STATISTICAL INFERENCE FOR UNIFORM STOCHASTIC ORDERING IN SEVERAL POPULATIONS

BY RICHARD DYKSTRA, SUBHASH KOCHAR AND TIM ROBERTSON

University of Iowa

Stochastic ordering between probability distributions is a widely studied concept. It arises in numerous settings and has useful applications. Since it is often easy to make value judgments when such orderings exist, it is desirable to recognize their occurrence and to model distributional structure under such orderings. Unfortunately, the necessary theory for statistical inference procedures has not been developed for many problems involving stochastic ordering and this development seems to be a difficult task. We show in this paper that the stronger notion of uniform stochastic ordering (which is equivalent to failure rate ordering for continuous distributions) is quite tractable in matters of statistical inference.

In particular, we consider nonparametric maximum likelihood estimation for $k$-population problems under uniform stochastic ordering restrictions. We derive closed-form estimates even with right-censored data by a reparameterization which reduces the problem to a well-known isotonic regression problem. We also derive the asymptotic distribution of the likelihood ratio statistic for testing equality of the $k$ populations against the uniform stochastic ordering restriction. This asymptotic distribution is of the chi-bar-square type as discussed by Robertson, Wright and Dykstra. These distributional results are obtained by appealing to elegant results from empirical process theory and showing that the proposed test is asymptotically distribution free. Recurrence formulas are derived for the weights of the chi-bar-square distribution for particular cases. The theory developed in this paper is illustrated by an example involving data for survival times for carcinoma of the oropharynx.

1. Introduction. Stochastic ordering between probability distributions is a widely studied concept. It arises in numerous settings and has useful applications. Since it is often easy to make value judgments when such orderings exist, it is desirable to recognize their occurrence and to model distributional structure under such orderings. Unfortunately, statistical inference procedures have not been developed for many problems involving stochastic ordering and the development of the necessary theory for these problems seems to be a difficult task.

Uniform stochastic ordering, as discussed in Keilson and Sumita (1982), is stronger than ordinary stochastic ordering but weaker than likelihood ratio ordering. In the continuous case, uniform stochastic ordering is equivalent to

Received July 1989; revised May 1990.


Key words and phrases. Stochastic ordering, uniform stochastic ordering, isotonic regression, empirical processes, Kaplan–Meier estimates, level probabilities, least squares projections, failure rate ordering, chi-bar-square distribution.
UNIFORM STOCHASTIC ORDERING

failure rate ordering. We show in this paper that uniform stochastic ordering is quite tractable in matters of statistical inference.

Definition 1.1. The univariate cdf $F_1$ is uniformly stochastically greater than the cdf $F_2$ ($F_1 \succeq F_2$) if

$\frac{F_1(x)}{F_2(x)}$ is nondecreasing for $x$ in $(-\infty, F_2^{-1}(1))$

(1.1)

where $F_i = 1 - F_i$ is the survival function corresponding to $F_i$.

If $F_1$ and $F_2$ are absolutely continuous with failure rates $r_1$ and $r_2$ defined by $r_i = F_i'/F_i$, $i = 1, 2$, then (1.1) is equivalent to saying that $r_1(x) \leq r_2(x)$ for all $x$. For this reason, the ordering $\succeq$ is sometimes called failure rate ordering.

Definition 1.1 is also equivalent to the inequalities

$P(X > s + t | X > t) \geq P(Y > s + t | Y > t)$ for all $s \geq 0, t$,

where $X$ has cdf $F_1$ and $Y$ has cdf $F_2$.

In other words, the conditional distributions, given that the random variables are at least of a certain size, are all stochastically ordered (in the standard sense) in the same direction. Thus, if $X$ and $Y$ represent the survival times of different models of an appliance that satisfy this ordering, one model is better (in the sense of stochastic ordering) when the appliances are new, the same appliance is better when both are one month old, and if fact is better no matter how much time has elapsed. It is clearly useful to know when this strong type of stochastic ordering holds since qualitative judgments are then easy to make.

This type of ordering is certainly of interest when populations correspond to survival times for different medical treatments. Even if the better of two treatments (better in the sense that its survival time stochastically dominates that of the other treatment) is administered initially, it may not be the better treatment when patients are examined at a later point in time. However, if the treatment populations are ordered in this stronger sense, there can be no doubt which treatment is preferred at any point in time.

It is easy to see that $F_1 \succeq F_2$ implies that $F_1$ is stochastically larger than $F_2$ in the sense that $F_1(x) \leq F_2(x)$ for all $x$. However, as shown in Ross (1983), this ordering is weaker than likelihood ratio ordering, that is, $f_1(x)/f_2(x)$ is nondecreasing in $x$. Various other relationships between uniform stochastic ordering and other partial orderings have been obtained by Bagai and Kocfar (1986). Recently, Capéráá (1988) has used this notion of ordering of probability distributions to compare asymptotic efficiencies of rank tests in two-sample problems. For applications of this ordering in queuing theory, see Ross (1983) and Stoyan (1983). Lynch, Mimmack and Proschan (1987) have discussed some closure properties of uniform stochastic ordering.

Tests for equality of distributions against ordered failure rates in the two-sample situation have been given by Kocfar (1979, 1981), Joe and Proschan (1984), Cheng (1985) and Aly (1988). However, none of these tests
allow for censored data. Proschlan and Singpurwalla (1980) discuss a Bayesian procedure for estimating several populations when the corresponding failure rates are presumed to be linearly ordered by \( r_1(x) \geq r_2(x) \geq \cdots \geq r_N(x) \) for all \( x \). Their procedure requires pooling adjacent violators, but is very different from what is done here.

The literature on estimation and hypothesis testing problems involving (ordinary) stochastic ordering is extensive. Brunk, Franck, Hanson and Hogg (1966) obtained nonparametric maximum likelihood estimates of two stochastically ordered cdf’s and studied their properties. Dykstra (1982) considered a similar problem with censored data. Dykstra and Feltz (1989) obtained maximum likelihood estimates of more than two cdf’s subject to stochastic ordering restrictions by using an iterative algorithm. Robertson and Wright (1974) have considered stochastic orderings in higher dimensions, as have Sampson and Whitaker (1989). Testing procedures which are based on maximum likelihood estimates of two stochastically ordered distributions are discussed in Robertson and Wright (1981), Lee and Wolfe (1976), Franck (1984) and Dykstra, Madsen and Fairbanks (1983). Closed-form algorithms for maximum likelihood estimates of more than two stochastically ordered distributions have not been found. Distribution theory for tests based on the likelihood principle have not been developed to a satisfactory conclusion to the best of our knowledge.

In this paper, we consider the multiple-sample problem when the observations are randomly censored on the right. Nonparametric maximum likelihood estimators of the survival functions are obtained under the assumptions that the various populations are uniformly stochastically ordered. We also discuss the likelihood ratio test for testing equality of distributions against the uniform stochastic ordering restrictions in a linear order situation. Asymptotic distributions are derived and related to the distribution of tests occurring in other situations. Finally, an example involving survival times for carcinoma of the oropharynx is discussed in Section 5.

We will use terminology as if the measured quantity of interest is survival time, although this need not be the case. Thus deaths will refer to complete observations and losses will refer to censored observations.

2. Maximum likelihood estimation. Suppose we have independent random samples from populations with corresponding cdf’s \( F_1, F_2, \ldots, F_N \). We allow the possibility that our observations may be censored on the right with respective censoring distributions \( G_1, \ldots, G_N \). We do assume that the censoring distributions are independent of the distributions of interest and that we are always dealing with nonnegative random variables.

The problem will be to construct nonparametric maximum likelihood estimates of the survival functions subject to the constraints \( F_1 \succeq F_2 \succeq \cdots \succeq F_N \). Similar procedures will work for other ordering configurations of the survival functions since a good deal is known about other ordering problems [Robertson, Wright and Dykstra (1988)]. For example, there exists a pool adjacent violators algorithm (PAVA) which will work nicely for a tree ordering \((F_1 \succeq F_i, i = 2, \ldots, N)\). However, we will restrict ourselves to the above linear
order since it is of considerable interest and is indicative of general behavior. We will assume notation as in Feltz and Dykstra (1985).

Suppose that complete observations from all \( N \) samples occur on a subset of the times \( S_1 < S_2 < \cdots < S_m, S_0 = -\infty, S_{m+1} = \infty \). We let

\[
d_{ij} = \text{number of complete observations from the } i \text{th population at } S_j;
\]

\[
l_{ij} = \text{number of observations from the } i \text{th population censored in } [S_j, S_{j+1}) \text{ (we assume these occur at } L_{r(i,j)}^r, r = 1, 2, \ldots, l_{ij};
\]

\[
n_{ij} = \sum_{r=1}^{l_{ij}} (d_{ir} + l_{ir}) = \text{number of observations from the } i \text{th population surviving to just prior to } S_j.
\]

We interpret maximum likelihood estimators in the generalized sense given by Kiefer and Wolfowitz (1956). Since these estimates of the cdf’s will put probability only on actual observation points and since censoring distributions are independent of survival distributions, the likelihood function is given by

\[
(2.1) \quad L(\overline{F}) = \prod_{i=1}^{N} \prod_{j=1}^{M} \left[ \frac{\overline{F}_i(S_j -) - \overline{F}_i(S_j)}{\overline{F}_i(S_j)} \right]^{d_{ij}} \prod_{r=1}^{l_{ij}} \frac{\overline{F}_i(L_{r(i,j)}^r)}{\overline{F}_i(L_{r(i,j)}^r)} H(G),
\]

where \( H \) will depend on the censoring distributions, but not on the vector of survival functions \( \overline{F} \). We wish to maximize \( L(\overline{F}) \) under the constraints that

\[
(2.2) \quad \frac{\overline{F}_i(t)}{\overline{F}_{i+1}(t)} \text{ is nondecreasing in } t \text{ for } i = 1, \ldots, N - 1.
\]

Note that if the survival functions \( \overline{F} \) satisfy the constraints (2.2), then the likelihood is not decreased and the constraints are still satisfied if \( \overline{F}_i(t) \) is replaced by \( \sum_{j=0}^{m} \overline{F}_i(S_j) I(S_j, S_{j+1}) \). Thus it will suffice to maximize \( L(\overline{F}) \) over this class or, equivalently, to maximize

\[
\prod_{i=1}^{N} \prod_{j=1}^{m} \left[ \frac{\overline{F}_i(S_{j-1}) - \overline{F}_i(S_j)}{\overline{F}_i(S_j)} \right]^{d_{ij}} \overline{F}_i(S_j)^{l_{ij}}
\]

subject to the constraints

\[
\frac{\overline{F}_i(S_j)}{\overline{F}_{i+1}(S_j)} \geq \frac{\overline{F}_i(S_{j-1})}{\overline{F}_{i+1}(S_{j-1})}.
\]

This is equivalent to maximizing

\[
(2.3) \quad \prod_{i=1}^{N} \prod_{j=1}^{m} \left[ 1 - \frac{\overline{F}_i(S_j)}{\overline{F}_i(S_{j-1})} \right]^{d_{ij}} \left[ \frac{\overline{F}_i(S_{j-1})}{\overline{F}_i(S_{j-2})} \frac{\overline{F}_i(S_{j-2})}{\overline{F}_i(S_{j-3})} \cdots \overline{F}_i(S_i) \right]^{l_{ij}}
\]

\[
\times \left[ \frac{\overline{F}_i(S_j)}{\overline{F}_i(S_{j-1})} \frac{\overline{F}_i(S_{j-1})}{\overline{F}_i(S_{j-2})} \cdots \overline{F}_i(S_1) \right]^{l_{ij}}
\]

subject to the constraints

\[
(2.4) \quad \frac{\overline{F}_i(S_j)}{\overline{F}_i(S_{j-1})} \geq \frac{\overline{F}_i(S_j)}{\overline{F}_{i+1}(S_{j-1})}, \quad i = 1, \ldots, N - 1; j = 1, \ldots, m.
\]
We reparameterize by setting \( \theta_{ij} = \bar{F}_i(S_j) / \bar{F}_i(S_{j-1}) \), \( i = 1, \ldots, N \), \( j = 1, \ldots, m \) [so that \( \bar{F}_i(S_j) = \prod_{r=1}^{j-1} \theta_{ir} \)]. Rearranging terms and noting that
\[
\prod_{j=1}^{m} \prod_{r=1}^{j-1} \theta_{irj}^{d_{irj} + l_{ij}} = \prod_{r=1}^{m-1} \prod_{j=r+1}^{m} \theta_{irj}^{d_{irj} + l_{ij}}
\]
\[
= \prod_{r=1}^{m-1} \theta_{irr}^{-d_{irr} - l_{ir}} = \prod_{r=1}^{m} \theta_{irr}^{-d_{irr} - l_{ir}},
\]
we see that maximizing (2.3) is equivalent to maximizing
\[
(2.5) \quad \prod_{j=1}^{m} \prod_{i=1}^{N} \theta_{ij}^{-d_{ij}} (1 - \theta_{ij})^{d_{ij}}
\]
subject to
\[
(2.6) \quad \theta_{1j} \geq \theta_{2j} \geq \cdots \geq \theta_{Nj} \quad \text{for all } j.
\]

The constraints in (2.6) do not relate \( \theta_{ij} \) for different values of \( j \). Thus the \( m \) factors \( \prod_{i=1}^{N} \theta_{ij}^{n_{ij} - d_{ij}} (1 - \theta_{ij})^{d_{ij}} \), \( j = 1, 2, \ldots, m \), can be maximized individually. For each fixed \( j \) this is a bioassay problem as discussed in Example 1.5.1 of Robertson, Wright and Dykstra (1988). The solution is the isotonic regression of the vector \((\hat{\theta}_{1j}, \hat{\theta}_{2j}, \ldots, \hat{\theta}_{Nj})\) with weights \((n_{1j}, n_{2j}, \ldots, n_{Nj})\), where
\[
(2.7) \quad \hat{\theta}_{ij} = \frac{n_{ij} - d_{ij}}{n_{ij}}.
\]
The “pool adjacent violators algorithm” (PAVA) provides an easy method for obtaining the solution, \((\theta_{1j}^*, \theta_{2j}^*, \ldots, \theta_{Nj}^*)\). The restricted mle’s of the survival functions under the ordering constraints are found by computing
\[
(2.8) \quad \bar{F}_i^*(t) = \prod_{j:S_j \leq t} \theta_{ij}^*.
\]

In order to maximize the likelihood under the constraint that all the survival functions are equal, we need only to set \( \hat{\theta}_{ij} = \theta_j \) for all \( i \) and solve the resulting optimization problems. This gives the estimates
\[
(2.9) \quad \bar{\theta}_j = \frac{\sum_{i=1}^{N} n_{ij} - \sum_{i=1}^{N} d_{ij}}{\sum_{i=1}^{N} n_{ij}}.
\]

The mle of the common survival function is then given as in (2.8) with \( \bar{\theta}_j \) replacing \( \theta_{ij}^* \).

The fact that this simple reparametrization makes tractable a seemingly complicated problem closely relates to the problem solved in Dykstra and Robertson (1982).

3. **Hypothesis testing.** Assume that all the random samples have a common censoring distribution and that censoring values are independent of theoretical lifetimes. We also assume that for every \( i \), \( F_i \) has support on the fixed set \( \{S_1, \ldots, S_m\} \) and that each point has positive probability, so that we
are concerned with discrete distributions with common support. In our asymptotic theory, \( m \) is fixed as the sample sizes change.

Consider the problem of testing the hypothesis \( H_0: F_1 = \cdots = F_N \) against the uniform stochastic ordering alternative \( H_1: F_1 \succeq \cdots \succeq F_N \) \((F_i \neq F_{i+1} \text{ for some } i)\). We base our test on the statistic \( Q = -2 \ln L \), where \( L \) is the likelihood ratio. Since we have constructed the ml’s in Section 2, it is straightforward to construct

\[
(3.1) \quad Q = -2 \ln L = 2 \sum_{j-1}^{m-1} \sum_{i=1}^{N} d_{ij} \ln \left( \frac{1 - \theta_{ij}^*}{1 - \bar{\theta}_j} \right) + (n_{ij} - d_{ij}) \ln \left( \frac{\theta_{ij}^*}{\bar{\theta}_j} \right)
\]

(since in this setting \( \bar{\theta}_m = \theta_{im}^* = 0 \) for all \( i \)).

If we expand \( \ln(1 - \theta_{ij}^*) \) and \( \ln \theta_{ij}^* \) about \( \bar{\theta}_j \), we can write

\[
(3.2) \quad Q = 2 \sum_{j=1}^{m-1} \sum_{i=1}^{N} \left\{ \frac{-d_{ij}}{(1 - \bar{\theta}_j)} \left( \theta_{ij}^* - \bar{\theta}_j \right) - \frac{d_{ij}}{2(1 - \alpha_{ij})} (\theta_{ij}^* - \bar{\theta}_j)^2 \right. \\
+ (n_{ij} - d_{ij}) \left( \frac{\theta_{ij}^* - \bar{\theta}_j}{\bar{\theta}_j} \right)^2
\]

where \( \max(\alpha_{ij}, \beta_{ij}) = \beta_{ij} = \beta_{ij} \leq \theta_{ij}^* - \bar{\theta}_j \). Combining the first and third parts of the previous sum, we obtain

\[
(3.3) \quad 2 \sum_{j=1}^{m-1} \frac{-\bar{\theta}_j}{(1 - \bar{\theta}_j)} \sum_{i=1}^{N} (\theta_{ij}^* - \bar{\theta}_j)n_{ij} + 2 \sum_{j=1}^{m-1} \sum_{i=1}^{N} \left( \frac{\theta_{ij}^* - \bar{\theta}_j}{1 - \bar{\theta}_j} \right) (n_{ij} - d_{ij}).
\]

However, by properties of the isotonic regression [Robertson, Wright and Dykstra (1988), Section 1.3]

\[
\sum_{i=1}^{N} \theta_{ij}^* n_{ij} = \sum_{i=1}^{N} \bar{\theta}_i n_{ij} = \sum_{i=1}^{N} \bar{\theta}_j n_{ij}
\]

so that the first part of (3.3) is 0. Moreover, the second part can be written as

\[
(3.4) \quad 2 \sum_{j=1}^{m-1} \frac{1}{(1 - \bar{\theta}_j)} \left[ \sum_{i=1}^{N} \theta_{ij}^* \hat{\theta}_{ij} n_{ij} - \bar{\theta}_j \sum_{i=1}^{N} \hat{\theta}_{ij} n_{ij} \right]
\]

\[
= 2 \sum_{j=1}^{m-1} \frac{1}{(1 - \bar{\theta}_j)} \left[ N \theta_{ij}^* \sum_{i=1}^{N} \hat{\theta}_{ij} n_{ij} - \bar{\theta}_j \sum_{i=1}^{N} \hat{\theta}_{ij} n_{ij} \right]
\]

\[
= 2 \sum_{j=1}^{m-1} \frac{N}{\sum_{i=1}^{N} (\theta_{ij}^* - \bar{\theta}_j) \theta_{ij} n_{ij}}
\]

\[
= 2 \sum_{j=1}^{m-1} \frac{(\theta_{ij}^* - \bar{\theta}_j)^2 n_{ij}}{(1 - \bar{\theta}_j)}
\]
Since we are concerned with asymptotic results and since $\theta_{ij}^*$ and $\bar{\theta}_j$ converge to $\theta_{ij}$ if the sample sizes go to $\infty$ and $H_0$ is true, we may use Slutsky’s theorem to replace $\alpha_{ij}$ and $\beta_{ij}$ by $\theta_{ij}$ in (3.2) without affecting the asymptotic behavior (recall we assume $\theta_{ij} > 0$). The second and fourth parts can then be combined and are asymptotically equivalent to

$$
\sum_{j=1}^{m-1} \sum_{i=1}^{N} \frac{-d_{ij}(\theta_{ij}^* - \bar{\theta}_j)^2}{(1 - \bar{\theta}_j)^2} - \frac{(n_{ij} - d_{ij})(\theta_{ij}^* - \bar{\theta}_j)^2}{\bar{\theta}_j^2}
$$

(3.5) = \sum_{j=1}^{m-1} \sum_{i=1}^{N} \frac{-\bar{\theta}_j^2 + (2\bar{\theta}_j - 1)\hat{\theta}_{ij} \left(\theta_{ij}^* - \bar{\theta}_j\right)^2}{(1 - \bar{\theta}_j)\bar{\theta}_j^2 n_{ij}}.

Since $|\hat{\theta}_{ij} - \bar{\theta}_j| \to p 0$ if $H_0$ is true, the first factor converges in probability to $-1$. By combining (3.4) and (3.5), we see that the asymptotic distribution of $Q$ is the same as that of

$$
Q' = \sum_{j=1}^{m-1} \sum_{i=1}^{N} \frac{(\theta_{ij}^* - \bar{\theta}_j)^2}{(1 - \bar{\theta}_j)\bar{\theta}_j} n_{ij}.
$$

(3.6)

This asymptotic distribution can be found by using the multivariate central limit theorem for multinomial distributions and the $\delta$ method. However, we appeal to some empirical processes results given in Shorack and Wellner (1987). To paraphrase the needed theory, assume that $X_1, X_2, \ldots, X_n$ are i.i.d. with common cdf $F$, and $Y_1, \ldots, Y_n$ are i.i.d. random variables (independent of the $X$’s) from the censoring distribution $G$. It is assumed that $Z_i = \min(X_i, Y_i)$ and $\delta_i = I_{X_i \leq Y_i}$ are observed for $i = 1, 2, \ldots, n$. We let

$$
H(t) = 1 - P(Z > t) = 1 - (1 - F(t))(1 - G(t)),
$$

$$
H^1(t) = P(Z \leq t, \delta = 1) = \int_{[0, t]} (1 - G_-) \, dF,
$$

$$
H^0(t) = P(Z \leq t, \delta = 0) = \int_{[0, t]} (1 - F) \, dG
$$

(where $G_-$ denotes the left-continuous version of $G$) and let their empirical counterparts be given as

$$
H_n(t) = \frac{1}{n} \sum_{i=1}^{N} I_{[0, t]}(Z_i),
$$

$$
H_n^1(t) = \frac{1}{n} \sum_{i=1}^{N} I_{[0, t]}(Z_i) \delta_i,
$$

$$
H_n^0 = \frac{1}{n} \sum_{i=1}^{N} I_{[0, t]}(Z_i)(1 - \delta_i).$$
The cumulative hazard function corresponding to \( F \) is given by

\[
\Lambda(t) = \int_{[0, t]} \frac{1}{1 - F_-} \, dF
\]

and the empirical version is

\[
\Lambda_n(t) = \int_{[0, t]} \frac{1}{1 - H_n^-} \, dH_n^1.
\]

Theorem 1 of Shorack and Wellner [(1987), page 307] states that if \( \Lambda(t) \) is the true cumulative hazard function of \( F \), then

\[
\sqrt{n} \left( \Lambda_n(t) - \Lambda(t) \right) \to_{w} W(C(t)),
\]

where \( W(t) \) denotes a Brownian motion process and the time transformation \( C \) is given by

\[
(3.7) \quad C(t) = \int_{[0, t]} (1 - H_-)^{-1} (1 - \Delta \Lambda) \, d\Lambda \quad \left(1 - \Delta \Lambda = \frac{1 - F}{1 - F_-}\right).
\]

For our problem where \( S_1, \ldots, S_m \) are the support points of the \( F_i \), the empirical cumulative hazard function for the \( i \)th population is given by

\[
\Lambda_{i,n}(t) = \sum_{j \leq t} (1 - \hat{\theta}_{ij}), \quad i = 1, \ldots, N,
\]

while the true cumulative hazard function is given by

\[
\Lambda_i(t) = \sum_{j \leq t} (1 - \theta_{ij}), \quad i = 1, \ldots, N.
\]

Thus

\[
\sqrt{n} \left( \hat{\theta}_{ij} - \theta_{ij} \right) = \sqrt{n} \left[ (1 - \theta_{ij}) - (1 - \hat{\theta}_{ij}) \right]
\]

\[
= \sqrt{n} \left[ (\Lambda_i(S_j) - \Lambda_i(S_{j-1})) - (\Lambda_{i,n}(S_j) - \Lambda_{i,n}(S_{j-1})) \right]
\]

\[
= -\sqrt{n} \left[ (\Lambda_{i,n}(S_j) - \Lambda_i(S_j)) - (\Lambda_{i,n}(S_{j-1}) - \Lambda_i(S_{j-1})) \right].
\]

Items censored before \( S_1 \) convey no information about the survival functions of interest, and we assume they do not occur. Thus we take \( (n_{11}, n_{21}, \ldots, n_{N1}) \) to be nonrandom and assume these are the number of items initially put on test. We let \( n = \sum_{i=1}^N n_{i1} \) and assume that \( \gamma_i = \lim_{n \to \infty} n_{i1}/n \) exists and is positive and finite. Then

\[
\frac{n_{ij}}{n} = \frac{n_{i1}}{n} \frac{n_{ij}}{n_{i1}} = \frac{n_{i1}}{n} \frac{1}{n_{i1}} \sum_{r=1}^{n_{i1}} I_{[x_{ir} \geq S_j] \cap [y_{ir} \geq S_j]} \to_p \gamma_i \bar{F}_i(S_j - ) \tilde{G}_i(S_j - )
\]

by the law of large numbers. Moreover, since the transfer function in (3.7) corresponding to the \( i \)th population has increments

\[
C_i(S_j) - C_i(S_{j-1}) = \theta_{ij} (1 - \theta_{ij}) \left[ F_i(S_j - ) \bar{G}_i(S_j - ) \right]^{-1}
\]

and since a Brownian motion process has independent increments, it follows
that
\[ W_{ij}^{(n)} = \frac{\sqrt{n} \left( \hat{\theta}_{ij} - \theta_{ij} \right)}{\left(1 - \bar{\theta}_j \right) \bar{\theta}_j}^{1/2} \to_d N \left( 0, \left[ \gamma_i \bar{F}_i(S_j -) - \bar{G}_i(S_j -) \right]^{-1} \right) \]
if \( H_0 \) is true, and that all the \( W_{ij}^{(n)} \) will be asymptotically independent. We let \( W_j^{(n)} \) indicate the vector which occurs when \( j \) is held fixed.

Suppose now that \( E_w(x|H_0) \) indicates the least squares projection with weights \( w \) of the vector \( x \) onto the set of constant vectors, while \( E_w(x|H_1) \) is the projection onto the set of nondecreasing vectors. We can then represent \( Q' \) as \( \sum_{J=1}^{m-1} Q'_J \), where
\[ Q'_J = \left\| E_w \left( W_j^{(n)}|H_1 \right) - E_w \left( W_j^{(n)}|H_0 \right) \right\|_w^2, \]
with \( w_j = (n_{1j}/n, \ldots, n_{N_j}/n) \) and \( \| \cdot \|_w \) indicates the usual least squares norm with weights \( w \).

After noting the \( E_w(x|H_i) \) is a continuous operator in both \( x \) and \( w \), we employ a continuity argument to express the asymptotic distribution of \( Q' \) (and \( Q \)) in terms of normal random variables. This leads to the following theorem.

**Theorem 3.1.** If \( F_1 = F_2 = \cdots = F_N \) and the common distribution puts positive probability on each of \( S_1, S_2, \ldots, S_m \); if there is a common censoring distribution \( G \) such that \( \bar{G}(S_m-) > 0 \); and if the sample sizes increase to \( \infty \) in such a manner that \( \gamma_i = \lim_{n \to \infty} n_{ij}/n \) exists, positive and finite, then \( Q' \) (and \( Q \)) has a limiting distribution which is the same as the distribution of

\[ \sum_{j=1}^{m-1} \sum_{i=1}^{N} \left( Z_{ij}^* - \bar{Z}_j \right)^2 \gamma_i, \]

where the \( \{Z_{ij}\} \) are independent, \( Z_{ij} \sim N(0, \gamma_i^{-1}) \),
\[ \bar{Z}_j = \sum_{i=1}^{N} Z_{ij} \gamma_i \]
and
\[ Z_{ij}^* = E_w \left( Z_j|H_1 \right) \]
is the \( i \)th element of the antitonic regression of \( \{Z_{1j}, Z_{2j}, \ldots, Z_{Nj}\} \) with weights \( \{\gamma_1, \ldots, \gamma_N\} \).

Note that the distribution of the inside sum in (3.8) does not depend on \( j \) so that the asymptotic distribution is that of a convolution of independent, identically distributed chi-bar-squared random variables as discussed in Robertson, Wright and Dykstra (1988).

If the numbers of items on test are initially nearly equal \( (n_{11} \sim n_{21} \sim \cdots \sim n_{N1}) \) and sufficiently large, the chi-bar-squared distributions of \( \sum_{i=1}^{N} (Z_{ij}^* - \bar{Z}_j)^2 \gamma_i \) can be taken to be of the equal weight variety. These
distributions are much more tractable than the unequal weight case and will
be discussed in Section 4.

If the censoring distributions are different but known for the various
populations, the asymptotic distribution still reduces to a convolution of
chi-bar-squared distributions. However, the summands need not be identically
distributed.

It is an appealing aspect of the proposed test that it is an asymptotically
similar test. That is, the asymptotic distribution of $Q$ under the null hypothe-
sis does not depend on the common $F_i$. This is in contrast to the likelihood
ratio test for standard stochastic ordering for two populations discussed in
Robertson and Wright (1981), where the asymptotic distribution varies with
the common $F_i$. Note that the test is invariant under a common, increasing
transformation of the samples.

Assuming that there is a constant censoring distribution $G$ [with $G(S_m) > 0$],
the hypotheses $H_0$ and $H_1$ are equivalent to

$$H_0' : 1 - \bar{F}_1 G = 1 - \bar{F}_2 G = \cdots = 1 - \bar{F}_N G$$

and

$$H_1' : 1 - \bar{F}_1 G \succeq 1 - \bar{F}_2 G \succeq \cdots \succeq 1 - \bar{F}_N G.$$ 

Of course, if one does not distinguish between censored and complete observa-
tions, but treats them all as being complete, the appropriate cdf’s would be
$1 - \bar{F}_1 G, \ldots, 1 - \bar{F}_N G$. Thus, if a test is constructed by treating censored
observations as if they were complete, the resulting test will still be testing the
correct hypotheses. Moreover, the asymptotic distribution of $Q'$ and $Q$ under
$H_{0'}$ will be the same as before, although, of course, the value of the test
statistic will be different. One might conjecture that this test would not be as
powerful, since it makes no use of the knowledge of whether an observation is
censored. We have not investigated the power properties of this test, however.

4. Equal weights case in hypothesis testing. The distribution of the
inner sum of (3.8) has been studied extensively. It is called a chi-bar-squared
distribution and its survival function is given by

$$\sum_{l=1}^{N} P_{\gamma}(l, N) P(\chi^2_{l-1} > x),$$

where $\chi^2_j$ denotes a central chi-squared random variable with $j$ degrees of
freedom ($\chi^2_0 = 0$). The weighting element, $P_{\gamma}(l, N)$ is the probability that the
vector, $Z_j^* = E_{\gamma}(Z_j | H_1)$ has exactly $l$ distinct values. The quantity $P_{\gamma}(l, N)$ is
called a level probability and is generally difficult to compute since it depends
on the variances of the $Z_{ij}$ [cf. Robertson, Wright and Dykstra (1988), Chapter
3]. If $\gamma_1 = \gamma_2 = \cdots = \gamma_N$, Bartholomew (1959) conjectured and Miles (1959)
proved the recurrence relation

$$P(l, N) = \frac{1}{N} P(l - 1, N - 1) + \frac{N - 1}{N} P(l, N - 1),$$
where \( P(0, N - 1) = P(N, N - 1) = 0 \). (It is customary to omit the weights when they are equal.) This recurrence relationship makes it easy to compute all necessary values of the \( P(l, N)'s \). Once these values are known, it is straightforward to compute \( p \) values and critical points for the distribution given in (4.1) when \( \gamma_1 = \gamma_2 = \cdots = \gamma_N \). (See Robertson, Wright and Dykstra (1988), Table 4.4.)

The moment generating function of the distribution associated with (4.1) is given by

\[
m_N(t) = \Theta_N(s) = \sum_{l=1}^{N} P_\gamma(l, N) s^{l-1}
\]

with \( s = (1 - 2t)^{-1/2} \) for \( t < 1/2 \) \( (s > 0) \). In the case \( \gamma_1 = \gamma_2 = \cdots = \gamma_N \), the recurrence relation, (4.2), implies that

\[
(4.3) \quad \Theta_N(s) = \frac{s + N - 1}{N} \Theta_{N-1}(s), \quad N = 2, 3, \ldots.
\]

Assuming henceforth that \( \gamma_1 = \gamma_2 = \cdots = \gamma_N \), the sum in (3.8) is the sum of \( m - 1 \) independent random variables each having moment generating function given by \( \Theta_N(s) \). Using (4.3), the moment generating function of the sum in (3.8) is given by

\[
\Phi_{N,m}(s) = \Theta_N(s)^{m-1} = \left[ \frac{s + N - 1}{N} \right]^{m-1} \left[ \Theta_{N-1}(s) \right]^{m-1} = \left[ \frac{s + N - 1}{N} \right]^{m-1} \Phi_{N-1,m}(s).
\]

By expanding \( ((s + N - 1)/N)^{m-1} \), expressing \( n \Phi_{N,m}(s) \) and \( \Phi_{N-1,m} \) as polynomials in \( s \) and equating coefficients on the two sides of (4.4), we can conclude that the random variable in (3.8) has a chi-bar-squared distribution whose level probabilities satisfy a recursive formula. This result is summarized in the following theorem.

**Theorem 4.1.** Assume the conditions of Theorem 3.1 and \( \gamma_1 = \gamma_2 = \cdots = \gamma_N \). Then the asymptotic survival function of \( Q \) and \( Q' \) under \( H_0 \) is given by

\[
(4.5) \quad \sum_{l=m-1}^{N(m-1)} R(l, N, m) P[\chi^2_l > x],
\]

where the \( R(l, N, m) \) satisfy the recurrence relationship

\[
(4.6) \quad R(l, N, m) = \sum_{u=0}^{m-1} \left( \begin{array}{c} m-1 \cr u \end{array} \right) \left( \frac{k-1}{N} \right)^{m-1-u} \left( \frac{1}{N} \right)^u R(l-u, N-1, m)
\]
for \( l = (m - 1), \ldots, N(m - 1) \). We take \( R(l, N, m) = 0 \) if \( l < m - 1 \) or \( l > N(m - 1) \).

If we fix \( m \) and arrange the \( R(l, N, m) \) in a “triangle” similar to Pascal’s triangle with \( R(l, N, m) \), \( l = m - 1, \ldots, N(m - 1) \), forming the \( N \)th row, then (4.6) says that each row is formed by taking a convex combination of the elements in the proceeding row [where \( R(m - 1, 1, m) = 1 \) for \( m \geq 2 \)]. This is analogous to Pascal’s triangle and can be used to generate as many of these values as necessary.

If \( \gamma_1 = \gamma_2 = \cdots = \gamma_m \), the moment generating function corresponding to the distribution given in (4.1) can be explicitly calculated as

\[
m_N(s) = \frac{(s + 1) \cdots (s + N - 1)}{N!},
\]

where \( s = (1 - 2t)^{-1/2} \) and \( t < 1/2 \) [Robertson, Wright and Dykstra (1988), page 81]. From this, the mean and variance for the corresponding distribution can be calculated as \( \sum_{j=2}^{N} j^{-1} \) and \( \sum_{j=2}^{N} (3j^{-1} - j^{-2}) \), respectively [Barlow, Bartholomew, Bremner and Brunk (1972), page 151].

If the weights \( R(l, N, m) \) are taken to be a probability mass function in \( l \), the respective mean and variance of this distribution will be

\[
(m - 1) \sum_{j=1}^{N} j^{-1} \quad \text{and} \quad (m - 1) \sum_{j=2}^{N} (j^{-1} - j^{-2}).
\]

Since the ratio of these two values converges to one as \( N \to \infty \), one might suspect that this distribution can be well approximated by a Poisson distribution if \( N \) and \( m \) are sufficiently large. Our experience indicates that a Poisson distribution with mean \( (m - 1) \sum_{j=2}^{N} (j^{-1} - j^{-2}) \) which has been translated \( m - 1 \) units to the right gives a significantly better approximation than a straight Poisson. This approximation to the weighting distribution is quite good even for \( N \) and \( m \) as small as 5. The approximating distribution for the asymptotic distribution of \( Q \) formed by replacing \( R(l, N, m) \) in (4.5) by its Poisson counterpart is generally quite accurate.

Of course, the central limit theorem ensures that the normal distribution may also be used as an approximation for the asymptotic distribution of \( Q \) if \( m \) is large enough. This can be easily implemented since the mean and variance of the asymptotic distribution of \( Q \) are given by

\[
(m - 1) \sum_{j=2}^{N} j^{-1} \quad \text{and} \quad (m - 1) \sum_{j=2}^{N} (3j^{-1} - j^{-2}).
\]

In the event that the \( \gamma_i \)'s are not all equal, matters become much more intractable. Although the asymptotic distribution of \( Q \) is still of the chi-bar-squared type, the correct weightings will generally be unknown. Robertson and Wright (1983) have investigated the chi-bar-square distribution and conclude that this distribution is quite insensitive to different values of the \( \gamma_i \)'s as long as their ratios do not differ from 1 by too large a factor. They conclude that as
long as all ratios stay between $1/4$ and 4, the equal weights chi-bar-squared distribution will serve as an adequate approximation. If this is not the case, they propose other approximations.

5. Example. To illustrate the methods discussed in earlier sections, we consider some data given in Data Set II from Kalbfleisch and Prentice (1980). These data consist of survival times for patients with carcinoma of the

<table>
<thead>
<tr>
<th>Group</th>
<th>Pop 0</th>
<th>Pop 1</th>
<th>Pop 2</th>
<th>Pop 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I</td>
<td>38</td>
<td>81</td>
<td>105</td>
<td>11</td>
</tr>
<tr>
<td>0–160</td>
<td>107</td>
<td>154</td>
<td>128</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>103</td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>74</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>89</td>
</tr>
<tr>
<td>Group II</td>
<td>167</td>
<td>216</td>
<td>170</td>
<td>162</td>
</tr>
<tr>
<td>161–260</td>
<td>172</td>
<td>254</td>
<td>184</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td>191</td>
<td>222</td>
<td>228</td>
<td>173</td>
</tr>
<tr>
<td></td>
<td>238</td>
<td>228</td>
<td>174</td>
<td>208</td>
</tr>
<tr>
<td></td>
<td>243</td>
<td>230</td>
<td>177</td>
<td>209</td>
</tr>
<tr>
<td>Group III</td>
<td>276*</td>
<td>343</td>
<td>275</td>
<td>338</td>
</tr>
<tr>
<td>261–360</td>
<td>296</td>
<td>351</td>
<td>301</td>
<td>347</td>
</tr>
<tr>
<td></td>
<td>324</td>
<td>324</td>
<td>310</td>
<td></td>
</tr>
<tr>
<td></td>
<td>336</td>
<td>328</td>
<td>346</td>
<td></td>
</tr>
<tr>
<td>Group IV</td>
<td>372</td>
<td>445*</td>
<td>382</td>
<td>395</td>
</tr>
<tr>
<td>361–540</td>
<td>374</td>
<td>446</td>
<td>532</td>
<td>407</td>
</tr>
<tr>
<td></td>
<td>376</td>
<td>446</td>
<td>465</td>
<td></td>
</tr>
<tr>
<td></td>
<td>404</td>
<td>498</td>
<td>477</td>
<td></td>
</tr>
<tr>
<td></td>
<td>432</td>
<td>525</td>
<td>518</td>
<td></td>
</tr>
<tr>
<td>Group V</td>
<td>541</td>
<td>553</td>
<td>546</td>
<td>544</td>
</tr>
<tr>
<td>541–700</td>
<td>545</td>
<td>575</td>
<td>608</td>
<td></td>
</tr>
<tr>
<td></td>
<td>560</td>
<td>599</td>
<td>661</td>
<td></td>
</tr>
<tr>
<td></td>
<td>561</td>
<td>600*</td>
<td>666</td>
<td></td>
</tr>
<tr>
<td></td>
<td>651*</td>
<td>631</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group VI</td>
<td>714</td>
<td>723*</td>
<td>751*</td>
<td>726</td>
</tr>
<tr>
<td>701–900</td>
<td>755</td>
<td>733*</td>
<td>822*</td>
<td>731*</td>
</tr>
<tr>
<td></td>
<td>763</td>
<td>825*</td>
<td>757</td>
<td>794*</td>
</tr>
<tr>
<td></td>
<td>854*</td>
<td></td>
<td>760*</td>
<td>800</td>
</tr>
<tr>
<td>Group VII</td>
<td>943*</td>
<td>1823*</td>
<td>929</td>
<td>915</td>
</tr>
<tr>
<td>901–1850</td>
<td>998*</td>
<td>933*</td>
<td>918*</td>
<td>1307*</td>
</tr>
<tr>
<td></td>
<td>1219*</td>
<td>1086*</td>
<td>928*</td>
<td>1312*</td>
</tr>
<tr>
<td></td>
<td>1234*</td>
<td>1092</td>
<td>932*</td>
<td>1455*</td>
</tr>
<tr>
<td></td>
<td>1460*</td>
<td>1317</td>
<td>1058*</td>
<td>1489*</td>
</tr>
<tr>
<td></td>
<td>1574</td>
<td>1317</td>
<td>1060*</td>
<td>1495*</td>
</tr>
<tr>
<td></td>
<td>1766*</td>
<td>1609*</td>
<td>1064*</td>
<td>1644*</td>
</tr>
</tbody>
</table>

*represents censored observation.
oropharynx and several covariates. A substantial portion of the survival time entries are censored.

These patients were classified into four populations, depending on the amount of lymph node deterioration upon entry into the study. Population 0 indicates no evidence of lymph node metastases, while populations 1, 2 and 3 indicate the presence of sequentially more serious tumors. [This classification is indicated under the variable N in Kalbfleisch and Prentice (1980).] Since this example is for illustrative purposes only, we ignore other concomitant information. The survival time data for the four populations are given in Table 1. Note that the sample size corresponding to population 3 is significantly larger than the sample sizes corresponding to the other populations.

It would seem reasonable that the survival times for the four populations should be stochastically ordered since lymph node deterioration is an indication of the seriousness of the carcinoma. It is not clear whether uniform stochastic ordering should hold, however, since this is a considerably stronger condition.

Figure 1 shows estimates of the survival function of the four populations obtained by the Kaplan–Meier (1958) approach. As we would expect, the

---

**Fig. 1.** Kaplan–Meier estimates (MLE's) of survival functions from data in Table 1.
survival function of population 3 lies substantially below that of the others. However, the survival functions of the other three populations cross a number of times and the survival function of population 0, surprisingly, lies below those of populations 1 and 2, particularly beyond 350 days.

The mle's of the survival functions of the four populations under uniform stochastic ordering ($F_0 \succeq F_1 \succeq F_2 \succeq F_3$) are given in Figure 2. Since uniform stochastic ordering is a rather stringent restriction, the estimates become substantially separated, especially in the far right tails of the distributions where there is little information.

To illustrate the testing procedure, the data are grouped into seven classes as indicated in Table 1. We treat the grouped exact data as occurring at the interval midpoints, and the grouped censored data in an interval as occurring after the midpoint. The $n_{ij}$'s and $d_{ij}$'s for the grouped populations (as discussed in Section 2) are given in Table 2. The groups were chosen rather arbitrarily. We tried to roughly balance the survival times corresponding to each group and to have at least a minimal number of observations from each
### Table 2

**Number of survivals and deaths for grouped data**

<table>
<thead>
<tr>
<th>Group</th>
<th>Interval</th>
<th>Pop 0</th>
<th>Pop 1</th>
<th>Pop 2</th>
<th>Pop 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$n$</td>
<td>$d$</td>
<td>$n$</td>
<td>$d$</td>
</tr>
<tr>
<td>I</td>
<td>0–160</td>
<td>39</td>
<td>3</td>
<td>28</td>
<td>2</td>
</tr>
<tr>
<td>II</td>
<td>161–260</td>
<td>36</td>
<td>5</td>
<td>26</td>
<td>2</td>
</tr>
<tr>
<td>III</td>
<td>261–360</td>
<td>31</td>
<td>5</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td>IV</td>
<td>361–540</td>
<td>25</td>
<td>9</td>
<td>18</td>
<td>2</td>
</tr>
<tr>
<td>V</td>
<td>541–700</td>
<td>15</td>
<td>4</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>VI</td>
<td>701–900</td>
<td>10</td>
<td>2</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>VII</td>
<td>901–1850</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>2</td>
</tr>
</tbody>
</table>

$n$ is the number surviving at the beginning of the interval; $d$ is the number of deaths in the interval.

---

**Fig. 3.** Kaplan–Meier estimates (MLE’s) of survival functions from grouped data in Table 1.
population in each group. The choice of groupings could have a considerable effect on the outcome of the analysis.

For the grouped data, the unrestricted mle's (Kaplan and Meier) are given in Figure 3, and the restricted (uniform stochastic ordering) mle's are given in Figure 4. As indicated previously, populations 0, 1 and 2 seem to be quite similar, while population 3 tends to be substantially smaller.

The \( p \) value of the likelihood ratio statistic discussed in Section 3, where equality versus uniform stochastic ordering is tested, is 0.04 [where the approximation in (4.5) is used]. A \( p \) value this small seems somewhat surprising since the graphs in Figure 3 would not appear to support stochastic ordering. However, they also do not support equality. Population 3, with its large number of observations, lies well below the other three populations. The fact that it nearly satisfies the uniform stochastic ordering constraints apparently has a large impact on the test statistic and is the main reason for the small \( p \) value.
REFERENCES


DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE
UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242