Inference for Likelihood Ratio Ordering in the Two-Sample Problem

Richard DYKSTRA, Subhash KOCHAR, and Tim ROBERTSON*

We obtain the maximum likelihood estimators of two multinomial probability vectors under the constraint that they are likelihood ratio ordered. We extend this estimation approach to the case of two univariate distributions and show strong consistency of the estimators. We also derive and study the asymptotic distribution of the likelihood ratio statistic for testing the equality of two discrete probability distributions against the alternative that one distribution is greater than the other in the likelihood ratio ordering sense. Finally, we examine a data set pertaining to average daily insulin dose from the Boston Collaborative Drug Surveillance Program and compare our testing procedure to testing procedures for other stochastic orderings.

KEY WORDS: Chi-bar square distribution; Hazard rate ordering; Isotonic regression; Multinomial distribution; Stochastic ordering; Strong consistency; Uniform stochastic ordering.

1. INTRODUCTION

Stochastic ordering of distributions is an important concept in the theory of statistical inference. Many different types of stochastic ordering have been defined in the literature, and in fact a comprehensive volume from Academic Press on this topic (Shaked, Shanthikumar, and collaborators 1994) is now available.

One of the earliest definitions of stochastic ordering was given by Lehmann (1955): A random variable X with distribution function F is said to be stochastically greater than a random variable Y with distribution function G if

$$F(x) \le G(x) \quad \text{for every } x. \tag{1}$$

This is called (usual) *stochastic ordering* and is typically denoted by $X \overset{\text{st}}{>} Y$.

In some cases a pair of distributions may satisfy a stronger condition called *likelihood ratio ordering.* If distributions F and G possess densities (or probability mass functions) f and g, then the condition required for likelihood ratio ordering is given by

$$\frac{f(x)}{g(x)} \quad \text{is nondecreasing in } x. \tag{2}$$

This ordering is denoted by $X \overset{\text{LR}}{>} Y$ and has the interpretation that (2) holds if and only if for every $a < b$, the conditional distribution of X given $X \in [a, b]$ is stochastically greater than that of Y given $Y \in [a, b]$. Keilson and Sumita (1982) called this ordering *local uniform ordering* and discussed many of its properties. They also gave many examples of stochastic processes where the underlying distributions are likelihood ratio ordered. Ross (1983) and Shanthikumar and Yao (1991) have observed the usefulness of this ordering in some stochastic scheduling, closed queueing network, and reliability problems. It is known that $X \overset{\text{LR}}{>} Y$ implies that $\bar{F}(x)/\bar{G}(x)$ is nondecreasing in x. This latter condition defines *uniform stochastic ordering* (or hazard rate ordering), and this in turn implies stochastic ordering.

Although Ross (1983, p. 268) has shown that $X \overset{\text{LR}}{>} Y$ implies that $2X + Y \overset{\text{st}}{>} X + 2Y$ for independent random variables X and Y, this conclusion will *not* be implied by the lesser condition $X \overset{\text{st}}{>} Y$. Shanthikumar and Yao (1991) have generalized this result and have given some bivariate functional characterizations of these stochastic order relations. In particular, they have shown that $X \overset{\text{LR}}{>} Y$ if and only if $Eg(X, Y) \ge Eg(Y, X)$ for all $g \in \mathcal{G}_{lr} = \{g(x, y) : g(x, y) \ge g(y, x), \text{ for all } x \ge y\}$. (Some other important references on likelihood ratio ordering are Karlin and Rubin 1956, Lehmann 1955, and Whitt 1980.)

There has been a considerable amount of work done on inference problems concerning (usual) stochastic ordering. Brunk, Frank, Hanson, and Hogg (1966) obtained nonparametric maximum likelihood estimates (MLE's) of two stochastically ordered distribution functions and studied their properties. Testing procedures based on MLE's of two stochastically ordered distribution functions have been discussed by Robertson and Wright (1981), Lee and Wolfe (1976), Franck (1984), and Dykstra, Madsen, and Fairbanks (1983). Of course, the literature contains several distribution-free tests for testing the equality of distributions against stochastically ordered alternatives.

Dykstra, Kochar, and Robertson (1991) obtained MLE's of the survival functions of k distributions under uniform stochastic ordering (i.e., with ordered hazard rates). They also derived the asymptotic null distribution of the likelihood ratio statistic for testing the equality of distributions against the alternative that their hazard rates are uniformly stochastically ordered in a discrete setting. Park (1992) studied the likelihood ratio test for testing uniform stochastic ordering as a null hypothesis.

It is surprising that very little attention has been given to the problem of developing inference procedures for likelihood ratio–ordered distributions. But because this ordering has many important theoretical implications (see Sec. 6 for the relevance of this ordering in comparing nonhomogeneous Poisson processes), we feel that it is a topic worthy of additional study. We are not aware of any tests in the literature

* Richard Dykstra and Tim Robertson are Professors of Statistics, Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242. Subhash Kochar is Associate Professor of Statistics, Indian Statistical Institute, New Delhi 100016, India. This work was done while Kochar was a visiting professor at the University of Iowa. Partial support was provided by National Science Foundation Grant DMS 91-04673. The authors thank the editor, associate editor and the referees for helpful comments and suggestions.

Journal of the American Statistical Association
September 1995, Vol. 90, No. 431, Theory and Methods

specifically designed for testing the equality of two probability distributions against the alternative of likelihood ratio ordering. In this article we consider this testing problem for the discrete case.

We assume that a random sample of size m is taken from a multinomial distribution with probability vector $\mathbf{p} = (p_1, \ldots, p_k)$ and denote the corresponding vector of observed frequencies by $\mathbf{m} = (m_1, \ldots, m_k)$. Similarly, we let $\mathbf{n} = (n_1, \ldots, n_k)$ be the observed frequencies of an independent random sample of size n from another multinomial distribution with probability vector $\mathbf{q} = (q_1, \ldots, q_k)$ ($m = m_1 + \cdots + m_k$ and $n = n_1 + \cdots + n_k$). We derive the nonparametric MLE's of the probability vectors $\mathbf{p}$ and $\mathbf{q}$ under the hypotheses

$$H_0: \quad \mathbf{p} = \mathbf{q} \tag{3}$$

and

$$H_1: \quad \mathbf{p} \overset{\text{LR}}{>} \mathbf{q} \tag{4}$$

(i.e., p_i/q_i is nondecreasing in i, for $i = 1, \ldots, k$,) and use these estimates to construct a likelihood ratio test.

These MLE's are obtained in Section 2 and are derived in the discrete setting. But they provide generalized MLE's, in the sense of Kiefer and Wolfowitz (1956), under the assumption that the family of interest is the collection of all pairs of univariate distributions. In Section 3 these estimates are shown to be strongly consistent. This result is particularly interesting, because the MLE's under the assumption of uniform stochastic ordering are not consistent (cf. Rojo and Samaniego 1991).

In Section 4 we derive the asymptotic distribution of the likelihood ratio statistic for testing H_0 against H_1 in the discrete setting. The asymptotic null distribution is shown to be of the chi-bar-squared type. In Section 5 we illustrate these estimation and testing procedures using a data set concerning the mean daily dose of insulin for patients with and without hypoglycemia. In Section 6 we discuss how the procedures developed in this article can be used to make inferences about two nonhomogeneous Poisson processes.

2. MAXIMUM LIKELIHOOD ESTIMATION

In this section we obtain the MLE's of $\mathbf{p}$ and $\mathbf{q}$ under H_0 and H_1. We begin by expressing the likelihood function of $(\mathbf{p}, \mathbf{q})$ as

$$L \propto \prod_{i=1}^{k} p_i^{m_i} q_i^{n_i}.$$

We reparameterize by letting

$$\theta_i = mp_i/(mp_i + nq_i), \qquad \phi_i = mp_i + nq_i, \tag{5}$$

to obtain

$$p_i = \theta_i\phi_i/m, \qquad q_i = \phi_i(1 - \theta_i)/n, \tag{6}$$

for $i = 1, \ldots, k$.

The basic restrictions on $\mathbf{p}$ and $\mathbf{q}$ are

(a) $p_i \geq 0$, $q_i \geq 0$, for $i = 1, \ldots, k$, and
(b) $\sum_{i=1}^{k} p_i = \sum_{i=1}^{k} q_i = 1$.

It is easy to see that (a) and (b) are equivalent to
(c) $0 \leq \theta_i \leq 1$, $\phi_i \geq 0$, for $i = 1, \ldots, k$,
(d) $\sum_{i=1}^{k} \phi_i = m + n$, and
(e) $\sum_{i=1}^{k} \theta_i\phi_i = m$.

It is straightforward to show that under the null hypothesis H_0, the MLE's of $p_i = q_i$ are given by

$$p_i^0 = q_i^0 = (m_i + n_i)/(m + n). \tag{7}$$

Therefore, the MLE's, θ_i^0 and ϕ_i^0, of θ_i and ϕ_i are given by

$$\theta_i^0 = mp_i^0/(mp_i^0 + nq_i^0) = m/(m + n)$$

and

$$\phi_i^0 = mp_i^0 + nq_i^0 = m_i + n_i,$$

for $i = 1, \ldots, k$.

We note that the unconstrained MLE of θ_i is given by $\hat{\theta}_i = m_i/(m_i + n_i)$. It is easily shown that the MLE of $\boldsymbol{\theta}$ under H_0 is equivalent to the least squares projection of the vector $\hat{\boldsymbol{\theta}}$ onto the cone $\mathcal{C} = \{(\theta_1, \ldots, \theta_k); \theta_1 = \theta_2 = \cdots = \theta_k\}$ of constant vectors with weights $\mathbf{w} = (w_1, \ldots, w_k)$, where $w_i = m_i + n_i$. We express this fact by writing

$$\boldsymbol{\theta}^0 = E_{\mathbf{w}}(\hat{\boldsymbol{\theta}} \mid \mathcal{C}). \tag{8}$$

We next consider the problem of finding the MLE's of the parameters under H_1. Observe that H_1 will hold if and only if (c), (d), and (e) are true together with the condition that the θ_i's are nondecreasing in i for $i = 1, \ldots, k$.

Rewriting L in terms of the θ's and ϕ's, we obtain

$$L \propto \prod_{i=1}^{k} \left(\frac{1}{m}\theta_i\phi_i\right)^{m_i} \left(\frac{1}{n}\phi_i(1 - \theta_i)\right)^{n_i} \tag{9}$$

$$= \left(\frac{1}{m}\right)^{m} \left(\frac{1}{n}\right)^{n} \prod_{i=1}^{k} \theta_i^{m_i}(1 - \theta_i)^{n_i} \prod_{i=1}^{k} \phi_i^{m_i+n_i}. \tag{10}$$

Thus the likelihood function factors into two parts, one involving only θ_i's and the other only ϕ_i's. First, consider maximizing L subject to (c), (d), and nondecreasing θ_i's. The first factor is a bioassay problem as discussed by Robertson, Wright, and Dykstra (1988, ex 1.5). The second factor is a straightforward multinomial MLE problem. The maximums are achieved at

$$\phi_i^* = m_i + n_i, \qquad i = 1, \ldots, k, \tag{11}$$

and

$$\boldsymbol{\theta}^* = (\theta_1^*, \ldots, \theta_k^*), \tag{12}$$

the isotonic regression of the unconstrained MLE, $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ with weights $\mathbf{w} = (w_1, \ldots, w_k)$ onto the cone $\mathcal{I} = \{(\theta_1, \ldots, \theta_k); \theta_1 \leq \theta_2 \leq \cdots \leq \theta_k\}$ of nondecreasing vectors, where $w_i = m_i + n_i$ and $\hat{\theta}_i = m_i/(m_i + n_i)$. In our earlier notation, we can write $\boldsymbol{\theta}^* = E_{\mathbf{w}}(\hat{\boldsymbol{\theta}} \mid \mathcal{I})$.

Moreover,

$$\sum_{i=1}^{k} \theta_i^* \phi_i^* = \sum_{i=1}^{k} E_{\mathbf{w}}(\hat{\boldsymbol{\theta}} \mid \mathcal{I})_i(m_i + n_i) = \sum_{i=1}^{k} \hat{\theta}_i w_i = m,$$

using Theorem 1.3.3 of Robertson et al. (1988), so that (e) is also satisfied. Thus $\boldsymbol{\theta}^*$ and $\boldsymbol{\phi}^*$ are the MLE's of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$

under H_1. Observe that the MLE of ϕ is the same under both H_0 and H_1. Using (7), (11), and (12), we obtain the MLE's of $\mathbf{p}$ and $\mathbf{q}$ under H_1, as reported in the following theorem.

Theorem 2.1. If $(m_i + n_i) > 0$, $i = 1, \ldots, k$, then the MLE of $(\mathbf{p}, \mathbf{q})$ subject to H_1: $\mathbf{p} \overset{\mathrm{LR}}{>} \mathbf{q}$ is given by $(\mathbf{p}^*, \mathbf{q}^*)$, where

$$p_i^* = \left(\frac{m_i + n_i}{m}\right) E_{(\mathbf{m}+\mathbf{n})}\left(\frac{\mathbf{m}}{\mathbf{m}+\mathbf{n}} \,\middle|\, \mathcal{I}\right)_i \tag{13}$$

and

$$q_i^* = \left(\frac{m_i + n_i}{n}\right) E_{(\mathbf{m}+\mathbf{n})}\left(\frac{\mathbf{n}}{\mathbf{m}+\mathbf{n}} \,\middle|\, \mathcal{A}\right)_i \tag{14}$$

for $i = 1, \ldots, k$, where $\mathcal{A} = \{(\theta_1, \ldots, \theta_k); \theta_1 \geq \theta_2 \geq \cdots \geq \theta_k\}$ is the cone of nonincreasing vectors.

The similarities between $(\mathbf{p}^*, \mathbf{q}^*)$ and the MLE of $(\mathbf{p}, \mathbf{q})$ under (usual) stochastic order $\mathbf{p} \overset{\mathrm{st}}{>} \mathbf{q}(\sum_{i=1}^{j} p_i \leq \sum_{i=1}^{j} q_i$, $j = 1, \ldots, k-1$, $\sum_1^k p_i = \sum_1^k q_i)$ is rather surprising. In particular, Robertson et al. (1988, pp. 252–253) showed that these MLE's are given by

$$\bar{p}_i = \hat{p}_i E_{\hat{\mathbf{p}}}\left(\frac{m\hat{\mathbf{p}} + n\hat{\mathbf{q}}}{(m+n)\hat{\mathbf{p}}} \,\middle|\, \mathcal{I}\right)_i \tag{15}$$

and

$$\bar{q}_i = \hat{q}_i E_{\hat{\mathbf{q}}}\left(\frac{m\hat{\mathbf{p}} + n\hat{\mathbf{q}}}{(m+n)\hat{\mathbf{q}}} \,\middle|\, \mathcal{A}\right)_i. \tag{16}$$

This similarity is even more apparent when the respective MLE's are expressed as

$$p_i^* = \hat{p}_i\left(\frac{m_i + n_i}{m_i}\right) E_{\mathbf{m}+\mathbf{n}}\left(\frac{\mathbf{m}}{\mathbf{m}+\mathbf{n}} \,\middle|\, \mathcal{I}\right)_i \tag{17}$$

and

$$\bar{p}_i = \hat{p}_i\left[\frac{m+n}{m} E_{\mathbf{m}+\mathbf{n}}\left(\frac{\mathbf{m}}{\mathbf{m}+\mathbf{n}} \,\middle|\, \mathcal{A}\right)_i\right]^{-1} \tag{18}$$

(with similar expressions for q_i^* and $\bar{q}_i$).

For the MLE of $(\mathbf{p}, \mathbf{q})$ under the restriction that $\mathbf{p}$ is greater than $\mathbf{q}$ according to uniform stochastic ordering (hazard rate ordering), see the work of Dykstra et al. (1991).

3. CONSISTENCY

If k is held fixed while letting $m, n \to \infty$, then it is easy to show that $p_i^* \to p_i$ and $q_i^* \to q_i$ if $p_i/q_i \nearrow$ in i by using properties of isotonic regression. If one interprets maximum likelihood in the generalized sense (Kiefer and Wolfowitz 1956), which puts probability only on observed values, then the MLE's given in Section 2 yield MLE's when the family of interest consists of all pairs of univariate distributions that are likelihood ratio ordered. By this we mean that there exist probability density functions f and g with respect to a dominating measure μ such that $f(x)/g(x) \nearrow$ in x. Thus many pairs of continuous and mixed distributions will be in our family. A natural question to ask is whether these MLE's are consistent in the sense that the associated cdf's converge pointwise to the true cdf's when $m, n \to \infty$ and the likelihood ratio order holds.

This question is especially pertinent in light of the fact that the MLE's under uniform stochastic ordering (which is implied by likelihood ratio ordering) are not consistent, as was shown in a special case by Rojo and Samaniego (1991). (But MLE's under usual stochastic ordering are at least weakly consistent [Dykstra 1982].) Because the condition of uniform stochastic order ($\bar{F}/\bar{G} \nearrow$) is similar in concept to likelihood ratio order ($f/g \nearrow$), it is difficult to anticipate the answer. We now show that the answer is in the affirmative.

We let $F(G)$ denote the cdf corresponding to the density $f(g)$ and assume that we have independent random samples of size m and n. We initially assume that $n/m \to \lambda$, $(0 < \lambda < \infty)$. We let $\tilde{F}_{m,n}(\cdot)$ denote the likelihood ratio–ordered MLE of F derived in Section 2 and let $\hat{F}_m(\cdot)$ denote the usual empirical cdf for the first sample. (We consider only $\tilde{F}_{m,n}(\cdot)$; similar results hold for $\tilde{G}_{m,n}(\cdot)$.)

We fix ω (arbitrarily in a set of probability 1) such that $\hat{F}_m(x,\omega) \to F(x)$ and $\hat{G}_n(x,\omega) \to G(x)$ *uniformly* in x. It will suffice to show for a fixed $\varepsilon > 0$ and t, there exists $m(\varepsilon, \omega)$ and $n(\varepsilon, \omega)$ such that

$$|\tilde{F}_{m,n}(t, w) - \hat{F}_m(t, w)| < \varepsilon \qquad \text{for} \quad m \geq m(\varepsilon, \omega), \qquad n \geq n(\varepsilon, \omega)$$

(we henceforth suppress the ω).

We let $s_1, s_2, \ldots, s_{k(n,m)}$ denote the collection of distinct values from the combined random samples and assume that $a_1, a_2, \ldots, a_\nu$ denote the upper end points of the level sets of $E_{\mathbf{m}+\mathbf{n}}[\mathbf{m}/(\mathbf{m}+\mathbf{n}) \mid \mathcal{I}]$ (see Robertson et al. 1988, chap. 2, for details). The level sets are those subsets of the s_i where the least squares projection has constant value. Of course, ν, the number of distinct level sets, will be a random variable (as will the a_i) depending on the random samples. We assume that $a_{r-1} < t \leq a_r$ (and $a_0 = -\infty$).

Then we can write

$$\begin{aligned}
\tilde{F}_{m,n}(t) &= \sum_{i; s_i \leq t} \frac{m_i + n_i}{m} E_{\mathbf{m}+\mathbf{n}}\left(\frac{\mathbf{m}}{\mathbf{m}+\mathbf{n}} \,\middle|\, \mathcal{I}\right)_i \\
&= \frac{1}{m} \sum_{j=1}^{r-1} \sum_{a_{j-1} < s_i \leq a_j} (m_i + n_i)\left[\frac{\sum_{a_{j-1} < s_i \leq a_j} m_i}{\sum_{a_{j-1} < s_i \leq a_j} (m_i + n_i)}\right] \\
&\quad + \left[\frac{\sum_{a_{r-1} < s_i \leq t} (m_i + n_i)}{\sum_{a_{r-1} < s_i \leq t} m_i} \cdot \frac{\sum_{a_{r-1} < s_i \leq a_r} m_i}{\sum_{a_{r-1} < s_i \leq a_r} (m_i + n_i)}\right] \\
&\quad \times \frac{1}{m} \sum_{a_{r-1} < s_i \leq t} m_i
\end{aligned} \tag{19}$$

$$= \hat{F}_m(a_{r-1}) + A(\hat{F}_m(t) - \hat{F}_m(a_{r-1})) \tag{20}$$

$$= \hat{F}_m(t) - (1 - A)(\hat{F}_m(t) - \hat{F}_m(a_{r-1})), \tag{21}$$

where A is the last entry in brackets.

By the minimum lower sets algorithm (Robertson et al. 1988, p. 24), $0 \leq A \leq 1$ so that $\tilde{F}_{m,n}(t) \leq \hat{F}_m(t)$. Moreover, because the empirical cdf's converge *uniformly* to the true cdf's, we can say there exist positive integers n_0 and m_0 such that

$$\left| A - \frac{F(t) - F(a_{r-1}) + \lambda[G(t) - G(a_{r-1})]}{F(t) - F(a_{r-1})} \cdot \frac{F(a_r) - F(a_{r-1})}{F(a_r) - F(a_{r-1}) + \lambda[G(a_r) - G(a_{r-1})]} \right| \quad (22)$$

is less than ε if $F(t) - F(a_{r-1}) \geq \varepsilon/2$ (because then the denominator is bounded away from zero) for $n \geq n_0$ and $m \geq m_0$.

Now

$$f/g \quad \text{is nondecreasing}$$

$$\Leftrightarrow \frac{f}{f+\lambda g} \quad \text{is nondecreasing,}$$

from which it follows that

$$\frac{F(a_r) - F(t)}{F(t) - F(a_{r-1})} \geq \frac{F(a_r) - F(t) + \lambda[G(a_r) - G(t)]}{F(t) - F(a_{r-1}) + \lambda[G(t) - G(a_{r-1})]}$$

or, equivalently,

$$1 \leq \frac{F(t) - F(a_{r-1}) + \lambda[G(t) - G(a_{r-1})]}{F(t) - F(a_{r-1})} \cdot \frac{F(a_r) - F(a_{r-1})}{F(a_r) - F(a_{r-1}) + \lambda[G(a_r) - G(a_{r-1})]}.$$

Thus

$$A > 1 - \varepsilon \quad \text{if } n \geq n_0,\ m \geq m_0 \quad \text{and} \quad F(t) - F(a_{r-1}) \geq \varepsilon/2.$$

Finally, if we select $m_1 \geq m_0$ such that

$$|(\hat{F}_m(t) - \hat{F}_m(a_{r-1})) - (F(t) - F(a_{r-1}))| < \frac{\varepsilon}{2} \quad \text{if } m \geq m_1,$$

then it follows that

$$|\tilde{F}_{m,n}(t) - \hat{F}_m(t)| < \varepsilon \quad \text{for} \quad m \geq m_1,\ n \geq n_0 \quad \text{by (21).} \quad (23)$$

Careful scrutiny will reveal that this result also holds when $\lambda = 0$ or ∞. Now suppose that $m \rightarrow \infty$ arbitrarily. Then it suffices to show that every subsequence $(m_k, n_k)_{k=1}^{\infty}$ contains a sub-subsequence that converges correctly. But it is always possible to choose a sub-subsequence of $(m_k, n_k)_{k=1}^{\infty}$ whose ratio converges (possibly to 0 or ∞), so that the general result holds. It easily follows that $\tilde{F}_{m,n}(\cdot)$ converges uniformly to $F(\cdot)$ a.s., as long as $m \rightarrow \infty$ (regardless of the behavior of n). Additional work will show that the convergence is of order $m^{-1/2}$ which is the best that we could hope for.

4. THE LIKELIHOOD RATIO TEST

We now consider the problem of testing the null hypothesis H_0 against the alternative H_1. In our asymptotic theory, k, the number of support points, is fixed. We initially assume that the sample sizes n and m increase to ∞ in such a way that $n/m \rightarrow \lambda$, $0 < \lambda < 1$, and $(m+n)^{1/2}(n/m - \lambda) \rightarrow 0$.

The likelihood ratio statistic is

$$\Psi = \frac{\sup_{(\mathbf{p},\mathbf{q}) \in H_0} L((\mathbf{p}, \mathbf{q}))}{\sup_{(\mathbf{p},\mathbf{q}) \in H_1} L((\mathbf{p}, \mathbf{q}))} = \frac{L(\mathbf{p}^0, \mathbf{q}^0)}{L(\mathbf{p}^*, \mathbf{q}^*)}$$

$$= \frac{\left(\frac{1}{m}\right)^m \left(\frac{1}{n}\right)^n \prod_{i=1}^k (\theta_i^0)^{m_i} (1-\theta_i^0)^{n_i} \prod_{i=1}^k (\phi_i^0)^{m_i+n_i}}{\left(\frac{1}{m}\right)^m \left(\frac{1}{n}\right)^n \prod_{i=1}^k (\theta_i^*)^{m_i} (1-\theta_i^*)^{n_i} \prod_{i=1}^k (\phi_i^*)^{m_i+n_i}}$$

$$= \prod_{i=1}^k \left(\frac{\theta_i^0}{\theta_i^*}\right)^{m_i} \left(\frac{1-\theta_i^0}{1-\theta_i^*}\right)^{n_i}, \quad (24)$$

because $\phi_i^0 = \phi_i^*$. Our test rejects H_0 for large values of $T = -2 \ln \Psi$; that is, for large values of

$$T = 2 \sum_{i=1}^k \{m_i \ln \theta_i^* + n_i \ln(1-\theta_i^*) - m_i \ln \theta_i^0 - n_i \ln(1-\theta_i^0)\}. \quad (25)$$

If we expand $\ln \theta_i^*$ and $\ln \theta_i^0$ about $\hat{\theta}_i$ and expand $\ln(1 - \theta_i^*)$ and $\ln(1 - \theta_i^0)$ about $(1 - \hat{\theta}_i)$ and use properties of isotonic regression, then we find that the linear terms in the expansion drop out. On simplification, T reduces to

$$T = \sum_{i=1}^k \left\{ (\hat{\theta}_i - \theta_i^0)^2 \left(\frac{m_i}{\beta_i^2} + \frac{n_i}{\delta_i^2} \right) - (\hat{\theta}_i - \theta_i^*)^2 \left(\frac{m_i}{\alpha_i^2} + \frac{n_i}{\gamma_i^2} \right) \right\}, \quad (26)$$

where

$$\max\{|\beta_i - \hat{\theta}_i|, |\delta_i - (1 - \hat{\theta}_i)|\} \leq |\theta_i^0 - \hat{\theta}_i| \quad (27)$$

and

$$\max\{|\alpha_i - \hat{\theta}_i|, |\gamma_i - (1 - \hat{\theta}_i)|\} \leq |\theta_i^* - \hat{\theta}_i|. \quad (28)$$

When we assume that H_0 is true, the right sides of (27) and (28) go to zero, which implies that T is asymptotically equivalent ($\approx$) to

$$\sum_{i=1}^k [(\hat{\theta}_i - \theta_i^0)^2 - (\hat{\theta}_i - \theta_i^*)^2](m_i + n_i) \frac{m_i + n_i}{m_i} \frac{m_i + n_i}{m_i} \cdot \frac{m_i}{n_i}$$

$$\approx \sum_{i=1}^k [\hat{\theta}_i^2 - 2\hat{\theta}_i \theta_i^0 + \theta_i^{02} - \hat{\theta}_i^2 + 2\hat{\theta}_i \theta_i^* - \theta_i^{*2}] \times (m_i + n_i)(\lambda + 1)^2/\lambda$$

$$= \sum_{i=1}^k \sum_{i=1}^k (\theta_i^* - \theta_i^0)^2 (m_i + n_i)(\lambda + 1)^2/\lambda \quad (29)$$

(because $\sum_{i=1}^k \hat{\theta}_i \theta_i^0 (m_i + n_i) = \sum_{i=1}^k \theta_i^* \theta_i^0 (m_i + n_i)$ and $\sum_{i=1}^k \hat{\theta}_i \theta_i^* (m_i + n_i) = \sum_{i=1}^k \theta_i^{*2} (m_i + n_i)$). Expression (29) can be written as

$$\sum_{i=1}^{k} [E_{\mathbf{w}}(\hat{\theta}|\mathcal{I})_i - E_{\mathbf{w}}(\hat{\theta}|\mathcal{C})_i]^2(m_i+n_i)(\lambda+1)^2/\lambda$$
$$\approx \sum_{i=1}^{k} [E_{\mathbf{w}}(\hat{\theta}|\mathcal{I})_i - E_{\mathbf{w}}(\hat{\theta}|\mathcal{C})_i]^2 \times (m+n)p_i(\lambda+1)^2/\lambda$$
$$= \sum_{i=1}^{k} \left[E_{\mathbf{w}}\left(\sqrt{n+m}\left(\hat{\theta} - \frac{1}{\lambda+1}\right)\Big|\mathcal{I}\right)_i - E_{\mathbf{w}}\left(\sqrt{n+m}\left(\hat{\theta} - \frac{1}{\lambda+1}\right)\Big|\mathcal{C}\right)_i\right]^2 \times p_i(\lambda+1)^2/\lambda.$$

But under our earlier assumptions,

$$\sqrt{m+n}\left[\begin{pmatrix} \frac{m}{m+n}\hat{\mathbf{p}} \\ \frac{n}{m+n}\hat{\mathbf{q}} \end{pmatrix} - \begin{pmatrix} \frac{1}{1+\lambda}\mathbf{p} \\ \frac{\lambda}{1+\lambda}\mathbf{q} \end{pmatrix}\right] \xrightarrow{\mathcal{L}} \mathrm{MVN}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{1+\lambda}\boldsymbol{\Sigma}_{\mathbf{p}} & 0 \\ 0 & \frac{\lambda}{1+\lambda}\boldsymbol{\Sigma}_{\mathbf{q}} \end{pmatrix}\right], \quad (30)$$

where

$$\boldsymbol{\Sigma}_{\mathbf{p}} = (\delta_{ij} - p_i p_j), \qquad 1 \le i, j \le k,$$

and

$$\boldsymbol{\Sigma}_{\mathbf{q}} = (\delta_{ij} - q_i q_j), \qquad 1 \le i, j \le k,$$

are standard multinomial covariance matrices (where δ_{ij} is the Kronecker delta; i.e., $\delta_{ij} = 0$ when $i \ne j$ and $\delta_{ij} = 1$ if $i = j$).

Straightforward but tedious application of the delta method (cf. Serfling 1980) will show that if H_0 is true, then

$$\sqrt{m+n}\left(\hat{\theta} - \frac{1}{1+\lambda}\right) \xrightarrow{\mathcal{L}} \mathrm{MVN}(\mathbf{0}, \boldsymbol{\Lambda}),$$

where

$$\boldsymbol{\Lambda} = \frac{\lambda}{(1+\lambda)^2}\begin{pmatrix} \frac{1}{p_1}-1 & -1 & \cdots & -1 \\ -1 & \frac{1}{p_2}-1 & \cdots & -1 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & -1 \\ -1 & \cdot & -1 & \frac{1}{p_k}-1 \end{pmatrix}. \quad (31)$$

Upon observing that this is the distribution of

$$(X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_k - \bar{X}), \quad (32)$$

where $X_1, \ldots, X_k$ are independent $N(0, \lambda/((1+\lambda)^2 p_i))$, $1 \le i \le k$, random variables with $\bar{X} = \sum_{i=1}^{k} p_i X_i$, we may use continuity properties (both in the argument and the weights) of a least squares projection to say that T is asymptotically distributed as

$$\sum_{i=1}^{k} [E_{\mathbf{p}}(\mathbf{X} - \bar{\mathbf{X}}|\mathcal{I})_i - E_{\mathbf{p}}(\mathbf{X} - \bar{\mathbf{X}}|\mathcal{C})_i]^2 p_i(\lambda+1)^2/\lambda$$
$$= \sum_{i=1}^{k} [E_{\mathbf{p}}(\mathbf{X}|\mathcal{I})_i - E_{\mathbf{p}}(\mathbf{X}|\mathcal{C})_i]^2[\mathrm{var}(X_i)]^{-1}. \quad (33)$$

But the exact distribution of (33), as obtained by Robertson et al. (1988, pp. 68–74), is a chi-bar-squared distribution, which is a mixture of chi-squared distributions, mixed over the degrees of freedom. Note that this distribution is free of λ. Additional work will show that the same asymptotic distribution holds under H_0 as long as both m and n go to infinity. We spare the reader the unpleasant details. A precise statement of the asymptotic distribution of T is given in the following theorem.

Theorem 3.1. If $p_i = q_i > 0$, $i = 1, \ldots, k$, and m and n go to ∞, then for all $t > 0$,

$$\lim_{m,n\to\infty} \mathrm{pr}[T \ge t] = \sum_{l=1}^{k} P(l, k; \mathbf{p})\mathrm{pr}(\chi^2_{l-1} \ge t) \quad (34)$$

where χ^2_ν denotes a chi-squared random variable with ν degrees of freedom and $P(l, k; \mathbf{p})$ is the probability that $E_{\mathbf{p}}(\mathbf{X}|\mathcal{I})$ takes on l distinct values, where $\mathbf{X} = (X_1, \ldots, X_k)$ consists of independent random variables and X_i is $N(0, 1/p_i)$.

We recommend that the quantity $\mathrm{pr}[T \ge t]$ be approximated by

$$\sum_{l=1}^{k} P(l, k; \hat{\mathbf{p}})\mathrm{pr}(\chi^2_{l-1} \ge t), \quad (35)$$

where $\hat{p}_i = (m_i + n_i)/(m+n)$. This expression has the same asymptotic distribution as T and generally provides a very good approximation to the distribution of T. Expressions for $P(l, k; \mathbf{p})$ have been given by Robertson et al. (1988, pp. 77–79) for k up to 5. Numerical simulations (or some other approximation, such as those discussed in chapter III of Robertson et al. 1988) are typically needed to approximate $P(l, k; \mathbf{p})$ for $k > 5$.

If the p_i's are of roughly the same magnitude (as evidenced by the values of $\hat{\mathbf{p}}$), then the equal weights $P(l, k)$ give an extremely robust approximation to the distribution of T. This approximation is remarkably good, as evidenced by the example in Section 5. Robertson and Wright (1983) recommended the equal weights approximation as long as $\sup_{i,j} p_i/p_j \le 4$. Robertson et al. (1988, chap. 2) gave a recursive formula (as well as tables for k up to 20 in appendix, table A.10) for these $P(l, k)$. Critical points for the accompanying distribution were also given in tables at the back of that book. A Fortran program for implementing a recommended pattern approximation was given by Pillers, Robertson, and Wright (1984).

Although the asymptotic least favorable distribution of T under H_0 is given by

$$\sup_{\mathbf{p}=\mathbf{q}} \lim_{n,m\to\infty} \operatorname{pr}[T \geq t] = \sum_{l=1}^{k} \binom{k-1}{l-1}\left(\frac{1}{2}\right)^{k-1} \operatorname{pr}(\chi^2_{l-1} \geq t), \quad (36)$$

this tends to be a very conservative bound and is not recommended except as a crude guideline.

5. EXAMPLE

To illustrate the estimation and testing procedures discussed in earlier sections, here we examine a data set discussed in a report from the Boston Collaborative Drug Surveillance Program (1974). The data set consists of observed values for the mean daily insulin dose from 80 subjects categorized as "hypoglycemia present" and 245 subjects from the population "hypoglycemia absent." The measurements are grouped into five ordered categories and are shown in Table 1.

One would expect that hypoglycemia (low blood sugar) would occur when large amounts of glucose are metabolized and hence would be consistent with higher levels of insulin dosage. This would suggest the hypothesis $F \overset{\text{LR}}{\geq} G$.

The computed value of the likelihood ratio statistic $T = -2 \ln \Lambda$ is 9.703. We computed the estimated chi-bar-squared weights $P(l, k; \hat{p})$ from the formulas of Robertson et al. (1988, pp. 77–79). These weights are given by $P(5, 5; \hat{p}) = .010$, $P(4, 5; \hat{p}) = .096$, $P(3, 5; \hat{p}) = .308$, $P(2, 5; \hat{p}) = .404$, and $P(1, 5; \hat{p}) = .182$. Despite the variability in $\hat{p}$, ($\sup_{i,j} \hat{p}_i/\hat{p}_j = 2.32$), these weights are remarkably similar to the equal weights $P(l, k)$ tabled by Robertson et al.: ($P(5, 5) = .0081$, $P(4, 5) = .083$, $P(3, 5) = .292$, $P(2, 5) = .417$, and $P(1, 5) = .200$). The p values are .006 for the first set of weights and .005 for the $P(l, k)$ approximation. Clearly there is strong evidence supporting the likelihood ratio ordering hypothesis over equality of distributions.

For comparison sake, we also computed the MLE's and likelihood ratio statistics for testing equality of distributions versus uniform stochastic ordering (implied by likelihood ratio ordering; see Dykstra et al. 1991) as well as the MLE's and likelihood ratio statistics for stochastic ordering (implied by uniform stochastic ordering; see Robertson and Wright 1981). Finally, we ran the standard likelihood ratio test for testing equality of distributions against all alternatives.

Rather surprisingly, the MLE's under uniform stochastic ordering are identical to the MLE's for likelihood ratio ordering and hence give the same test statistic value of 9.703. Because there are only two populations, the asymptotic chi-bar-squared weights can be expressed as binomial (4, 1/2)

Table 1. Mean Daily Insulin Dose and Maximum Likelihood Estimates Under Stochastic Orderings

Insulin level	1	2	3	4	5
	<.25	.25–49	.50–.74	.75–.99	≥1.0
Pop. I (F)—Hypoglycemia present	4	21	28	15	12
Pop. II (G)—Hypoglycemia absent	40	74	59	26	46

MLE's $F \overset{LR}{\geq} G$[a] likelihood ratio order

	$\hat{p}_i = \frac{m_i}{m}$	$\hat{q}_i = \frac{n_i}{n}$	$\hat{\hat{p}}_i = \frac{m_i + n_i}{m + n}$	$\hat{\theta}_i = \frac{m_i}{m_i + n_i}$	θ_i^*	p_i^*	q_i^*	$\theta_i^0 = \frac{m}{m + n}$
1	.050	.163	.135	.091	.091	.050	.163	.246
2	.262	.302	.292	.221	.221	.262	.302	.246
3	.350	.241	.268	.322	.296	.322	.250	.246
4	.188	.106	.126	.366	.296	.152	.118	.246
5	.150	.188	.178	.207	.296	.215	.167	.246

MLE's $F \overset{ust}{\geq} G$[b] $F \overset{st}{\geq} G$[c]

	$F \overset{ust}{\geq} G$ uniform stochastic order							$F \overset{st}{\geq} G$ stochastic order	
	$\hat{\theta}_{1i} = \frac{n_{1i} - d_{1i}}{n_{1i}}$	$\hat{\theta}_{2i} = \frac{n_{2i} - d_{2i}}{n_{2i}}$	$\theta_i^0 = \frac{n_{1i} + n_{2i} - d_{1i} - d_{2i}}{n_{1i} + n_{2i}}$	$\tilde{\theta}_{1i}$	$\tilde{\theta}_{2i}$	$\tilde{p}_i$	$\tilde{q}_i$	$\tilde{p}_i$	$\tilde{q}_i$
1	.950	.837	.865	.950	.837	.050	.163	.048	.165
2	.724	.639	.662	.724	.639	.262	.302	.254	.306
3	.401	.550	.532	.532	.532	.322	.250	.338	.244
4	.444	.639	.586	.586	.586	.152	.118	.181	.107
5	0	0	0	0	0	.215	.167	.178	.178

[a] p_i/q_i increasing in i

[b] $\sum_{j=i}^{5} p_j / \sum_{j=i}^{5} q_j$ increasing in i

[c] $\sum_{j=i}^{5} p_j \geq \sum_{j=i}^{5} q_j$ for all i

probabilities giving a p value of .012. Moreover, this implies that the likelihood ratio statistic for testing likelihood ratio order against uniform stochastic order is zero and hence provides no support for choosing (less restrictive) uniform stochastic order over (more restrictive) likelihood ratio order.

The MLE's under stochastic order (st) and uniform stochastic order (ust) are also given in Table 1. The likelihood ratio statistic value for testing equality versus stochastic order is 12.742, and the asymptotic chi-bar-squared weights are expressible from the likelihood ratio ordering weights as $P_{st}(l, k; p) = P_{LR}(k + 1 - l, k; p)$, $l = 1, \ldots, k$. In this case the p values are again .005 for the equal weights approximation ($P(l, k)$) and .006 when p is estimated by $\hat{p}$. The conservative binomial bound (36) gives a p value of .012 for the likelihood ratio ordering case. A least favorable bound gives .009 for stochastic ordering. The likelihood ratio test statistic value for testing equality of distributions against all alternatives is 13.268, which gives a p value of .010 from the chi-squared (4) distribution.

6. POISSON PROCESSES

The methods developed in this article can also be used to compare trends in Poisson processes (cf. Boyett and Saw 1980; Lee 1982, 1982). Let $N_1(t)$ and $N_2(t)$ be two nonhomogeneous Poisson processes with mean value functions $\Lambda_1(t)$ and $\Lambda_2(t)$ and let $\lambda_i(t) = d\Lambda_i(t)/dt$ denote the intensity function corresponding to $N_i(t)$; $i = 1, 2$.

Suppose that we observe these two processes up to a predetermined time T_0, and that we let $0 < t_{i,1} < t_{i,2} < \ldots t_{i,n(i)} < T_0$, $i = 1, 2$ be the observed times of occurrence for these two processes. Then it is well known that conditional on $N_i(T_0) = n(i)$, the observation times $t_1, t_2, \ldots t_{n(i)}$ have the same distribution as the order statistics of a random sample of size $n(i)$ from a distribution with density, $f_i(t) = \lambda_i(t)/\Lambda_i(T_0)$, $i = 1, 2$. Because λ_1/λ_2 is proportional to f_1/f_2, the procedures developed here can be adapted to make inferences about λ_1/λ_2 based on the foregoing data when collected in group form. Specifically, we can estimate the intensity functions (and thus the mean functions) subject to the restriction that their ratio is monotone in t using the methods developed here. We can also test the null hypothesis that this ratio is a constant against the alternative that it is monotone in t, using an adapted version of our likelihood ratio test in Section 4. Obviously, the resulting test will be conditional on $N_i(T_0) = n(i)$, $i = 1, 2$.

7. CONCLUSION

It is shown in this article that MLE's for distributions that are likelihood ratio ordered can be obtained in a form that is similar to MLE's under (usual) stochastic ordering constraints. These MLE's can be neatly characterized in terms of least squares projections onto isotonic cones. It is also shown that these MLE's are strongly consistent in a general setting and that they converge at the rate $n^{-1/2}$.

The asymptotic distribution of the likelihood ratio statistic for testing equality of distributions against the alternative that the distributions are likelihood ratio ordered is also derived and shown to be of the chi-bar-squared type. Moreover, this chi-bar-squared distribution involves the same weighting coefficients (reversed in order) as the likelihood ratio test for (usual) stochastic order.

[Received September 1993. Revised December 1994.]

REFERENCES

Boston Collaborative Drug Surveillance Program (1974), "Relation of Bodyweight and Insulin Dose to the Frequency of Hypoglycemia," *Journal of the American Medical Association*, 228, 192–194.

Boyett, J., and Saw, J. (1980), "On Comparing Two Poisson Intensity Functions," *Communications in Statistics, Part A—Theory and Methods*, 9, 943–948.

Brunk, H. D., Frank, W. E., Hanson, D. L., and Hogg, R. V. (1966), "Maximum Likelihood Estimation of the Distributions of Two Stochastically Ordered Random Variables," *Journal of the American Statistical Association*, 61, 1067–1080.

Dykstra, R. L. (1982), "Maximum Likelihood Estimation of the Survival Functions of Stochastically Ordered Random Variables," *Journal of the American Statistical Association*, 77, 621–628.

Dykstra, R., Kochar, S., and Robertson, T. (1991), "Statistical Inference for Uniform Stochastic Ordering in Several Populations," *The Annals of Statistics*, 19, 870–888.

Dykstra, R., Madsen, R. W., and Fairbanks, K. (1983), "A Nonparametric Likelihood Ratio Test," *Journal of Statistical Computations and Simulations*, 18, 247–264.

Franck, W. E. (1984), "A Likelihood Ratio Test for Stochastic Ordering," *Journal of the American Statistical Association*, 79, 686–691.

Karlin, S., and Rubin, H. (1956), "The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio," *Annals of Mathematical Statistics*, 27, 272–299.

Keilson, J., and Sumita, U. (1982), "Uniform Stochastic Ordering and Related Inequalities," *Canadian Journal of Statistics*, 10, 181–198.

Kiefer, J., and Wolfowitz, J. (1956), "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters," *Annals of Mathematical Statistics*, 27, 887–906.

Lee, L. (1981), "A Graphical Method for Comparing Trends in Series of Events," *Communications in Statistics, Part A—Theory and Methods*, 10, 827–848.

——— (1982), "Distribution-Free Tests for Comparing Trends in Poisson Series," *Stochastic Processes and Their Applications*, 12, 107–113.

Lee, Y. J., and Wolfe, D. A. (1976), "A Distribution-Free Test for Stochastic Ordering," *Journal of the American Statistical Association*, 71, 722–727.

Lehmann, E. L. (1955), "Ordered Families of Distributions," *Annals of Mathematical Statistics*, 26, 399–419.

Park, C. G. (1992), "Statistical Inferences for Uniform Stochastic Ordering," unpublished Ph.D. dissertation, University of Iowa, Dept. of Statistics.

Pillers, C. T., Robertson, T., and Wright, F. T. (1984), "A FORTRAN Program for the Level Probabilities of Order-Restricted Inference," *Journal of the Royal Statistical Association*, Ser. C, 33, 115–119.

Robertson, T., and Wright, F. T. (1981), "Likelihood Ratio Tests for and Against Stochastic Ordering Between Multinomial Populations," *The Annals of Statistics*, 9, 1248–1257.

——— (1983), "On Approximation of the Level Probabilities and Associated Distributions in Order-Restricted Inference," *Biometrika*, 70, 597–606.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988), *Order-Restricted Statistical Inference*, Chichester, U.K.: John Wiley.

Rojo, J., and Samaniego, F. J. (1991), "On Nonparametric Maximum Likelihood Estimation of a Distribution Uniformly Stochastically Smaller Than a Standard," *Statistics and Probability Letters*, 11, 267–271.

Ross, S. (1983), *Stochastic Processes*, New York: John Wiley.

Shaked, M., Shanthikumar, J. G., and collaborators (1994), *Stochastic Orders and Their Applications*, San Diego: Academic Press.

Shanthikumar, J. G., and Yao, D. D. (1991), "Bivariate Characterization of Some Stochastic Order Relations," *Advances in Applied Probability*, 23, 642–659.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: John Wiley.

Whitt, W. (1980), "Uniform Conditional Stochastic Order," *Journal of Applied Probability*, 17, 112–123.