The Annals of Statistics
1982, Vol. 10, No. 1. 302–306

BOUNDS ON MIXTURES OF DISTRIBUTIONS ARISING IN ORDER RESTRICTED INFERENCE[1]

BY TIM ROBERTSON AND F. T. WRIGHT

University of Iowa and University of Missouri-Rolla

In testing hypotheses involving order restrictions on a collection of parameters, distributions arise which are mixtures of standard distributions. Since tractable expressions for the mixing proportions generally do not exist even for parameter collections of moderate size, the implementation of these tests may be difficult. Stochastic upper and lower bounds are obtained for such test statistics in a variety of these kinds of problems. These bounds are also shown to be tight. The tightness results point out some situations in which the bounds could be used to obtain approximate methods. These results can also be applied to obtain the least favorable configuration when testing the equality of two multinomial populations versus a stochastic ordering alternative.

1. Introduction.

Mixtures of standard distributions play an intrinsic role in the distribution theory of tests for order restricted hypotheses; cf. Barlow et al. (1972), particularly Chapter 3. For example, Bartholomew (1959) studied a likelihood ratio statistic, T_{01}, for testing the homogeneity null hypothesis, H_0, for a collection of k normal means when the alternative is restricted by the trend $H_1: \mu_1 \leq \mu_2 \leq \cdots \leq \mu_k$. Under H_0,

$$P(T_{01} \geq c) = \sum_{\ell=1}^{k} P(\ell, k) P(\chi^2_{\ell-1} \geq c), \tag{1}$$

where χ^2_ν denotes a standard Chi squared variable with ν degrees of freedom ($\chi^2_0 \equiv 0$) and $P(\ell, k)$ denotes the probability, under H_0, that there are exactly ℓ distinct values (levels) when the sample means are smoothed to obtain the maximum likelihood estimates satisfying H_1.

The $P(\ell, k)$'s depend on the sample sizes, $n_1, n_2, \cdots, n_k$, and on the population variances, $\sigma_1^2, \sigma_2^2, \cdots, \sigma_k^2$, through the weights $w_i = n_i/\sigma_i^2$, $i = 1, 2, \cdots, k$. Their values are discussed in detail in Barlow et al. (1972), who give explicit formulas for the $P(\ell, k)$'s for $k \leq 4$ and a recursion formula for arbitrary k; cf. their (3.23). However, computation via the recursion formula may be virtually impossible since $P(j, j)$ must first be computed, and for $j \geq 5$ no closed form expression exists for $P(j, j)$; one can use the table in Abrahamson (1964) to obtain $P(5, 5)$. For the case of equal weights, the $P(\ell, k)$'s can be found recursively and are given in the Appendix A.5 for $k \leq 12$.

It has been suggested that the $P(\ell, k)$'s are fairly robust to the weights (Siskind, 1976; Grove, 1980) and that the values for equal weights give reasonable approximations except in unusual cases. Siskind (1976) felt that the pattern of weights was important and that one of the unusual cases was a U-shaped configuration of weights, that is, a set of weights in which the weights indexed by values near one and k are relatively large. This supposition seems to be substantiated by the sharpness analysis of the bounds for (1) which are presented below in Section 2.

Siskind (1976) obtained approximate critical values for a test based upon T_{01} with

Received July 1980; revised July 1981.

[1] This research was partially sponsored by the Office of Naval Research under Contracts ONR N00014-80-C0322 and ONR N00014-80-C-0321.

AMS 1970 *subject classifications.* Primary, 62E15; secondary, 62G10.

Key words and phrases. Order restricted inference, tests for and against a trend, Chi-bar-squared distribution, E-bar-squared distribution, tail probability bounds, least favorable configurations.

unequal weights. However, his results only apply for $k \leq 8$. Chase (1974) studied the $P(\ell, k)$'s in the case in which w_1 is much larger than the other w_i.

In Theorem 1, we give upper and lower stochastic bounds for the Chi-bar-squared distributions of (1), and in Remarks 2 and 3 these bounds are shown to be tight by considering sequences of weights which are either U-shaped or have an inverted U-shape. These bounds are, for even moderate k, far enough apart that one would not want to use the upper bound to provide conservative tests for an arbitrary set of weights. However, the results given here do point out some cases in which an experimenter would not want to use the equal-weights $P(\ell, k)$ as an approximation and they also provide approximations for these cases. The bounds also determine least favorable configurations in hypothesis testing problems involving a stochastic ordering between two multinomial populations. These applications are mentioned briefly at the end of Section 2.

2. Tail probability bounds. The vector $\bar{\mu}_k = (\bar{\mu}_{k1}, \bar{\mu}_{k2}, \cdots, \bar{\mu}_{kk})$ is defined to be the isotonic regression of the vector $\bar{X} = (\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_k)$ of sample means with weights $w_1, w_2, \cdots, w_k$. The computation algorithms referred to in this section are described in detail in Barlow et al. (1972). A basic idea in this section is that since $P(\chi^2_\nu \geq c) \leq P(\chi^2_{\nu+1} \geq c)$ for $\nu = 0, 1, 2, \cdots$ and for all c, one obtains an upper (lower) bound for (1) by making the distribution of the number of level sets in $\bar{\mu}_k$ as large (small) as possible.

THEOREM 1. *Assume that* $\mu_1 = \mu_2 = \cdots = \mu_k$. *If* $a_1 \leq a_2 \leq \cdots \leq a_k$, *then for* $k = 2, 3, \cdots$

$$(a_1 + a_2)/2 \leq \sum_{\ell=1}^k P(\ell, k) a_\ell \leq \sum_{\ell=1}^k \binom{k-1}{\ell-1} 2^{-k+1} a_\ell. \tag{2}$$

If $a_1 \geq a_2 \geq \cdots \geq a_k$, then the inequalities in (2) are reversed.

PROOF. The second conclusion follows easily from the first. For the first conclusion, observe that if $\sum_{\ell=1}^j b_\ell \leq \sum_{\ell=1}^j b_\ell^*$ for $j = 1, 2, \cdots, k$, with equality for $j = k$, then, using Abel's method of summation, we see that $\sum_{\ell=1}^k a_\ell b_\ell \geq \sum_{\ell=1}^k a_\ell b_\ell^*$. Thus, it suffices to show that $P(1, k) \leq 1/2$ and

$$\sum_{\ell=1}^j \binom{k-1}{\ell-1} 2^{-k+1} \leq \sum_{\ell=1}^j P(\ell, k); \qquad j = 1, 2, \cdots, k-1. \tag{3}$$

For the first inequality, observe that

$$P(1, k) = P[\min_{1 \leq \alpha \leq k} \text{Av}(\{1, \cdots, \alpha\}) = \text{Av}(\{1, \cdots, k\})]$$
$$\leq P(\text{Av}(\{1, \cdots, k-1\}) \geq \text{Av}(\{1, \cdots, k\})] = 1/2$$

where $\text{Av}(B) = \sum_{j \in B} w_j \bar{X}_j / \sum_{j \in B} w_j$ for $\phi \neq B \subset \{1, \cdots, k\}$.

In order to establish (3) we use induction on k and the "pool adjacent violators" algorithm (PAVA). For the case $k = 2$, it is easy to see that $P(1, 2) = P(2, 2) = 1/2$ independent of the values of w_1 and w_2. Assume (3) holds for k and let $L_k(\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_k; w_1, w_2, \cdots, w_k)$ denote the number of level sets in the isotonic regression of $\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_k$ with weights $w_1, w_2, \cdots, w_k$. Now, by the PAVA, the isotonic regression, $\bar{\mu}_{k+1}$, may be formed by first constructing $\bar{\mu}_k$ and then combining $\bar{\mu}_k$ with $\bar{X}_{k+1}$ in the appropriate way. Thus, using the obvious notational abuses, either $L_{k+1} = L_k + 1$ or $L_{k+1} \leq L_k$ and the former case is characterized by $\bar{\mu}_{kk} < \bar{X}_{k+1}$. Thus, $L_{k+1} \leq L_k + I_{[\bar{X}_{k+1} > \bar{\mu}_{kk}]}$ and for $j = 1, 2, \cdots, k$,

$$P(L_{k+1} \geq j+1) \leq P(L_k = j \text{ and } \bar{X}_{k+1} > \bar{\mu}_{kk}) + P(L_k \geq j+1).$$

Using the notation in Barlow et al. (1972) and the proof of their (3.23), we can express $P(L_k = j \text{ and } \bar{X}_{k+1} > \bar{\mu}_{kk})$ as a sum of the form

$$\sum_{\mathscr{S}_{j,k}} P\{\text{Av}(B_1) < \text{Av}(B_2) < \cdots < \text{Av}(B_j) < \bar{X}_{k+1}\} \prod_{i=1}^j P(1, C_{B_i}; w(B_i)).$$

Applying their inequality (3.20) to the first factor in each term yields the following upper bound:

$$\sum_{\mathscr{L}_{j,k}} P\{\mathrm{Av}(B_j) < \bar{X}_{k+1}\} P(j, j;\ W_{B_1}, W_{B_2}, \cdots, W_{B_j}) \prod_{i=1}^{j} P(1, C_{B_i}; w(B_i)).$$

The first factor in each term is ½ and using (3.23) again yields $P(L_k = j \text{ and } \bar{X}_{k+1} > \bar{\mu}_{kk}) \le P(j, k)/2$. Thus, applying the induction hypothesis, we obtain

$$P(L_{k+1} \ge j+1) \le \{P(j, k) + 2P(L_k \ge j+1)\}/2$$

$$= \{P(L_k \ge j) + P(L_k \ge j+1)\}/2 \le 2^{-k} \sum_{\ell=j+1}^{k+1} \binom{k}{\ell-1},$$

which is the desired result.

The lower bound in (2) was obtained by Perlman (1969) for a related problem.

The bounds given in (2) cannot, in general, be improved. Specifically, there exist sequences $w_n = (w_{n1}, w_{n2}, \cdots, w_{nk})$ and $w_n' = (w_{n1}', w_{n2}', \cdots, w_{nk}')$ with

$$\lim_{n\to\infty} P(\ell, k; w_n) = \binom{k-1}{\ell-1} 2^{-k+1} \tag{4}$$

and

$$\lim_{n\to\infty} P(\ell, k; w_n') = 2^{-1} I_{[\ell=1,2]} \tag{5}$$

for $\ell = 1, 2, \cdots, k$. We first consider the shifted binomial distribution in (4). If we let

$$w(\varepsilon) = \begin{cases} (\varepsilon^H, \varepsilon^{H-1}, \cdots, \varepsilon, 1, \varepsilon, \cdots, \varepsilon^H) & \text{for } k \text{ odd}, \\ (\varepsilon^{H-1}, \varepsilon^{H-2}, \cdots, \varepsilon, 1, 1, \varepsilon, \cdots, \varepsilon^{H-1}) & \text{for } k \text{ even}, \end{cases}$$

where $H = [k/2]$, then (4) can be obtained by letting $\varepsilon \to 0$. The proof, however, is complicated and we offer a less involved proof by induction.

REMARK 2. Let $\mu_1 = \mu_2 = \cdots = \mu_k$. For each $k \ge 2$ and for each $\varepsilon > 0$ there exist positive weights $w_1, w_2, \cdots, w_k$ such that

$$\left| P(\ell, k; w_1, w_2, \cdots, w_k) - \binom{k-1}{\ell-1} (1/2)^{k-1} \right| < \varepsilon, \qquad \ell = 1, 2, \cdots, k.$$

PROOF. The result is obvious for $k = 2$. Assume that $k \ge 3$ and that $\varepsilon > 0$. By the induction hypothesis, there exist weights $w_1, w_2, \cdots, w_{k-1}$ such that

$$\left| P(\ell, k-1) - \binom{k-2}{\ell-1} (1/2)^{k-2} \right| < \varepsilon; \qquad \ell = 1, 2, \cdots, k-1.$$

Let $w' = (w_1, w_2, \cdots, w_{k-1})$ and let $w = (w_1, w_2, \cdots, w_k)$ where w_k is to be specified. For computing the $P(\ell, k)$ we may assume that $\bar{X}_i = Z_i/\sqrt{w_i}$ with $Z_1, Z_2, \cdots, Z_k$ independent standard normal variables defined on some probability space. With w' already chosen, we consider fixed ω in the probability space. If $Z_k(\omega) > 0$, then for w_k sufficiently small $\bar{X}_k > \bar{\mu}_{k-1,k-1}$ and so by the PAVA, $L_k = L_{k-1} + 1$ for such ω and w_k. (L_k was defined in the proof of Theorem 1.) If $Z_k(\omega) < 0$, then for w_k sufficiently small $\bar{X}_k < \bar{\mu}_{k-1,k-1}$ and so in obtaining $\bar{\mu}_k$ by using the PAVA, $\bar{X}_k$ and $\bar{\mu}_{k-1,k-1}$ need to be pooled. However, if $\{j, \cdots, k-1\}$ denotes the last level set in $\bar{\mu}_{k-1}$, then $\mathrm{Av}(\{j, \cdots, k\}) \to \mathrm{Av}(\{j, \cdots, k-1\})$ as $w_k \to 0$. So for such ω and w_k sufficiently small $L_k = L_{k-1}$. Hence, with probability one, $L_k \to L_{k-1} + I_{[Z_k>0]}$ as $w_k \to 0$. Thus,

$$\lim_{w_k\to 0} P(\ell, k; w) = \{P(\ell-1, k-1; w') + P(\ell, k-1; w')\}/2.$$

Using our induction hypothesis, the desired result is obtained.

REMARK 3. Let $\mu_1 = \mu_2 = \cdots = \mu_k$. If $w_n' \to (a, 0, 0, \cdots, b)$ with a and b positive, then $P(\ell, k; w_n') \to 2^{-1} I_{[\ell=1,2]}$ for $\ell = 1, 2, \cdots, k$.

PROOF. As in the proof of Remark 2, we assume that $\bar{X}_i = Z_i/\sqrt{w'_{ni}}$ for $i = 1, 2, \cdots, k$ with Z_i fixed. Clearly, the random vector $(\sqrt{w'_{ni}}\,Z_1, \sqrt{w'_{n2}}\,Z_2, \cdots, \sqrt{w'_{nk}}\,Z_k, w'_{n1}, \cdots, w'_{nk})$ converges weakly to $(\sqrt{a}\,Z_1, 0, 0, \cdots, 0, \sqrt{b}\,Z_k, a, 0, \cdots, 0, b)$ and so $\text{Av}(\{1, 2, \cdots, k\}) - \min_{1\le i\le k-1} \text{Av}(\{1, 2, \cdots, i\})$ converges weakly to $b\{(Z_k/\sqrt{b}) - (Z_1/\sqrt{a})\}/(a+b)$. Hence,

$$P(1, k; w'_n) = P[\text{Av}(\{1, 2, \cdots, k\}) \le \min_{1\le i\le k-1} \text{Av}(\{1, 2, \cdots, i\})] \to 1/2.$$

Next consider $P(\ell, k)$ for $\ell \ge 3$. Write $P(\ell, k)$ using (3.23) in Barlow et al. (1972) and bound the general term by

$$P\{\text{Av}(B_1) < \text{Av}(B_2) < \cdots < \text{Av}(B_\ell)\}$$
$$\le P\{\text{Av}(B_1) < \text{Av}(B_2 \cup B_3 \cup \cdots \cup B_{\ell-1}) < \text{Av}(B_\ell)\}.$$

(This inequality follows from the Cauchy Mean Value Property of averages.) Now this bound can be written as $P(\sigma_1 Z_1 < \sigma_2 Z_2 < \sigma_3 Z_3) = P\{(\sigma_1/\sigma_2)Z_1 < Z_2 < (\sigma_3/\sigma_2)Z_3\}$ where $\sigma_1/\sigma_2 \to 0$ and $\sigma_3/\sigma_2 \to 0$. It follows that $\lim_{n\to\infty} P(\ell, k; w'_n) = 0$ for $\ell \ge 3$.

If one uses T_{01} to test H_0 vs. H_1 with the α level equal weights critical value, the true significance level may differ from α considerably. For instance, with $k = 9$ and $\alpha = .05$ the true significance level may be as large as .1941 (determined by the upper bound) and as small as .0059 (determined by the lower bound). If the weight set has a pronounced U shape (inverted U shape) it is clear that the approximation obtained by using the lower (upper) bound given here would be better than the equal weights approximation. Siskind's comment that his approximation is not adequate for U-shaped weights is also interesting in this light. (In his comparison, weights with an inverted U shape were not considered.)

For unknown variances, but of the form $a_i\sigma^2$ with the a_i known, the likelihood ratio test is given in Barlow et al. (1972). The results given here provide upper and lower bounds with the same weight-configuration producing the bounds.

Robertson and Wegman (1978) considered testing H_1 against all alternatives. The least favorable, null hypothesis distribution for the likelihood ratio statistic is given by $\sum_{\ell=1}^{k} P(\ell, k)P(\chi^2_{k-\ell} \ge c)$, provided the variances are known. The second conclusion in Theorem 1 provides upper and lower bounds for this distribution. These bounds are sharper than those for T_{01} and as k increases the rate of increase of the difference between the upper bound and the equal weights probability decreases dramatically. The same authors obtained an $\bar{E}^2$ test for the case of unknown variances and similar remarks can be made about the test.

Robertson (1978) discusses tests of trend in a multinomial population and Theorem 1 and the sharpness results apply in that setting too.

Suppose we have independent samples from two multinomial populations with parameters $p = (p_1, \cdots, p_k)$ and $q = (q_1, \cdots, q_k)$. If one wishes to test $H_0: p = q$ vs. $H_1 - H_0$ with $H_1: \sum_{j=1}^{i} p_j \ge \sum_{j=1}^{i} q_j$; $i = 1, \cdots, k-1$ or H_i against all alternatives, then the null hypothesis is not simple. These problems are discussed in Robertson and Wright (1981) and the results given here provide least favorable configurations within their null hypotheses.

REFERENCES

ABRAHAMSON, I. G. (1964). Orthant probabilities for the quadrivariate normal distribution. *Ann. Math. Statist.* **35** 1685–1703.

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inferences Under Order Restrictions.* Wiley, New York.

BARTHOLOMEW, D. J. (1959). A test for homogeneity for ordered alternatives I and II. *Biometrika* **46** 36–48, 328–335.

CHASE, G. R. (1974). On testing for ordered alternatives with increased sample size for a control. *Biometrika* **61** 569–578.

GROVE, D. M. (1980). A test of independence against a class of ordered alternatives in a 2 × *C* contingency table. *J. Amer. Statist. Assoc.* **75** 454–459.

PERLMAN, M. D. (1969) One-sided problems in multivariate analysis. *Ann. Math. Statist.* **40** 549–567.

ROBERTSON, TIM and WEGMAN, EDWARD J. (1978). Likelihood ratio tests for order restrictions in exponential families. *Ann. Statist.* **6** 485–505.

ROBERTSON, TIM (1978). Testing for and against an order restriction on multinomial parameters. *J. Amer. Statis. Assoc.* **73** 197–202.

ROBERTSON, TIM and WRIGHT, F. T. (1981). Likelihood ratio tests for and against a stochastic ordering between multinomial populations. *Ann. Statist.* **9** 1248–1257.

SISKIND, V. (1976). Approximate probability integrals and critical values for Bartholomew's test for ordered means. *Biometrika* **63** 647–654.

DEPARTMENT OF STATISTICS
UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242

UNIVERSITY OF MISSOURI
DEPARTMENT OF MATHEMATICS
326 MATHEMATICS-COMPUTER SCIENCES BUILDING
ROLLA, MO 65401