LIKELIHOOD RATIO TESTS FOR SYMMETRY
AGAINST ONE-SIDED ALTERNATIVES

RICHARD DYKSTRA[1], SUBHASH KOCHAR[2] AND TIM ROBERTSON[1]

[1] *Department of Statistics and Actuarial Science, The University of Iowa,*
Iowa City, IA 52242, U.S.A.
[2] *Indian Statistical Institute, 7, S.J.S. Sansanwal Marg, New Delhi - 110016, India*

Abstract. A random variable X is said to have a symmetric distribution (about 0) if and only if X and $-X$ are identically distributed. By considering various types of partial orderings between the distributions of X and $-X$, one obtains various notions of skewness or one-sided bias. In this paper we study likelihood ratio tests for testing the symmetry of a discrete distribution about zero against the alternatives, (i) X is stochastically greater than $-X$; and (ii) $\mathrm{pr}(X = j) \geq \mathrm{pr}(X = -j)$ for all $j > 0$. In the process, we obtain maximum likelihood estimators of the distribution function under the above alternatives. The asymptotic null distributions of the test statistics have been obtained and are of the chi-bar square type. A simulation study was performed to compare the powers of these tests with other tests.

Key words and phrases: Chi-bar square distribution, chi square test for goodness of fit, isotonic regression, positive biasedness, skewness, stochastic ordering.

1. Introduction

A common assumption underlying many statistical analyses is that the underlying distribution is symmetric. The validity of some commonly used procedures depends heavily upon this assumption. This is particularly true of several non-parametric procedures, such as the Wilcoxon signed rank test. Moreover, it is well known that many statistical procedures based on normal theory are robust to the normality assumption provided that the underlying distribution is symmetric. For example, in an intensive simulation study Chaffin and Rhiel (1993) have confirmed that the one sample t-test is approximately valid even for non-normal distributions provided they are symmetric. However, this approximation may fail very badly if the underlying distribution is skewed.

Testing for the symmetry of a distribution about a specified or unspecified point θ has always been an important topic of interest in statistics. In this paper we shall focus on the case when the point of symmetry θ is specified.

A large number of nonparametric tests are available in the literature for this problem against both one-sided and two-sided alternatives. However, most of the research on one-sided alternatives has focused on developing nonparametric tests for possible changes in the location of a symmetric distribution. Many of the suggested tests for symmetry can be described as variations of either sign tests, Wilcoxon tests, Komogorov-Smirnov tests or Cramer-von Mises tests. Since symmetry (about 0) is equivalent to $F(t-) + F(-t) = 1\forall t$, many tests are phrased in terms of the empirical CDF and take advantage of the wealth of knowledge concerning the empirical CDF. A good reference for this and related problems is Chapter 22 of Shorack and Wellner (1986).

It is our point that if symmetry is to be rejected, then it is important to know the distributional structure that led to its rejection. Since X has a symmetric distribution iff X and $-X$ have the same distribution, one can consider different one-sided alternatives to symmetry by considering various types of stochastic ordering between the random variables X and $-X$. Certainly the most common and well understood notion of positive bias (skewness to the right) would be to have X stochastically greater than $-X$. This would imply that $E(g(X)) \geq E(g(-X))$ for all nondecreasing functions g.

In this paper we discuss this problem when the data is discrete or grouped and assume without loss of generality that $\theta = 0$. We let X take on the $(2k + 1)$ values $-k, -(k-1), \ldots, -1, 0, 1, \ldots, (k-1), k$ with corresponding probabilities $p_{-k}, p_{-(k-1)}, \ldots, p_{-1}, p_0, p_1, \ldots, p_{k-1}, p_k$ although any finite set which equals its negative would work as well. We shall let $\boldsymbol{p}$ denote the $2k + 1$ dimensional vector of p_i's.

Assume that we have a random sample of size n from our population and let n_i be the number of times that X takes the value i for $i = 0, \pm 1, \ldots, \pm k$ so that $\sum_{i=-k}^{k} n_i = n$. Based on this data we consider the problem of testing the null hypothesis of symmetry about 0

$$(1.1) \qquad\qquad H_0 : p_j = p_{-j}; \qquad j = 1, 2, \ldots, k$$

against the alternatives $H_1 - H_0$ and $H_2 - H_0$ where

$$(1.2) \qquad\qquad H_1 : \sum_{i=j}^{k} p_i \geq \sum_{i=j}^{k} p_{-i}, \qquad j = 1, 2, \ldots, k$$

and

$$(1.3) \qquad\qquad H_2 : p_j \geq p_{-j}, \qquad j = 1, 2, \ldots, k.$$

The criteria of positive bias as represented by H_1 and H_2 have been discussed by Yanagimoto and Sibuya (1972). If (1.2) holds, we say that X is positively biased according to criterion $\mathcal{B}_1$ (type I bias) and if (1.3) holds, we say that X is positively biased according to criterion $\mathcal{B}_2$ (type II bias).

In the next section we obtain the maximum likelihood estimators (MLE's) of $\boldsymbol{p}$ under the hypotheses H_0 and H_1 and then use these estimators to obtain the likelihood ratio statistic for testing H_0 against $H_1 - H_0$. The asymptotic null distribution of this test statistic is obtained and shown to be of the chi-bar

square type. The corresponding problem of testing H_0 against $H_2 - H_0$ is treated in Section 3. In Section 4, we show that the maximum likelihood estimator of F, the distribution function of X under H_1, as obtained in Section 2, is strongly consistent when considered in the class of all univariate distributions. Surprisingly, the maximum likelihood estimator of F under H_2 obtained in Section 3 is *not* consistent when considered in the general setting of all univariate distributions. In the last section we perform a short Monte Carlo study to compare the powers of the various tests. For this purpose, we focus on the shifted binomial distribution for which the uniformly most powerful (UMP) test for testing symmetry against either $H_1 - H_0$ or $H_2 - H_0$ is to reject the null hypothesis when the unrestricted MLE of the parameter, p, is too large. We use this UMP test as a benchmark and show that our nonparametric tests are quite powerful.

2.   Testing H_0 against $H_1 - H_0$

First we obtain the maximum likelihood estimators (MLE) of the vector $\boldsymbol{p}$ under both H_0 and H_1. The likelihood function is proportional to

$$(2.1) \qquad L(\boldsymbol{p} \mid \boldsymbol{n}) = p_0^{n_0} \prod_{i=1}^{k} [p_{-i}^{n_{-i}} p_i^{n_i}].$$

The unrestricted MLE of p_i is, of course, $\hat{p}_i = n_i/n$, $i = 0, \pm 1, \ldots, \pm k$ and the MLE of $\boldsymbol{p}$ under H_0 is given by

$$(2.2) \qquad \hat{p}_{-i}^{(0)} = \hat{p}_i^{(0)} = (n_{-i} + n_i)/2n, \qquad i = 1, 2, \ldots, k$$

and

$$\hat{p}_0^{(0)} = \hat{p}_0 = n_0/n.$$

Under H_1, the constraints on the vector $\boldsymbol{p}$ are

$$\sum_{i=j}^{k} p_i \geq \sum_{i=j}^{k} p_{-i}, \qquad j = 1, 2, \ldots, k.$$

These constraints are equivalent to

$$(2.3) \qquad \sum_{i=j}^{k} p_i \geq \sum_{i=j}^{k} p_{-i}, \qquad j = -k, -k+1, \ldots, k$$

which we write as $\boldsymbol{p} \gg \boldsymbol{p}'$, using the notation of Robertson, Wright and Dykstra (1988) (to be written RWD henceforth). Here $\boldsymbol{p}'$ denotes the reversed $(2k + 1)$ dimensional vector $(p_k, p_{k-1}, \ldots, p_{-k})^T$, (note that "prime" here means "reversed" and not vector transposition).

Now, maximizing the likelihood product, $L(\boldsymbol{p} \mid \boldsymbol{n})$ under H_1 is equivalent to maximizing

$$(2.4) \qquad L^2(\boldsymbol{p} \mid \boldsymbol{n}) = \prod_{i=-k}^{k} p_i^{n_i} \prod_{i=-k}^{k} p_{-i}^{n_{-i}}$$

subject to (2.3). The two-sample problem has been studied by Brunk *et al.* (1966) and Barlow and Brunk (1972) and its solution can be used here (throughout this paper, we assume multiplication and division of vectors is done coordinate-wise).

THEOREM 2.1.   *If $\hat{p}_i > 0$ for $i = -k, \ldots, k$ then the MLE of p subject to the restriction, H_1, is given by*

$$(2.5) \qquad \hat{p}^{(1)} = \hat{p} E_{\hat{p}} \left(\frac{\hat{p} + \hat{p}'}{2\hat{p}} \mid I \right)$$

where $E_w(x \mid I)$ denotes the least squares projection with weights w of the vector x onto the cone I of nondecreasing vectors.

PROOF.   Since

$$\sup_{p \gg p'} L^2(p \mid n) \leq \sup_{p \gg q} L(p \mid n) L(q \mid n')$$

and the solution to the right side is given by

$$\hat{p}^{(1)} = \hat{p} E_{\hat{p}} \left(\frac{\hat{p} + \hat{p}'}{2\hat{p}} \mid I \right)$$

and

$$\hat{q}^{(1)} = \hat{p}' E_{\hat{p}'} \left(\frac{\hat{p}' + \hat{p}}{2\hat{p}'} \mid A \right)$$

where A is the cone of nonincreasing vectors (RWP, Theorem 5.4.4), the result directly follows by verifying that $\hat{q}^{(1)} = \hat{p}^{(1)'}$.

There are several algorithms available in the literature for computing $E_w(x \mid I)$. The easiest to implement is the pool adjacent violators algorithm (PAVA) as discussed in Section 1.2 of RWD.

Using (2.2) and (2.5), we can obtain the MLE's of the distribution function of X under the restrictions imposed by H_0 and H_1. It easily follows that the likelihood ratio test for testing H_0 against $H_1 - H_0$ rejects for sufficiently large values of

$$(2.6) \qquad T_1 = 2 \sum_{i=-k}^{k} n_i \{ \ln \hat{p}_i^{(1)} - \ln \hat{p}_i^{(0)} \}.$$

The asymptotic null distribution of T_1

Expanding $\ln \hat{p}_i^{(1)}$ and $\ln \hat{p}_i^{(0)}$ about $\hat{p}_i$ with a second degree remainder term and using the fact that $\sum_{i=-k}^{k} \hat{p}_i^{(1)} = \sum_{i=-k}^{k} \hat{p}_i^{(0)} = 1$, it follows that under H_0, (assuming $\hat{p}_i > 0$, $\forall_i$)

$$T_1 = \sum_{i=-k}^{k} n_i [\alpha_i^{-2} (\hat{p}_i^{(0)} - \hat{p}_i)^2 - \beta_i^{-2} (\hat{p}_i^{(1)} - \hat{p}_i)^2]$$

where α_i and β_i come from the Taylor's expansion and converge almost surely to p_i. Moreover, since $\hat{p}^{(0)} - \hat{p} = (\hat{p}' - \hat{p})/2$ and $\hat{p}^{(1)} - \hat{p} = \hat{p}E_{\hat{p}}(\frac{\hat{p}'-\hat{p}}{2\hat{p}} \mid I)$, we can write

$$T_1 = \left(\frac{1}{4}\right) \sum_{i=-k}^{k} \hat{p}_i[\alpha_i^{-2}\hat{p}_i^2\psi_i^2 - \beta_i^{-2}\hat{p}_i^2 E_{\hat{p}}(\psi \mid I)_i^2]$$

where $\psi = \sqrt{n}(\hat{p}' - \hat{p})/\hat{p}$. By the multinomial central limit theorem the random vector, $\sqrt{n}(\hat{p} - p)$ converges in law to $p(U - \bar{U}E)$ where $U_{-k}, U_{-k+1}, \ldots, U_k$ are independent normal random variables with mean zero and respective variances $p_{-k}^{-1}, \ldots, p_k^{-1}$, $\bar{U} = \sum_{i=-k}^{k} p_i U_i$, and $E = (1, 1, \ldots, 1)^T$.

Since we are assuming that H_0 is true we can write

$$\psi = \frac{\sqrt{n}[(\hat{p} - p)' - (\hat{p} - p)]}{\hat{p}}$$

$$\xrightarrow{\mathcal{L}} \frac{p'(U - \bar{U}E)' - p(U - \bar{U}E)}{p}$$

$$= (U' - U) = V.$$

Thus, by the continuity of $E_{\hat{p}}(\psi \mid I)$ in $\hat{p}$ and ψ, we have that

(2.7) $$T_1 \xrightarrow{\mathcal{L}} \left(\frac{1}{4}\right) \sum_{i=-k}^{k} p_i[V_i^2 - E_p(V \mid I)_i^2].$$

Note that $V_0 = 0$ and $V_{-i} = -V_i$. It follows from this, the assumption that $p_i = p_{-i}$ under H_0, and the maxmin formula for $E_p(V \mid I)$ in Section 1.4 of RWD that $E_p(V \mid I)_0 = 0$ and $E_p(V \mid I)_{-i} = -E_p(V \mid I)_i$.

The isotonic regression $E_p(V \mid I)$ can be computed as follows. Let V_r and p_r be the restrictions of V and p to $\{1, 2, \ldots, k\}$ and let $J = \{x = (x_1, x_2, \ldots, x_k) : 0 \leq x_1 \leq x_2 \leq \cdots \leq x_k\}$. Compute the isotonic regression, $E_{p_r}(V_r \mid J)$. Then

(2.8) $$E_p(V \mid I)_i = \begin{cases} E_{p_r}(V_r \mid J)_i, & i = 1, 2, \ldots, k \\ 0, & i = 0 \\ -E_{p_r}(V_r \mid J)_{-i}, & i = -k, \ldots, -1. \end{cases}$$

Thus

(2.9) $$T_1 \xrightarrow{\mathcal{L}} \sum_{i=1}^{k} [V_i^2 - E_{p_r}(V_r \mid J)_i^2]\left(\frac{p_i}{2}\right)$$

$$= \sum_{i=1}^{k} [V_i - E_{p_r}(V_r \mid J)_i]^2 \left(\frac{p_i}{2}\right)$$

by (1.3.7) of RWD. Note that $V_1, V_2, \ldots, V_k$ are independent and that V_i is normal with mean zero and variance $(2/p_i)$. It follows from an easy generalization of Theorem 2.3.1 in RWD that this limiting random variable has a chi-bar square

distribution. This distribution depends on the cone J and the unknown values of $p_0, p_1, p_2, \ldots, p_k$ through its level probabilities. The least favorable distribution can be found using the theory derived in Section 3.4 of RWD. These results are summarized in the following theorem.

THEOREM 2.2.   *If $\boldsymbol{p}$ satisfies H_0 and if $p_i > 0$, $i = -k, \ldots, k$, then for any real number t*

$$(2.10) \qquad \lim_{n \to \infty} \mathrm{pr}_{\boldsymbol{p}}[T_1 \geq t] = \sum_{\ell=0}^{k} p(\ell, k, \boldsymbol{p}_r) \mathrm{pr}[\chi^2_{k-\ell} \geq t]$$

where $p(0, k, \boldsymbol{p}_r)$ is the probability that $E_{\boldsymbol{p}_r}(\boldsymbol{V}_r \mid J)$ is identically zero and $p(\ell, k, \boldsymbol{p}_r)$ for $\ell = 1, 2, \ldots, k$ is the probability that $E_{\boldsymbol{p}_r}(\boldsymbol{V}_r \mid J)$ has ℓ distinct non-zero values. Furthermore,

$$(2.11) \qquad \sup_{\boldsymbol{p}} \lim_{n \to \infty} \mathrm{pr}_{\boldsymbol{p}}[T_1 \geq t] = \frac{1}{2} \mathrm{pr}[\chi^2_{k-1} \geq t] + \frac{1}{2} \mathrm{pr}[\chi^2_k \geq t].$$

A test based upon the least favorable distribution given above is likely to be conservative (depending upon the true values of $p_1, p_2, \ldots, p_k$). There is considerable evidence that if the values of $p_1, p_2, \ldots, p_k$ do not vary too much (say the ratio of the largest p_i to the smallest p_i is less than 3) then a test based upon the equal weights ($p_1 = p_2 = \cdots = p_k$) critical value will have a significance level reasonably close to the reported value. These equal weights level probabilities are discussed in Section 3.3 of RWD and are tabled in A.12 of RWD. (When using these tables, the value of k should be increased by 1 to account for 0 as a lower bound. It is like having $k + 1$ normal variables indexed by $0, 1, 2, \ldots, k$ with the weight associated with the variable indexed by 0 being ∞.)

If we have the additional information that $p_1 \geq p_2 \geq \cdots \geq p_k$ then the least favorable distribution is given by

$$(2.12) \qquad \sum_{\ell=0}^{k} \binom{k}{\ell} \left(\frac{1}{2}\right)^k \mathrm{pr}[\chi^2_\ell \geq t]$$

(see Lee *et al.* (1993)). A critical value chosen from this distribution would result in a much less conservative test than if we choose our critical value from the least favorable distribution (2.11).

Another alternative is to approximate $\mathrm{pr}_{\boldsymbol{p}}[T_1 \geq t]$ by $\sum_{\ell=0}^{k} p(\ell, k, \hat{\boldsymbol{p}}^{(0)}) \cdot \mathrm{pr}[\chi^2_{k-\ell} \geq t]$ where $\hat{\boldsymbol{p}}^{(0)}$ is as given by (2.2). This expression has the same asymptotic distribution as T_1 under H_0 and provides a good approximation. For $k \leq 4$, the level probabilities in this approximation can be computed using the formulas in Section 2.4 of RWD, again letting the weight associated with the variable indexed by 0 go to infinity. For $k > 4$ no closed form expressions for these level probabilities exist. One could approximate them using Monte-Carlo methods. Another approach would be to use a pattern approximation such as the one developed in

Section 3.4 of RWD. This would use approximate level probabilities which are obtained by interpolating between the equal weights level probabilities mentioned above and level probabilities obtained by letting the weights associated with the large p_i go to infinity. These limiting level probabilities would have the generating function

$$P_{00}(s) = \binom{s+I-1}{I} \binom{\frac{1}{2}s+B-1}{B}$$

where I is the number of large weights and B is the number of small weights having indices larger than that of the largest index corresponding to a large weight. The interpolation uses the $\frac{1}{3}$ power of the ratio of the average of the small weights to the average of the large weights. Pillers $et\ al.$ (1984) provide a Fortran program for computing these level probabilities.

3. Testing H_0 against H_2

Recall that the likelihood function is proportional to

$$L(\boldsymbol{p} \mid \boldsymbol{n}) = p_0^{n_0} \prod_{i=1}^{k} p_i^{n_i} p_{-i}^{n_{-i}}.$$

To find the MLE's of $\boldsymbol{p}$ under H_0 and H_2, we reparametrize as follows. Let

(3.1) $\theta_i = p_i/(p_{-i} + p_i)$ and $\varphi_i = (p_i + p_{-i})$, $i = 1, 2, \ldots, k$

so that

(3.2) $p_i = \theta_i \varphi_i$ and $p_{-i} = \varphi_i(1 - \theta_i)$, $i = 1, 2, \ldots, k.$

The likelihood function in terms of the new parameters is proportional to

(3.3) $L_0(\boldsymbol{\theta}, \boldsymbol{\varphi} \mid \boldsymbol{n}) = \left[\prod_{i=1}^{k} \theta_i^{n_i}(1 - \theta_i)^{n_{-i}} \right] \left[\prod_{i=1}^{k} \varphi_i^{n_i + n_{-i}} \left(1 - \sum_{i=1}^{k} \varphi_i \right)^{n_0} \right].$

The MLE's under H_0 are

(3.4) $\hat{\theta}_i^{(0)} = \dfrac{1}{2}$ and $\hat{\varphi}_i^{(0)} = \dfrac{n_i + n_{-i}}{n}$, $i = 1, 2, \ldots, k.$

Under the H_2 alternative, $\theta_i \geq \frac{1}{2}$, $i = 1, 2, \ldots, k$ and there is no restriction on the values of the φ_i's. Thus, the MLE of $\boldsymbol{\varphi}$ is the same as the estimate under H_0. The MLE of $\boldsymbol{\theta}$ under H_2 is given by

(3.5) $\hat{\theta}_i^{(2)} = \left(\dfrac{n_i}{n_i + n_{-i}} \right) \vee \dfrac{1}{2}$

where $a \vee b$ $(a \wedge b)$ denotes the maximum (minimum) of a and b.

Using (3.4) and (3.5), the MLE of p under H_2 is given in the following theorem.

THEOREM 3.1.   *The MLE of p subject to H_2 is given by*

$$(3.6) \qquad \hat{p}_i^{(2)} = \begin{cases} \left(\dfrac{n_i + n_{-i}}{n}\right)\left(\dfrac{n_i}{n_i + n_{-i}} \vee \dfrac{1}{2}\right), & i = 1, 2, \ldots, k \\[2ex] \left(\dfrac{n_i + n_{-i}}{n}\right)\left(\dfrac{n_i}{n_i + n_{-i}} \wedge \dfrac{1}{2}\right), & i = -k, -(k-1), \ldots, -1 \\[2ex] \dfrac{n_0}{n}, & i = 0 \end{cases}$$

The asymptotic null distribution of T_2

The negative of twice the log of the likelihood ratio statistic for testing H_0 against $H_2 - H_0$ is given by

$$T_2 = 2 \sum_{i=1}^{k} \left[n_i \left\{ \ln \hat{\theta}_i^{(2)} - \ln\left(\frac{1}{2}\right) \right\} + n_{-i} \left\{ \ln(1 - \hat{\theta}_i^{(2)}) - \ln\left(\frac{1}{2}\right) \right\} \right].$$

By following the techniques used in Section 2, it can be shown that

$$(3.7) \qquad T_2 \xrightarrow{\mathcal{L}} \sum_{i=1}^{k} [Z_i \vee 0]^2$$

where $Z_1, Z_2, \ldots, Z_k$ are independent standard normal variables. By Theorem 5.3.1 of RWD the following result follows.

THEOREM 3.2.   *Under H_0*

$$(3.8) \qquad \lim_{n \to \infty} \mathrm{pr}[T_2 \geq t] = \sum_{\ell=0}^{k} \binom{k}{\ell} \left(\frac{1}{2}\right)^k \mathrm{pr}[\chi_\ell^2 \geq t]$$

for all real t ($\chi_0^2 \equiv 0$).

A nice feature of this test is that its null, asymptotic distribution is free of p.

4.   Consistency of the MLE under type I bias

Strong consistency of the maximum likelihood estimators obtained under the constraints of type I bias (as described in Section 1) will guarantee that the test developed in Section 2 will be a consistent test. Actually, a more interesting question is whether the restricted maximum likelihood estimator of the CDF will be consistent when the type I bias constraint is generalized to the class of all univariate distributions by the restriction $F(x-) + F(-x) \leq 1$, $\forall x$. We take our definition of the maximum likelihood estimator to be the generalized version of

Kiefer and Wolfowitz (1956) which effectively allows probability to only be placed on observation points (and hence reduces to the discrete case).

We first prove a stronger theorem of independent interest; namely that the maximum likelihood estimates of the CDF's under the constraints of stochastic order and independent samples are strongly consistent when the underlying distributions satisfy the stochastic ordering constraint. This is not at all obvious since Rojo and Sameniego (1991) have shown that maximum likelihood estimates under the more stringent condition of uniform stochastic ordering need not even be weakly consistent. Moreover, maximum likelihood estimates under the imposed constraints of type II bias also need not be consistent in the class of all univariate distributions. Our proof is a refinement of one given by Brunk et al. (1966). Note that strong consistency actually holds without the condition of independent samples for these estimators.

To set notation, we assume that $X_1, \ldots, X_m$ is a random sample from a distribution with CDF F, that $Y_1, \ldots, Y_n$ is a second random sample from a distribution with CDF G and that $F(x) \leq G(x)$ for all x. We let $\hat{F}_m$ indicate the usual empirical CDF from the first sample, and let $\hat{F}_{m,n}$ denote the (generalized) MLE for the first population CDF under the constraints of the given stochastic ordering and independent samples. In addition, we let $S = \{s_1, \ldots, s_k\}$ denote the distinct values for the F sample and let $m_i(n_i)$ denote the number of $F(G)$ observations at s_i. The theorem we shall prove is stated as follows:

THEOREM 4.1.   The constrained MLE $\hat{F}_{m,n}$ converges uniformly to F a.s. as $m \to \infty$ if $F(x) \leq G(x)$ for all x.

PROOF.   Since the generalized MLE of F only puts probability on values in the set S (and possibly at observation points from the second sample that are larger than all elements of S), we can use a variant of Theorem 5.4.4 of RWD to state

$$\hat{F}_{m,n}(t) - \hat{F}_m(t) = \sum_{i; s_i \leq t} \frac{m_i}{m} \left\{ E_m \left[\left(\frac{m}{m+n} \right) \frac{m+n}{m} \mid I \right]_i - 1 \right\}$$

for m sufficiently large and $0 < F(t) < 1$. It then follows that $\hat{F}_{m,n}(t) \leq \hat{F}_m(t)$ by the monotonicity of the least squares projection. Thus for m sufficiently large,

$$\hat{F}_{m,n}(t) - \hat{F}_m(t) = \sum_{i; s_i \leq t} \frac{m_i}{m} \left(\frac{m}{m+n} \right) E_m \left[\frac{m+n}{m} - \frac{m+n}{m} \mid I \right]_i$$

$$= \frac{n}{n+m} \sum_{i; s_i \leq t} m_i E_m \left[\frac{n/n - m/m}{m} \mid I \right]_i$$

$$= \frac{n}{n+m} \left\{ \sum_{j=1}^{r-1} \sum_{a_{j-1} < s_i \leq a_j} m_i \right.$$

$$\left. \cdot \left[\frac{(\hat{G}_n(a_j) - \hat{G}_n(a_{j-1})) - (\hat{F}_m(a_j) - \hat{F}_m(a_{j-1}))}{\sum_{a_{j-1} < s_i \leq a_j} m_i} \right] \right.$$

$$+ \left[\frac{\sum_{a_{r-1} < s_i \le t} m_i}{\sum_{a_{r-1} < s_i \le a_r} m_i} \right]$$

$$\left. \cdot \left[(\hat{G}_n(a_r) - \hat{G}_n(a_{r-1})) - (\hat{F}_m(a_r) - \hat{F}_m(a_{r-1})) \right] \right\}$$

where $a_1 < a_2 < \cdots < a_r$ are upper endpoints ($a_0 = -\infty$) of level sets (see RWD, Section 1.4) and $a_{r-1} < t \le a_r$.

Labeling the two factors of the last term as R and Q, we can write

$$\hat{F}_{m,n}(t) - \hat{F}_m(t) = \frac{n}{n+m} [\hat{G}_n(a_{r-1}) - \hat{F}_m(a_{r-1}) + RQ], \qquad (0 \le R \le 1)$$

$$\ge \frac{n}{n+m} \min[\hat{G}_n(a_{r-1}) - \hat{F}_m(a_{r-1}), \hat{G}_n(a_r) - \hat{F}_m(a_r)]$$

(if $Q < 0$, replace R by 1, if $Q > 0$, replace R by 0). If m and $n \to \infty$, then $\hat{G}_n \to G$ and $\hat{F}_m \to F$ uniformly (a.s.) and hence the limit of the right side is nonnegative (a.s.). The result now easily follows by noting that $\hat{F}_m(t) \to F(t)$ (a.s.). Note that if $m \to \infty$ but n remains finite, we still have strong consistency of $\hat{F}_{m,n}$.

For the type I bias situation, we have only to take as our G sample the negatives of the values from the F random sample, and then compute the MLE $\hat{F}_{m,n}$ as in the two sample problem.

Type II bias extends naturally to arbitrary distributions and is equivalent to $\mathrm{pr}(0 < X < x) - \mathrm{pr}(-x < X < 0)$ being nondecreasing in x for $x > 0$. A generalized MLE for the CDF can be easily found for this restriction, but unfortunately this estimate need not be consistent except in the case of a discrete distribution. The problem is that an observation at a negative x must be "shared" with $-x$ to preserve monotonicity while an observation at a positive x need not be. This inconsistency is rather surprising in that the constraints for the two types of bias are really quite similar being $F(t-) + F(-t) \le 1$ for $t \ge 0$ for type I and $F(t-) + F(-t)$ nondecreasing in t for $t \ge 0$ for type II.

5. A simulation power study

In this section we perform a simulation study to compare the powers of the newly proposed tests T_1 and T_2 with two other tests. For this purpose, we focus on the shifted binomial distribution

$$p_j = \binom{2k}{j+k} p^{j+k} (1-p)^{k-j}, \qquad j = 0, \pm 1, \ldots, \pm k,$$

with $k = 3$ so that there are a total of 7 cells. This distribution is symmetric when $p = .5$ and it satisfies the alternatives H_1 and H_2 when $p > .5$. In our study, the sample size is fixed at $n = 100$ and the replication size is 10,000. We take $\alpha = .05$ and let $X_1, \ldots, X_n$ denote the binomial random variables.

For the shifted binomial distribution, the UMP test is to reject H_0 in favor of H_2 (and H_1) if $\sum_1^n X_i$ is too large. We use this UMP test as a benchmark.
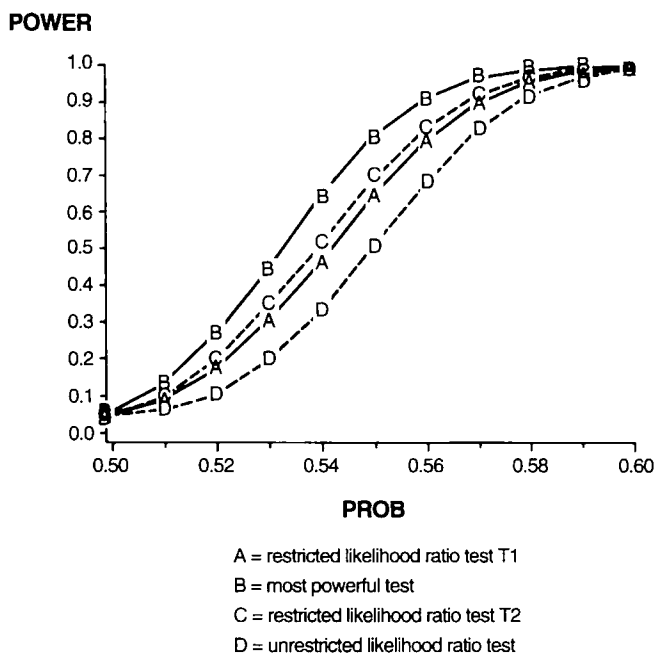
POWER

PROB

A = restricted likelihood ratio test T1
B = most powerful test
C = restricted likelihood ratio test T2
D = unrestricted likelihood ratio test

Fig. 1. Power curve for $k = 3$, $n = 100$, ALPHA $= 0.05$.

For comparison purposes we also consider the *unrestricted* likelihood ratio test for testing symmetry vs. non-symmetry (which is essentially equivalent to the usual chi square goodness of fit test for symmetry). Let us denote these two tests by T_3 and T_4, respectively. Using the asymptotic 5% critical points, we find that under the null hypothesis the p-values of these tests are .0374, .0597, .0486 and .0610, respectively. Thus it appears that the T_1 test is slightly conservative, while the T_2 test is slightly liberal, asymptotically. To put matters on an even footing, we used simulated critical values for power comparisons. The power functions of the four tests are shown in Fig. 1. It is clear from this figure that the newly proposed tests are much better than the unrestricted likelihood ratio test and that they perform favorably with respect to the UMP test. The T_2 test performs better than the T_1 test as expected since the H_2 alternatives are more restrictive than the H_1 alternatives. We expect similar behavior for other alternatives.

Acknowledgements

REFERENCES

Barlow, R. E. and Brunk, H. D. (1972). The isotonic regression problem and its dual, *J. Amer. Statist. Assoc.*, **67**, 140–147.

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

Chaffin, W. W. and Rhiel, G. S. (1993). The effect of skewness and kurtosis on the one-sample *t* test and the impact of the knowledge of the population standard deviation, *J. Statist. Comput. Simulation*, **46**, 79–90.

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters, *Ann. Math. Statist.*, **27**, 887–906.

Lee, C., Robertson, T. and Wright, T. (1993). Bounds for distributions arising in order restricted inference with restricted weights, *Biometrika*, **80**, 405–416.

Pillers, C., Robertson, T. and Wright, F. (1984). A FORTRAN program for the level probabilities of order restricted inference, *J. Roy. Statist. Soc. Ser. C*, **33**, 115–119.

Robertson, T., Wright, T. and Dykstra, R. (1988). *Order Restricted Statistical Inference*, Wiley, New York.

Rojo, J. and Samaniego, F. J. (1991). On nonparametric maximum likelihood estimation of a distribution function uniformly stochastically smaller than a standard, *Statist. Probab. Lett.*, **11**, 267–271.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*, Wiley, New York.

Yanagimoto, T. and Sibuya, M. (1972). Test of symmetry of a one-dimensional distribution against positive biasedness, *Ann. Inst. Statist. Math.*, **24**, 423–434.