The Annals of Statistics
1997, Vol. 25, No. 4, 1622–1645

A CHARACTERIZATION OF MARGINAL DISTRIBUTIONS OF (POSSIBLY DEPENDENT) LIFETIME VARIABLES WHICH RIGHT CENSOR EACH OTHER[1]

BY TIM BEDFORD AND ISAAC MEILIJSON

Delft University of Technology and Tel Aviv University

It is well known that the joint distribution of a pair of lifetime variables X_1 and X_2 which right censor each other cannot be specified in terms of the subsurvival functions

$$P(X_2 > X_1 > x), \qquad P(X_1 > X_2 > x) \quad \text{and} \quad P(X_1 = X_2 > x)$$

without additional assumptions such as independence of X_1 and X_2. For many practical applications independence is an unacceptable assumption, for example, when X_1 is the lifetime of a component subjected to maintenance and X_2 is the inspection time. Peterson presented lower and upper bounds for the marginal distributions of X_1 and X_2, for given subsurvival functions. These bounds are sharp under nonatomicity conditions. Surprisingly, not every pair of distribution functions between these bounds provides a feasible pair of marginals. Crowder recognized that these bounds are not functionally sharp and restricted the class of functions containing all feasible marginals. In this paper we give a complete characterization of the possible marginal distributions of these variables with given subsurvival functions, without any assumptions on the underlying joint distribution of (X_1, X_2). Furthermore, a statistical test for an hypothesized marginal distribution of X_1 based on the empirical subsurvival functions is developed.

The characterization is generalized from two to any number of variables.

1. Introduction. We are given the *subdistribution* functions $P(X_1 \leq x,\ X_1 < X_2)$, $P(X_2 \leq x,\ X_2 < X_1)$ and $P(X_1 \leq x,\ X_1 = X_2)$ of a pair (X_1, X_2) of random variables. What are the possible marginal distribution functions $P(X_1 \leq x)$ and $P(X_2 \leq x)$?

This question is motivated by competing risks. The random variable X_1 is the lifetime of a component of some system, and the random variable X_2 is the time at which the system's life is interrupted due to other possible causes. Let $T = \min(X_1, X_2)$ be the lifetime of the system, and let I indicate whether T is equal only to X_1, only to X_2 or to both. The pair (T, I), whose joint distribution is described by the three subdistribution functions above is the natural observable data.

It has been of major interest to study conditions under which the joint distribution of (X_1, X_2) is uniquely determined by that of (T, I). It is well

Received June 1994; revised November 1996.

[1]This research was partially supported by ESTEC study contract number QSS/CCN.4/8.92/CP. This paper appeared in the preprint series of the Faculty of Technical Mathematics and Informatics, Preprint 93-116.

AMS 1991 *subject classifications*. 62E15, 62G15, 62N05, 90C39.

Key words and phrases. Competing risk, dependent censoring, identifiability, survival analysis, Kolmogorov–Smirnov test.

known (see [11]) that the marginal and joint distributions of (X_1, X_2) are in general "nonidentifiable"; that is, there are many different distributions of (X_1, X_2) which are compatible with the data (i.e., yield identical subdistribution functions). Is it also well known (see [3], [7] and [6]) that if X_1 and X_2 are independent, are nonatomic and share essential suprema, their marginal distributions are identifiable. (See also [4] for distributions with atoms.) In many applications, however, the assumption of independence is too strong. Some authors (see [1], [8] and [5]) have considered other kinds of assumptions on the joint distribution of (X_1, X_2).

Accepting nonidentifiability as a fact of life(time), Peterson [9] presents bounds for this joint distribution as well as for its marginals, assuming that $P(X_1 = X_2) = 0$. Peterson further proves that these bounds are sharp. However, this statement and its proof hold only under the additional assumption (not stated by Peterson) of continuity of the two subsurvival functions.

We study the class of pairs of marginal distributions $P(X_1 \le x)$, $P(X_2 \le x)$ for which there exists a joint distribution giving rise to a given triplet of subdistribution functions as described above, and we show how to estimate the set of feasible $P(X_1 \le x)$ or test for the feasibility of a given distribution function, from an i.i.d. sample of (T, I) data.

The Peterson bounds, expanded to cover the case $P(X_1 = X_2) \ge 0$, can be expressed in terms of sums of subdistribution functions as

$$\begin{aligned} P(X_1 \le x,\ X_1 \le X_2) &\le F_1(x) \\ &\le P(X_1 \le x,\ X_1 \le X_2) + P(X_2 \le x,\ X_2 < X_1) \\ &= P(T \le x) \end{aligned} \tag{1}$$

and

$$\begin{aligned} P(X_2 \le x,\ X_2 \le X_1) &\le F_2(x) \\ &\le P(X_2 \le x,\ X_2 \le X_1) + P(X_1 \le x,\ X_1 < X_2) \\ &= P(T \le x), \end{aligned} \tag{2}$$

where $F_1(x) = P(X_1 \le x)$ and $F_2(x) = P(X_2 \le x)$.

However, the following more stringent bounds hold too:

$$\begin{aligned} P(X_1 \le x,\ X_1 \le X_2) &\le F_1(x) \\ &\le P(X_1 \le x,\ X_1 \le X_2) + P(X_2 < x,\ X_2 < X_1) \\ &= P(T \le x) - P(X_2 = x,\ X_2 < X_1) \end{aligned} \tag{3}$$

and

$$\begin{aligned} P(X_2 \le x,\ X_2 \le X_1) &\le F_2(x) \\ &\le P(X_2 \le x,\ X_2 \le X_1) + P(X_1 < x,\ X_1 < X_2) \\ &= P(T \le x) - P(X_1 = x,\ X_1 < X_2), \end{aligned} \tag{4}$$

showing the Peterson bounds not to be sharp when the subdistribution functions have atoms, even under the assumption $P(X_1 = X_2) = 0$. We shall see that the improved bounds (3) and (4) are *pointwise* sharp, but not *functionally* sharp. That is, for any x and any of the four inequalities in (3) and (4) there exist joint distributions for (X_1, X_2) with the given subdistribution functions, for which that inequality holds as an equality, but not every pair of distribution functions F_1 and F_2 bounded as in (3) and (4) can serve as marginal distribution functions of X_1 and X_2. Indeed, considering any $x < y$,

$$\begin{aligned} P(X_1 \le y) - P(X_1 \le x) &= P(x < X_1 \le y) \\ &\ge P(x < X_1 \le y,\ X_1 \le X_2) \\ &= P(X_1 \le y,\ X_1 \le X_2) - P(X_1 \le x,\ X_1 \le X_2) \end{aligned}$$

or

$$P(X_1 \le x) - P(X_1 \le x,\ X_1 \le X_2) \le P(X_1 \le y) - P(X_1 \le y,\ X_1 \le X_2),$$

which may be rephrased as saying that $P(X_1 \le x) - P(X_1 \le x,\ X_1 \le X_2)$ is a nonnegative, *nondecreasing* function. In other words, the gap between the left and middle terms of inequalities (3) and (4) must be nondecreasing in x.

This functional inequality was first found by Crowder [2]. In attempting to add further conditions to obtain a characterization, Crowder gives up on necessity by requiring a technically convenient but unnecessary condition, and unfortunately rules out sufficiency as well by failing to notice a pathological aspect of the upper Peterson bound. We shall see that the above simple necessary conditions are almost sufficient but that a rather subtle additional measure-theoretic condition is required. This condition, which holds automatically if either X_1 or X_2 is a discrete random variable, asserts that F_1 and F_2 may only "lightly touch" their upper bounds, a notion which will be made precise in Theorem 1.

As an illustration of the nature or extent of nonidentifiability, we show that if X_1 is assumed to be exponentially distributed, the set of feasible failure rates λ for this distribution always constitute a (possibly empty, possibly open from above) interval.

The definition of (T, I) for $m > 2$ competing risks and the characterization of the possible m-tuples $(F_1, F_2, \ldots, F_m)$ is presented after the proof of this theorem.

In the second part of this paper we construct a statistical test for a hypothesized marginal distribution for X_1 given the empirical subdistribution functions.

Let

$$\hat{P}(X_1 \le x,\ X_1 \le X_2) = \frac{1}{n} \operatorname{card}\left\{ i | X_1^{(i)} \le x,\ X_1^{(i)} \le X_2^{(i)} \right\}$$

be the empirical subdistribution function of X_1 based on a sample of n observations, and let H be any distribution function. We show how to construct,

among all functions H_1' such that H_1' and $H - H_1'$ are nonnegative and nondecreasing, one (denoted by H_1) which minimizes

$$\sup_x |\hat{P}(X_1 \le x,\ X_1 \le X_2) - H_1'(x)|$$

and prove that

$$\sup_x |\hat{H}(x) - H(x)| \ge \sup_x |\hat{P}(X_1 \le x,\ X_1 \le X_2) - H_1(x)| \tag{5}$$

for any discrete distribution function $\hat{H}$ whose jumps contain those of $\hat{P}(X_1 \le x,\ X_1 \le X_2)$. In particular this holds for any possible empirical distribution function of the partially unseen X_1-sample. Hence, if the RHS of (5) exceeds some critical point, so does its LHS. Since the LHS of (5) is the regular Kolmogorov–Smirnov statistic evaluated at the empirical distribution of the X_1-sample, this permits the application of a *conservative* Kolmogorov–Smirnov test of the hypothesized distribution H based on subdistribution functions.

This method is illustrated by restricting H to the class of exponential distributions. As expressed earlier, the feasible failure rates λ constitute a (possibly empty) interval J. The Kolmogorov–Smirnov test gives rise rather naturally to a sort of confidence interval $\hat{J}^{(n)}$: letting C be the critical point of the Kolmogorov–Smirnov test, for n observations, with some preassigned confidence coefficient $1 - \alpha$, define $\hat{J}^{(n)}$ to be the set (in fact, interval) of λ values for H under which the RHS of (5) does not exceed C. The interval-statistic $\hat{J}$ is a *confidence interval* in the sense that, under every $\lambda \in J$, $P_\lambda(\lambda \in \hat{J}) \ge 1 - \alpha$. Furthermore, the confidence interval $\hat{J}^{(n)}$ is *consistent* in the sense that J equals the closure of the interior of $\limsup_n \hat{J}^{(n)}$, with probability 1.

2. Marginal distributions with given subdistribution functions. Let X_1 and X_2 be random variables taking values in $\mathbb{R}$. We define five functions as follows:

$$F_1(x) = P(X_1 \le x); \tag{6}$$

$$F_2(x) = P(X_2 \le x); \tag{7}$$

$$G_{12}(x) = P(X_1 \le x,\ X_1 = X_2); \tag{8}$$

$$G_1(x) = P(X_1 \le x,\ X_1 < X_2); \tag{9}$$

$$G_2(x) = P(X_2 \le x,\ X_2 < X_1). \tag{10}$$

Let

$$\begin{aligned} \underline{F}_1(x) &= G_{12}(x) + G_1(x) = P(X_1 \le x,\ X_1 \le X_2), \\ \underline{F}_2(x) &= G_{12}(x) + G_2(x) = P(X_2 \le x,\ X_2 \le X_1) \end{aligned} \tag{11}$$

denote the lower bounds [see (3) and (4)] for F_1 and F_2, with the nonnegative and nondecreasing gaps

$$\begin{aligned} F_1(x) - \underline{F}_1(x) &= P(X_1 \le x,\ X_2 < X_1), \\ F_2(x) - \underline{F}_2(x) &= P(X_2 \le x,\ X_1 < X_2). \end{aligned} \tag{12}$$

In terms of these functions, the upper bounds [see (3) and (4)] for F_1 and F_2 are

$$\begin{aligned} \overline{F}_1(x) &= \underline{F}_1(x) + G_2(x-), \\ \overline{F}_2(x) &= \underline{F}_2(x) + G_1(x-). \end{aligned} \tag{13}$$

DEFINITION 1. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a nondecreasing function. A choice of n nondecreasing functions $f_1, \dots, f_n$ such that $f = f_1 + \cdots + f_n$ is called a *comonotone representation* of f, and the set of such choices $(f_1, \dots, f_n)$ is denoted $\mathscr{C}(f)$. If f and each f_i are nonnegative, then $f = f_1 + \cdots + f_n$ is called a *nonnegative comonotone representation* of f, and the set of such $(f_1, \dots, f_n)$ is denoted $\mathscr{C}^+(f)$. If $n = 2$, then we will write $f_1 \in \mathscr{C}^+(f)$ for short.

LEMMA 1. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be right continuous and nondecreasing, and let $f = f_1 + \cdots + f_n$ be a comonotone representation of f. Then $f_1, \dots, f_n$ are right continuous.*

PROOF. Express $f = f_1 + (f_2 + \cdots + f_n)$ to see that it is enough to consider the case $n = 2$.

We show below that if f_1 is not right continuous at a point x, then the set $\{y | y > x \text{ and } f_2(y) < f_2(x)\}$ is nonempty. This will show that f_1 is right continuous. By symmetry, so is f_2.

If the nondecreasing function f_1 is not right continuous at x, then

$$\varepsilon = \frac{f_1(x+) - f_1(x)}{2} > 0.$$

By right continuity of f there is a $\delta > 0$ such that $f(y) - f(x) < \varepsilon$ whenever $0 < y - x < \delta$. But then

$$\begin{aligned} f_2(x) - f_2(y) &> f_2(x) - f_2(y) - \big(\varepsilon - (f(y) - f(x))\big) \\ &= f_1(y) - f_1(x) - \varepsilon \\ &\ge f_1(x+) - f_1(x) - \varepsilon = \varepsilon > 0, \end{aligned}$$

as claimed. □

We require one further definition.

DEFINITION 2. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a nondecreasing function. We say that $g\colon \mathbb{R} \to \mathbb{R}$ is *the left continuous version of* f, and we write $g = \tilde{f}$, if g is left continuous and $g(x) = f(x)$ for every continuity point x of f.

Note that $\tilde{f}$ is uniquely defined, by monotonicity of f.

THEOREM 1. (i) *Let X_1 and X_2 be random variables. Then, using the notation established above*:

1. $F_i = \underline{F}_i + (F_i - \underline{F}_i)$ *is a nonnegative comonotone representation of a nondecreasing right continuous function, for* $i = 1, 2$;
2. $F_i(x) \leq \overline{F}_i(x)$ *for all* x, *and the Lebesgue measure of the range set* $\{(F_i - \underline{F}_i)(x) | F_i(x) = \overline{F}_i(x)\}$ *is zero, for* $i = 1, 2$;
3. (a) $F_i(-\infty) = 0$ *and* $F_i(\infty) = 1$, *for* $i = 1, 2$; (b) $(F_i - \underline{F}_i)(\infty) = (\overline{F}_i - \underline{F}_i)(\infty)$, *for* $i = 1, 2$; (c) $G_1(\infty) + G_2(\infty) + G_{12}(\infty) = 1$.

(ii) *If nondecreasing right continuous functions F_1, F_2, G_{12}, G_1 and G_2 satisfy conditions* 1–3 *of* (i), *then there are random variables X_1 and X_2 such that* (6)–(10) *hold*.

PROOF. (i) It is clear that $\underline{F}_1 + (F_1 - \underline{F}_1)$ and $\underline{F}_2 + (F_2 - \underline{F}_2)$ are nonnegative comonotone representations of the nondecreasing and right continuous functions F_1 and F_2. It is also obvious that condition 3(a) holds, since F_1 and F_2 are distribution functions. Furthermore,

$$(F_2 - \underline{F}_2)(\infty) = P(X_1 < X_2) = G_1(\infty),$$
$$(F_1 - \underline{F}_1)(\infty) = P(X_2 < X_1) = G_2(\infty)$$

and

$$G_1(\infty) + G_2(\infty) + G_{12}(\infty) = P(X_1 < X_2) + P(X_2 < X_1) + P(X_1 = X_2) = 1$$

which demonstrates 3(b) and 3(c).

We shall now prove property 2 for the case $i = 2$. The other case follows by a similar argument.

Since $X_2 \leq x$ and $X_1 < X_2$ imply that $X_1 < x$ and $X_1 < X_2$, it is certainly true that $(F_2 - \underline{F}_2)(x) \leq \tilde{G}_1(x)$, or, equivalently, $F_2(x) \leq \overline{F}_2(x)$.

We make the following definitions (see Figure 1):

$$A = \{x | (F_2 - \underline{F}_2)(x) = \tilde{G}_1(x)\} = \{x | F_2(x) = \overline{F}_2(x)\} \subseteq \mathbb{R};$$
$$K_x = \{(u, v) | u < x < v\} \subset \mathbb{R}^2;$$
$$K = \bigcup_{x \in A} K_x \subset \mathbb{R}^2.$$

Since $\tilde{G}_1(x) - (F_2 - \underline{F}_2)(x) = P(X_1 < x < X_2)$, it is clear that $P((X_1, X_2) \in K_x) = 0$ if and only if $x \in A$, and that A is closed. It is easy to see that if $(x_n)_{n \geq 0}$ is dense in A then $\bigcup_n K_{x_n} = K$. This implies immediately that $P((X_1, X_2) \in K) = 0$.

As A is closed we can write A^c as a countable union of disjoint open intervals (a_n, b_n). The mass $P(X_1 < X_2)$ is supported by the complement of K, which is the disjoint union of the (possibly unbounded) triangles

$$U_n = \{(x_1, x_2) \colon a_n \leq x_1 < x_2 \leq b_n\}.$$

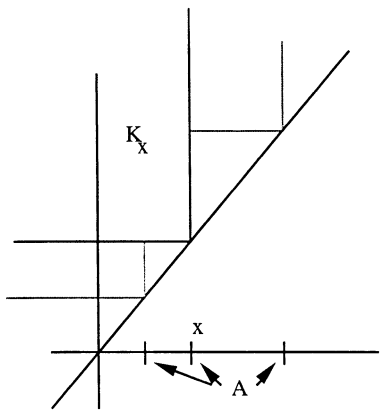

FIG. 1. *Some sets defined in the* (X_1, X_2) *plane.*

Consider the artificial purely atomic subdistribution function L [see (14)], supported by the (countable) set $\{b_n | n = 1, 2, \ldots\}$, that assigns to b_n the point mass $P((X_1, X_2) \in U_n)$:

$$L(x) = \sup\{(F_2 - \underline{F}_2)(y) | y \leq x, \ y \in A\}. \tag{14}$$

Since $x \in A$ implies that $L(x) = (F_2 - \underline{F}_2)(x)$, we have

$$\{F_2(x) - \underline{F}_2(x) | x \in A\} = \{L(x) | x \in A\} \subseteq \{L(x) | x \in \mathscr{R}\}.$$

In other words, the range set $\{F_2(x) - \underline{F}_2(x) | x \in A\}$, which we are trying to prove to be a Lebesgue null subset of $[0, 1]$, is a subset of the range set of some purely atomic subdistribution function. The proof will be finished if we show that range sets of purely atomic distribution functions are always Lebesgue null sets. However, this is clear, since the jumps of such a distribution function add up to unity and map on the y-axis to disjoint open intervals contained in the complement of the range set of the distribution function.

This completes the proof of (i).

REMARK. The range set (or set of values) of a purely atomic distribution function may seem countable at first (one value per atom)—thus, obviously a Lebesgue null set—but it need not be, for if the atoms are dense (e.g., a distribution supported by the rationals), then the distribution function is strictly increasing. As such, it is a one-to-one mapping from the real line onto its range, so this range set is necessarily uncountable. The intuitive picture of countability is misleading; Cantor-like sets are a more accurate description.

(ii) Given functions F_1, F_2, G_{12}, G_1 and G_2 satisfying conditions 1–3 of (i), we explicitly construct a pair of random variables X_1 and X_2 such that the interpretation (6)–(10) of these five functions holds. This construction is illustrated in Figure 2.

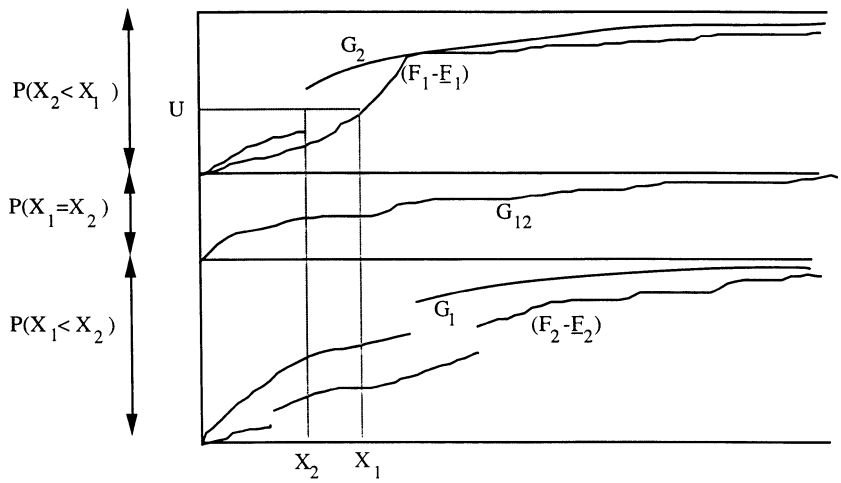

FIG. 2. *The construction of* X_1 *and* X_2.

First note that, by Lemma 1, Condition 1 implies that G_1, $F_1 - \underline{F}_1$, G_2, $F_2 - \underline{F}_2$ and G_{12} are nondecreasing right continuous functions. Furthermore, by 3(a), we have $J(-\infty) = 0$ whenever J is any of these functions.

Let U be uniformly distributed on $[0, 1]$. We shall distinguish between three cases.

Using the standard convention for inverting a right continuous function

$$F^{-1}(u) = \inf\{x | F(x) \geq u\},$$

we have the following:

Case 1. If $U < G_1(\infty)$, then define $X_1 = G_1^{-1}(U)$ and $X_2 = (F_2 - \underline{F}_2)^{-1}(U)$ [X_2 is well defined since $(F_2 - \underline{F}_2)(\infty) = G_1(\infty)$].

Case 2. If $G_1(\infty) < U < G_1(\infty) + G_{12}(\infty)$, then define $U' = U - G_1(\infty)$ and set $X_1 = X_2 = G_{12}^{-1}(U')$.

Case 3. If $G_1(\infty) + G_{12}(\infty) < U < G_1(\infty) + G_{12}(\infty) + G_2(\infty) = 1$, then write $U' = U - (G_1(\infty) + G_{12}(\infty))$ and define $X_1 = (F_1 - \underline{F}_1)^{-1}(U')$ and $X_2 = G_2^{-1}(U')$ [X_1 is well defined since $(F_1 - \underline{F}_1)(\infty) = G_2(\infty)$].

The three cases we have distinguished correspond to $X_1 < X_2$, $X_1 = X_2$ and $X_1 > X_2$, respectively. For, conditional on $U < G_1(\infty)$ (i.e., in Case 1) either U is a continuity value of G_1 or the distribution of X_1 has an atom at $G_1^{-1}(U)$. In the first situation, except for a Lebesgue null set of U values, by property 2, and because $\tilde{G}_1 - (F_2 - \underline{F}_2) = \overline{F}_2 - F_2$,

$$U = G_1(X_1) = \tilde{G}_1(X_1) > (F_2 - \underline{F}_2)(X_1)$$

so that, by right continuity and monotonicity of $F_2 - \underline{F}_2$, $(F_2 - \underline{F}_2)^{-1}(U) = X_2 > X_1$. In the second situation (idem) we have

$$G_1(X_1) > U > \tilde{G}_1(X_1) \geq (F_2 - \underline{F}_2)(X_1)$$

and again we get $X_2 > X_1$. The argument for $X_1 > X_2$ is similar, and for $X_1 = X_2$ is obvious. This shows that (8)–(10) hold.

We now just have to check that X_1 and X_2 have the right marginals. By symmetry consider only X_1. The construction shows that

$$\begin{aligned} P(X_1 \le x) &= P(X_1 \le x,\ X_1 = X_2) + P(X_1 \le x,\ X_1 < X_2) \\ &\quad + P(X_1 \le x,\ X_2 < X_1) \\ &= G_{12}(x) + G_1(x) + (F_1 - \underline{F}_1)(x) \\ &= F_1(x) \end{aligned}$$

by condition 1. This completes the proof. □

REMARK (Product form of the characterization). The theorem says that there are no joint conditions to be satisfied by the marginal distribution functions.

As a corollary of Theorem 1 we recover the improved Peterson bounds (3) and (4), a similar bound on the joint survival function of X_1 and X_2, and the result that these bounds are pointwise sharp. This corollary is essentially the same as Theorem 1 in [9], but with two improvements. First, we take account of the possibility of atoms in the subdistribution functions (Peterson implicitly assumes continuity, as we noted in the Introduction) and of positive probability mass on the diagonal. Second, we show that the bounds can be achieved pointwise (a slight improvement on [9], where distributions are given getting arbitrarily close to the bounds).

COROLLARY 1. *The following inequalities hold for the joint survival function of X_1 and X_2*:

$$\begin{aligned} P(X_1 > x_1,\ X_2 > x_2) \le{}& P(X_1 > x_1,\ X_1 < X_2) \\ &+ P(X_2 > x_2,\ X_2 < X_1) \\ &+ P(X_1 > \max(x_1, x_2), X_1 = X_2) \end{aligned} \tag{15}$$

and

$$\begin{aligned} &P(X_1 > x_1,\ X_2 > x_2) \\ &\quad \ge \begin{cases} P(X_1 \ge x_2,\ X_1 < X_2) + P(X_2 > x_2,\ X_2 \le X_1), & \text{if } x_1 < x_2, \\ P(X_1 > x_1,\ X_1 \le X_2) + P(X_2 \ge x_1,\ X_2 < X_1), & \text{if } x_2 < x_1, \\ P(X_1 > x_1,\ X_1 < X_2) + P(X_2 > x_1,\ X_2 \le X_1), & \text{if } x_1 = x_2. \end{cases} \end{aligned} \tag{16}$$

The following inequalities [*a rephrasing of inequalities* (3) *and* (4)] *hold for the marginals*:

$$\begin{aligned} &P(X_1 > x_1,\ X_1 \le X_2) + P(X_2 \ge x_1,\ X_2 < X_1) \\ &\quad \le P(X_1 > x_1) \\ &\quad \le P(X_1 > x_1,\ X_1 \le X_2) + P(X_2 < X_1), \end{aligned} \tag{17}$$

$$
\begin{aligned}
&P(X_2 > x_2,\ X_2 \le X_1) + P(X_1 \ge x_2,\ X_1 < X_2) \\
&\qquad \le P(X_2 > x_2) \\
&\qquad \le P(X_2 > x_2,\ X_2 \le X_1) + P(X_1 < X_2).
\end{aligned} \tag{18}
$$

Furthermore, given subdistribution functions G_1, G_2 and G_{12}, and given any x_1' and x_2', considering any of the six inequalities in (15)–(18)*, there exist joint distributions for which that inequality holds as an equality at $x_1 = x_1'$ and $x_2 = x_2'$, and which satisfy all six inequalities for all x_1 and x_2.*

PROOF. First of all write

$$
\begin{aligned}
P(X_1 > x_1,\ X_2 > x_2) = {} & P(X_1 > x_1,\ X_2 > x_2,\ X_1 = X_2) \\
& + P(X_1 > x_1,\ X_2 > x_2,\ X_1 < X_2) \\
& + P(X_1 > x_1,\ X_2 > x_2,\ X_2 < X_1)
\end{aligned}
$$

to get inequality (15) and the first two cases of (16). The third inequality in the lower bound, (16), holds always as an equality.

We now show that each bound is sharp at arbitrarily chosen points x_1' and x_2'. Theorem 1 may be used to construct joint distributions for given functions G_1, G_2 and G_{12}. It is only necessary for us to specify functions F_1 and F_2 satisfying the conditions of the theorem. We do this by defining $F_1 - \underline{F}_1$ and $F_2 - \underline{F}_2$ (recall $\underline{F}_1 = G_1 + G_{12}$).

Take $F_1 - \underline{F}_1$ to be any nondecreasing and right continuous function such that $(F_1 - \underline{F}_1)(\infty) = G_2(\infty)$, $(F_1 - \underline{F}_1)(y) = 0$ if either $y \le \max(x_1', x_2')$ or $\tilde{G}_2(y) = 0$, and $(F_1 - \underline{F}_1)(y) < \tilde{G}_2(y)$ otherwise. Similarly, let $F_2 - \underline{F}_2$ be any nondecreasing and right continuous function such that $(F_2 - \underline{F}_2)(\infty) = G_1(\infty)$, $(F_2 - \underline{F}_2)(y) = 0$ if either $y \le \max(x_1', x_2')$ or $\tilde{G}_1(y) = 0$, and $(F_2 - \underline{F}_2)(y) < \tilde{G}_1(y)$ otherwise. Such choices of F_1 and F_2 clearly satisfy the conditions of Theorem 1.

By construction, $P(X_2 \le \max(x_1', x_2'),\ X_1 < X_2) = 0 = P(X_1 \le \max(x_1', x_2'), X_2 < X_1)$, and so there is equality in (15). Equalities in the right-hand bounds of (17) and (18) hold also for this joint distribution.

A similar construction will show that the remaining inequalities can actually be equalities: take $F_1 - \underline{F}_1$ and $F_2 - \underline{F}_2$ nondecreasing and right continuous such that

$$
(F_1 - \underline{F}_1)(y) \begin{cases} = G_2(\infty), & \text{if } y = \infty, \\ = 0, & \text{if } y < \max(x_1', x_2'), \\ = \tilde{G}_2(\max(x_1', x_2')), & \text{if } \tilde{G}_2(y) = \tilde{G}_2(\max(x_1', x_2')), \\ < \tilde{G}_2(y), & \text{otherwise} \end{cases}
$$

and

$$
(F_2 - \underline{F}_2)(y) \begin{cases} = G_1(\infty), & \text{if } y = \infty, \\ = 0, & \text{if } y < \max(x_1', x_2'), \\ = \tilde{G}_1(\max(x_1', x_2')), & \text{if } \tilde{G}_1(y) = \tilde{G}_1(\max(x_1', x_2')), \\ < \tilde{G}_1(y), & \text{otherwise.} \end{cases}
$$

Now, if $x_1' < x_2'$, then there is equality in (16) if and only if $P(x_1' < X_1 < x_2', X_2 > x_2', X_1 < X_2) = 0$. This holds if $P(X_1 < x_2' < X_2) = 0$, which in turn is equivalent to $\tilde{G}_1(x_2') = (F_2 - \underline{F}_2)(x_2')$. Hence equality does indeed hold. Similar arguments show that equality holds in (16) if $x_2' < x_1'$. Equality always holds if $x_1' = x_2'$.

This choice of joint distribution shows that the lower bound on the joint survival function is pointwise sharp. It also attains equalities in the left-hand bounds of (17) and (18). □

It is well known that the pairs (F_1, F_2) of distribution functions for which there exist random variables X_1 and X_2, with $X_1 \sim F_1$, $X_2 \sim F_2$ and $X_1 \leq X_2$ a.s., are precisely those with $F_1 \geq F_2$. Theorem 1 provides as an immediate corollary a characterization of the pairs (F_1, F_2) admitting a joint distribution with $X_1 < X_2$ a.s.

COROLLARY 2. *The pair (F_1, F_2) of distribution functions admits a joint distribution of two random variables X_1 and X_2 with $X_1 \sim F_1$, $X_2 \sim F_2$ and $X_1 < X_2$ a.s. if and only if* (i) $F_1(x-) \geq F_2(x)$ *for all* $x \in \mathbb{R}$ *and* (ii) *the Lebesgue measure of the range set* $\{F_2(x) | F_1(x-) = F_2(x)\}$ *is zero.*

2.1. *Examples.* Our first example is an application of the above results to families of distributions ordered by monotone likelihood ratio.

2.1.1. *Identifiability for MLR families of distributions.*

DEFINITION 3. A point t is a *calibrator* of a distribution $F(\cdot; \theta_0)$, in the context of a parametric family $\{F(\cdot; \theta);\ \theta \in \Theta\}$, if the value θ_0 is a maximum likelihood estimate of the parameter θ when the (single) observation is t.

For example, if the family is exponential type with density $f(x; \theta) = h(x)\Psi(\theta)\exp\{x\theta\}$ and support $[0, \infty)$, then the unique calibrator of a member of the family is its mean.

For families of distributions ordered by monotone likelihood ratio there is the following interpretation of the calibrator. Suppose that the family is smooth enough and that $f(x; \theta'')/f(x; \theta')$ is strictly increasing in x for every pair $\theta' < \theta''$ in Θ. Then the calibrator of $F(\cdot; \theta_0)$ is the limit as $\theta \to \theta_0$ of the point where the two densities $f(\cdot; \theta)$ and $f(\cdot; \theta_0)$ cross each other.

With this interpretation, the following characterization of identifiability becomes clear:

Assume that X_1 is distributed according to some member $F(\cdot; \theta_0)$ of a family of distributions ordered by monotone likelihood ratio as above, and let X_2 have conditional distribution satisfying $P(X_2 = 0 | X_1) = 1$ on $\{X_1 \leq \eta\}$ and $P(X_2 > X_1 | X_1) = 1$ on $\{X_1 > \eta\}$. Then

$$G_1(x) = \begin{cases} 0, & \text{if } x \leq \eta, \\ F_1(x) - F_1(\eta), & \text{if } x > \eta. \end{cases}$$

Now, hypothesizing that the distribution of X_1 is in the family $F(\cdot;\theta)$, the parameter θ_0 is identifiable if and only if the calibrator of $F(\cdot;\theta_0)$ is in $[\eta,\infty)$.

This is so because the upper Peterson bound is sharp and claims that $\theta \geq \theta_0$, while comonotonicity admits some $\theta < \theta_0$ only if the two densities $f(\cdot;\theta)$ and $f(\cdot;\theta_0)$ cross each other strictly to the left of the cutoff point η.

If the hypothesized family of distributions is smooth enough, the lower Peterson bound can be sharp only in the trivial case where $\eta = 0$.

As a particular example, consider the family of exponential distributions and let $\theta_0 = \eta = 1$. Comonotonicity identifies the failure rate as being equal to 1, while the lower Peterson bound

$$1 - e^{-\lambda x} \geq e^{-1} - e^{-x}$$

only implies that $\lambda \geq 0.1355$. In other words, the Peterson bound claims $E(X_1)$ to be between 1 and 7.38 while comonotonicity identifies this mean as 1.

This is an illustration of the potentially significant difference between the pointwise Peterson and the comonotone bounds.

As an illustration of a case where comonotonicity contributes next to nothing beyond the pointwise Peterson bounds, consider the following example.

2.1.2. *Independent censoring.* Let X_1 and X_2 be independent, exponentially distributed with parameters λ_1 and λ_2, respectively. Then

$$G_1(x) = \frac{\lambda_1}{\lambda_1+\lambda_2}(1 - \exp[-(\lambda_1+\lambda_2)x]),$$

$$G_2(x) = \frac{\lambda_2}{\lambda_1+\lambda_2}(1 - \exp[-(\lambda_1+\lambda_2)x]).$$

Hypothesizing exponentiality of X_1, the Peterson bounds claim that its failure rate λ satisfies $\lambda \in [\lambda_1, \lambda_1+\lambda_2]$. [This is easy to see after realizing that $\lim_{x\downarrow 0} G_1(x)/x = \lambda_1$.] Comonotonicity improves the result by merely ruling out $\lambda = \lambda_1+\lambda_2$. The next subsection generalizes the previous example by showing, as announced in the Introduction, that if X_1 is hypothesized to be exponentially distributed (without assuming that the censored data admits independent censoring with exponential marginals), then the feasible values of the failure rate λ of X_1 constitute an interval.

2.2. *An exponential marginal.* In this subsection the exponential distribution with parameter λ will be denoted by $H_\lambda(x) = 1 - e^{-\lambda x}$. If one is given subdistribution functions G_1 and G_2 and a comonotone representation $H_\lambda(x) = G_1(x) + (H - G_1)(x)$ with $(H - G_1) \leq \tilde{G}_2$ (with strict inequality except for a set of values of Lebesgue measure zero), then according to Theorem 1, there is a joint distribution for (X_1, X_2) with subdistribution functions G_1 and G_2, and such that X_1 has marginal distribution H_λ. In this case we say that H_λ is a *possible marginal distribution* given G_1 and G_2.

PROPOSITION 1. *Given subdistribution functions G_1 and G_2 as above, with $G_1(\infty) > 0$, the set of compatible failure rates*

$$\Lambda = \{\lambda | H_\lambda \text{ is a possible marginal of } X_1 \text{ given } G_1, G_2\}$$

is a bounded interval (possibly empty or one point).

PROOF. As the pointwise Peterson bounds clearly imply that λ belongs to half-lines, it is enough to show that the comonotonicity condition implies that λ belongs to an interval. The set of compatible failure rates is now the intersection of all of these intervals and is therefore also an interval.

Define the function $f_1(x)$ to be the upper Dini derivative of G_1 at x,

$$f_1(x) = \limsup_{y \to x} \frac{G_1(y) - G_1(x)}{y - x}.$$

It is easy to see that if $G_1 \in \mathscr{C}(H_\lambda)$, then $0 \le f_1(x) \le \lambda e^{-\lambda x}$ for all $x > 0$. For any such fixed x we have

$$\log f_1(x) \le \log \lambda - \lambda x,$$

which is satisfied by a (possibly unbounded) interval of λ values. Since $f_1(x) > 0$ for some x, the intersection of these intervals over all x gives us the bounded interval of λ values for which H_λ satisfies the comonotonicity condition. □

As evidence that the lower bounds may not be too wasteful in practice, it is interesting to notice that, whenever the independent feasible solution involves only exponential marginals, none of the competing risks could possibly be exponential with a lower failure rate.

2.3. *The case of multiple competing risks.* A number of authors have considered competing risk problems with more than two risks. Given m random variables $\{X_1, \ldots, X_m\}$, let $T = \min\{X_1, \ldots, X_m\}$ and set $I = \{i\colon X_i = T\}$. For any nonempty subset $K \subseteq \{1, \ldots, m\}$ define the subdistribution function

$$G_K(x) = P(T \le x, I = K).$$

The Peterson-type lower bound for the distribution function of the random variable X_i is

$$\underline{F}_i(x) = \sum_{K\colon i \in K} G_K(x),$$

while the improved Peterson-type upper bound [cf. (3)] is

$$\overline{F}_i(x) = \underline{F}_i(x) + \sum_{K\colon i \notin K} \tilde{G}_K(x).$$

Theorem 1 may be generalized to the following theorem.

THEOREM 2. (i) *Let $X_1, X_2, \ldots, X_m$ be random variables. Then, using the notation established above, we have the following*:

1. $F_i = \underline{F}_i + (F_i - \underline{F}_i)$ *is a nonnegative comonotone representation of a right continuous function, for* $i = 1, \ldots, m$;
2. $F_i(x) \leq \overline{F}_i(x)$ *for all* x, *and the Lebesgue measure of the range set* $\{\tilde{G}_K(x) | F_i(x) = \overline{F}_i(x)\}$ *is zero for every* K *with* $i \notin K$, *for* $i = 1, \ldots, m$;
3. (a) $F_i(-\infty) = 0$ *and* $\underline{F}_i(\infty) = 1$, *for* $i = 1, \ldots, m$;
   (b) $(F_i - \underline{F}_i)(\infty) = (\overline{F}_i - \underline{F}_i)(\infty)$, *for* $i = 1, \ldots, m$;
   (c) $\sum_K G_K(\infty) = 1$.

(ii) *If nondecreasing right continuous functions* $(F_i,\ i = 1, \ldots, m)$ *and* $(G_K,\ K \subseteq \{1, \ldots, m\})$ *satisfy the conditions* 1–3 *of* (i), *then there are random variables* $X_1, \ldots, X_m$ *for which these are, respectively, their marginal distributions and their subdistribution functions.*

SKETCH OF PROOF. Necessity follows as in the proof of Theorem 1.

Sufficiency of these conditions may be proven by a construction analogous to the strip construction shown in Figure 2, with one strip for every nonempty $K \subseteq \{1, \ldots, m\}$, except that not all needed functions are specified by the subdistribution functions: while X_i is properly defined for $i \in K$ on the strip corresponding to the subset K by inverting G_K, this is not the case for $i \notin K$ because the events $\{X_i \leq x,\ I = K\}$ are not (T, I)-measurable. The gap $F_i(x) - \underline{F}_i(x)$ is equal to $\sum_{K: i \notin K} P(X_i \leq x,\ I = K)$, but this sum involves a single summand only if $m = 2$. In the general case, these summands must be created in a consistent way from their sums. We skip most details but present the main lemma which allows the construction. □

LEMMA 2. *Let* $f_i: \mathbb{R} \to \mathbb{R}^+$, *for* $i = 1, \ldots, n$, *and* $g: \mathbb{R} \to \mathbb{R}^+$, *with* $g \leq \sum_{i=1}^n f_i$, *be nondecreasing, with* $f_i(-\infty) = g(-\infty) = 0$. *Then there exists a nonnegative comonotone representation* $(g_1, g_2, \ldots, g_n)$ *of* g *such that* $g_i \leq f_i$ *for* $i = 1, \ldots, n$, *with the further property that if* $g(x) < \sum_{i=1}^n f_i(x)$ *and all the given functions are continuous at* x, *then, for every* $i = 1, \ldots, n$, *either* $g_i(x) < f_i(x)$ *or* f_i *is constant on some nonempty interval* $]x - \varepsilon, x]$.

SKETCH OF A PROOF OF LEMMA 2. First replace the upper functions f_i by upper functions ϕ_i with $0 \leq \phi_i \leq f_i,\ g \leq \sum \phi_i = \phi$, such that if $\phi(x) < \sum_{i=1}^n f_i(x)$ and all the given functions are continuous at x, then, for every $i = 1, \ldots, n$, either $\phi_i(x) < f_i(x)$ or f_i is constant on some nonempty interval $]x - \varepsilon, x]$.

A choice of such functions is given by

$$\phi_i(x) = \sup_{y \leq x} \left\{ f_i(y) \frac{g(y)}{\sum_{j=1}^n f_j(y)} \right\}, \qquad 1 \leq i \leq n$$

(with the convention that $0/0 = 0$).

As a second step, prove the existence of functions g_i, without worrying about the further property: let

$$g_{(2)}(x) = \sup_{y \leq x} (g(y) - \phi_1(y))^+,$$
$$g_1(x) = g(x) - g_{(2)}(x).$$

Then $(g_1, g_{(2)})$ solves the problem for $n = 2$ (with $\sum_2^n g_i$ playing the role of g_2) and what remains is the original problem with a smaller number of variables (with $g_{(2)}$ playing the role of g). $\square$

3. Kolmogorov–Smirnov tests and confidence sets. For the rest of the paper we shall consider the problem of testing a hypothesized distribution function H for the marginal distribution of X_1. A confidence set of feasible choices of H is defined as the class of H *not* rejected by the test.

We assume from now on that $P(X_1 = X_2) = 0$ and make no assumptions about the marginal distribution of X_2. The distribution function H should be assumed to be continuous for the application of asymptotic results about the Kolmogorov–Smirnov statistic. We will otherwise not assume continuity explicitly, except in the subsections of this section.

Let $(X_1^{(1)}, X_2^{(1)}), (X_1^{(2)}, X_2^{(2)}), \ldots, (X_1^{(n)}, X_2^{(n)})$ be i.i.d. random vectors with $X_1^{(i)}$ distributed according to some distribution F_1. Suppose further that only $T^{(i)} = \min(X_1^{(i)}, X_2^{(i)})$ and $I(X_1^{(i)} < X_2^{(i)})$ are observed, and consider the empirical subdistribution function of X_1,

$$\hat{G}_1(x) = \frac{1}{n}\sum_{i=1}^{n} I\big(X_1^{(i)} \le x,\ X_1^{(i)} < X_2^{(i)}\big), \tag{19}$$

contrasted with the unseen empirical distribution function of X_1,

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_1^{(i)} \le x). \tag{20}$$

Clearly, $\hat{G}_1 \in \mathscr{C}^+(\hat{F})$. If we could observe $\hat{F}$, then we could perform a Kolmogorov–Smirnov test for H, by calculating

$$\sup_x |\hat{F}(x) - H(x)|. \tag{21}$$

As we do not observe $\hat{F}$ this is impossible. We can, however, use the Peterson bounds for $\hat{F}$, which now are

$$\hat{G}_1(x) \le \hat{F}(x) \le \hat{G}_1(x) + \hat{\tilde{G}}_2(x), \tag{22}$$

in order to estimate (21) or its one-sided versions. First observe that

$$\inf_x(\hat{F}(x) - H(x)) \le \inf_x(\hat{G}_1(x) + \hat{\tilde{G}}_2(x) - H(x)), \tag{23}$$

so if the right-hand side of (23) is negative and below a critical point for the one-sided Kolmogorov–Smirnov test, the hypothesized H can be rejected. Similarly,

$$\sup_x(\hat{G}_1(x) - H(x)) \le \sup_x(\hat{F}(x) - H(x)), \tag{24}$$

so if the left-hand side of (24) is positive and above a critical value of the one-sided Kolmogorov–Smirnov test, then H can be rejected.

There is, however, an important improvement that can be made to the second estimate (24), for we have not yet used the fact that $\hat{F} = \hat{G}_1 + (\hat{F} - \hat{G}_1)$ must be a comonotone representation. Example 2.1.1 shows that this approach may reject more hypothesized distributions H than by using only the Peterson bounds.

Rather than using the estimate (24), we try to find a nonnegative comonotone representation $H = H_1 + (H - H_1)$ for which H_1 is close to $\hat{G}_1$. Lemma 3 below asserts the existence of an H_1 which fits $\hat{G}_1$ pointwise better than H fits $\hat{F}$. Therefore, the best fit

$$\inf_{H_1 \in \mathscr{C}^+(H)} \sup_x |\hat{G}_1(x) - H_1(x)|$$

over *all possible* comonotone representations provides, whatever $\hat{F}$ is, a lower bound on $\sup_x |\hat{F}(x) - H(x)|$.

Part (i) of the next theorem summarizes this discussion and is clearly a corollary of the following Lemma 3. The proof of part (ii) follows Lemma 3.

THEOREM 3. *Given an empirical subdistribution function*

$$\hat{G}_1(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_1^{(i)} \le x,\ X_1^{(i)} < X_2^{(i)})$$

and a distribution function H, *let*

$$D = \inf_{H_1 \in \mathscr{C}^+(H)} \sup_x |\hat{G}_1(x) - H_1(x)|. \tag{25}$$

Then we have the following:

(i) *For any distribution function* $\hat{F}$ *such that* $\hat{G}_1 \in \mathscr{C}^+(\hat{F})$,

$$\sup_x |\hat{F}(x) - H(x)| \ge D.$$

(ii) *There exist* $H_1 \in \mathscr{C}^+(H)$ *such that*

$$\sup_x |\hat{G}_1(x) - H_1(x)| = D.$$

The lemma referred to above is the following.

LEMMA 3. *Let* f, h: $\mathbb{R} \to \mathbb{R}$ *be nondecreasing and right continuous*, *and let* $f = g_1 + (f - g_1)$ *be a comonotone representation of* f. *Then there exists a comonotone representation* $h = h_1 + (h - h_1)$ *of* h *such that*

$$0 \le g_1(x) - h_1(x) \le f(x) - h(x) \quad \text{if } h(x) \le f(x) \tag{26}$$

and

$$f(x) - h(x) \le g_1(x) - h_1(x) \le 0 \quad \text{if } f(x) \le h(x). \tag{27}$$

PROOF. We first modify f slightly. Define the set $B \subset \mathbb{R}$ to be those discontinuity points x of f for which $h(x-) \geq f(x-)$ and $h(x) \leq f(x)$. Define $f'(x) = h(x)$ for $x \in B$ and $f'(x) = f(x)$ otherwise. It is clear that f' is nondecreasing. Modify g_1 so as to have a comonotone representation of f': set

$$g_1'(x) = \begin{cases} g_1(x-) + \dfrac{h(x) - f(x-)}{f(x) - f(x-)}(g_1(x+) - g_1(x-)), & \text{if } x \in B, \\ g_1(x), & \text{otherwise.} \end{cases}$$

It is then clear that f', g_1' and $f' - g_1'$ are nondecreasing, and that f' is right continuous except on B.

Let

$$r^+(x) = \inf_{y \geq x}\{g_1'(y) + (h(y) - f'(y))^+\} \tag{28}$$

and

$$r^-(x) = \sup_{y \leq x}\{g_1'(y) - (h(y) - f'(y))^-\}, \tag{29}$$

where

$$z^+ = \max(z, 0), \qquad z^- = (-z)^+ = -\min(z, 0). \tag{30}$$

We claim that the function

$$h_1(x) = \begin{cases} r^+(x), & \text{if } h(x) \geq \tilde{f}(x), \\ r^-(x), & \text{if } h(x) \leq f'(x), \end{cases} \tag{31}$$

satisfies all requirements, and we proceed now to prove this statement.

The inequalities

$$r^+(x) = \inf_{y \geq x}\{g_1'(y) + (h(y) - f'(y))^+\} \geq \inf_{y \geq x}\{g_1'(y)\} = g_1'(x)$$

and

$$r^+(x) = \inf_{y \geq x}\{g_1'(y) + (h(y) - f'(y))^+\} \leq g_1'(x) + (h(x) - f'(x))^+$$

can be summarized as

$$g_1'(x) \leq r^+(x) \leq g_1'(x) + (h(x) - f'(x))^+. \tag{32}$$

Similarly,

$$g_1'(x) - (h(x) - f'(x))^- \leq r^-(x) \leq g_1'(x). \tag{33}$$

Inequalities (32) and (33) clearly prove (26) and (27) for $x \in B^c$, a dense subset of $\mathbb{R}$. Since f, g_1, h and h_1 are all right continuous, (26) and (27) hold on the whole of $\mathbb{R}$.

It remains to prove that h_1 and $(h - h_1)$ are nondecreasing functions. Let $x < z$. We will compare $h_1(x)$ with $h_1(z)$ and $(h - h_1)(x)$ with $(h - h_1)(z)$ in each of four (somewhat overlapping) cases covering all possibilities.

Case 1 [$h(x) \le f'(x)$ and $h(z) \le f'(z)$]. In this case, $h_1(x) \le h_1(z)$ because $h_1 = r^-$ in these two points, and r^- is nondecreasing by construction, as a supremum over an increasing class of sets. As for $(h - h_1)$, if $r^-(x) = r^-(z)$, there is nothing to prove. Otherwise, $r^-(z)$ can be approximated arbitrarily closely by values of the function $g_1'(y) - (h(y) - f'(y))^-$ in the interval $[x, z)$. Consider such y:

$$\begin{aligned} h_1(z) - h_1(x) &\approx \{g_1'(y) - (h(y) - f'(y))^-\} - h_1(x) \\ &\le \{g_1'(y) - (h(y) - f'(y))^-\} - \{g_1'(x) - (h(x) - f'(x))^-\} \\ &= \{g_1'(y) - (h(y) - f'(y))^-\} - \{f'(x) + (h(x) - f'(x))\} \\ &\le \{g_1'(y) + (h(y) - f'(y))\} - \{g_1'(x) + (h(x) - f'(x))\} \\ &\le h(y) - h(x) \le h(z) - h(x). \end{aligned}$$

Thus,

$$(h - h_1)(z) - (h - h_1)(x) \ge 0.$$

Case 2 [$h(x) \ge f'(x)$ and $h(z) \ge f'(z)$]. This case can be handled analogously to the previous case, working with r^+ (rather than with r^-), which is nondecreasing by construction as well, as the infimum over a decreasing class of sets.

Case 3 [$h(x) \le f'(x)$ and $h(z) \ge f'(z)$]. In this case,

$$h_1(x) = r^-(x) \le g_1'(x) \le g_1'(z) \le r^+(z) = h_1(z).$$

As for $(h - h_1)$, we proceed as follows:

$$\begin{aligned} h_1(z) - h_1(x) &= r^+(z) - r^-(x) \\ &\le \{g_1'(z) + (h(z) - f'(z))^+\} - \{g_1'(x) - (h(x) - f'(x))^-\} \\ &= g_1'(z) + h(z) - f'(z) - g_1'(x) - h(x) + \tilde{f}(x), \end{aligned}$$

and the proof proceeds as in the first case.

Case 4 [$h(x) > f'(x)$ and $h(z) < f'(z)$]. Consider

$$x_0 = \sup\{y \in [x, z] | h(u) > f'(u) \text{ for all } u \in [x, y]\}. \tag{34}$$

Clearly $x \notin B$ and so f' is right continuous at x, which implies (together with right continuity of h) that $x_0 > x$. By the definition of f', $h(z) < f'(z)$ implies that $h(z-) < f'(z-)$. Hence $x_0 < z$. We now have $h(x_0-) \ge f'(x_0-)$ and $h(x_0) \le f'(x_0+)$. By the definition of f', this implies that $h(x_0) = f'(x_0)$. The problem now splits into two subproblems (in one of which x_0 takes the role of z and in the other the role of x) handled in full by the first two cases. □

PROOF OF THEOREM 3(ii). The expression $\sup_x |\hat{G}_1(x) - H_1(x)|$ is a function of H_1 only via its left- and right-hand limits at the discontinuities of $\hat{G}_1$ and at $\pm\infty$, and it is a continuous function of these variables. Let $u^+(x_0), u^-(x_1), u^+(x_1), u^-(x_2), \ldots, u^+(x_k), u^-(x_{k+1})$ be these limit values,

with the obvious notation, including $x_0 = -\infty$ and $x_{k+1} = +\infty$. The comonotonicity conditions may be phrased in terms of these variables as follows:

$$u^+(-\infty) = 0$$

and

$$0 \le u^-(x_{i+1}) - u^+(x_i) \le H(x_{i+1}-) - H(x_i), \qquad 0 \le i \le k,$$
$$0 \le u^+(x_i) - u^-(x_i) \le H(x_i) - H(x_i-), \qquad 1 \le i \le k.$$

Since these inequalities describe a closed subset of a $(2k+1)$-dimensional unit cube, it is clear that the infimum over H_1 is actually a minimum. □

3.1. *A dynamic programming approach to constructing an optimal* H_1. With Kolmogorov–Smirnov applications as motivation and the proof of Theorem 3(ii) as a guideline, we shall now assume for simplicity that the hypothesized distribution H is continuous with $H(0) = 0$ and give a simple dynamic programming algorithm which enables a fast (linear complexity) computation of the distance D [see (25)].

Using the notation (x_i) from the previous section with a reinterpretation of x_0 as 0, we may define functions $V_i\colon \mathbb{R}^+ \to \mathbb{R}$ by

$$V_i(t) = \inf\left\{\sup_{x \ge x_i} |\hat{G}_1(x) - H_1(x)| \;\middle|\; H_1(x_i) = t\right\}$$

for $i = 0, \ldots, k$. Since $H_1(0) = 0$, the required optimal distance is just $V_0(0)$. Now, although the definition of V_i does not look very promising from a computational point of view, we can give a simple inductive formula:

PROPOSITION 2. *The functions* $V_i(t)$ *are given by*

$$V_i(t) = \begin{cases} \max\{|\hat{G}_1(x_i) - t|, |\hat{G}_1(x_i-) - t|, \min\limits_{t \le y \le t+\Delta_i} V_{i+1}(y)\}, & \text{if } 0 \le i < k, \\ \max\{|\hat{G}_1(x_n) - t|, |\hat{G}_1(x_n-) - t|\}, & \text{if } i = k, \end{cases}$$

where $\Delta_i = H(x_{i+1}) - H(x_i)$.

PROOF. For $i = k$ the formula is clear since an optimal H_1 may be taken to be constant on the interval $[x_k, \infty)$. For $i < k$ we may use the fact that the supremum distance between $\hat{G}_1$ and H_1 on the interval $[x_i, x_{i+1}]$ is achieved as a left or right limit at one of x_i and x_{i+1}. Hence $V_i(t)$ equals the maximum of $|\hat{G}_1(x_i) - H_1(x_i)|$, $|\hat{G}_1(x_i-) - H_1(x_i)|$ and the optimal distance on $[x_{i+1}, \infty)$. This optimal distance equals the last term in the claimed expression for $V_i(t)$ since a function $H_1 \in \mathscr{C}(H)$ with $H_1(x_i) = t$ may pass through any point of $[t, t + \Delta_i]$ at $x = x_{i+1}$. □

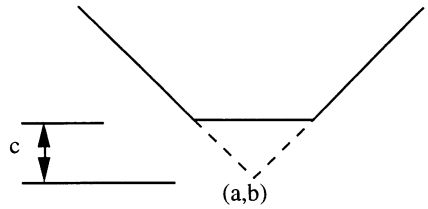

FIG. 3. *A trough function.*

DEFINITION 4. A function $V: \mathbb{R} \to \mathbb{R}$ is a *trough function* if there are reals a, b and c with $c \geq 0$ such that

$$V(y) = \begin{cases} -y + (b + a), & \text{if } y \leq a - c, \\ b + c, & \text{if } a - c \leq y \leq a + c, \\ y + (b - a), & \text{if } y \geq a + c \end{cases}$$

(see Figure 3). The trough has *vertex* (a, b) and a *base* at height c above the vertex.

A little thought shows that the maximum of two trough functions is again a trough function, and since V_n is certainly a trough function we have the following corollary.

COROLLARY 3. *Each function V_i is a trough function.*

The importance of this result is that the recurrence formula for the V_i given by Proposition 2 can easily be rewritten in terms of a recurrence relation for the corresponding parameters (a_i, b_i, c_i), which enables a linear-complexity computation of $D = V_0(0)$.

4. Consistency of the confidence set. As expressed in the Introduction, the conservative Kolmogorov–Smirnov test based on (25) may be used to build conservative confidence sets. The collection of distributions H for which $\sqrt{n}D$ does not exceed some critical point constitute a set that we rightfully term *confidence set*, because under any theoretically feasible distribution the probability that D exceeds the critical point (chosen to be a critical point for the Kolmogorov–Smirnov statistic) does not exceed the probability that the Kolmogorov–Smirnov statistic [lhs of (25)] exceeds that point, that is, the preassigned confidence coefficient.

In this section we show that these confidence sets have a consistency property.

Write $\mathbf{z} = (z_1, \ldots, z_n, \ldots)$ for an infinite sequence of realizations $z_i = (x_1^{(i)}, x_2^{(i)})$ from the unknown joint distribution of (X_1, X_2). This joint distribution determines subdistribution functions (that can be estimated from observable data), which in turn via Theorem 1 specify a feasible set J of

marginal distributions for X_1. The unknown true marginal distribution, F, of X_1 is a member of J.

For every n and α we obtain a confidence set that we denote $\hat{J}_\alpha^{(n)}(\mathbf{z})$. We consider the closure of the lim sup of this sequence, as $n \to \infty$, and denote this set $\hat{J}_\alpha(\mathbf{z})$. Such a set is called a *limit confidence set*. Note that the limit confidence sets satisfy a nesting property, for $\hat{J}_\alpha(\mathbf{z}) \subset \hat{J}_{\alpha'}(\mathbf{z})$ when $\alpha > \alpha'$.

The confidence set satisfies $P(F \in \hat{J}_\alpha^{(n)}(\mathbf{z})) \geq 1 - \alpha$, for every n (with "$\geq$" instead of the usual "$=$" because of the conservativeness of our criterion).

For distribution functions H not in J we have the following proposition.

PROPOSITION 3. *For a distribution function $H \notin J$, with probability* 1, *there is an N so that, for $n > N$, $H \notin \hat{J}_\alpha^{(n)}(\mathbf{z})$.*

PROOF. If $H \notin J$, then $H = G_1 + (H - G_1)$ is not a comonotone representation, so $H - G_1$ is not monotone increasing. Hence there are points $x_1 < x_2$ such that $H(x_1) - G_1(x_1) = \beta_1$ and $H(x_2) - G_1(x_2) = \beta_2$ with $\beta_1 > \beta_2$.

Now, given $\beta_1 > \beta_1' > \beta_2' > \beta_2$, with probability 1 we have

$$H(x_1) - \hat{G}_1^{(n)}(x_1) > \beta_1', \qquad H(x_2) - \hat{G}_1^{(n)}(x_2) < \beta_2',$$

for large enough n.

Set $\gamma = (\beta_1' - \beta_2')/2$. We claim that if $H_1 \in \mathscr{C}^+(H)$, then

$$|H_1(x_1) - \hat{G}_1^{(n)}(x_1)| < \gamma$$

implies

$$|H_1(x_2) - \hat{G}_1^{(n)}(x_2)| > \gamma.$$

This holds because if $H_1(x_1) - \hat{G}_1^{(n)}(x_1) < \gamma$, then $H(x_1) - H_1(x_1) > \beta_1' - \gamma$ so comonotonicity of $H = H_1 + (H - H_1)$ implies that $H(x_2) - H_1(x_2) > \beta_1' - \gamma$. But then

$$\hat{G}_1^{(n)}(x_2) - H_1(x_2) > -\beta_2' + \beta_1' - \gamma = \gamma.$$

This proves the claim.

The claim shows that

$$\inf_{H_1 \in \mathscr{C}^+(H)} \sup_x |H_1(x) - \hat{G}_1^{(n)}(x)| \geq \gamma,$$

for large enough n, with probability 1. Hence the Kolmogorov–Smirnov test will reject H, with probability 1, for any α. □

Having shown that any $H \notin J$ is eventually rejected with probability 1 for large enough n, we have to show that any $F \in J$ is not rejected infinitely often.

Let ζ_i be a sequence of i.i.d. random variables having the uniform distribution on $(0, 1)$. The corresponding empirical process is

$$U_n(t) = n^{-1/2} \sum_{i=1}^{n} [\mathbf{1}\{\zeta_i \leq t\} - t]$$

for $0 \le t \le 1$. The empirical central limit theorem [10] states that the sequence U_n converges in distribution, as a sequence of random cadlag functions on $[0, 1]$ to the Brownian bridge.

PROPOSITION 4. *Let F be in the feasible set J. Given any $D > 0$ there exists with probability one a sequence of integers $n_1, n_2, \ldots$ such that*

$$\sqrt{n_i} \sup_x |\hat{F}^{(n_i)}(x) - F(x)| < D$$

for every i. That is, with probability 1,

$$F \in \hat{J}_\alpha(\underline{z})$$

for any α for which $D \le D_\alpha$.

PROOF. The sequence of i.i.d. realizations $(x_1^{(1)}, x_1^{(2)}, \ldots)$ gives rise to a sequence of independent uniform variables $\zeta_i = F(x_i)$. Letting U_n be the corresponding empirical process one has

$$\sqrt{n} \sup_x |\hat{F}^{(n)}(x) - F(x)| = \sup_{t \in (0,1)} |U_n(t)|.$$

Now, by the empirical central limit theorem, with probability 1 we can find n_1 such that

$$\sup_{t \in (0,1)} |U_{n_1}(t)| < D.$$

Denote by $U_m^{(1)}$ the empirical process constructed from the sequence ζ_{n_1+1}, $\zeta_{n_1+2}, \ldots$. This is clearly independent of U_{n_1}. With probability 1, we can find m such that

$$\sup_{t \in (0,1)} |U_m^{(1)}(t)| < D/2,$$

and, additionally, $\sqrt{n_1}/\sqrt{n_1+m} < \frac{1}{4}$ and $\sqrt{m}/\sqrt{n_1+m} - 1 < \frac{1}{4}$. Writing $n_2 = n_1 + m$, we then have

$$\begin{aligned}
\sup |U_{n_2}(t)| &\le \sup |U_m^{(1)}(t)| + \sup |U_{n_2}(t) - U_m^{(1)}(t)| \\
&\le \frac{D}{2} + \sup \left| \frac{1}{\sqrt{n_2}} \left[\sqrt{n_1} U_{n_1}(t) + \sqrt{m} U_m^{(1)}(t) \right] - U_m^{(1)}(t) \right| \\
&\le \frac{D}{2} + \sup \frac{\sqrt{n_1}}{\sqrt{n_2}} |U_{n_1}(t)| + \sup \left(\frac{\sqrt{m}}{\sqrt{n_2}} - 1 \right) |U_m^{(1)}(t)| \\
&\le \frac{D}{2} + \frac{D}{4} + \frac{D}{4} = D.
\end{aligned}$$

Continuing in this way we construct a sequence $n_1, n_2, n_3, \ldots$ as claimed. □

THEOREM 4. *For any* $\alpha > 0$, *with probability* 1:

(i) *The feasible set* J *is contained in* $\hat{J}_\alpha(\underline{z})$.
(ii) *The difference* $\hat{J}_\alpha(\underline{z}) - J$ *contains no open set of distribution functions*.

PROOF. (i) Take a countable dense subset of the feasible set J. For each distribution function in this dense subset we know that it lies in $\hat{J}_\alpha(\underline{z})$ with probability 1. Hence with probability 1 they all lie in that set. However, since $\hat{J}_\alpha(\underline{z})$ is by definition closed, it must also contain J.

(ii) Work as in (i) but now with a countable set that is dense in the complement of J. □.

REMARK. The random set $\hat{J}_\alpha(\underline{z}) - J$ may be some nonempty nowhere dense set.

5. Applications to the exponential family of marginals. The exponential family is parameterized by the single real parameter λ, so if we hypothesize that the underlying distribution of X_1 is exponential, we can construct a confidence interval of λ values.

Since the set J of feasible values is necessarily an interval, it is by Theorem 4 equal to the closure of the interior of the limit confidence interval $\hat{J}_\alpha(\underline{z})$, with probability 1.

5.1. *Simulation results.* The particular example given in Section 2.1.1 was simulated 100 times using sample size 1000. The Kolmogorov–Smirnov statistic for the lower Peterson bound [see (24)] gives, at the 90% confidence level, a mean lower bound of $\underline{\lambda} = 0.122$, with standard deviation 0.0077. This may be compared with the theoretical lower bound $\lambda = 0.1355$ achieved via the Peterson lower bound, as reported in Section 2.

The Kolmogorov–Smirnov statistic yielding lower bounds at the 90% confidence level via comonotonicity gives a mean lower bound estimate of 0.2929, with standard deviation 0.0429. While sharper than the above, given that the theoretical value is $\lambda = 1$, this bounding technique still seems coarse. However, when the cutoff point for the monitoring of failures is reduced from $x = 1$ to $x = 0.5$, the mean lower bound estimate is 0.5983, with standard deviation 0.0606. If the cutoff point is further reduced to $x = 0$ (i.e., all failures are observed), the mean lower bound estimate becomes 0.9492, with standard deviation 0.0342. The latter corresponds to the conventional construction of bounds on λ via the Kolmogorov–Smirnov distance between empirical and hypothesized distribution functions.

Typically, the upper bound on λ obtained via comonotonicity does not improve the estimates obtained via the Peterson upper bounds.

The example given in Section 2.1.2 was simulated 100 times using a sample size of 1000, letting $\lambda_1 = 1$ and $\lambda_2 = 0.5$. As shown above, comonotonicity and pointwise bounds agree except at the endpoints, and claim that $\lambda_1 \in [1, 1.5)$. The Kolmogorov–Smirnov statistics built upon comonotonicity and upon the

lower Peterson bound coincided in the range relevant for the construction of 90% confidence intervals, and yielded a mean lower bound of 0.8050 with standard deviation 0.051. The mean upper bound was 1.5649 with standard deviation 0.057.

Acknowledgments. The authors would like to thank Roger Cooke and Ludolf Meester for helpful discussions. The second author is grateful for the hospitality of the TU Delft.

REFERENCES

[1] COOKE, R. M. (1993). The total time on test statistic and age-dependent censoring. *Statist. Probab. Lett.* **18**.
[2] CROWDER, M. (1991). On the identifiability crisis in competing risks analysis. *Scand. J. Statist.* **18** 223–233.
[3] KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.
[4] LANGBERG, N., PROSCHAN, F. and QUINZI, A. J. (1978). Converting dependent models into independent ones, preserving essential features. *Ann. Probab.* **6** 174–181.
[5] LANGBERG, N., PROSCHAN, F. and QUINZI, A. J. (1981). Estimating dependent life lengths, with applications to the theory of competing risks. *Ann. Statist.* **9** 157–167.
[6] MILLER, D. R. (1977). A note on independence of multivariate lifetimes in competing risk models. *Ann. Statist.* **5** 576–579.
[7] NÁDAS, A. (1970). On estimating the distribution of a random vector when only the smallest coordinate is observable. *Technometrics* **12** 923–924.
[8] NÁDAS, A. (1971). The distribution of the identified minimum of a normal pair determines the distribution of the pair. *Technometrics* **13** 201–202.
[9] PETERSON, A. V. (1976). Bounds for a joint distribution function with fixed subdistribution functions: application to competing risks. *Proc. Nat. Acad. Sci. U.S.A.* **73** 11–13.
[10] POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.
[11] TSIATIS, A. (1975). A nonidentifiability aspect in the problem of competing risks. *Proc. Nat. Acad. Sci. U.S.A.* **72** 20–22.

DEPARTMENT OF TECHNICAL MATHEMATICS
AND INFORMATICS
DELFT UNIVERSITY OF TECHNOLOGY
MEKELWEG 4, 2628 CD DELFT
THE NETHERLANDS
E-MAIL: t.j.bedford@twi.tudelft.nl

SCHOOL OF MATHEMATICAL SCIENCES
RAYMOND AND BEVERLY SACKLER
FACULTY OF EXACT SCIENCES
TEL AVIV UNIVERSITY
TEL AVIV 69978
ISRAEL
E-MAIL: isaco@math.tau.ac.il