Estimation of two ordered mean residual life functions

Xiaomi Hu[a], Subhash C. Kochar[b,1], Hari Mukerjee[a,*], Francisco J. Samaniego[c]

[a]*Department of Mathematics and Statistics, Wichita State University, 1845 Fairmount, 67260-0033 Wichita, KS, USA*
[b]*Indian Statistical Institute, India*
[c]*University of California, Davis, USA*

Abstract

If X is a life distribution with finite mean then its mean residual life function (MRLF) is defined by $M(x)=E[X-x|X>x]$. It has been found to be a very intuitive way of describing the aging process. Suppose that M_1 and M_2 are two MRLFs, e.g., those corresponding to the control and the experimental groups in a clinical trial. It may be reasonable to assume that the remaining life expectancy for the experimental group is higher than that of the control group at all times in the future, i.e., $M_1(x) \leqslant M_2(x)$ for all x. Randomness of data will frequently show reversals of this order restriction in the empirical observations. In this paper we propose estimators of M_1 and M_2 subject to this order restriction. They are shown to be strongly uniformly consistent and asymptotically unbiased. We have also developed the weak convergence theory for these estimators. Simulations seem to indicate that, even when $M_1=M_2$, both of the restricted estimators improve on the empirical (unrestricted) estimators in terms of mean squared error, uniformly at all quantiles, and for a variety of distributions.

MSC: primary 62P10; 62G05; 62E10

Keywords: Mean residual life; Order restricted inference; Asymptotic theory

1. Introduction

The mean residual life (MRL) of a unit or a subject at age x is the average remaining life among those population members who have survived until time x. If lifelengths of the population are described by a random variable X with survival function (s.f.) $S(x)$ and a finite mean, then the MRL function (MRLF) is defined by

$$M(x) = E[X - x|X > x] = \frac{\int_x^\infty S(u)\,\mathrm{d}u}{S(x)} I[S(x) > 0]. \tag{1.1}$$

An MRLF is right continuous with left hand limits, and has the same set of discontinuities as the s.f., except that a MRLF is always continuous at the right endpoint of the support of the s.f., if finite. A distribution is characterized by its MRLF by the relation

$$S(x) = \frac{M(0)}{M(x)}\,\mathrm{e}^{-\int_0^x (1/M(u))\,\mathrm{d}u} I[M(x) > 0]. \tag{1.2}$$

In some cases, particularly in health sciences and actuarial sciences, the MRLF gives a more intuitive picture of survival or aging than the s.f. or the hazard rate function $r(x) = f(x)/S(x)$, where $f(x)$ is the density.

Let X and Y be random variables with finite means representing the lifetimes of two populations with s.f.s S_1 and S_2 and MRLFs M_1 and M_2, respectively. These could be patients undergoing two different treatments or the times to recurrence of cancer after the patients have been treated with different kinds of therapies. In the industrial engineering context, X and Y could represent the lifetimes of two different brands of an appliance. Suppose that we are confronted with the problem of comparing two populations to see which one has longer life. A naive approach would be to just compare the two means, i.e., $M_1(0)$ and $M_2(0)$. Rather than basing the decision on two single points, one could compare X and Y under a stochastic ordering (SO) restriction, i.e., $S_1(x) \leqslant (\geqslant) S_2(x)$ for all x. However, both of these measures compare the two systems when they are *new*. They do not say anything about their survival as time passes and the systems age. One way to do this would be to compare X and Y under a uniformly stochastic ordering (USO) restriction, i.e., under SO of the conditional distributions of X and Y given survival till time x. This is a very strong ordering restriction. A more meaningful and intuitive way of comparing X and Y would be to compare their MRLFs. The review article by Guess and Proschan (1988) gives a nice summary of the theory of MRLF.

There is a substantial literature on the nonparametric maximum likelihood estimators (NPMLEs) on two distributions under SO (Brunk et al., 1966; Huang and Praestgaard, 1996) and USO (Dykstra et al., 1991; Rojo and Samaniego, 1991), and on a projection type estimator for SO (Rojo and Ma, 1996; Rojo, 1995) and USO (Rojo and Samaniego, 1993; Mukerjee, 1996; Arcones and Samaniego, 2000), the latter often proving to be superior to the NPMLEs.

Yang (1978) studied the properties of an empirical estimator of the MRLF. Hall and Wellner (1979) and Csörgő and Zitikis (1996) have extended some of her results. In this paper we propose estimators of M_1 and M_2 subject to the constraint $M_1(x) \leqslant (\geqslant) M_2(x)$ for all x when M_2 is known (1-sample problem) and unknown (2-sample problem). The estimators could be extended to the case where the order restriction holds only on an interval $[t_1, t_2)$. These are simple intuitive projection type estimators, paralleled after estimators that have proven to be excellent in the stochastic and uniform stochastic ordering cases. Ebrahimi (1993) has also considered these problems with the restriction $M_1 \geqslant M_2$ on $[t_1, t_2]$ only, and provided an excellent real-life example in his Figs. 1 and 2 (p. 414). His estimators are similar to ours, but have to be slightly modified to assure that they are indeed MRLFs, and we show how this could be done. We provide a rigorous proof of asymptotic unbiasedness, since Ebrahimi's (1993) arguments regarding this property were largely heuristic. We also derive the weak convergence of our estimators that provide confidence bands for our estimators. We have conducted extensive simulations under a variety of conditions, and some of these results are presented in Section 4. As is to be expected, for small sample sizes the estimators are biased; however, the mean squared errors (MSEs) of both estimators appear to be uniformly smaller than those of the empiricals. The same has been observed for the SO and USO cases. These outcomes are intriguing and worthy of further study. We should also mention that Berger et al. (1988) have considered the problem of testing $M_1 \leqslant M_2$, but they do not consider the estimation problem.

In Section 2 we describe our estimators. In Section 3 we prove strong uniform consistency and asymptotic unbiasedness of our estimators. In Section 4 we provide some of our simulation results. In Section 5 we consider the asymptotic distributions and the weak convergence of our estimators, and provide formulas for simultaneous confidence intervals and confidence bands, and provide an example. In Section 6 we make some concluding remarks.

2. The estimators

Suppose that X and Y are nonnegative random variables (r.v.s) representing lifetimes with s.f.s S_1 and S_2, MRLFs M_1 and M_2, and right endpoints of their supports, if finite, b_1 and b_2, respectively. Assume that we have independent random samples of sizes n_1 and n_2 from S_1 and S_2, respectively. Let $\hat{S}_1$ and $\hat{S}_2$ denote the usual empirical estimators of the s.f.s, and define the empirical estimators of M_1 and M_2 (Yang, 1978) by

$$\hat{M}_i(x) = \frac{\int_x^\infty \hat{S}_i(u)\,\mathrm{d}u}{\hat{S}_i(x)} I(\hat{S}_i(x) > 0), \quad i = 1,2, \tag{2.1}$$

where the dependence on sample sizes has been suppressed to simplify the notation. Note that $\hat{M}_i$ is a right-continuous function with upward jumps only and a derivative equal to -1 wherever it exists.

2.1. The 1-sample case

Suppose that M_2 is known and $M_1(x) \leqslant M_2(x)\ \forall x$. Then our estimator of M_1 is given by

$$M_1^*(x) = \hat{M}_1(x) \wedge M_2(x). \tag{2.2}$$

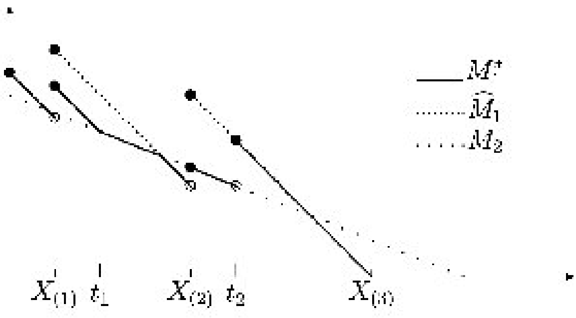

Fig. 1. Illustration of M_1^*, $\hat{M}_1$ and M_2.

Now suppose that the order restriction is $M_1(x) \leqslant M_2(x)$ on $[t_1, t_2)$ only. Since an MRLF cannot have a jump down, in fact, $M'(x) \geqslant -1$ where it exists for any MRLF, M, we have to be careful in defining $M_1^*(x)$ for $x < t_1$. We propose the estimator given by

$$M_1^*(x) = \begin{cases} \hat{M}_1(x), & x < t_1 \text{ and } \hat{M}_1(t_1) \leqslant M_2(t_1), \\ \hat{M}_1(x), & x < c \text{ and } \hat{M}_1(t_1) > M_2(t_1), \\ M_2(t_1) + (t_1 - x), & c \leqslant x < t_1 \text{ and } \hat{M}_1(t_1) > M_2(t_1), \\ \hat{M}_1(x) \wedge M_2(x), & t_1 \leqslant x < t_2, \\ \hat{M}_1(x), & x \geqslant t_2, \end{cases} \tag{2.3}$$

where $c = \max\{X_i : \hat{M}_1(X_i^-) \leqslant M_2(t_1) + (t_1 - X_i)\}$, and 0 if no such i exists. By this definition, if $\hat{M}_1(t_1) > M_2(t_1)$ then M_1^* is extended to the left by a straight line with a slope of -1 from $M_2(t_1)$ until this line is above $\hat{M}_1(X_i^-)$ for the first time for some i, or all the way to 0 if no such i exists (see Fig. 1).

For the reverse order restriction, the estimators are the obvious parallels to those in (2.2) and (2.3) with the reverse ordering, noting that, if $\hat{M}_1(t_2) < M_2(t_2)$, then

$$M_1^*(x) = \begin{cases} M_2(t_2) - (x - t_2), & t_2 \leqslant x < c, \\ \hat{M}_2(x), & x \geqslant c, \end{cases}$$

where $c = \min\{X_i: \hat{M}_1(X_i) \geqslant M_2(t_2) - (X_i - t_2)\}$, and $c = t_2 + M_2(t_2)$ if no such i exists. Note that $c = t_2 + M_2(t_2)$ implies $\hat{M}_1(c) = 0$.

To check if the various estimators are indeed MRLFs, we use the fact that (Hall and Wellner, 1981) a function M is a MRLF of a nondegenerate life distribution if and only if (i) $M : [0,\infty) \to [0,\infty)$, (ii) $M(0) > 0$, (iii) if for some $x_0 < \infty$, $M(x_0^-) = 0$, then $M(x) = 0\ \forall x \geqslant x_0$, and $\int_0^\infty 1/m(x)\,\mathrm{d}x = \infty$ if such an x_0 does not exist, (iv) $M(x) + x$ is nondecreasing in x, and (v) M is right continuous. The first three conditions are easily verified in all cases. Condition (iv) follows from the fact that for any two nondecreasing functions, f and g, $f \wedge g$ and $f \vee g$ are also nondecreasing, and from the way M_1^* was defined on $[t < t_1)$ when $M_1 \leqslant M_2$ and on $[t \geqslant t_2)$ when $M_1 \geqslant M_2$. Condition (v) is clearly satisfied on $[t_1, t_2)$ by the right-continuity of $\hat{M}_1$ and M_2. When $M_1 \geqslant M_2$, this condition is also satisfied on $[0, c)$ and on $[t_2, \infty)$ by right continuity of $\hat{M}_1$, and on $[c, t_1)$ by continuity of M_1^*. A similar argument applies when $M_1 \geqslant M_2$.

2.2. The 2-sample case

The motivation for this estimation stems from the NPMLEs for two stochastically ordered unknown s.f.s where one first estimates the common s.f. by pooling both samples, and then estimates each s.f. under the proper ordering restriction as a 1-sample problem, using the respective empirical s.f. and this common s.f. We give formulas only for the case where $M_1 \leqslant M_2$ everywhere; extension to the case of order restriction on an interval only can be done exactly as in the 1-sample case since our estimation procedure reduces to two separate 1-sample cases.

For any fixed but arbitrary n_1 and n_2, define $\hat{S} = (n_1\hat{S}_1 + n_2\hat{S}_2)/(n_1 + n_2)$ to be the estimator of the common s.f. $S = (n_1S_1 + n_2S_2)/(n_1 + n_2)$ by pooling both samples. The MRLF of S is given by

$$\begin{aligned} M(x) &= \frac{\int_x^\infty [n_1S_1(u) + n_2S_2(u)]\,\mathrm{d}u}{n_1S_1(x) + n_2S_2(x)} I[x < b_2] \\ &= \frac{n_1S_1(x)M_1(x) + n_2S_2(x)M_2(x)}{n_1S_1(x) + n_2S_2(x)} I[x < b_2] \\ &\equiv w_1(x)M_1(x) + w_2(x)M_2(x), \end{aligned} \tag{2.4}$$

where $b_1 \leqslant b_2$ are the right endpoints of the supports of S_1 and S_2, respectively, and

$$w_i(x) = \frac{n_iS_i(x)}{n_1S_1(x) + n_2S_2(x)} I[S_2(x) > 0], \quad i = 1, 2.$$

Note that for each fixed x, $M(x)$ is a convex combination of $M_1(x)$ and $M_2(x)$. Thus, $M_1 \leqslant M_2 \Rightarrow M_1 \leqslant M \leqslant M_2$. Substituting the empirical estimates in (2.4) using an obvious notation, we estimate M_1 and M_2 as two separate 1-sample problems as above, with the restrictions $M_1 \leqslant \hat{M}$ and $M_2 \geqslant \hat{M}$ as if $\hat{M}$ is known. These estimators are given by

$$\begin{aligned} M_1^*(x) &= \hat{M}_1(x) \wedge \hat{M}(x) \\ &= \hat{w}_1(x)\hat{M}_1(x) + \hat{w}_2(x)[\hat{M}_1(x) \wedge \hat{M}_2(x)] \\ &= \hat{M}_1(x) - \hat{w}_2(x)[\hat{M}_1(x) - \hat{M}_2(x)]I[\hat{M}_1(x) > \hat{M}_2(x)] \end{aligned} \tag{2.5}$$

and

$$\begin{aligned} M_2^*(x) &= \hat{M}_2(x) \vee \hat{M}(x) \\ &= \hat{w}_2(x)\hat{M}_2(x) + \hat{w}_1(x)[\hat{M}_1(x) \vee \hat{M}_2(x)] \\ &= \hat{M}_2(x) + \hat{w}_1(x)[\hat{M}_1(x) - \hat{M}_2(x)]I[\hat{M}_1(x) > \hat{M}_2(x)]. \end{aligned} \tag{2.6}$$

When the order restriction holds only on $[t_1, t_2)$, the estimators are exactly as in (2.3) and the material following, with $\hat{M}$ replacing M_2. Note that when $\hat{M}_1(x) > \hat{M}_2(x)$, the weights used to average them are proportional to the number of items alive after time x. The 1-sample estimator in (2.2) can be seen as the limit of (2.5) as $n_2 \to \infty$. Ebrahimi (1993) mentions the estimator in (2.6), and uses it for the example in his Figs. 1 and 2. However, for his problem of estimating M_2 only, he defines a different estimator that minimizes the MSE if certain functionals of the distributions are known. He uses the asymptotic distributions to compute and then estimate these functionals. It is not clear how large the sample sizes must be before reliable estimates could be obtained, especially in the right tail where the variance of $\hat{M}_i(x)$ is very large.

3. Consistency and asymptotic unbiasedness

Let $\|f\|_a^b$ denote $\sup_{a \leqslant x \leqslant b} |f(x)|$. Yang (1978) has shown that, under the first moment assumption only,

$$\|\hat{M}_i - M_i\|_0^b \to 0 \quad \text{a.s. for any } b < b_i, \quad i = 1, 2.$$

Now consider the last expression in (2.5) for M_1^* and let $0 < b < b_1$ be arbitrary. If $n_1 \to \infty$ and n_2 is finite then $\|\hat{w}_2\|_0^b \to 0$ a.s. If $n_1, n_2 \to \infty$, then

$$|\hat{M}_1(x) - \hat{M}_2(x)| \to 0 \quad \text{a.s. uniformly on } \{x \in [0, b]\colon M_1(x) = M_2(x)\}$$

and

$$I[\hat{M}_1(x) > \hat{M}_2(x)] \to 0 \quad \text{a.s. uniformly on } \{x \in [0, b]\colon M_1(x) < M_2(x)\}.$$

Thus $\|M_1^* - M_1\|_0^b \to 0$ a.s. The strong uniform consistency of M_2^* can be derived similarly.

3.1. Asymptotic unbiasedness

Under the first moment assumption only, Yang (1978) has shown that

$$E\hat{M}_i(x) = M_i(x)P[S_i(x) > 0], \quad i = 1, 2.$$

This shows that the unrestricted estimators are asymptotically unbiased. For the restricted estimators we need a stronger moment assumption in our proof.

Theorem 3.1. *Assume that X and Y have finite variances. Then M_1^* and M_2^* are asymptotically unbiased as $n_1, n_2 \to \infty$.*

Proof. We first note that

$$\hat{M}_1(x) = \frac{n_1 \int_x^\infty \hat{S}_1(u)\,\mathrm{d}u}{n_1 \hat{S}_1(x)} I[\hat{S}_1(x) > 0]$$

$$= \frac{\sum_j (X_j - x) I[X_j > x]}{\sum_j I[X_j > x]} I\left[\sum_j I[X_j > x] > 0\right]$$

with a similar expression for $\hat{M}_2(x)$. Using the last expression in (2.5) for $M_1^*(x)$, its asymptotic unbiasedness will be proven if we can show that the expected value of

$$\Delta \equiv \hat{w}_2(x)[\hat{M}_1(x) - \hat{M}_2(x)] I[\hat{M}_1(x) > \hat{M}_2(x)]$$

converges to 0 as $n_1, n_2 \to \infty$. Fix x. To simplify the notation we write $L_i = n_i \hat{S}_i(x)$ for $i = 1, 2$. As pointed out by Yang (1978), given $L_1 = l_1, \hat{M}_1$ has the distribution of the average of l_1 i.i.d. random variables, $U_1, U_2, \ldots, U_{l_1}$, say, with s.f. $S_1(u)/S_1(x)$ for $u \geqslant x$, $EU_i = M_1(x)$, and

$$\mathrm{Var}(U_i) = \mathrm{Var}[X - x | X > x] \equiv \sigma_1^2(x).$$

Similarly, given $L_2 = l_2$, $\hat{M}_2$ has the distribution of the average of l_2 i.i.d. random variables, $V_1, V_2, \ldots, V_{l_2}$, say, with s.f. $S_2(u)/S_2(x)$ for $u \geqslant x$, $EV_i = M_2(x)$, and

$$\mathrm{Var}(V_i) = \mathrm{Var}[Y - x | Y > x] \equiv \sigma_2^2(x).$$

The expected value of Δ is given by

$$E\Delta = E[E[\Delta | L_1, L_2]]$$

$$= \sum_{l_1} \sum_{l_2} \frac{l_2 I[l_1 + l_2 > 0]}{l_1 + l_2} E[(\bar{U}_{l_1} - \bar{V}_{l_2}) I(\bar{U}_{l_1} - \bar{V}_{l_2} > 0)]$$

$$\times P[L_1 = l_1] P[L_2 = l_2],$$

where $\bar{U}_{l_1}$ and $\bar{V}_{l_2}$ denote the averages. Note that

$$\sum_{l_1} \sum_{l_2} \frac{l_2 I[l_1 + l_2 > 0]}{l_1 + l_2} P[L_1 = l_1] P[L_2 = l_2] = E\hat{w}_2(x).$$

Now $E[\bar{U}_{l_1} - \bar{V}_{l_2}] = M_1(x) - M_2(x) \leqslant 0$ and $\mathrm{Var}[\bar{U}_{l_1} - \bar{V}_{l_2}] = \sigma_1^2(x)/l_1 + \sigma_2^2(x)/l_2 = \sigma_{l_1 l_2}$, say, and $\sigma_{l_1 l_2} \to 0$ if $l_1, l_2 \to \infty$. Let $Z = (\bar{U}_{l_1} - \bar{V}_{l_2})/\sigma_{l_1 l_2}$. Then $EZ = (M_1(x) - M_2(x))/\sigma_{l_1 l_2}$ and $\mathrm{Var}\, Z = 1$. Thus

$$E[(\bar{U}_{l_1} - \bar{V}_{l_2})I(\bar{U}_{l_1} - \bar{V}_{l_2} > 0)] = \sigma_{l_1 l_2} E[ZI(Z > 0)] \to 0 \text{ if } l_1, l_2 \to \infty.$$

Since $\sum_{l_1 \leqslant m_1} \sum_{l_2 \leqslant m_2} P[L_1 = l_1]P[L_2 = l_2]$ can be made arbitrarily small for any given m_1 and m_2 by choosing n_1 and n_2 large enough, we see that $E\Delta \to 0$ as $n_1, n_2 \to \infty$. This concludes the proof of the theorem. □

Remarks. The second moment assumption in the theorem could be relaxed to that of a moment of order $1 + \delta$ by using truncation arguments. We do not know if the first moment assumption is sufficient.

4. Simulations

We have carried out a quite extensive simulation using the following decreasing, constant, and increasing MRLFs:

$$M_i(x) = a_i(1 - x/b_i)I[x \leqslant b_i],\ b_i > a_i, \text{ with } S_i(x) = (1 - x/b_i)^{b_i/a_i - 1},$$

which corresponds to the $U(0,1)$ distribution when $a_i = 0.5$ and $b_i = 1$;

$$M_i(x) = \theta_i, \text{ corresponding to the } \mathrm{Exp}(\theta_i) \text{ distribution; and}$$

$$M_i(x) = a_i x + b_i,\ a_i,\ b_i > 0, \text{ with } S_i(x) = (a_i/b_i)(1 + a_i x/b_i)^{1/a_i - 1},\ x \geqslant 0.$$

We were particularly interested in comparing the estimated bias and MSE of the various estimators, especially when $M_1 = M_2$. Typically the sample sizes chosen were small to moderate, but we also chose some very large sample sizes to check on the asymptotic unbiasedness. Some of these results are presented in Tables 1 and 2.

In Table 2 we consider the exponential distribution only. Two of these simulations use small, unequal sample sizes and different means. The other simulation uses equal means but with very large sample sizes.

It is clear that in all cases the restricted estimators have more bias, as is to be expected, but the MSE is uniformly smaller than the unrestricted empirical. For the exponential case, the bias is seen to be steadily decreasing with the sample size according to Theorem 3.1.

5. Asymptotic distributions and weak convergence

In this section we consider the joint asymptotic distribution of $M_1^* - M_1$ and $M_2^* - M_2$ at a point x, and also their weak convergence on $[0, b]$ for $b < b_1 \wedge b_2 = b_1$. We also construct simultaneous confidence intervals and confidence bands for M_1 and M_2 using the asymptotic distributions.

Table 1
Comparison of bias (B) and MSE of $\hat{M}_1$, M_1^*, $\hat{M}_2$, and M_2^* at various q-quantiles

q	$B(\hat{M}_1(x))$	$B(M_1^*(x))$	$\frac{\mathrm{MSE}(\hat{M}_1(x))}{\mathrm{MSE}(M_1^*(x))}$	$B(\hat{M}_2(x))$	$B(M_2^*(x))$	$\frac{\mathrm{MSE}(\hat{M}_2(x))}{\mathrm{MSE}(M_2^*(x))}$
$M_1(x)=M_2(x)=1$, $n_1=10$, $n_2=10$, #*iterations* $=10,000$						
0.1	−0.0056	−0.0990	1.5175	−0.0028	+0.0906	1.1959
0.2	−0.0061	−0.1050	1.5200	−0.0063	+0.0926	1.2013
0.5	−0.0026	−0.1224	1.6261	−0.0120	+0.1073	1.1921
0.8	−0.0067	−0.1486	1.7688	−0.0174	+0.1252	1.1751
0.9	−0.0135	−0.1589	1.7293	−0.0201	+0.1312	1.1701
$M_1(x)=M_2(x)=1$, $n_1=20$, $n_2=20$, #*iterations* $=10,000$						
0.1	−0.0010	−0.0673	1.4326	−0.0049	+0.0613	1.2583
0.2	−0.0030	−0.0728	1.4502	−0.0058	+0.0638	1.2546
0.5	−0.0040	−0.0862	1.4983	−0.0074	+0.0746	1.2420
0.8	−0.0072	−0.1015	1.5216	−0.0069	+0.0882	1.2233
0.9	−0.0071	−0.1061	1.5255	−0.0078	+0.0927	1.2123
$M_1(x)=M_2(x)=0.5x+1$, $n_1=10$, $n_2=10$, #*iterations* $=10,000$						
0.1	−0.0195	−0.1571	1.7857	−0.0152	+0.1224	1.0744
0.2	−0.0245	−0.1766	1.8031	−0.0169	+0.1364	1.0619
0.5	−0.0323	−0.2402	2.0108	−0.0134	+0.1947	1.0362
0.8	−0.0471	−0.3157	2.2385	−0.0402	+0.2370	1.0182
0.9	−0.0616	−0.3469	2.2215	−0.0516	+0.2547	1.0120
$M_1(x)=M_2(x)=0.5x+1$, $n_1=20$, $n_2=20$, #*iterations* $=10,000$						
0.1	−0.0151	−0.1185	1.6247	−0.0233	+0.0797	1.1397
0.2	−0.0196	−0.1350	1.6448	−0.0223	+0.0863	1.1339
0.5	−0.0205	−0.1785	1.7388	−0.0367	+0.1200	1.1059
0.8	−0.0210	−0.2303	1.8875	−0.0447	+0.1612	1.0762
0.9	−0.0267	−0.2528	1.9287	−0.0508	+0.1727	1.0717
$M_1(x)=M_2(x)=\frac{1-x}{2}$, $n_1=10$, $n_2=10$, #*iterations* $=10,000$						
0.1	−0.0008	−0.0251	1.3199	−0.0003	+0.0239	1.3358
0.2	−0.0014	−0.0244	1.3197	−0.0004	+0.0225	1.3392
0.5	−0.0016	−0.0209	1.3555	−0.0005	+0.0186	1.3645
0.8	−0.0112	−0.0210	1.1825	−0.0101	+0.0091	1.7010
0.9	−0.0179	−0.0208	1.0589	−0.0170	−0.0030	1.5827
$M_1(x)=M_2(x)=\frac{1-x}{2}$, $n_1=20$, $n_2=20$, #*iterations* $=10,000$						
0.1	−0.0006	−0.0177	1.3311	+0.0001	+0.0173	1.3313
0.2	−0.0003	−0.0167	1.3388	−0.0005	+0.0160	1.3529
0.5	−0.0004	−0.0136	1.3642	−0.0001	+0.0131	1.3548
0.8	−0.0010	−0.0099	1.3505	−0.0009	+0.0091	1.4810
0.9	−0.0061	−0.0108	1.1646	−0.0059	+0.0044	1.7665

Throughout this section we assume that X and Y have finite variances. We also assume that S_1 and S_2, and hence M_1 and M_2 have common discontinuities on the intersection of their supports. The latter assumption is automatically satisfied if the s.f.s are continuous, the appropriate model for life distributions. However, sampling

Table 2
Comparison of the bias and MSE of $\hat{M}_1$, M_1^*, $\hat{M}_2$, and M_2^* for the exponential case at various q-quantiles, ξ_q, of M_1,—two with different means and (small) sample sizes, and one with equal means and sample sizes equal to 5000

q	$B(\hat{M}_1(x))$	$B(M_1^*(x))$	$\frac{\text{MSE}(\hat{M}_1(x))}{\text{MSE}(M_1^*(x))}$	$B(\hat{M}_2(x))$	$B(M_2^*(x))$	$\frac{\text{MSE}(\hat{M}_2(x))}{\text{MSE}(M_2^*(x))}$
$M_1(x)=1,\ M_2(x)=1.1,\ n_1=7,\ n_2=10,\ \#iterations=10,000$						
0.1	−0.0129	−0.1076	1.5706	−0.0046	+0.0608	1.1676
0.2	−0.0162	−0.1175	1.5881	−0.0060	+0.0631	1.1673
0.5	−0.0145	−0.1466	1.7748	−0.0137	+0.0712	1.1666
0.8	−0.0322	−0.1890	1.8578	−0.0166	+0.0849	1.1448
0.9	−0.0422	−0.2080	1.8781	−0.0206	+0.0879	1.1392
$M_1(x)=1,\ M_2(x)=1.1,\ n_1=10,\ n_2=7,\ \#iterations=10,000$						
0.1	−0.0023	−0.0690	1.3563	+0.0006	+0.0953	1.2540
0.2	−0.0029	−0.0752	1.3883	+0.0001	+0.1016	1.2447
0.5	−0.0010	−0.0922	1.4726	−0.0050	+0.1207	1.2302
0.8	−0.0058	−0.1183	1.5437	−0.0166	+0.1409	1.2251
0.9	−0.0074	−0.1276	1.5668	−0.0257	+0.1449	1.2266
$M_1(x)=M_2(x)=1,\ n_1=5000,\ n_2=5000,\ \#iterations=5000$						
0.1	−0.0036	−0.0077	1.1285	−0.0034	+0.0007	1.5613
0.2	−0.0050	−0.0094	1.0698	−0.0049	−0.0006	1.6311
0.5	−0.0067	−0.0117	1.0370	−0.0066	−0.0015	1.6543
0.8	−0.0072	−0.0131	1.0515	−0.0071	−0.0013	1.6252
0.9	−0.0065	−0.0127	1.0839	−0.0065	−0.0004	1.6008

Here $B(\cdot)=\text{Bias}(\cdot)$.

schemes might render them discrete. The assumption is to assure that the continuous mapping theorem (Billingsley, 1968) applies to some functions of the MRL processes defined below.

We first review the weak convergence of the unrestricted estimators. Let

$$Z_{in_i}=\sqrt{n_i}(\hat{M}_i-M_i)\text{ on }[0,b_i)\text{ for }i=1,2$$

denote the two independent MRL processes. Yang (1978) showed that Z_{in_i}, when composed with S_i^{-1} (the left-continuous inverse of S_i) converges weakly to a Gaussian process on $[0,d]$ for any $d<1$ under the assumption that S_i has a density. Hall and Wellner (1979) pointed out that the density assumption is unnecessary if we consider the convergence on the domain of S_i, and that the convergence on all intervals of the form $[0,b]$, $b<b_i$ implies convergence on $[0,b_i)$. Thus

$$Z_{in_i}\overset{w}{\Rightarrow}Z_i\text{ on }[0,b_i),\tag{5.1}$$

where Z_i is a mean-zero Gaussian process with

$$\text{Cov}[Z_i(x),Z_i(y)]=\frac{\sigma_i^2(y)}{S_i(x)}\text{ for }0\leqslant x\leqslant y<b_i.\tag{5.2}$$

This weak convergence is in $D[0, b_i)$ endowed with the Skorohod topology. By using stronger, weighted metrics Hall and Wellner (1979) and Csörgő and Zitikis (1996) have extended the above results using stronger assumptions on S_i. Since the weights do not apply universally to all s.f.s, we have not considered these generalizations. Our results for the restricted estimators are based on (5.1) and (5.2). Let

$$Z^*_{in_i} = \sqrt{n_i}(M^*_i(x) - M_i(x)), \quad i = 1, 2.$$

In the proofs below we frequently use the continuous mapping theorem (Billingsley, 1968), which we mention, and Slutsky's theorem, which we do not mention.

5.1. The 1-sample case

Consider the estimator M^*_1 in (2.2) when M_2 is known. By (5.1) and (5.2),

$$\begin{aligned} Z^*_{1n_1}(x) &= \sqrt{n_1}(M^*_1(x) - M_1(x)) \\ &= \sqrt{n_1}(\hat{M}_1(x) \wedge M_2(x) - M_1(x)) \\ &= \sqrt{n_1}[\hat{M}_1(x) - M_1(x)] \wedge \sqrt{n_1}[M_2(x) - M_1(x)] \\ &\xrightarrow{\mathrm{d}} Z^*_1(x), \end{aligned} \tag{5.3}$$

where $Z^*_1(x) \stackrel{\mathrm{d}}{=} Z_1(x)$, if $M_1(x) < M_2(x)$, and $Z^*_1(x) \stackrel{\mathrm{d}}{=} [Z_1(x)] \wedge 0$, with a point mass of $\frac{1}{2}$ at 0, if $M_1(x) = M_2(x)$.

For the weak convergence of M^*_1 we have the following theorem.

Theorem 5.1. *Let $b < b_1$ be fixed. Consider the estimator M^*_1 in (2.2) when M_2 is known.*

(i) *If $M_1 < M_2$ on $[0, b]$, then*

$$Z^*_{1n_1} \stackrel{\mathrm{w}}{\Rightarrow} Z_1 \text{ on } [0, b]. \tag{5.4}$$

(ii) *If $M_1(x_0) = M_2(x_0)$ for some $x_0 \in (0, b)$ and $M_1 < M_2$ on $(x_0, s_0]$, $s_0 < b$, or on $[s_0, x_0)$, $s_0 > 0$, then $Z^*_{1n_1}$ does not converge weakly.*

(iii) *If $M_1 = M_2$ on $[0, b]$, then*

$$Z^*_{1n_1} \stackrel{\mathrm{w}}{\Rightarrow} Z_1 \wedge 0 \text{ on } [0, b]. \tag{5.5}$$

Proof. (i) Since $\sqrt{n_1}(M_2 - M_1) \to \infty$ uniformly on $[0, b]$ by our assumption, the result follows from the third expression on the r.h.s. of (5.3) for $Z^*_{1n_1}$, (5.1), and the continuous mapping theorem.

(ii) This proof is similar to that of Rojo (1995, Theorem 2.1). We assume that $M_1 < M_2$ on $(x_0, s_0]$, $s_0 < b$; the proof of the other case is similar. We first note that under our assumptions,

$$M_1^*(x_0) - M_1(x_0) = [\hat{M}_1(x_0) - M_1(x_0)] \wedge 0 \text{ from (5.3)}.$$

Now, for any $x_0 < t \leqslant s_0$ and $\varepsilon > 0$, we have
(1) $\hat{M}_1(t) < M_2(t)$ eventually w.p.1, and
(2) on this event, $M_1^*(t) = \hat{M}_1(t)$. Thus,

$$\begin{aligned}
&\lim_{n_1\to\infty} P\{|Z_{1n_1}^*(t) - Z_{1n_1}^*(x_0)| \geqslant \varepsilon\} \\
&\quad = \lim_{n_1\to\infty} P\{\sqrt{n_1}|[\hat{M}_1(t) - M_1(t)] - [\hat{M}_1(x_0) - M_1(x_0)] \wedge 0| \geqslant \varepsilon\} \\
&\quad \geqslant \lim_{n_1\to\infty} P\{\sqrt{n_1}[\hat{M}_1(t) - M_1(t)] - [\hat{M}_1(x_0) - M_1(x_0)] \wedge 0 \geqslant \varepsilon\} \\
&\quad \geqslant \lim_{n_1\to\infty} P\{\sqrt{n_1}[\hat{M}_1(t) - M_1(t)] \geqslant \varepsilon\} = 1 - \Phi(\varepsilon\sqrt{S_1(t)}/\sigma_1(t))
\end{aligned}$$

from (5.1), where Φ is the standard normal c.d.f. This can be easily shown to violate the necessary tightness condition for weak convergence in Theorem 15.2 in Billingsley (1968).

(iii) When $M_1 = M_2$ on $[0, b]$, we have $Z_{1n_1}^* = \sqrt{n_1}[(\hat{M}_1 - M_1) \wedge 0]$ on $[0, b]$, and the result follows from (5.1) and the continuous mapping theorem. □

5.2. The 2-sample case

First we consider the asymptotic distribution of the vector $(Z_{1n_1}^*(x), Z_{2n_2}^*(x))'$ for a fixed $0 < x < b_1 \wedge b_2 = b_1$ with M_1^* and M_2^* as defined in (2.5) and (2.6), noting that $(Z_{1n_1}(x), Z_{2n_2}(x))' \xrightarrow{d} (Z_1(x), Z_2(x))'$ with independent components. We write

$$\begin{aligned}
Z_{1n_1}^*(x) &= \sqrt{n_1}(M_1^*(x) - M_1(x)) \\
&= \sqrt{n_1}\{\hat{w}_1(x)(\hat{M}_1(x) - M_1(x)) + \hat{w}_2(x)[(\hat{M}_1(x) - M_1(x)) \\
&\qquad \wedge [(\hat{M}_2(x) - M_2(x)) + (M_2(x) - M_1(x))]]\} \\
&= Z_{1n_1}(x) + \hat{w}_2(x)[0 \wedge (\sqrt{n_1/n_2}\, Z_{2n_2}(x) - Z_{1n_1}(x) \\
&\qquad + \sqrt{n_1}(M_2(x) - M_1(x))] \qquad (5.6)
\end{aligned}$$

and

$$
\begin{aligned}
Z^*_{2n_2}(x) &= \sqrt{n_2}(M^*_2(x) - M_2(x)) \\
&= \sqrt{n_2}\{\hat{w}_2(x)(\hat{M}_2(x) - M_2(x)) + \hat{w}_1(x)[(\hat{M}_2(x) - M_2(x)) \\
&\quad \vee [(\hat{M}_1(x) - M_1(x)) - (M_2(x) - M_1(x))]]\} \\
&= Z_{2n_2}(x) + \hat{w}_1(x)[0 \vee (\sqrt{n_2/n_1}\, Z_{1n_1}(x) - Z_{2n_2}(x) \\
&\quad - \sqrt{n_2}(M_2(x) - M_1(x))],)].
\end{aligned} \tag{5.7}
$$

If $n_1 \to \infty$ and $n_2 < \infty$, then $\hat{w}_1(x) \to 1$ and $\sqrt{n_1}\hat{w}_2(x) \to 0$ a.s., $Z^*_{1n_1}(x) \xrightarrow{d} Z_1(x)$, and $Z^*_{2n_2}(x)$ has no limiting distribution. The same is true if the subscripts 1 and 2 are switched.

If $n_1 \to \infty, n_2 \to \infty$, and $n_2/n_1 \to 0$, then $\hat{w}_1(x) \to 1$ and $\sqrt{n_1}\hat{w}_2(x) \to 0$ a.s. Again $Z^*_{1n_1}(x) \xrightarrow{d} Z_1(x)$. Now $\sqrt{n_2}(\hat{M}_1(x) - M_1(x)) \xrightarrow{p} 0$ and $\sqrt{n_2}(M_2(x) - M_1(x)) = 0$ or converges to ∞ depending on whether $M_2(x)$ is equal to or more than $M_1(x)$. Thus

$$
Z^*_{2n_2}(x) \xrightarrow{d} \begin{cases} Z_2(x), & \text{if } M_1(x) < M_2(x), \\ Z_2(x) \vee 0, & \text{if } M_1(x) = M_2(x) \end{cases}
$$

and $Z^*_{1n_1}(x)$ and $Z^*_{2n_2}(x)$ are asymptotically independent. Similarly it can be seen that, when $n_1 \to \infty, n_2 \to \infty$, and $n_2/n_1 \to \infty$, $Z^*_{2n_2} \xrightarrow{d} Z_2(x)$ and

$$
Z^*_{1n_1}(x) \xrightarrow{d} \begin{cases} Z_1(x), & \text{if } M_1(x) < M_2(x), \\ Z_1(x) \wedge 0, & \text{if } M_1(x) = M_2(x) \end{cases}
$$

and the two are asymptotically independent.

If $n_1 \to \infty, n_2 \to \infty$, and $n_2/n_1 \to \alpha$ with $0 < \alpha < \infty$, then $\hat{w}_1(x) \to w_1(\alpha, x)$ and $\hat{w}_2(x) \to w_2(\alpha, x)$ a.s., where

$$
w_1(\alpha, x) = \frac{S_1(x)}{S_1(x) + \alpha S_2(x)} \quad \text{and} \quad w_2(\alpha, x) = \frac{\alpha S_2(x)}{S_1(x) + \alpha S_2(x)}. \tag{5.8}
$$

Since $(Z_{1n_1}(x), Z_{2n_2}(x))' \xrightarrow{d} (Z_1(x), Z_2(x))'$ and $\{w_i(x)\}$ converge to constants a.s., by the continuous mapping theorem we have

$$
(Z^*_{1n_1}(x), Z^*_{2n_2}(x))' \xrightarrow{d} (W_1(x), W_2(x))'
$$

with the limiting distributions given as follows. If $M_1(x) < M_2(x)$, then $(W_1(x), W_2(x))' \stackrel{d}{=} (Z_1(x), Z_2(x))'$ with independent components. If $M_1(x) = M_2(x)$, then the marginal distributions of $(W_1(x), W_2(x))'$ are given by

$$W_1(x) \stackrel{d}{=} w_1(\alpha, x)Z_1(x) + w_2(\alpha, x)[Z_1(x) \wedge (Z_2(x)/\sqrt{\alpha})]$$

$$= Z_1(x) + w_2(\alpha, x)[0 \wedge (Z_2(x)/\sqrt{\alpha} - Z_1(x))]$$

and

$$W_2(x) \stackrel{d}{=} w_2(\alpha, x)Z_2(x) + w_1(\alpha, x)[Z_2(x) \vee (\sqrt{\alpha}Z_1(x))]$$

$$= Z_2(x) + w_1(\alpha, x)[0 \vee (\sqrt{\alpha}Z_1(x) - Z_2(x))].$$

Since $\{Z_i(x)\}$ are independent mean-zero normals with variances $\{\sigma_i^2(x)/S_i(x)\}$, it is easy to compute the means of $\{W_i(x)\}$. These are given by

$$E[W_1(x)] = -\frac{w_2(\alpha, x)}{\sqrt{2\pi}}\sqrt{\frac{\sigma_1^2(x)}{S_1(x)} + \frac{\sigma_2^2(x)}{\alpha S_2(x)}}$$

and

$$E[W_2(x)] = \frac{w_1(\alpha, x)}{\sqrt{2\pi}}\sqrt{\frac{\alpha\sigma_1^2(x)}{S_1(x)} + \frac{\sigma_2^2(x)}{S_2(x)}}.$$

The covariances do not have closed form expressions, but they can be computed in a straightforward manner numerically. For example, writing τ_i^2 for the variance of $Z_i(x)$, so that $(Z_1(x), Z_2(x))' \stackrel{d}{=} (\tau_1 U, \tau_2 V)'$, where U and V are i.i.d. standard normals, $E[Z_1(x)(0 \vee (\sqrt{\alpha}Z_1(x) - Z_2(x)))]$ may be written as

$$\alpha\tau_1^2 \int\int u(u - (\tau_2/\tau_1\sqrt{\alpha})v) I(u > (\tau_2/\tau_1\sqrt{\alpha})v)\, \phi(u)\, \phi(v)\, du\, dv,$$

where ϕ is the density of the standard normal. It may be noted that in the limit as $\alpha \to 0$ or ∞, $(W_1(x), W_2(x))'$ has the same distributions as derived above.

We now consider the weak convergence of the bivariate process $(Z_{1n_1}^*, Z_{2n_2}^*)'$, as given by (5.6) and (5.7), on $[0, b] \times [0, b]$, with $b < b_1$.

Theorem 5.2. *Assume that* $n_1 \to \infty, n_2 \to \infty$, *and* $n_2/n_1 \to \alpha$ *with* $0 < \alpha < \infty$.

(i) *If* $M_1 < M_2$ *on* $[0, b]$, *then*

$$(Z_{1n_1}^*, Z_{2n_2}^*)' \stackrel{w}{\Rightarrow} (Z_1, Z_2)' \text{ with independent components on } [0, b] \times [0, b].$$

(ii) *If* $M_1(x_0) = M_2(x_0)$ *for some* $x_0 \in (0, b)$ *and* $M_1 < M_2$ on $(x_0, s_0]$, $s_0 < b$, *or on* $[s_0, x_0)$, $s_0 > 0$, *then* $(Z_{1n_1}^*, Z_{2n_2}^*)'$ *does not converge weakly.*

(iii) *If* $M_1 = M_2$ *on* $[0,b]$, *then*

$$(Z^*_{1n_1}, Z^*_{2n_2})' \stackrel{w}{\Rightarrow} (W_1, W_2)' \text{ on } [0,b]\times[0,b], \text{ where}$$

$$W_1 \stackrel{d}{=} \frac{1}{1+\alpha} Z_1 + \frac{\alpha}{1+\alpha}[Z_1 \wedge (Z_2/\sqrt{\alpha})] \text{ and}$$

$$W_2 \stackrel{d}{=} \frac{\alpha}{1+\alpha} Z_2 + \frac{1}{1+\alpha}[Z_2 \vee (\sqrt{\alpha} Z_1].$$

Proof. We first note that, since $\hat{S}_1$ and $\hat{S}_2$ are strongly uniformly convergent, the bivariate weight process

$$(\hat{w}_1(\cdot), \hat{w}_2(\cdot))' \to (w_1(\alpha,\cdot) w_2(\alpha,\cdot))' \text{ a.s., uniformly on } [0,b]\times[0,b].$$

(i) Since Z_{1n_1} and Z_{2n_2} are independent, (5.1) implies that

$$(Z_{1n_1}, Z_{2n_2})' \stackrel{w}{\Rightarrow} (Z_1, Z_2)' \text{ on } [0,b]\times[0,b], \tag{5.9}$$

where the convergence is in $D[0,b]\times D[0,b]$ in the product Skorohod topology. Using this and the fact that $\sqrt{n_i}[M_2 - M_1] \to \infty$, uniformly on $[0,b]$, the conclusion follows from the definitions in (5.6) and (5.7) and the continuous mapping theorem.

(ii) We assume that $M_1 < M_2$ on $(x_0, s_0]$, $s_0 < b$; the proof of the other case is similar. Note that

$$Z^*_{1n_1}(x_0) = Z_{1n_1}(x_0) + \hat{w}_2(x_0)[0 \wedge (\sqrt{n_1/n_2} Z_{2n_2}(x_0) - Z_{1n_1}(x_0))]$$

$$\equiv Z_{1n_1}(x_0) + U(x_0),$$

where $U(x_0)$ has the limiting distribution of

$$0 \wedge w_2(\alpha, x_0)[Z_2(x_0)/\sqrt{\alpha} - Z_1(x_0)] \stackrel{d}{=} 0 \wedge V,$$

where $V \sim N(0,\tau^2)$ with $\tau^2 = w_2^2(\alpha, x_0)[\sigma_2^2(x_0)/\alpha S_2(x_0) + \sigma_1^2(x_0)/S_1(x_0)]$. Also, note that for any $x_0 < t \leqslant s_0$, and $\eta > 0$,

$$P[Z^*_{1n_1}(t) \neq Z_{1n_1}(t)] \leqslant \eta \text{ for all sufficiently large } n_1 \text{ and } n_2. \tag{5.10}$$

Suppose that $\{Z^*_{1n_1}\}$ is tight. Then, for each $\varepsilon > 0$ and $\eta > 0$, there exists $x_0 < t_1 \leqslant s_0$ such that

$$P\left[\sup_{x_0 \leqslant x \leqslant t_1} |Z^*_{1n_1}(x) - Z^*_{1n_1}(x_0)| \geqslant \varepsilon\right] \leqslant \eta \text{ for all sufficiently large } n_1 \text{ and } n_2 \tag{5.11}$$

and, since $\{Z_{1n_1}\}$ is tight, there exists $x_0 < t_2 \leqslant s_0$ such that

$$P\left[\sup_{x_0 \leqslant x \leqslant t_2} |Z_{1n_1}(x) - Z_{1n_1}(x_0)| \geqslant \varepsilon\right] \leqslant \eta \text{ for all sufficiently large } n_1 \text{ and } n_2. \tag{5.12}$$

Let $t_0 = t_1 \wedge t_2$ and assume that the sample sizes are large enough that (5.10)–(5.12) hold. Then, using $|a-b| < \varepsilon \;\Rightarrow\; |a| \geqslant \varepsilon$ or $|b| < 2\varepsilon$,

$$\begin{aligned} P[|Z^*_{1n_1}(t_0) - Z^*_{1n_1}(x_0)| < \varepsilon] &\leqslant P[|Z_{1n_1}(t_0) - Z^*_{1n_1}(x_0)| < \varepsilon] + \eta \\ &= P[|Z_{1n_1}(t_0) - Z_{1n_1}(x_0) - U(x_0)| < \varepsilon] + \eta \\ &\leqslant P[|Z_{1n_1}(t_0) - Z_{1n_1}(x_0)| \geqslant \varepsilon] + P[|U(x_0)| < 2\varepsilon] + \eta \\ &\leqslant P[|U(x_0)| < 2\varepsilon] + 2\eta. \end{aligned}$$

Now $P[|U(x_0)| < 2\varepsilon] = P[|0 \wedge V| < 2\varepsilon] = P[0 \vee (-V) < 2\varepsilon] \to \Phi(2\varepsilon/\tau)$. Thus (5.11) cannot hold for η sufficiently small, and $\{Z^*_{1n_1}\}$ cannot be tight.

(iii) If $M_1 = M_2$ on $[0,b]$, then $S_1 = S_2$ on $[0,b]$. Thus

$$(\hat{w}_1(\cdot), \hat{w}_2(\cdot))' \to (1/(1+\alpha), \alpha/(1+\alpha))' \text{ a.s., uniformly on } [0,b] \times [0,b].$$

The conclusion now follows from (5.6), (5.7), and (5.9) with an application of the continuous mapping theorem. □

Remark. Following the same arguments it may be seen that, if $M_1 \leqslant M_2$ on $[t_1, t_2]$ only, then Theorems 5.1 and 5.2 hold on $[a,b]$ with $a > t_1$ and $b < t_2$ for the estimators described in Section 2.

5.3. Asymptotic confidence intervals and confidence bands

In this section we consider construction of $(1-\gamma)$-coefficient (asymptotic) confidence intervals for $M_1(x)$ and $M_2(x)$ and $(1-\gamma)$-coefficient (asymptotic) confidence bands for M_1 and M_2 over intervals. Let X_M and Y_M denote the largest order statistics from the respective samples, and let $\tau_i^2(x) = \sigma_i^2(x)/S_i(x)$ denote the variance of $\hat{M}_i(x)$, $i = 1,2$. Denote the sample estimates by $\hat{\tau}_i^2(x) = \hat{\sigma}_i^2(x)/\hat{S}_i(x)$, where $\hat{\sigma}_i^2(x)$ is the sample variance (we recommend using $(n_i\hat{S}_i(x)-1)$ as the denominator) of the remaining $n_i\hat{S}_i(x)$ sample after time x. Let $L_i(\cdot)$ and $U_i(\cdot)$, $i=1,2$, denote the lower and upper confidence intervals (bands), respectively. For confidence intervals at a point, although we use the quantiles of the standard normal distribution in the formulas based on asymptotic distributions, a more conservative approach, recommended by Berger et al. (1988) for their testing problem, will be to use the same quantiles of a t-distribution with the degrees of freedom chosen by Welch's approximation for unequal variances, if the remaining sample size is small. For confidence bands over intervals we use Corollary 3 of Hall and Wellner (1979) which we state below as a theorem. Let B be a standard Brownian motion, let $\hat{\sigma}_i(0)$ denote the sample standard deviation of the entire sample, $i=1,2$, and for any $\beta \in (0,1)$ let $a = a(\beta)$ be such that $P(\|B\|_0^1) \leqslant a) = \beta$. Let $d_i(\cdot) = \hat{\sigma}_i(0)/\sqrt{n_i}\hat{S}_i(\cdot)$ for $i = 1,2$.

Theorem 5.3 (Hall and Wellner, 1979). *If the sampling variables have moments of the order of r for some $r > 2$, then*

$$\lim_{n_i \to \infty} P[|\hat{M}_i(x) - M_i(x)| \leqslant a d_i(x)\ \forall x \geqslant 0] \geqslant \beta \tag{5.13}$$

with equality for continuous s.f.s.

The probability $P(a) \equiv P(\|B\|_0^1) \leqslant a)$ has an infinite series expansion in the standard normal c.d.f. (Billingsley, 1968). Hall and Wellner (1979) show that for $a > 1.4$, the approximation $P(a) = 4\Phi(a) - 3$ gives a 3-place accuracy. They also provide a short table of values that we reproduce below.

For the confidence bands below we assume that the moment conditions of Theorem 5.3 hold for X and Y.

Let $\tau_i^2(x) = \sigma_i^2(x)/S_i(x)$, and let $\hat{\tau}_i^2(x)$ denote its sample estimate for $i = 1, 2$. In the 1-sample case, for $x < X_M$ and the restriction $M_1(x) \leqslant M_2(x)$, using (5.3) we define

$$L_1(x) = 0 \vee [\hat{M}_1(x) - z_{\gamma/2}\hat{\tau}_1(x)/\sqrt{n_1}] \wedge M_2(x),$$

$$U_1(x) = [\hat{M}_1(x) + z_{\gamma/2}\hat{\tau}_1(x)/\sqrt{n_1}] \wedge M_2(x).$$

Note that there is a positive probability of getting the degenerate interval $\{M_2(x)\}$. Using (5.4) and (5.13) we define the upper and lower bounds of a $(1-\gamma)$-coefficient confidence band for M_1 over an interval $[0, b]$ for a $b < X_M$ by

$$L_1^{\text{HW}}(\cdot) = 0 \vee [\hat{M}_1(\cdot) - a d_1(\cdot)] \wedge M_2(\cdot) \quad \text{and}$$

$$U_1^{\text{HW}}(\cdot) = [\hat{M}_1(\cdot) + a d_1(\cdot)] \wedge M_2(\cdot),$$

where a is found from Table 3 using $P(a) = 1 - \gamma$ (the superscript stands for Hall and Wellner). A similar interval or band could be defined for the reverse ordering.

In the 2-sample case our interest is in constructing $(1-\gamma)$-coefficient simultaneous confidence intervals and bands for M_1 and M_2. Since there are only two of these, we use the Bonferroni procedure. We note that a simultaneous confidence region of the form $[L_1, U_1] \times [L_2, U_2]$ could be possibly reduced by intersecting it with the set $A = \{(x_1, x_2) \in \mathscr{R}^2\colon x_1 \leqslant x_2\}$. For rectangular confidence regions we could use

$$[L_1, U_1 \wedge U_2] \times [L_1 \vee L_2, U_2], \tag{5.14}$$

which is what we employ. Of course we need to be careful that we do not define empty intervals.

We note that M_2^* is positive on $[0, X_M \vee Y_M)$, while M_1^* is positive on $[0, X_M)$. For an $x < X_M$, if $\hat{M}_1(x) < \hat{M}_2(x)$, we define the simultaneous confidence intervals for $M_1(x)$ and $M_2(x)$ by (5.14) with

$$L_i = \hat{M}_i(x) - z_{\gamma/4}\hat{\tau}_i(x)/\sqrt{n_i} \quad \text{and} \quad U_i = \hat{M}_i(x) + z_{\gamma/4}\hat{\tau}_i(x)/\sqrt{n_i}, \quad i = 1, 2. \tag{5.15}$$

If $\hat{M}_1(x) \geqslant \hat{M}_2(x)$, we have $M_1^*(x) = M_2^*(x) = \hat{M}(x) = \hat{w}_1(x)\hat{M}_1(x) + \hat{w}_2(x)\hat{M}_2(x)$. In this case we propose the following confidence procedure. Since our point estimates

Table 3
Approximate values of $P(a)$ for some values of a

a	2.807	2.241	1.960	1.534	1.149	0.871
$P(a)$	0.99	0.95	0.90	0.75	0.50	0.25

coincide, we use the distribution of the pooled estimator $\hat{M}(x)$ to construct a common confidence interval. It is clear that, when $M_1(x) = M_2(x) = M(x)$, say,

$$\sqrt{n_1}[\hat{M}(x) - M(x)] \xrightarrow{d} w_1(\alpha, x)Z_1(x) + w_2(\alpha, x)Z_2(x)/\sqrt{\alpha}$$
$$\sim N(0, w_1^2(\alpha, x)\tau_1^2(x) + w_2^2(\alpha, x)\tau_2^2(x)/\alpha)$$
$$\equiv N(0, \rho^2(x)). \tag{5.16}$$

We then use the common confidence interval by

$$L_i = 0 \vee [\hat{M}(x) - z_{\gamma/4}\hat{\rho}(x)/\sqrt{n_i}] \quad \text{and} \quad U_i = \hat{M}(x) + z_{\gamma/4}\hat{\rho}(x)/\sqrt{n_i},$$
$$i = 1, 2, \tag{5.17}$$

where $\hat{\rho}(x)$ is the sample estimate of $\rho(x)$, using $\hat{w}_i(x)$ to estimate $\hat{w}_i(\alpha, x)$. For a $(1-\gamma)$-coefficient simultaneous confidence band on $[0, b]$ for a $b < X_M$, we first compute a^* defined by the a that corresponds to $P(a) = 1 - \gamma/2$ in Table 3. If $\hat{M}_1 < \hat{M}_2$ on $[0, b]$, then we use (5.14) (extended to functions on $[0, b]$) to define the confidence bands with

$$L_i^{\mathrm{HW}}(\cdot) = 0 \vee [\hat{M}_i(\cdot) - a^* d_i(\cdot)] \quad \text{and} \quad U_i^{\mathrm{HW}}(\cdot) = \hat{M}_i(\cdot) + a^* d_i(\cdot),$$
$$i = 1, 2. \tag{5.18}$$

If $\hat{M}_1 \geqslant \hat{M}_2$ over some regions, we would have liked to have defined our confidence bands as in (5.18) on $\{\hat{M}_1 < \hat{M}_2\}$ and some generalization of (5.16) on $\{\hat{M}_1 \geqslant \hat{M}_2\}$. Theorem 5.3, which is applicable for a single $\hat{M}_i$, is based on the distribution of the random variable $\|B\|_0^1$ which is well known. For a comparable result involving $\hat{M}$ we need to know the distribution of the sup of the sum of two independent and (differently) scaled Brownian motions, and we have not been able to derive that result. Lacking this distribution theory, we define the confidence bands by (5.18) and (5.14) on $\{\hat{M}_1 < \hat{M}_2\}$ and a common confidence band, $[L(\cdot), U(\cdot)]$, by

$$L(\cdot) = L_1^{\mathrm{HW}}(\cdot) \vee L_2^{\mathrm{HW}}(\cdot) \quad \text{and} \quad U(\cdot) = U_1^{\mathrm{HW}}(\cdot) \wedge U_2^{\mathrm{HW}}(\cdot) \text{ on } \{\hat{M}_1 \geqslant \hat{M}_2\}, \tag{5.19}$$

which is based on the heuristic that both confidence bands are simultaneously valid for a common MRLF, which is the way we make our point estimation. However, our point estimate, $\hat{M}$, may not be in the band always. If $\hat{M}_1 \geqslant \hat{M}_2$ on all of $[0, b]$, we could provide a common confidence band, essentially with the assumption that $M_1 = M_2$, and using $\hat{M}$ as the estimator of a single sample problem. However, it will probably be wiser to revise our opinion about the order restriction.

We end this section by verifying the asymptotic coverage probabilities. This is obvious for the 1-sample case from (5.1) for the confidence interval, and (5.13) for the confidence band. For the simultaneous confidence intervals for M_1 and M_2, we note that (5.15) provides the correct (conservative) asymptotic coverage probability always, since each individual does with confidence coefficient $1-\gamma/2$ (the possible shortening of the intervals using (5.14) comes free of charge under the order restriction). Note that, if $M_1(x) < M_2(x)$, then eventually this is the only formula that will apply wp1. If $M_1(x) = M_2(x)$, then (5.17) also provides the correct asymptotic coverage from (5.16), thus providing a shorter confidence interval in case of violation of the ordering. Similarly, the confidence bands given by (5.18) and (5.14) always provide the correct (conservative) asymptotic coverage probability by Theorem 5.3, again noting that this is the only formula that will eventually apply wp1 if $M_1 < M_2$ on $[0,b]$ from the strong uniform consistency. On the set $M_1 = M_2$ (5.19) also provides the correct asymptotic coverage since both individual confidence bands given by the (5.13) in Theorem 2.3 provide a coverage probability $\geqslant 1-\gamma/2$ for the same MRLF.

5.4. An example

Bjerkedal (1960, p. 140) reports on two studies of survival time (in days) of guinea pigs infected with different dosages of tubercle bacilli. We compare the MRLFs for Regimens 4.3 (M_2) and 5.5 (M_1), assuming that the higher dosage corresponds to a smaller MRL. The data is complete, i.e., there was no censoring, with $n_1 = n_2 = 72$. Fig. 2 presents a graph of $\hat{M}_1$, $\hat{M}_2$, M_1^*, and a 90% confidence band for M_1 alone on

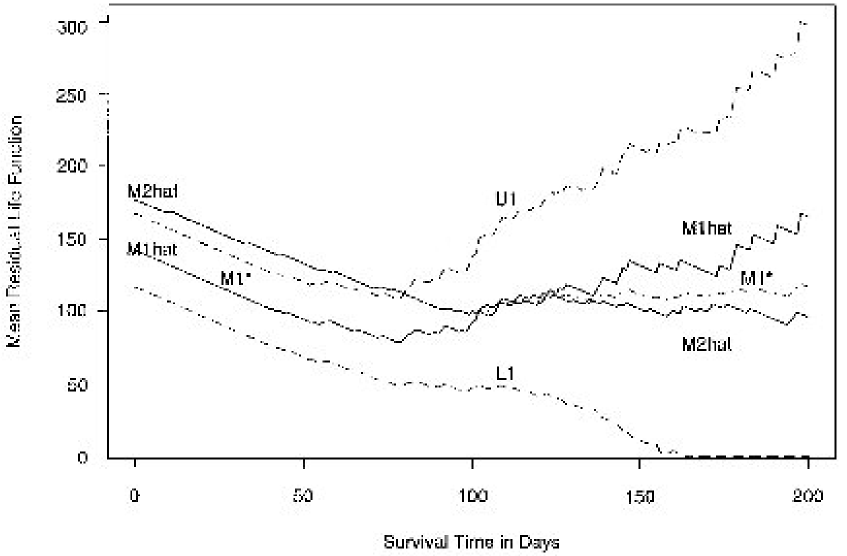

Fig. 2. Restricted and unrestricted estimators of MRLF of M_1 (regimen 5.5) and M_2 (regimen 4.3) and 90% confidence band for M_1. M1hat $=\hat{M}_1$, M2hat $=\hat{M}_2$, M1* $=M_1^*$, M2* $=M_2^*$, U1 $=U_1^{HW}$ and L1 $=L_1^{HW}$.

[0, 200] only since the lower bound of the confidence band becomes 0 at approximately 160 days.

6. Concluding remarks

In this paper we have provided estimators for two MRLFs, M_1 and M_2, under the order restriction that $M_1 \leqslant (\geqslant) M_2$, on their entire ranges or on a closed interval, when M_2 is known or unknown. Ebrahimi (1993) initiated this study, and has provided an excellent example (his Figs. 1 and 2) using a 2-sample problem with real data. We have shown that they are strongly uniformly consistent and asymptotically unbiased. We have also derived their (joint) asymptotic distributions, both at a point, and their (joint) weak convergence on an interval. We have provided formulas for confidence intervals and confidence bands in the 1-sample case and for simultaneous confidence intervals and bands for the 2-sample case. The confidence intervals or bands are always of the same lengths or shorter than those in the unrestricted case (using the Bonferroni procedure in the 2-sample case). However, these confidence procedures do not employ the distribution theory developed under order restriction; these are useful in testing for and against the order restrictions, a problem we propose to pursue in the future. We have also carried out an extensive simulation, and have presented some of the results. A surprising outcome of these simulations is that the restricted estimators appear to be superior to the unrestricted empirical ones in terms of MSE, uniformly at all quantiles of the distributions we have investigated. We do not completely understand this phenomenon, and it is worth further studies.

Acknowledgements

The comments of the referees and the associate editor resulted in an improved paper. The example in Section 5.4 and the explanation to show that the restricted estimators in Section 2 are indeed mean residual life functions are due to their prodding.

References

Arcones, M.A., Samaniego, F.J., 2000. On the asymptotic distribution theory of a class of consistent estimators of a distribution satisfying a uniform stochastic ordering constraint. Ann. Statist. 28, 116–150.

Berger, R.L., Boos, D.D., Guess, F.M., 1988. Tests and confidence sets for comparing two mean residual functions. Biometrics 44, 103–115.

Billingsley, P., 1968. Convergence of Probability Measures. Wiley, New York.

Bjerkedal, T., 1960. Acquisition of resistance in guinea pigs infected with different dosages of virulent tubercle bacilli. Amer. J. Hygiene 72, 130–148.

Brunk, H.D., Franck, W.E., Hanson, D.L., Hogg, R.V., 1966. Maximum likelihood estimation of the distributions of two stochastically ordered random variables. J. Amer. Statist. Assoc. 61, 1067–1080.

Csörgő, M., Zitikis, R., 1996. Mean residual life processes. Ann. Statist. 24, 1717–1739.

Dykstra, R., Kochar, S., Robertson, T., 1991. Statistical inference for uniform stochastic ordering in several populations. Ann. Statist. 19, 870–888.

Ebrahimi, N., 1993. Estimation of two ordered mean residual life functions. Biometrics 49, 409–417.

Guess, F., Proschan, F., 1988. Mean residual life: theory and applications. In: Krishnaiah, P.R., Rao, C.R. (Eds.), Handbook of Statistics, Vol. 7, pp. 215–224.

Hall, W.J., Wellner, J.A., 1979. Estimation of mean residual life. Unpublished manuscript.

Hall, W.J., Wellner, J.A., 1981. Mean residual life. In: Csörgő, M., Dawson, D.A., Rao, J.N.K., Saleh, A.K.Md.E. (Eds.), Statistics and Related Topics. North-Holland, Amsterdam, pp. 169–184.

Huang, J., Praestgaard, J.T., 1996. Asymptotic theory of nonparametric estimation of survival curves under order restrictions. Ann. Statist. 24, 1679–1716.

Mukerjee, H., 1996. Estimation of survival functions under uniform stochastic ordering. J. Amer. Statist. Assoc. 91, 1684–1689.

Rojo, J., 1995. On the weak convergence of certain estimators of stochastically ordered survival functions. Nonparametric Statist. 4, 349–363.

Rojo, J., Ma, Z., 1996. On the estimation of stochastically ordered survival functions. J. Statist. Comput. Simul. 55, 1–21.

Rojo, J., Samaniego, F.J., 1991. On nonparametric maximum likelihood estimation of a distribution uniformly stochastichally smaller than a standard. Statist. Probab. Lett. 11, 267–271.

Rojo, J., Samaniego, F.J., 1993. On estimating a survival curve subject to a uniform stochastic ordering constraint. J. Amer. Statist. Assoc. 88, 566–572.

Yang, G.L., 1978. Estimation of a biometric function. Ann. Statist. 6, 112–116.