses the case of censored data and proves the weak convergence of the processes defined by the proposed estimators. As a consequence, confidence bands for F and G can be obtained using standard results available for the empirical process and for the Kaplan-Meier process. Section 4 develops the asymptotic theory of the estimators when $P(x) = Q(x)$ for all x. The results can be used for testing for stochastic order. Section 5 reports on simulation results which suggest that the new estimators perform well in terms of bias and mean squared error when compared to the estimators of Lo (1987) and the non-parametric maximum likelihood estimators.

2. New Estimators and Their Strong Uniform Consistency

Let $X_1,\ldots,X_m$ and $Y_1,\ldots,Y_n$ be independent random samples from the distribution F and G respectively, and let F_m and G_n be the empirical distribution functions based on $X_1,\ldots,X_m$ and $Y_1,\ldots,Y_n$.

Suppose that it can be assumed that $F(x) \leq G(x)$ for all x. Let $P(x) = 1 - F(x)$ and $Q(x) = 1 - G(x)$ denote their respective survival functions. Lo (1987) considered the problem of the estimation of F and G under the stochastic order constraint and derived the asymptotic minimax estimators with respect to a wide class of loss functions. The estimators proposed by Lo (1987) were given as

$$\hat{P}_{mn} = \max(P_m, Q_n) \quad \text{and} \quad \hat{Q}_{mn} = \min(P_m, Q_n),$$

where P_m and Q_n represent the empirical survival functions. It is not difficult to see that

$$\sup_x |\hat{P}_{mn} - P| \leq \max(\sup_x |P_m - P|, \sup_x |Q_n - Q|)$$

and similarly for $\hat{Q}_{mn}$. Therefore as $m, n \to \infty$, $\hat{Q}_{mn}$ and $\hat{P}_{mn}$ converge uniformly with probability one to Q and P respectively. Rojo (1995) demonstrated the weak convergence of $\hat{P}_{mn}$ and $\hat{Q}_{mn}$ under the assumption that $P(x) > Q(x)$ for all x. However, if $P(x_0) = Q(x_0)$ for some x_0, then weak convergence does not hold for the process $\{\sqrt{m}(\hat{P}_{mn}(x) - P(x)), x > 0\}$.

One possible drawback of these estimators is that the asymptotic minimaxity and strong uniform consistency require that both m and n tend to infinity. In fact, it is not difficult to see that $\hat{P}_{mn}$ is not consistent if only m or n tend to infinity. This can easily be seen as follows. Since $\hat{P}_{mn}(x) - P(x) = \max(P_m(x) - P(x), Q_n(x) - P(x))$, then

$$\lim_{m\to\infty} P\{(\hat{P}_{mn}(x) - P(x)) > \varepsilon\} \geq P\{Q_n(x) - P(x) > \varepsilon\} > 0.$$

Similarly,

$$\lim_{n\to\infty} P\{(\hat{Q}_{mn}(x) - Q(x)) > \varepsilon\} \geq P\{P_m(x) - P(x) > \varepsilon\} > 0.$$

Also, since the influence of Q_n on $\hat{P}_{mn}$ does not vanish as $m \to \infty$, the estimator $\hat{P}_{mn}$ will tend to have a large bias. To correct these problems the following estimators are proposed. Define R_{m+n} to be the empirical cumulative survival function of the combined samples (see, e.g., Dykstra (1982)). That is

$$R_{m+n}(x) = \frac{n}{m+n} Q_n + \frac{m}{m+n} P_m. \tag{2.1}$$

Now define the estimators

$$\tilde{P}_{mn}(x) = \max(P_m(x), R_{m+n}(x)), \text{ and} \tag{2.2}$$

$$\tilde{Q}_{mn}(x) = \min(Q_n(x), R_{m+n}(x)). \tag{2.3}$$

Clearly, $\tilde{P}_{mn}$ and $\tilde{Q}_{mn}$ satisfy the stochastic order constraint. Since

$$\min(P_m, Q_n) \leq R_{m+n} \leq \max(P_m, Q_n),$$

it follows easily that $\tilde{P}_{mn} \leq \hat{P}_{mn}$ and $\tilde{Q}_{mn} \geq \hat{Q}_{mn}$. Consequently, $\tilde{P}_{mn}$ ($\tilde{Q}_{mn}$) has less positive (negative) bias than $\hat{P}_{mn}$ ($\hat{Q}_{mn}$). The strong uniform consistency of $\tilde{P}_{mn}$ and $\tilde{Q}_{mn}$ is given in the following result.

Theorem 2.1.- Suppose that $P(x) \geq Q(x)$ for all x. The estimator $\tilde{P}_{mn}$ converges uniformly with probability one to the underlying survival function P, as $m \to \infty$. Similarly, $\tilde{Q}_{mn}$ converges uniformly with probability one to Q as $n \to \infty$.

Proof.- Only the result for $\tilde{P}_{mn}$ is shown here. The corresponding proof for $\tilde{Q}_{mn}$ is similar.
Using (2.1), note that

$$\begin{aligned}
|\tilde{P}_{mn} - P| &= |P_m - P + \frac{n}{n+m}\max(0, Q_n - P_m)| \\
&\leq |P_m - P| + \frac{n}{n+m}\max(0, Q_n - P_m) \\
&\leq |P_m - P| + \frac{n}{n+m}\max(0, Q_n - Q + Q - P + P - P_m) \\
&\leq |P_m - P| + \frac{n}{n+m}(|Q_n - Q| + |P_m - P|). \qquad (2.4)
\end{aligned}$$

Now, it is clear from (2.4) that $|\tilde{P}_{mn} - P|$ converges uniformly to zero with probability one as $m \to \infty$.

We now turn our attention to the weak convergence of the processes $\{\sqrt{m}(\tilde{P}_{mn}(t) - P(t)), t > 0\}$ and $\{\sqrt{n}(\tilde{Q}_{mn}(t) - Q(t)), t > 0\}$. To try to motivate some of the results below, perhaps some intuitive remarks are in order. It is not difficult to see that if $P(x_0) = Q(x_0)$ for some x_0, the asymptotic distribution of $\tilde{P}_{mn}$ when $m/n \to 0$, suitably normalized, puts 1/2 of mass at zero while "spreading" the other half of the mass on the positive real axis according to a normal distribution. That is, if $m/n \to 0$, then

$$\sqrt{m}(\tilde{P}_{mn}(x_0) - P(x_0)) \xrightarrow{\mathcal{D}} \max(0, T)$$

where T is a normal random variable with mean zero and variance $F(x_0)P(x_0)$. If $m/n \to \infty$, then $\sqrt{m}(\tilde{P}_{mn}(x_0) - P(x_0))$ converges in distribution to a mean zero normal distribution with variance $F(x_0)P(x_0)$. If $m/n \to c, 0 < c < \infty$, then the asymptotic distribution of $\sqrt{m}(\tilde{P}_{mn}(x_0) - P(x_0))$ is somewhat more complex. The following theorem provides the precise statement for the asymptotic distribution of $\sqrt{m}(\tilde{P}_{mn}(x_0) - P(x_0))$ in the case that $P(x_0) = Q(x_0)$. The proof follows from a direct application of the continuous mapping theorem for weak convergence and Slutsky's theorem.

Theorem 2.2.- Let $\tilde{P}_{mn}$ be defined by (2.2) and suppose that $P(x_0) = Q(x_0)$ for some x_0. Let $m/n \to c$, where $0 \leq c \leq \infty$. Then,

$$\sqrt{m}(\tilde{P}_{mn}(x_0) - P(x_0)) \xrightarrow{\mathcal{D}} \max(W_1, \frac{\sqrt{c}\, W_2 + c\, W_1}{1+c}),$$

where W_1 and W_2 are two independent and identically distributed mean zero normal random variables with variance $F(x_0)P(x_0)$, and where the asymptotic limit is interpreted as W_1 when $c = \infty$, and as $\max(W_1, 0)$ when $c = 0$.

Proof.- To facilitate the notation, we drop the argument x_0 throughout, and hence write, e.g., P instead of $P(x_0)$. By the central limit theorem and the independence of P_m and Q_n, it follows that

$$\{\frac{n}{n+m}\sqrt{\frac{m}{n}}\sqrt{n}(Q_n - P), \frac{m}{n+m}\sqrt{m}(P_m - P)\} \xrightarrow{\mathcal{D}} (\frac{\sqrt{c}}{1+c}W_2, \frac{c}{1+c}W_1),$$

where W_1 and W_2 are *iid*, mean zero normal random variables with variance $P(x_0)F(x_0)$. By the continuity of the mapping $h(x, y) = \left(\frac{1+c}{c}y, x + y\right)$, for $0 < c < \infty$, it then follows by the continuous mapping theorem that, since

$$\sqrt{m}(\tilde{P}_{mn} - P) = \max(\sqrt{m}(P_m - P), \frac{n}{n+m}\sqrt{\frac{m}{n}}\sqrt{n}(Q_n - Q) + \frac{m}{m+n}\sqrt{m}(P_m - P)), \tag{2.5}$$

then, for $0 < c < \infty$,

$$\sqrt{m}(\tilde{P}_{mn} - P) \xrightarrow{\mathcal{D}} \max(W_1, \frac{\sqrt{c}\, W_2 + c\, W_1}{1+c}).$$

Note that the result holds for $c = 0$ since then, as $m \to \infty$ with $m/n \to 0$, the second term on the right side of (2.5) converges in probability to zero. On the other hand, when $m/n \to \infty$ the second term on the right side of (2.5) converges in distribution to W_1. Thus interpreting $\max(W_1, \frac{\sqrt{c}W_2 + cW_1}{1+c})$ to mean W_1 when $c = \infty$ the result follows.

On the other hand, if $P(x_0) > Q(x_0)$, note that writing

$$\sqrt{m}(\tilde{P}_{mn}(x_0) - P(x_0)) = \sqrt{m}(P_m - P) + \max(0, \frac{n}{n+m}\sqrt{m}(Q_n - Q + Q - P + P_m - P)),$$

it is clear that for $\frac{m}{n} \to c, 0 \leq c \leq \infty$, the second term on the right side of the previous expression converges to zero in probability and hence $\sqrt{m}(\tilde{P}_{mn}(x_0) - P(x_0)) \xrightarrow{\mathcal{D}} N(0, F(x_0)P(x_0))$. Therefore, if there is an interval, say $(x_0, x_0+a]$, where $P(x) > Q(x)$, while $P(x_0) = Q(x_0)$, the asymptotic probabilistic behavior of the process at x_0 is totally different than that of the process at $x_0 + \varepsilon$, for every sufficiently small $\varepsilon > 0$, and hence tightness can not hold on $[x_0, x_0 + a]$, for some $a > 0$, and hence weak convergence can not hold. These arguments are made precise in the following Theorem.

Theorem 2.3.- Let W_1 and W_2 denote independent Gaussian processes with $E(W_1(t)) = 0$, $E(W_2(t)) = 0$, $E(W_1(t)W_1(s)) = F(s)P(t)$, and $E(W_2(s)W_2(t)) = G(s)Q(t)$, respectively, for $s \leq t$, and suppose that F and G are continuous distributions functions with the same support.

(i) If $P(x) \geq Q(x)$ for all x, then as $m \to \infty$, with $m/n \to \infty$, $(n \to \infty$ and $n/m \to \infty)$

$$\sqrt{m}(\tilde{P}_{mn} - P) \xrightarrow{\mathcal{D}} W_1 \quad (\sqrt{n}(\tilde{Q}_{mn} - Q) \xrightarrow{\mathcal{D}} W_2).$$

(ii) If $P(x_0) = Q(x_0)$ for some x_0, $P \neq Q$, and $m/n \to c$, $(n/m \to c)$, $0 \leq c < \infty$ then the process $\sqrt{m}(\tilde{P}_{mn} - P)$ $(\sqrt{n}(\tilde{Q}_{mn} - Q))$ does not converge weakly as $m \to \infty$ $(n \to \infty)$.

If $P(x) > Q(x)$ for all x, then as $m \to \infty$, $(n \to \infty)$

$$\sqrt{m}(\tilde{P}_{mn} - P) \xrightarrow{\mathcal{D}} W_1 \quad (\sqrt{n}(\tilde{Q}_{mn} - Q) \xrightarrow{\mathcal{D}} W_2).$$

Proof.- To prove *(i)* note that

$$\tilde{P}_{mn} = P_m + \frac{n}{n+m} \max(0, Q_n - P_m) \tag{2.6}$$

and hence, the weak limit of $\sqrt{m}(\tilde{P}_{mn} - P)$, as $m \to \infty$, is the weak limit of $\sqrt{m}(P_m - P)$. To prove *(ii)*, suppose that $P(x_0) = Q(x_0)$. Recall that if $\sqrt{m}(\tilde{P}_{mn} - P)$ converges weakly, then the sequence $\sqrt{m}(\tilde{P}_{mn} - P)$ must be tight on $[x_0, x_0 + a]$, for example, for any $a > 0$. It is now demonstrated that if $P(x_0) = Q(x_0)$ for some x_0, then the sequence $\sqrt{m}(\tilde{P}_{mn} - P)$ is not tight when $0 \leq c < \infty$ and hence can not converge weakly as $m \to \infty$. Consider first the case $m/n \to 0$. Without loss of generality suppose that the support of F and G is $[0, 1]$, and that there is $\gamma > 0$ such that $P > Q$

on $(x_0, x_0+\gamma)$. Define $Z_{m,n}(x) = \sqrt{m}(\tilde{P}_{m,n}(x) - P(x))$, and consider for $\delta > 0$

$$\begin{aligned}\sup_{x_0 \le s \le x_0+\delta} |Z_{m,n}(s) - Z_{m,n}(x_0)| &= \sqrt{m} \sup_{x_0 \le s \le x_0+\delta} |\max(P_m(x_0) - P(x_0), R_{m,n}(x_0) - P(x_0)) \\ &\quad - \max(P_m(s) - P(s), R_{m,n}(s) - P(s))| \\ &= \sqrt{m} \sup_{x_0 \le s \le x_0+\delta} |\max(P_m(x_0) - P(x_0), \\ &\quad \frac{n}{n+m}(Q_n(x_0) - Q(x_0)) + \frac{m}{n+m}(P_m(x_0) - P(x_0)) \\ &\quad - (P_m(s) - P(s)) - \frac{n}{n+m}\max(0, Q_n(s) - P_m(s))|,\end{aligned} \tag{2.7}$$

where the sup in the above expressions is taken over $[x_0, 1]$ if $x_0 + \delta > 1$. For $s_0 = x_0 + \frac{\min(\delta,\gamma)}{2}$, eventually with probability one, $P_m(s_0) > Q_n(s_0)$ so that, eventually with probability one,

$$\begin{aligned}&\sup_{x_0 \le s \le x_0+\delta} |Z_{m,n}(s) - Z_{m,n}(x_0)| \\ \ge\ & |Z_{m,n}(s_0) - Z_{m,n}(x_0)| \\ \ge\ & \sqrt{m}|\max(P_m(x_0) - P(x_0), \frac{n}{n+m}(Q_n(x_0) - Q(x_0)) \\ &+ \frac{m}{n+m}(P_m(x_0) - P(x_0))) - (P_m(s_0) - P(s_0))| \\ =\ & |\max(\sqrt{m}(P_m(x_0) - P(x_0)), \frac{\sqrt{mn}}{n+m}\sqrt{n}(Q_n(x_0) - Q(x_0)) \\ &+ \frac{m}{n+m}\sqrt{m}(P_m(x_0) - P(x_0))) - \sqrt{m}(P_m(s_0) - P(s_0))|.\end{aligned} \tag{2.8}$$

Since $\frac{m}{n}$, $\frac{\sqrt{mn}}{m+n}$, and $\frac{m}{m+n}$ go to zero, it follows from (2.8) that, for $\varepsilon > 0$

$$\begin{aligned}&\lim_{m,n\to\infty} \Pr\{\sup_{x_0 \le s \le x_0+\delta} |Z_{m,n}(s) - Z_{m,n}(x_0)| \ge \varepsilon\} \\ &\ge \lim_{m,n\to\infty} \Pr\{\max(\sqrt{m}(P_m(x_0) - P(x_0)), \tfrac{\sqrt{mn}}{n+m}\sqrt{n}(Q_n(x_0) - Q(x_0)) \\ &+ \tfrac{m}{n+m}\sqrt{m}(P_m(x_0) - P(x_0))) - \sqrt{m}(P_m(s_0) - P(s_0)) \ge \varepsilon\} \\ &\ge \lim_{m,n\to\infty} \Pr\{\tfrac{\sqrt{mn}}{n+m}\sqrt{n}(Q_n(x_0) - Q(x_0)) \\ &+ \tfrac{m}{m+n}\sqrt{m}(P_m(x_0) - P(x_0)) - \sqrt{m}(P_m(s_0) - P(s_0)) \ge \varepsilon\}\end{aligned}$$

$$= \Phi\left(-\frac{\varepsilon}{P(s_0)(1-P(x_0))}\right), \tag{2.9}$$

where Φ denotes the standard normal distribution. It follows from (2.9) that (15.8) in Theorem 15.3 in Billingsley (1968) does not hold for the interval $[x_0, x_0 + a]$. Consequently, the sequence $\{Z_{m,n}\}_{m,n=1}^{\infty}$ is not tight on $[x_0, x_0 + a]$ and hence the process can not converge weakly.

Now consider the case $m/n \to c$, $0 < c < \infty$. Without loss of generality suppose that $F(t) = t$ and that there is $\gamma > 0$ such that $P(x) > Q(x)$ on $(x_0, x_0 + \gamma)$. Define $Z_{m,n}(t) = \sqrt{m}(\tilde{P}_{mn}(t) - (1-t))$ and consider, for $\delta > 0$,

$$\sup_{x_0 \le s \le x_0+\delta} |Z_{m,n}(s) - Z_{m,n}(x_0)| =$$

$$\sup_{x_0 \le s \le x_0+\delta} |\sqrt{m}\max(x_0 - F_m(x_0), \frac{n}{n+m}(x_0 - G_n(x_0)) + \frac{m}{n+m}(x_0 - F_m(x_0))$$

$$- \sqrt{m}\max(s - F_m(s), \frac{n}{n+m}(s - G_n(s)) + \frac{m}{m+n}(s - F_m(s)))|.$$

Now, for $s_0 = x_0 + \min(\delta, \gamma)/2$, eventually with probability one, $G_n(s_0) > F_m(s_0)$, so that eventually with probability one,

$$\sup_{x_0 \le s \le x_0+\gamma} |Z_{m,n}(s) - Z_{m,n}(x_0)| \ge |Z_{m,n}(s_0) - Z_{m,n}(x_0)|$$

$$= \sqrt{m}|x_0 - F_m(x_0) + \max(0, \frac{n}{m+n}(F_m(x_0) - x_0) + \frac{n}{n+m}(x_0 - G_n(x_0))) - s_0 + F_m(s_0)|$$

$$= |\sqrt{m}(x_0 - F_m(x_0)) + \sqrt{m}(F_m(s_0) - s_0) + \max(0, \frac{n}{n+m}\sqrt{m}(F_m(x_0) - x_0)$$

$$+ \frac{\sqrt{nm}}{m+n}\sqrt{n}(x_0 - G_n(x_0)))|.$$

Therefore, $\lim_{m,n\to\infty} P(\sup_{x_0 \le s \le x_0+\delta} |Z_{m,n}(s) - Z_{m,n}(x_0)| \ge \varepsilon) \ge$

$$\lim_{m,n\to\infty} \int_0^{\infty}\int_{\varepsilon}^{\infty} P\{\max(0, \frac{n}{n+m}z_{x_0} + \frac{n}{m+n}\sqrt{\frac{m}{n}}Y_n) \ge \varepsilon - z_{x_0} - z_{s_0}\} dH_m(z_{x_0}, z_{s_0}), \tag{2.10}$$

where $H_m(z_{x_0}, z_{s_0})$ denotes the joint distribution of $\sqrt{m}(F_m(s_0) - s_0)$ and $\sqrt{m}(x_0 - F_m(x_0))$ and $Y_n = \sqrt{n}(x_0 - G_n(x_0))$. Now note that the integrand in (2.10) for $z_{x_0} > \varepsilon$ and $z_{x_0} > 0$, equals one. Consequently,

$$\lim_{m,n\to\infty} \Pr\{\sup_{x_0 \le s \le x_0+\delta} |Z_{m,n}(s) - Z_{m,n}(x_0) \ge \varepsilon\} \tag{2.11}$$

$$\geq \lim_{m,n\to\infty} \Pr\{\sqrt{m}(F_m(s_0) - s_0) \geq 0, \sqrt{m}(x_0 - F_m(x_0) \geq \varepsilon\}.$$

Now, as $n \to \infty$, the right side of (2.11) converges to Pr $\{X \geq 0,\ Y \geq \varepsilon\}$, where (X, Y) has a bivariate distribution with mean vector (0,0) and $Var(X) = F(s_0)(1 - F(s_0))$, $Var(Y) = F(x_0)$ $(1-$ $F(x_0))$, and $Cov(X, Y) = x_0 s_0 - \min(x_0, s_0)$. Therefore,

$$\lim_{m,n\to\infty} \Pr\{\sup_{x_0 \leq s \leq x_0+\delta} |Z_{m,n}(s) - Z_{m,n}(x_0)| \geq \varepsilon\} \geq \Pr\{X \geq 0, Y \geq \varepsilon\}. \qquad (2.12)$$

It follows from 2.12 that (15.8) in Theorem 15.3 in Billingsley (1968) does not hold for the interval $[x_0, x_0 + a]$, and therefore the sequence $\{Z_{m,n}\}_{m,n=1}^{\infty}$ is not tight on $[x_0, x_0 + a]$ and hence can not converge weakly.

To prove *(iii)*, note that because of *(i)*, it is enough to consider the case where $m = O(n)$. Assume without loss of generality that $F(x) = x, \quad 0 \leq x \leq 1$. Let $0 < x_1 <, \ldots, < x_k < 1$, and consider the random vector with components

$$\sqrt{m}(\tilde{P}_{m,n}(x_i) - (1 - x_i)) = \sqrt{m}(P_m(x_i) - (1 - x_i) + \frac{n}{m+n}\max(0, Q_n(x_i) - P_m(x_i))),$$

for $i = 1, \ldots, k$. Since $G(x_i) > x_i$ for $i = 1, \ldots, k$, it follows that with probability one, $\frac{\sqrt{mn}}{m+n}(Q_n(x_i) - (1 - x_i))$ converges to $-\infty$, and hence, eventually, with probability one, $\sqrt{m}(\tilde{P}_{mn}(x_i) - (1 - x_i)) = \sqrt{m}(P_m(x_i) - (1 - x_i))$ for $i = 1, \ldots, k$. Therefore, the finite-dimensional distributions of the process $\{\sqrt{m}(\tilde{P}_{mn}(x) - (1 - x)),\ \ 0 \leq x \leq 1\}$ converge to the finite-dimensional distributions of W°, where W° denotes Brownian bridge. It remains to show the tightness of $\{Z_{m,n}\}$. Since $Z_{m,n}(0) = 0$, and using Theorem 15.2 and inequality (14.9) in Billingsley (1968), it is enough to prove that for each positive ε and η, there exists a δ, $0 < \delta < 1/2$, and integers n_0, m_0 such that

$$P\{\sup_{t \leq s \leq t+\delta} |Z_{m,n}(s) - Z_{m,n}(t)| > \varepsilon\} \leq \eta \text{ for } n \geq n_0,\ \ m \geq m_0, \text{ and all } t. \qquad (2.13)$$

Note that using expression (2.5),

$$\begin{aligned}
&\sup_{t\le s\le t+\delta}|Z_{m,n}(s)-Z_{m,n}(t)| \\
&=\sqrt{m}\sup_{t\le s\le t+\delta}|P_m(t)-(1-t)+\frac{n}{n+m}\max(0,Q_n(t)-P_m(t))-P_m(s)+(1-s) \\
&-\frac{n}{n+m}\max(0,Q_n(s)-P_m(s))| \\
&\le\sqrt{m}\sup_{t\le s\le t+\delta}|P_m(t)+t+P_m(s)-s|+\sqrt{m}\sup_{t\le s\le t+\delta}\frac{n}{n+m}\max(0,Q_n(t)-P_m(t)) \\
&+\sqrt{m}\sup_{t\le s\le t+\delta}\frac{n}{n+m}\max(0,Q_n(s)-P_m(s)).
\end{aligned}$$

Since $\{P_m(t)-(1-t), 0<t<1, m=1,2,\ldots\}$ is known to be tight, and $\max(0,Q_n(t)-P_m(t))=0$ eventually with probability one, (2.13) will follow if it can be shown that

$$\sup_{t\le s\le t+\delta}\sqrt{m}\max(0,Q_n(s)-P_m(s)) \tag{2.14}$$

converges to zero in probability for all t and all $\delta>0$. In fact, it is now demonstrated that

$$\sup_{0\le t\le 1}\sqrt{m}\max(0,Q_n(t)-P_m(t))\xrightarrow{\mathcal{P}}0. \tag{2.15}$$

The proof of (2.15) follows along the lines of Rojo (1995). Let β_m be a decreasing sequence of positive real numbers, such that $\beta_m\to 0$ and $\sqrt{m}\beta_m\to\infty$. Define $k_m\downarrow 0$ such that $\inf_{k_m\le x\le 1-k_m}(G(x)-F(x))=\beta_m$, and let $A_m=\{x: k_m\le x\le 1-k_m\}$ and $B_{mn}=\{x: F_m(x)>G_n(x)\}$. Note that for $x\in B_{mn}$, $\max(0,Q_n(x)-P_m(x))=Q_n(x)-P_m(x)$. Therefore,

$$\begin{aligned}
\sqrt{m}\sup_{0\le s\le 1}\max(0,Q_n(s)-P_m(s))&=\sqrt{m}\max\{\sup_{s\le k_m}\max(0,Q_n(s)-P_m(s)), \\
&\sup_{s\in A_m}\max(0,Q_n(s)-P_m(s)),\ \sup_{s\ge 1-k_m}\max(0,Q_n(s)-P_m(s))\}.
\end{aligned}$$

Now,

$$\begin{aligned}
&\Pr\{\sqrt{m}\sup_{0\le s\le 1}\max(0,Q_n(s)-P_m(s))\ge\varepsilon\} \\
&=\Pr\{\max\{\sqrt{m}\sup_{s\in A_m^c}\max(0,Q_n(s)-P_m(s)),\sqrt{m}\sup_{s\in A_m}\max(0,Q_n(s)-P_m(s))\}\ge\varepsilon\} \\
&\le\Pr\{\sqrt{m}\sup_{s\in A_m^c\cap B_{mn}}(Q_n(s)-P_m(s))\ge\varepsilon\}+\Pr\{\sqrt{m}\sup_{s\in A_m\cap B_{mn}}(Q_n(s)-P_m(s))\ge\varepsilon\},
\end{aligned}$$

since $\max(0, Q_n(s) - P_m(s)) = 0$ on B^c_{mn}. Now, for $s \in A_m$, $P(s) - Q(s) \geq \beta_m$, and therefore,

$$\begin{aligned}
&\Pr\{\sqrt{m} \sup_{s\in A_m\cap B_{mn}} (Q_n(s) - P_m(s)) \geq \varepsilon\} \\
&= \Pr\{\sqrt{m} \sup_{s\in A_m\cap B_{mn}} (Q_n(s) - Q(s) + Q(s) - P(s) + P(s) - P_m(s)) \geq \varepsilon\} \\
&\leq \Pr\{\sqrt{m} \sup_{s\in A_m\cap B_{mn}} (Q_n(s) - Q(s) + P(s) - P_m(s)) \geq \varepsilon + \sqrt{m}\beta_m\} \to 0,
\end{aligned}$$

since $\sqrt{m}\beta_m \to \infty$ and $m = O(n)$.
Also,

$$\begin{aligned}
&\Pr\{\sqrt{m} \sup_{s\in A^c_m\cap B_{mn}} (Q_n(s) - P_m(s)) \geq \varepsilon\} \leq \Pr\{\sqrt{m} \sup_{s\in A^c_m\cap B_{mn}} (G_m(s) - G(s)) \geq \varepsilon/2\} \\
&+ \Pr\{\sqrt{m} \sup_{s\in A^c_m\cap B_{mn}} (G(s) - F_n(s)) \geq \epsilon/2\}. \qquad (2.16)
\end{aligned}$$

Now consider

$$\Pr\{\sqrt{m} \sup_{s\leq k_m} (G_n(s) - G(s)) \geq \varepsilon/2\} + \Pr\{\sqrt{m} \sup_{s\leq k_m} (G_n(s) - F_m(s)) \geq \varepsilon/2\}. \qquad (2.17)$$

Since $G(s) \leq F(s)$ for all s, and using inequality 1 on page 134 of Shorack and Welner (1986) with $g(t) \equiv 1$, for $0 \leq t \leq 1$, it follows that (2.17) is bounded above by

$$\begin{aligned}
&4G(k_m)/\varepsilon^2 + \Pr\{\sqrt{n} \sup_{s\leq k_m} (F(s) - F_n(s)) \geq \sqrt{n}\varepsilon/2\sqrt{m}\} \\
&\leq 4G(k_m)/\varepsilon^2 + 4mF(k_m)/n\varepsilon^2 \to 0 \ \text{ since } \ m = O(n).
\end{aligned}$$

By symmetry then $\Pr\{\sqrt{m}\sup_{s\geq 1-k_m}(Q_n(s) - P_m(s)) \geq \varepsilon\} \to 0$ and hence (2.13) follows.

3. The Case of Censored Data

The case of censored data is now considered. As before, let $X_1, \ldots, X_m$ be a random sample from F and let $Y_1, \ldots, Y_n$ be a random sample from G. Following Csörgő and Horváth (1983), random samples $X'_1, \ldots, X'_m$ and $Y^*_1, \ldots, Y^*_n$ with left-continuous distributions H' and H^*, respectively, censor $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ on the right. Therefore, the observations available consist of the pairs $(Z'_i, \delta'_i), (Z^*_j, \delta^*_j)$, $i = 1, \ldots, m$, $j = 1, \ldots, n$, where $Z'_i = \min(X_i, X'_i)$, $Z^*_j = \min(Y_j, Y^*_j)$, and δ'_i and δ^*_j are the indicators of the events $\{Z'_i = X_i\}$ and $\{Z^*_j = Y_j\}$ respectively. Let P_m and Q_n denote the

Kaplan-Meier (1958) estimators of F and G respectively. That is,

$$P_m(t) = \begin{cases} \prod_{1\leq i\leq m: Z_i^{'}<t} \left(\frac{m - M_{i:m} - 1}{m - M_{i:m}}\right)^{\delta_i}, & t \leq Z_{(m)}^{'} \\ 0, & t > Z_{(m)}^{'} \end{cases} \tag{3.1}$$

where $Z_{(m)}^{'} = \max(Z_1, \ldots, Z_m)$ and $M_{i:m} = \sum_{k=1}^{m} I_{\{Z_k^{'}<Z_i^{'}\}}$. Replacing P_m by Q_n, X_i and Y_i, $M_{i:m}$ by $N_{j:n} = \sum_{k=1}^{n} I_{\{Z_k^{*}<Z_j^{*}\}}$, and $\delta_i^{'}$ by δ_j^{*} a similar expression for Q_n is obtained. Note that the distribution $W^{'}$ and W^{*} of $Z_i^{'}$ and Z_j^{*} are given, respectively, by $\bar{W}^{'}(t) = \bar{F}(t)\bar{H}^{'}(t)$ and $\bar{W}^{*}(t) = \bar{G}(t)\bar{H}^{*}(t)$ for each t. Csörgő and Horváth (1983), considered the rate of convergence of P_m (Q_n) to $\bar{F}$ $(\bar{G})$.

For a distribution F, let $T_F = \inf\{t : F(t) = 1\}$. The weak convergence of the Kaplan-Meier estimator F_m as defined by (3.1) is provided by the following Theorem of Breslow and Crowley (1974). Define $T^{'} = \min(T_F, T_{H^{'}})$.

Theorem 3.1.- Let F and $H^{'}$ be continuous and consider $T < T^{'}$ with $W^{'}(T) < 1$. Then the process $\{Z_m^{'}(t) = \sqrt{m}(P_m(t) - \bar{F}(t)), 0 \leq t \leq T\}$, converges weakly to a zero mean Gaussian process Z with covariance function

$$\mathrm{Cov}(Z(s), Z(t)) = C(s)\bar{F}(s)\bar{F}(t), s \leq t,$$

where

$$C(s) = \int_0^s \frac{dF(t)}{\bar{F}^2(t)\bar{H}^{'}(t)}, \quad s < t.$$

With appropriate changes and similar assumptions, the corresponding weak convergence of $\{Z_n^{*}(t) = \sqrt{n}(Q_n(t) - Q(t)), 0 \leq t \leq T\}$, $T < T^{*} = \min(T_G, T_H^{*}), W^{*}(T) < 1$, follows.

Now suppose $F(x) \leq G(x)$ for all x. To estimate F and G based on $(Z_i^{'}, \delta_i^{'})$ and (Z_j^{*}, δ_j^{*}) define, as in (2.1),

$$R_{m+n}(x) = \frac{n}{m+n}Q_n(x) + \frac{m}{m+n}P_m \tag{3.2}$$

$$\tilde{P}_{mn}(x) = \max(P_m(x), R_{m+n}(x)) \tag{3.3}$$

$$\tilde{Q}_{mn}(x) = \min(Q_n(x), R_{m+n}(x)). \tag{3.4}$$

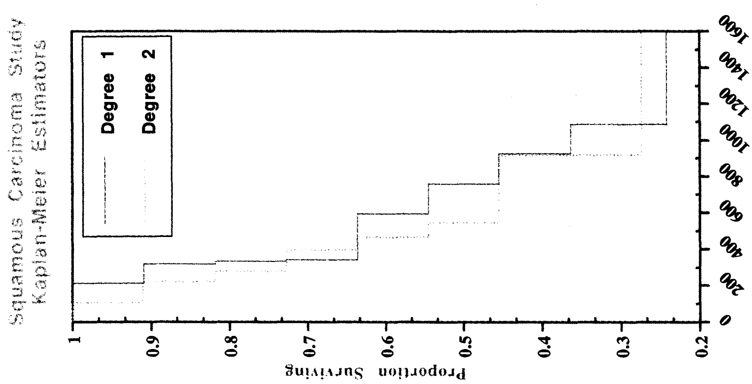


Figure 2.1

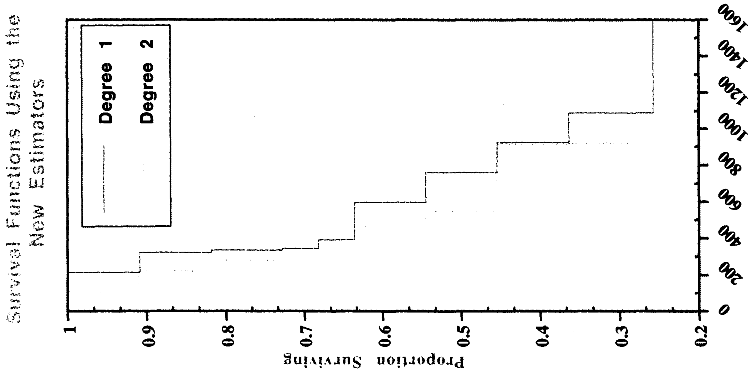


Figure 2.2

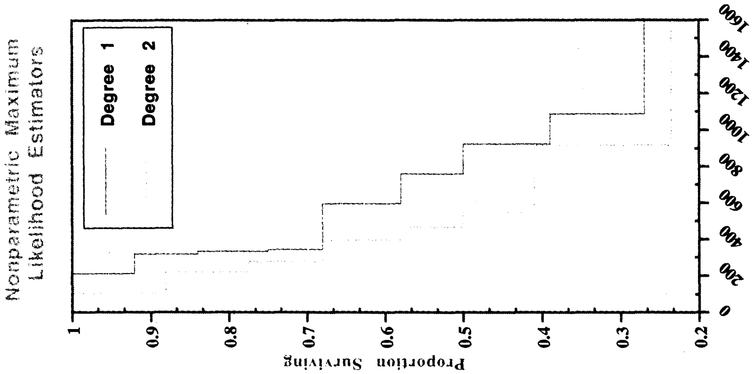


Figure 2.3

As it was the case in section 2, $\tilde{P}_{m,n}$ and $\tilde{Q}_{mn}$ are stochastically ordered with $\tilde{P}_{mn} \geq R_{m+n} \geq \tilde{Q}_{mn}$. Figures 2.1, 2.2, and 2.3, show the Kaplan-Meier estimators, the estimators defined by (3.3) and (3.4) and their Nonparametirc Maximum Likelihood counterparts, for the Squamous Carcinoma data. Note that the modification of the Kaplan-Meier estimators provided by (3.3) and (3.4) is substantially smaller than that provided by the Nonparametirc Maximum Likelihood estimators. Thus, the new estimators only modify the Kaplan-Meier estimators where the stochastic order constraint is violated, while the Nonparametric Maximum Likelihood estimators, "push" both Kaplan-Meier estimators apart until the violation of the order disappears. This, therefore, causes the NPMLE to have a larger bias than the estimators proposed here.

The strong uniform convergence of $\tilde{P}_{mn}$ and $\tilde{Q}_{mn}$ is discussed next. Th following result hinges on the work of Stute and Wang (1993). Note that in particular, the following result does not require F nor H' to be continuous, and the strong convergence is demonstrated on $(-\infty, T_{M'}]$, where $T_{M'}$=min$(T_{W'}, T)$.

Theorem 3.2.- Suppose that F and H', and G and H^* do not have any jumps in common, and that $F(x) \leq G(x)$ for all x. Let $T_{W'}=\inf\{t : W'(t) = 1\}$, and $T_{W^*}=\inf\{t : W^*(t) = 1\}$, and $\bar{W}'=\bar{F}\bar{H}'$, $\bar{W}^*=\bar{G}\bar{H}^*$. Then,

$$\sup_{-\infty<t<\leq T_{M'}} |\tilde{P}_{mn} - P| \rightarrow 0, \tag{3.5}$$

with probability one as $m \rightarrow \infty$, if and only if $F(T_{W'}) = 0$ or $F(T_{W'}) > 0$ but $H'(T^-_{W'}) < 1$, and $G(T^*_W) = 0$ or, $G(T^*_W) > 0$ but $H^*(T^-_{W^*}) < 1$.

Proof.- The result follows immediately from Corollary 1.3 in Stute and Wang (1993) after writing

$$\begin{aligned}
|\tilde{P}_{mn} - P| &= |P_m - P + \frac{n}{n+m} \max(0, Q_n P_m)| \\
&\leq |P_m - P| + \frac{n}{n+m} \max(0, Q_n - Q + Q - P + P - P_m) \\
&\leq |P_m - P| + \frac{n}{n+m} (|Q_n - Q| + |P_m - P|).
\end{aligned}$$

A similar result holds for $\tilde{Q}_{mn}$.

We now turn our attention to the weak convergence of the processes $\{\tilde{P}_{mn}, 0 \leq t \leq T_1\}$ and $\{\tilde{Q}_{mn}, 0 \leq t \leq T_2\}$, where $T_1 < T', W'(T_1) < 1$ and $T_2 < T^*, W^*(T_2) < 1$. In what follows, $\{Z(t), 0 \leq t \leq T\}$ denotes the Gaussian process in Theorem 3.1. Only the case of $\tilde{P}_{mn}$ is considered. Similar results hold for $\tilde{Q}_{mn}$.

Theorem 3.3.- Suppose that F, G, H', and H^* are continuous and F and G have the same support.

(i) If $P(x) \geq Q(x)$ for all x, then as $m \to \infty$, with $m/n \to \infty$, $\{\sqrt{m}(\tilde{P}_{mn} - P), 0 \leq t \leq T_1\}$ converges weakly to the process $\{Z(t), 0 \leq t \leq T_1\}$.

(ii) If $P(x_0) = Q(x_0)$ for some x_0, $P \neq Q$, and $m/n \to c$, $0 \leq c \leq \infty$, then the process, $\{\sqrt{m}(\tilde{P}_{mn} - P), 0 \leq t \leq T_1\}$ does not converge weakly as $m \to \infty$.

(iii) If $P(x) > Q(x)$ for all x, then as $m \to \infty$, the process $\{\sqrt{m}(\tilde{P}_{mn} - P), 0 \leq t \leq T_1\}$ converges weakly to $\{Z(t), 0 \leq t \leq T_1\}$.

Proof.- The proofs of (i) and (ii) are similar to those of (i) and (ii) in Theorem 2.2 by replacing the empirical distribution functions by the corresponding Kaplan-Meier estimators. The proof of (iii), although similar to that of (iii) in Theorem 2.2, requires some more detail and hence is given here. Note that because of (i) only the case of $m = O(n)$ is considered here. Consider the process $S_m(t) = m^{1/2}(\tilde{P}_{mn}(t) - P(t)), 0 \leq t \leq T_1$, and select, for arbitrary k, $0 = t_1 < t_2 \ldots < t_k = T_1$. Now consider the random vector with components $(S_m(t_1), \ldots, S_m(t_k))$ and $(Z'_m(t_1), \ldots, Z'_m(t_k))$ where Z'_m is as defined in Theorem 3.1. By a result of Földes and Rejtő (1981),

$$\sup_{0 \leq t \leq T_1} |P_m(t) - P(t)| = O(m^{-1/2}(\log m)^{1/2}) \tag{3.6}$$

and

$$\sup_{0 \leq t \leq T_2} |Q_n(t) - Q(t)| = O(n^{-1/2}(\log n)^{1/2}). \tag{3.7}$$

Since $m = O(n)$ and $S_m(t) = m^{1/2}(P_m(t) - P(t)) + \frac{n\sqrt{m}}{n+m} \max(0, Q_n(t) - P_m(t))$, and $Q(x) < P(x)$ for all x, it follows that almost surely, eventually,

$$(S_m(t_1), \ldots, S_m(t_k)) = (Z'_m(t_1), \ldots, Z'_m(t_k)).$$

Therefore, Theorem 3.1 implies that the finite-dimensional distributions of the process $\{S_m(t), 0 \leq t \leq T_1\}$ converge to those of the process $\{Z(t), 0 \leq t \leq T_1\}$.

It remains to show that $\{S_m\}$ is tight. For that purpose, for arbitrary $\varepsilon > 0$, consider for $\delta > 0$, note that, by using arguments similar to those used in the proof of Theorem

2.2

$$\begin{aligned}
\sup_{t\le s\le t+\delta} |S_m(s)-S_m(t)| &= \sup_{t\le s\le t+\delta} |\tilde{P}_{mn}(s)-\tilde{P}_{mn}(t)-P(s)+P(t)| \\
&\le \sup_{t\le s\le t+\delta} |Z_m^{'}(s)-Z'_m(t)| \\
&+ \frac{\sqrt{mn}}{m+n} \sup_{t\le s\le t+\delta} |\max(0, Q_n(s)-P_m(s)) \\
&- \max(0, Q_n(t)-P_m(t))|
\end{aligned} \tag{3.8}$$

where $Z_m^{'}(\cdot)=\sqrt{m}(P_m(\cdot)-P(\cdot))$.

Since $Z_m^{'}$ converges weakly, it follows that $Z_m^{'}$ is tight on $[t, t+\delta]$ for $t+\delta \le T_1$, and hence

$$\lim_{\delta\to 0}\overline{\lim_{m\to\infty}} Pr\{\sup_{t\le s\le t+\delta} |Z_m^{'}(s)-Z_m^{'}(t)|>\varepsilon\}=0.$$

Also since $m=O(n)$, it follows from the strong uniform convergence of Q_n and P_m to Q and P respectively, that the second term on the right side of the inequality (3.8) equals zero with probability one, and hence

$$\lim_{\delta\to 0}\overline{\lim_{m\to\infty}} \Pr\{\sup_{t\le s\le t+\delta} |S_m(s)-S_m(t)|>\varepsilon\}=0$$

and therefore $\{S_m\}$ is tight and hence converges weakly.

Consequently, under the conditions of (i) and (ii) in Theorem 3.3, since the weak limit of the process $\{\sqrt{m}(\tilde{P}_{mn}-P), 0\le t\le T_1\}$ is the same as the weak limit of the Kaplan-Meier process, the standard results of Hall and Wellner (1980) may be applied to construct confidence bands for the survival functions of interest. Note that, although Hall and Wellner (1980) conjecture that the confidence bands hold for all t up to the last failure time, recently Chen and Ying (1996), have provided a counterexample to the conjecture.

4. Testing the Hypothesis of Equality of Distributions Against the Alternative of Stochastic Ordering

For the purpose of testing the null hypothesis $H_0: F(x)=G(x)$ for all x against the alternative that $H_A: F(x)\le G(x)$ for all x, with $F(x)<G(x)$ for at least some x, it is of interest to study the asymptotic behavior of the estimator $\tilde{P}_{mn}$ defined by (2.2). Only the case of complete data will be considered, as a parallel argument will yield

similar results for the case of censored data by replacing the empirical distribution functions by their Kaplan-Meier counterparts.

Because of part (i) in Theorem 2.2, it follows that for the case $\frac{m}{n} \to \infty$, the process $\{\sqrt{m}(\tilde{P}_{mn} - P), -\infty < t < \infty\}$ converges weakly to the Gaussian process W_1 defined in Theorem 2.2. Therefore it is sufficient to consider the case when $\frac{m}{n} \to c$, $0 \leq c < \infty$. For this purpose, write, as before,

$$\sqrt{m}(\tilde{P}_{mn} - P) = \max(\sqrt{m}(P_m - P), \frac{m}{m+n}\sqrt{m}(P_m - P) + \frac{\sqrt{mn}}{m+n}\sqrt{n}(Q_n - Q)). \quad (4.1)$$

It follows immediately from (4.1), and the continuity of the map $(x)^+ = \max(0, x)$ with respect to the sup norm, that if $\frac{m}{n} \to 0$ then

$$\sqrt{m}(\tilde{P}_{mn} - P) \xrightarrow{\mathcal{D}} W_1^+, \quad (4.2)$$

where W_1 is the Gaussian process defined in Theorem 2.2. Thus, in these two cases, Kolmogorov-Smirnov-type test statistics may be constructed to test the null hypothesis. The asymptotic distributions of such statistics have been discussed, for example, in Shorack and Wellner (1986).

It remains to consider the case where $\frac{m}{n} \to c$, with $0 < c < \infty$. In this case, consider the sequence

$$\{\sqrt{m}(P_m(t) - P(t)), \sqrt{n}(Q_n(t) - Q(t)), -\infty < t < \infty\} \quad (4.3)$$

and note that the "component" processes $\{\sqrt{m}(P_m(t) - P(t)), -\infty < t < \infty\}$ and $\{\sqrt{n}(Q_n(t) - Q(t)), -\infty < t < \infty\}$ are independent, each converging weakly to the Gaussian processes W_1* and W_2 of Theorem 2.2 respectively. It follows from example 1.4.6 in van der Vart and Wellner (1996), that the sequence defined by (4.3) converges weakly to the process $(W_1(t), W_2(t), -\infty < t < \infty)$ where W_1 and W_2 are as in Theorem 2.2 and are also independent.

Now, for f and $g \in D[-\infty, \infty]$, the set of all right-continuous functions with left limits everywhere, the map $(f, g) \to \max(f, af + bg)$ is continuous with respect to the norm defined by $\max(\sup_{-\infty<x<\infty} |f(x)|, \sup_{-\infty<x<\infty} |g(x)|)$. Therefore, by the continuous mapping theorem for weak convergence,

$$\sqrt{m}(\tilde{P}_{mn} - P) \xrightarrow{\mathcal{D}} \max(W_1, \frac{c}{1+c}W_1 + \frac{\sqrt{c}}{1+\sqrt{c}}W_2). \quad (4.4)$$

5. Simulation Work

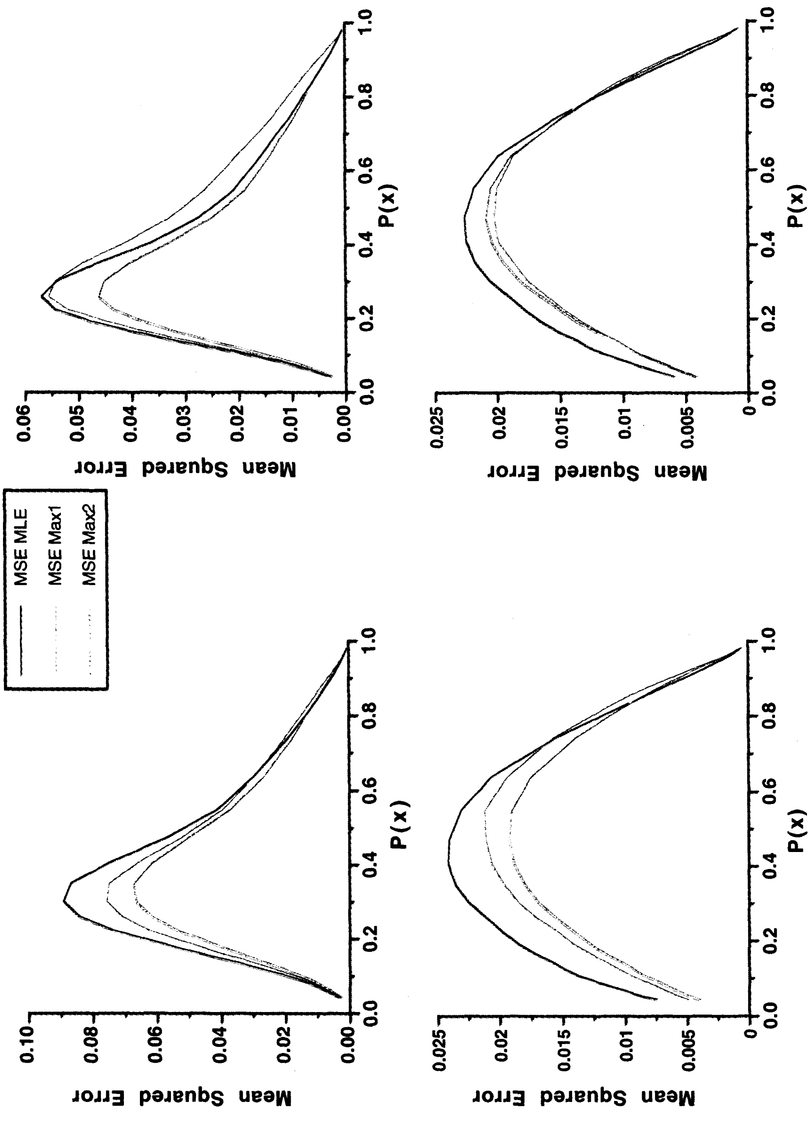
Monte Carlo simulations were performed to study the finite-sample properties of the estimators defined by (2.2) and (3.3). Figures 3.1-3.8 are representative of the results obtained from the simulations. Each simulation consisted of 10,000 replications. Figures 3.1-3.4 show the mean squared error functions for the case of censored and uncensored data, while figures 3.5-3.8 show the bias functions. The distributions and sample sizes used for the simulations are as follows:

	Underlying Distributions	Sample sizes	Censoring Distributions
Figures 3.1 and 3.5:	$F(x) = 1 - e^{-x}$	$m = 10$	$H'(x) = e^{-1.5x}$
	$G(x) = 1 - e^{-1.2x}$	$n = 10$	$H^*(x) = e^{-1.5x}$
Figures 3.2 and 3.6:	$F(x) = 1 - e^{-x}$	$m = 20$	$H'(x) = e^{-1.5x}$
	$G(x) = 1 - e^{-1.2x}$	$n = 10$	$H^*(x) = e^{-1.5x}$
Figures 3.3 and 3.7:	$F(x) = 1 - e^{-x}$	$m = 10$	No censoring
	$G(x) = 1 - e^{-1.2x}$	$n = 10$	No censoring
Figures 3.4 and 3.8:	$F(x) = 1 - e^{-x}$	$m = 20$	No censoring
	$G(x) = 1 - e^{-1.5x}$	$n = 20$	No censoring

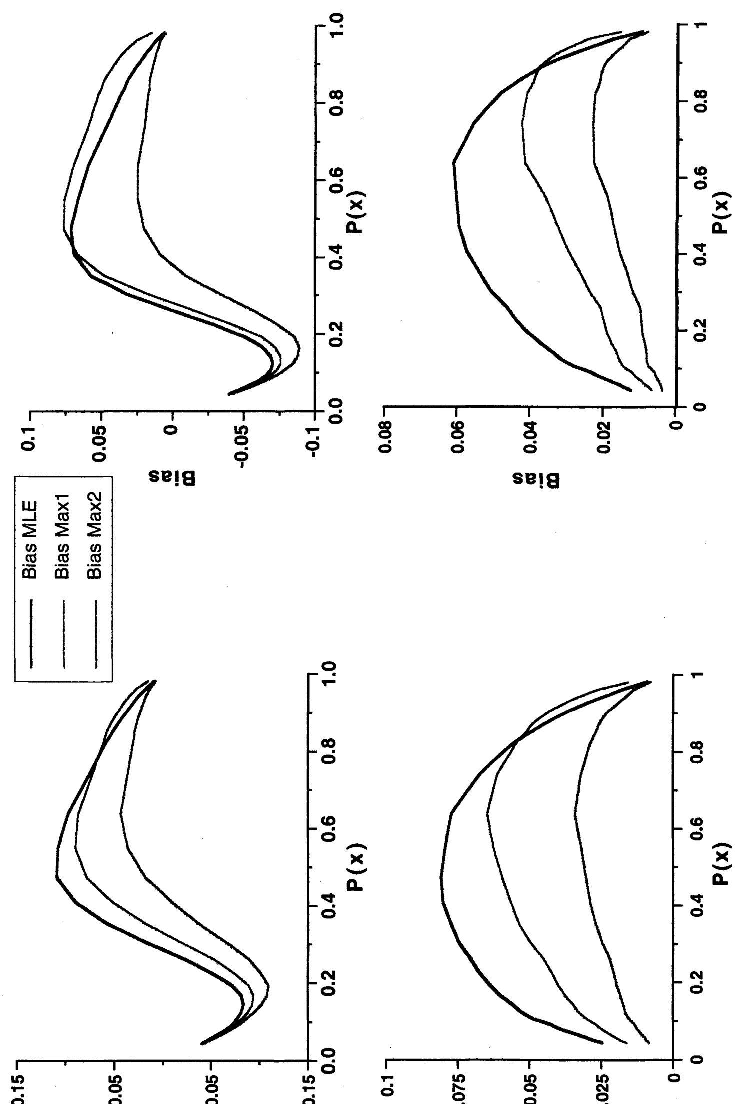
In the plots, the Nonparametric Maximum Likelihood estimator is denoted by NPMLE, Lo's estimator defined by $\hat{P}_{mn}$=max(P_m, Q_n) is denoted by Max1, and the estimator defined by (3.3) is denoted by Max2.

As can be seen from the representative results presented in Figures 3.1-3.8, the new estimators seem to behave better than the estimator $\max(P_m, Q_n)$ and the nonparametric maximum likelihood estimator in terms of the mean squared error, in both the censored and uncensored case. When dealing with bias, the estimator defined by (2.2) behaves better than the NPMLE and $\max(P_m, Q_n)$ in the uncensored case, and the estimator defined by (3.3) behaves well in terms of squared error and bias when compared to the NPMLE in the case of censored data.

Figures 3.1 - 3.4: Simulated Mean Squared Error Functions



Figures 3.5 - 3.8: Simulated Bias Functions



6. Acknowledgements

This work was partially supported by the National Institutes of Health under grant 1 G12 RR08124-04, and NSF through the MRCE program grant no. RII-8802973. Some of this work was conceived while the author was visiting the Statistics Department at the University of California at Berkeley, and Centro de Investigaciones Matemáticas in Guanajuato, México. I thank all institutions for their support.

References

[1] Arenaz, P., L. Bitticks, K. H. Panell and S. García (1992).- Genotoxic potential of crown ethers in mammalian cells; induction of sister chromatid exchanges, *Mutation Research,* 280, 109-115.

[2] Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product-limit estimators under random censorship, *Ann Statist.*, 2, 437-453.

[3] Chen, K., and Ying, Z. (1996). A counterexample to a conjeture concerning the Hall-Wellner band, *Ann Statist.,* Vol 24, 641-646.

[4] Csörgő and Horváth (1983). The rate of strong uniform consistency for the product-limit estimator, *Z. Wahrsch. verw. Gebiete,* 62, 411-426.

[5] Dykstra, R. L., (1982). Maximum likelihood estimation of the survival functions of two stochastically ordered random variables, *J. Amer. Statist. Assoc.,* 77, 621-628.

[6] Dykstra, R. L., Kochar, S., and Robertson, T. (1991). Statistical inference for uniform stochastic ordering in several populations, *Ann Statist.,* 19, 870-888.

[7] Dykstra, R. L. and Feltz, C. J. (1989). Nonparametric maximum likelihood estimation of survival functions with a general stochastic ordering and its dual, *Biometrika*, 76, 331-341.

[8] Embury, S. H., Elias, L., Heller, P. H., Hood, C. E., Greenberg, P. L., and Schrier, S. L. (1977). Remission maintenance therapy in acute myelogenous lukemia, *Western Journal of Medicine,* 126, 267-272.

[9] Földes, A. and Rejtő, L. (1981). A LIL type result for the product limit estimator, *Z. Wahrsch. verw. Gebiete,* 56, 75-86.

[10] Hall, W. J. and Wellner, J. (1980). Confidence bands for a survival curve from censored data, *Biometrika*, 67, 133-143.

[11] Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*, John Wiley & Sons, New York.

[12] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations, *J. Am. Statist. Assoc.*, 53, 457-481.

[13] Lehmann, E. L. (1955). Ordered families of distributions, *Ann Statist.,* 26, 399-419.

[14] Lehmann, E. L. and Rojo, J. (1992). Invariant directional orderings, *Ann Statist.*, 20, 2100-2110.

[15] Lo, S. H. (1987). Estimation of distribution functions under order restrictions, *Statistics and Decisions,* 5, 251-262.

[16] Ma, Z. (1991). Estimation of the survival functions of stochastically ordered random variables, Unpublished Master's Thesis, Mathematical Sciences Dept., Univ. of Texas, El Paso.

[17] Praestgaard, J. T. and Huang, J. (1996). Asymptotic Theory for Estimation of Survival Curves under Order Restrictions, *Ann. Statist.,* 24, 679-712.

[18] Puri, P. and Singh, H. (1992). Estimation of a distribution function dominating stochastically a known distribution function, *Austral. J. Statist.*, 34, 31-38.

[19] Rojo, J. (1995). On the weak convergence of certain estimators of stochastically ordered survival functions, *J. Nonparam. Statist.,* Vol 4, 349-363.

[20] Rojo, J. and Ma, Z. (1993). On the estimation of stochastically ordered survival functions, Technical Report, The Biostaistical Laboratory, Univ. of Texas, El Paso.

[21] Shibata, T. and Takeyama, T. (1977). Stochastic theory of pitting corrosion, *Corrosion,* 33, No. 7, 243.

[22] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with applications to Statistics,* John Wiley and Sons, New York.

[23] Stute, W. and Wang, J. L. (1993). The strong law under random censorship, *Ann. Statist.,* Vol. 21, 1591-1607.

[24] van der Vaart, A. W., and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes,* Springer, New York.

Javier Rojo
Department of Statistics, MS-138
Rice University
6100 Main Street
Houston, TX 77005

jrojo@stat.rice.edu