BAYES TWO-DECISION PROCEDURES FOR DISTRIBUTION-FREE TWO-SAMPLE PROBLEMS AND RANK-ORDER DATA

By J. C. Baskerville
University of Western Ontario

and

D. L. Solomon
Cornell University

BU-464-M Rev.

SUMMARY

A two-sample problem for rank-order data is formulated as a two-decision problem. Using the general Bayes solution, Bayes procedures are derived for several configurations of the set of states of nature for which the problem is distribution-free. It is shown that for given prior distributions these procedures reduce to certain classical LMP rank tests. Some devices for selection of prior distributions are suggested. The admissibility of the proposed procedures is considered, and it is shown that their Bayes risk tends to zero as sample sizes increase.

Some key words: Bayes procedures; Two-sample problem; Locally most powerful rank tests; Lehmann alternatives.

1. INTRODUCTION

The problem of incorporating prior information into a statistical decision procedure when no parametric family of distributions is specified or when the only data are rank orders has received little attention (see Ferguson, 1973 and Saxena, 1965). In this article such procedures are derived for the two-sample problem and their relationship to classical rank tests considered.

The classical rank tests for the two-sample problem of testing the hypothesis $G(x) \geq F(x)$ (or $G(x) = F(x)$) against the alternative $G(x) < F(x)$ have been shown to be optimal (locally most powerful; LMP) only for certain parametric families within the above alternative (see Hajek and Šidák, 1967). It therefore seems desirable to search for other admissible procedures which are Bayes with respect to a prior distribution on a parameter that measures departure from the hypothesis in a class of alternatives that render the problem distribution-free. Such classes of alternatives are those introduced by Lehmann (1953). Membership of a pair (F,G) in such classes has meaningful interpretations in terms of the indexing parameter, and these can be used in the selection of prior distributions.

2. FORMULATION AND BAYES SOLUTION

2.1. Formulation of the two-decision problem

The two-sample problem for rank-order data can be formulated as a two-decision problem as in Saxena (1965). Suppose we have independent samples $X = (X_1, X_2, \cdots, X_m)$ and $Y = (Y_1, Y_2, \cdots, Y_n)$ from continuous CDF's F and G respectively. Let $S = (S_1, S_2, \cdots, S_n)$ be the vector of ordered ranks of the Y's in the combined sample of $m + n = N$ observations, and $s = (s_1, s_2, \cdots, s_n)$

a realization of S. Only S is assumed to be observable and we denote the sample space by

$$\mathcal{S} = \{(s_1, s_2, \cdots, s_n) | s_1 < s_2 < \cdots < s_n, \quad s_i \text{ an integer in } [1,N]\} \quad .$$

The set of states of nature, Ω, consists of two disjoint sets:

$$\Omega_0 = \{(F,G) | G = \psi_\delta(F); \quad \delta \in A_0\}, \quad \Omega_1 = \{(F,G) | G = \psi_\delta(F); \; \delta \in A_1\}$$

where ψ is an absolutely continuous nondecreasing function from the interval [0,1] onto [0,1] and A_0 and A_1 are disjoint indexing sets of real numbers.

The decision space is $D = \{d_0, d_1\}$, where d_0 corresponds to the decision $(F,G) \in \Omega_0$ (or $\delta \in A_0$) and d_1 to $(F,G) \in \Omega_1$ (or $\delta \in A_1$). Thus taking d_0 corresponds to accepting the classical "null" hypothesis and d_1 to rejecting that hypothesis.

The following loss function is assumed throughout:

$$L(d_i, \delta) = \begin{cases} a(>0) & \text{if } i = 1, \quad \delta \in A_0 \\ b(>0) & \text{if } i = 0, \quad \delta \in A_1 \\ 0 & \text{otherwise} \end{cases} \quad .$$

2.2. Bayes solution

Suppose we consider the parameter δ, which indexes Ω, as the realization of a random variable Δ with CDF W. Let $\mathcal{T}$ be the class of all nonrandomized decision rules $t: \mathcal{S} \to D$. If we denote the probability of observing $S = s$ when $\Delta = \delta$ by $P\{s|\delta\}$, then the risk incurred by using procedure t when $\Delta = \delta$ is given by $r(t,\delta) = EL\big(t(S),\delta\big)$. The Bayes risk is $R(t,W) = Er(t,\Delta)$, and the Bayes solution of the decision problem defined in section 2.1 is a procedure t_0 such that $R(t_0,W) = \inf_{t \in \mathcal{T}} R(t,W)$.

For the two-decision problem, a simple way of determining the Bayes procedure is to calculate the posterior risk associated with each decision. For a given rank order s, the Bayes procedure takes d_1 if its posterior risk is less than that of d_0, i.e., if

$$B(s) = \int_{A_1} P\{s|\delta\}dW(\delta) \Big/ \int_{A_0} P\{s|\delta\}dW(\delta) > a/b . \tag{1}$$

B(s) will be referred to as the Bayes statistic.

2.3. Probabilities of rank orders

The approach used here will be to derive Bayes statistics explicitly by the application of a result of Hoeffding (1950) which states that whenever densities f from F and g from G are such that $f(x) = 0$ implies $g(x) = 0$, $P\{s|\delta\} = \binom{N}{n}^{-1} E \prod_{i=1}^{n} \{g(V(s_i))/f(V(s_i))\}$, where $V(1), V(2), \cdots, V(N)$ are the order statistics of a sample of size N from a population with density f. Under the aforesaid conditions we may write $G = \psi_\delta(F)$, so that $g = f \cdot \psi'_\delta(F)$, and this followed by application of the probability-integral transformation yields

$$P\{s|\delta\} = \binom{N}{n}^{-1} E \prod_{i=1}^{n} \psi'_\delta(U(s_i)) \tag{2}$$

where $U(1), U(2), \cdots, U(N)$ are the order statistics of a sample of size N from a population with uniform distribution on [0,1].

3. BAYES PROCEDURES

When $\psi_\delta(y)$ does not depend on F the probability $P\{s|\delta\}$ is independent of the underlying distributions and the problem is distribution-free. In this section two forms of ψ are considered, each of which results in a class of alternatives with a meaningful interpretation.

3.1. Mixture of Lehmann alternatives

Let $\psi_{\delta,K}(y) = (1-\delta)y + \delta y^K$, where K is a specified integer greater than 1. Suppose $A_0 = \{0\}$ and $A_1 = \{\delta | 0 < \delta \le 1\}$. We will consider Δ as a random variable with CDF W of the mixed type. More precisely, $W(\delta) = pM(\delta) + (1-p)H(\delta)$, where $0 < p \le 1$, $M(\delta) = 0(1)$ for $\delta < 0$ $(\delta \ge 0)$, and H an absolutely continuous or discrete CDF with support in $(0,\infty)$. Let h denote the continuous probability density for H in the former case and the discrete density for H in the latter. Then the density of Δ, say w, can be written as

$$w(\delta) = \begin{cases} p & \text{if} \quad \delta = 0 \\ (1-p)h(\delta) & \text{if} \quad \delta > 0 \end{cases} .$$

From expression (1) of section 2.2, the Bayes procedure takes d_1 if

$$\int_0^\infty P\{s|\delta\}h(\delta)d\delta > ap/b(1-p) \binom{N}{n} . \tag{3}$$

It should be noted that $G = \psi_{\delta,K}(F)$ implies that Y_i comes from F $(Y_i \sim X_j)$ with probability $(1-\delta)$ or Y_i comes from F^K $(Y_i \sim \max(X_1, X_2, \cdots, X_K))$ with probability δ, where the mixing parameter δ will be considered as the realization of a random variable Δ.

Methods of estimating K from rank orders are discussed in Baskerville and Solomon (1973). For this discussion we assume that K is known.

Applying a general result of David and Johnson (1954) for calculating mixed moments of uniform order statistics to the expression resulting from (2) when $(F,G) \in \Omega$ yields

$$\begin{aligned} P\{s|\delta,K\} = \binom{N}{n}^{-1} \Big\{ &(1-\delta)^n + \frac{K(1-\delta)^{n-1}\delta N!}{(N+K-1)!} \sum_{i=1}^{n} \frac{(s_i+K-2)!}{(s_i-1)!} \\ &+ \frac{K^2(1-\delta)^{n-2}\delta^2 N!}{(N+2K-2)!} \sum_{i<j} \frac{(s_i+K-2)!(s_j+2K-3)!}{(s_i-1)!(s_j+K-2)!} + \cdots \\ &+ \frac{K^n\delta^n N!}{(N+nK-n)!} \prod_{i=1}^{n} \frac{(s_i+i(K-1)-1)!}{(s_i+(i-1)(K-1)-1)!} \Big\} \end{aligned} \tag{4}$$

Suppose that h is a beta density, $Be(\nu_1,\nu_2)$. Then

$$\int_0^1 P\{s|\delta,K\}h(\delta)d\delta = \frac{\binom{N}{n}^{-1}\Gamma(\nu_1+\nu_2)N!}{\Gamma(\nu_1)\Gamma(\nu_2)\Gamma(\nu_1+\nu_2+n)}\left\{\frac{\Gamma(\nu_1)\Gamma(\nu_2+n)}{N!}\right.$$

$$+\frac{\delta\Gamma(\nu_1+1)\Gamma(\nu_2+n-1)}{(N+K-1)!}\sum_{i=1}^{n}\frac{(s_i+K-2)!}{(s_i-1)!}+\frac{K^2\Gamma(\nu_1+2)\Gamma(\nu_2+n-2)}{(N+2K-2)!}\sum_{i<j}\frac{(s_i+K-2)!(s_j+2K-3)!}{(s_i-1)!(s_j+K-2)!}$$

$$\left.+\cdots+\frac{\delta^n\Gamma(\nu_1+n)\Gamma(\nu_2)}{(N+nK-n)!}\prod_{i=1}^{n}\frac{(s_i+i(K-1)-1)!}{(s_i+(i-1)(K-1)-1)!}\right\}$$

Multiplying the above expression by $\binom{N}{n}$ and denoting the resulting quantity by $B_M(s,K)$, the Bayes procedure takes d_1 if $B_M(s|K) > ap/b(1-p)$.

<u>Example 1</u>. If $K = 2$ and $n = 2$,

$$B_M(s|2) = \frac{1}{(\nu_1+\nu_2+1)(\nu_1+\nu_2)}\left[(\nu_2+1)\nu_2 + \frac{2\nu_1\nu_2}{N+1}(s_1+s_2) + \frac{4(\nu_1+1)\nu_1 s_1(s_2+1)}{(N+2)(N+1)}\right].$$

Table 1 gives values of $B_M(s|2)$ for $m = 3$ and selected prior distributions $Be(\nu_1,\nu_2)$.

Table 1. Values of $B_M(s|2)$ for $n = 2$, $m = 3$ and various $Be(\nu_1,\nu_2)$ priors

(s_1,s_2)	$\sum_{i=1}^{2} s_i$	Be(1,20)	Be(1,1)	Be(3,3)	Be($\frac{1}{4}$,$\frac{1}{4}$)	Be(20,1)
(4,5)	9	1.049	1.595	1.582	1.619	2.212
(3,5)	8	1.032	1.349	1.347	1.353	1.678
(3,4)	7	1.016	1.198	1.194	1.206	1.404
(2,5)	7	1.015	1.103	1.112	1.087	1.144
(2,4)	6	.999	.984	.986	.980	.957
(2,3)	5	.984	.865	.861	.873	.769
(1,5)	6	.998	.857	.878	.821	.610
(1,4)	5	.983	.770	.779	.754	.509
(1,3)	4	.968	.683	.680	.686	.408
(1,2)	3	.954	.595	.582	.619	.307

Prior densities that weight small values of δ heavily, favor the hypothesis $G = F$ while those that put high probability on values of δ near 1 favor $G = F^2$. This is reflected in table 1 by the fact that rank orders least supportive of $G = F^2$, such as (1,2) and (1,3), have larger values of $B_M(s|2)$ for Be(1,20) than for Be(20,1) .

For $a = b$ and $p = .4$ rank orders corresponding to values of $B_M(s|2)$ above the line in table 1 lead to d_1. This shows that since the prior Be(20,1) favors $G = F^2$, the occurrence of rank orders supportive of $G = F$ leads to d_0 while the Bayes procedure with respect to Be(1,20) always takes d_1.

It is of interest to note that as $\delta \to 0$, $\psi_{\delta,K}(F) \to F$ so the LMP rank test of $G = F$ is the one which is uniformly most powerful for δ near zero. From (4) it can be seen that for small values of δ, $P\{s|\delta,K\}$ is large for those rank orders s which lead to large values of $\sum_{i=1}^{n}(s_i+K-2)!/(s_i-1)!$. Therefore, the LMP rank test takes d_1 for large values of this sum. For $K = 2$ this is the Wilcoxon rank sum. A comparison of the rank sums with the values of $B_M(s|2)$ for a Be(1,20) prior in table 1 shows that both statistics order the ranks in the same way. Thus for $a = b$ and an appropriate choice of p, to completely specify the prior, the rank sum test at level α is seen to be equivalent to this Bayes procedure.

Example 2. For a $\alpha = .2$ the Wilcoxon test rejects if $\sum_{i=1}^{2} s_i \geq 8$. An equivalent Bayes procedure takes d_1 if $B_M(s|2) \geq 1.032$. Taking $a = b$, the Bayes procedure with respect to a Be(1,20) prior with $p = .508$ is seen to be equivalent.

The generalization to $\psi(y) = \sum_{i=1}^{k} \delta_i y^{a_i}$ with a Dirichlet prior for $(\Delta_2,\cdots,\Delta_K)$ is straightforward.

3.2. Lehmann alternatives

In this section tests of alternatives of the form $G = F^{\delta}$, as introduced by Lehmann (1953) are discussed. Here $\psi_{\delta}(y) = y^{\delta}$ and we take $A_0 = \{\delta | 0 < \delta \le 1\}$ and $A_1 = \{\delta | \delta > 1\}$. Lehmann shows that for $(F,G) \in \Omega$, equation (2) yields $P\{s|\delta\} = \binom{N}{n}^{-1} \delta^n \prod_{i=1}^{n} \Gamma(s_i+i\delta-i)\Gamma(s_{i+1})/\Gamma(s_i)\Gamma(s_{i+1}+i\delta-i)$ where we put $s_{n+1} = N + 1$. This is of course equivalent to the last term in equation (4).

It will be convenient to adopt the following notation used by Savage (1956). Let $Z_i = 1$ if $S_j = i$ for some $j=1,2,\cdots,n$ and $Z_i = 0$ otherwise, $i=1,2,\cdots,N$. Put $Z = (Z_1,\cdots,Z_N)$ and let $z = (z_1,\cdots,z_N)$ be a realization of Z. If we also take $v_i = \sum_{j=1}^{i} z_j$ and $u_i = i - v_i$, then Savage shows that $P\{z|\delta\} = m!n!\delta^n / \prod_{i=1}^{N}(u_i+\delta v_i)$.

If Δ is a continuous random variable with values in $A_0 \cup A_1$, and $(F,G) \in \Omega$, the fact that $P\{X < Y|\delta\} = \delta/(\delta+1)$ suggests the re-parameterization $\gamma = \delta/(\delta+1)$. Then $A_0 = \{\gamma | 0 < \gamma \le \frac{1}{2}\}$, $A_1 = \{\gamma | \frac{1}{2} < \gamma < 1\}$, and the Bayes procedure takes d_1 if

$$\int_{\frac{1}{2}}^{1} \frac{\gamma^n w(\gamma)d\gamma}{(1-\gamma)^n \prod_{i=1}^{N} [u_i + \frac{\gamma}{1-\gamma} v_i]} \Bigg/ \int_{0}^{\frac{1}{2}} \frac{\gamma^n w(\gamma)d\gamma}{(1-\gamma)^n \prod_{i=1}^{N} [u_i + \frac{\gamma}{1-\gamma} v_i]} > a/b \,. \qquad (5)$$

Example. Suppose γ is considered a realization of a random variable Γ, and that Γ has the Beta distribution, $Be(\nu_1,\nu_2)$ symmetric about $\gamma = \frac{1}{2}$ (i.e., $\nu_1 = \nu_2 = \nu$). Then the left side of (5) becomes

$$\int_{0}^{1} \frac{\gamma^{\nu+n-1}(1-\gamma)^{\nu-n-1}d\gamma}{\prod_{i=1}^{N} [u_i + \frac{\gamma}{1-\gamma} v_i]} \Bigg/ \int_{0}^{\frac{1}{2}} \frac{\gamma^{\nu-n-1}(1-\gamma)^{\nu+n-1}}{\prod_{i=1}^{N} [u_i + \frac{1-\gamma}{\gamma} v_i]} \,.$$

For specified ν, n, and z these integrals can be evaluated by the method of partial fractions.

If we consider Δ as discrete it will be more meaningful to take $A_0 = \{\delta = 1\}$ and $A_1 = \{\delta | \delta > 1\}$. Then w will be discrete with support 1, 2, 3, $\cdots$. The resulting Bayes procedure takes d_1 if

$$B_L(z) = \sum_{\delta=2}^{\infty} \frac{\delta^n w(\delta)}{\prod_{i=1}^{N} (u_i+\delta v_i)} > ap/bN! \quad (6)$$

If w has finite support the computation of $B_L(z)$ presents no great problems. In such cases the assessment of subjective probabilities may be aided by the following device. Since $G = F^\delta$ implies $P\{X < Y|\delta\} = \delta/(\delta+1)$, the prior distribution for Δ might be constructed from the experimenter's feelings about $P\{X < Y|\delta\}$ for various values of δ.

If previous samples from F and G are available, the following method might be used for selection of prior probabilities. Suppose r pairs of samples of sizes m_i and n_i, $i=1,2,\cdots,r$, are available from F and G. It is desired to estimate what proportion, $w(\delta)$, of the samples of Y's come from F^δ. This suggests that we estimate δ for each of the r samples and take $w(\delta) = \#(\hat{\delta} \approx \delta)/r$. Methods of estimating δ from such data are considered in Baskerville and Solomon (1973).

Table 2 gives values of $B_L(z)$ for $m = 2$, $n = 3$, and various prior probability vectors $\big(w(1), w(2), \cdots, w(K)\big)$. Values of the Savage statistic (see Savage, 1956) are included for the sake of comparison.

Table 2. Values of the Savage statistic $\Sigma v_i/i$ and the Bayes statistic $B_L(z)$ for various prior probability vectors $\big(w(1),\cdots,w(K)\big)$ for m = 2 and n = 3.

z	$\sum_{i=1}^{5}\frac{v_i}{i}$	(.5,.5)	(.5,0,.5)	(.5,0,0,.5)	(.5,0,0,0,.5)	(.5,.3,.1,.1)	(.4,.3,.2,.1)
11100	4.350	.00149	.00076	.00046	.00031	.00114	.00129
11010	4.017	.00179	.00097	.00061	.00042	.00139	.00158
11001	3.767	.00208	.00122	.00079	.00056	.00165	.00190
10110	3.717	.00238	.00146	.00098	.00070	.00192	.00221
10101	3.267	.00278	.00183	.00127	.00093	.00229	.00265
10011	2.933	.00347	.00256	.00191	.00146	.00298	.00349
01110	2.517	.00476	.00438	.00391	.00348	.00452	.00539
01101	2.267	.00556	.00548	.00508	.00464	.00544	.00654
01011	1.933	.00694	.00767	.00762	.00729	.00722	.00876
00111	1.433	.01041	.01534	.01905	.02188	.01313	.01620

The rank orders corresponding to values of $B_L(z)$ below the line are those that would lead to d_1 for the given prior and a = b = 1. It is interesting to note that although rank orders most supportive of $(F,G) \in \Omega_1$ lead to larger values of $B_L(z)$, as prior probability shifts to larger δ (for p = .5) the set of rank orders leading to d_1 decreases in size. This phenomenon was also noted in table 1. The following observation provides an explanation for this and a check on the tabled values.

Consider the joint distribution of (Z,Δ) and let ℓ denote the marginal density of Z. Then

$$\ell(z) = \sum_{\delta} P\{z|\delta\}P\{\Delta = \delta\} = \sum_{\delta=1}^{K} \frac{m!n!w(\delta)}{\prod_{i=1}^{N}(u_i+\delta v_i)} = m!n![p/N!+B_L(z)] \quad . \qquad (7)$$

Therefore, $B_L(z) = \ell(z)/m!n! - p/N!$, so that the $B_L(z)$'s are adjusted probabilities.

It is therefore reasonable that a rank order such as (0,1,0,1,1), although supportive of $(F,G) \in \Omega_1$, may have higher likelihood of occurrence out of the joint distribution of Z with (.5,0,.5) than with (.5,0,0,0,.5). As a constant mass of prior probability is moved further from $\delta = 1$ a conflicting dichotomy of prior belief is expressed. Rank orders (e.g., 01110) that were initially supportive of d_1 become supportive of d_0 as the dichotomy becomes more pronounced. That is rank orders that were "extreme" enough to support d_1 are no longer "extreme" under a prior placing mass further from the null hypothesis.

From (7) it follows that $\sum_z \ell(z) = p + m!n! \sum_z B_L(z) = 1$ so that $\sum_z B_L(z) = (1-p)/m!n!$, providing a check on the tabled values of $B_L(z)$ through the column sums.

Savage (1956) shows that for δ sufficiently close to 1, the statistic $\sum_{i=1}^{N} v_i/i$ orders the rank orders inversely as $P\{z|\delta\}$. Thus the LMP procedure takes d_1 for $\sum_{i=1}^{N} v_i/i < C_\alpha$. It should be noted that for many of the priors in table 2, $B_L(z)$ gives precisely the opposite ordering to the rank orders as does the Savage statistic. Thus each is equivalent to the LMP test at some α level.

5. ADMISSIBILITY

In this section two general results (see Ferguson, 1967) are adapted to establish the admissibility of the Bayes procedures of section 3. The proofs of these results are routine and are omitted (see Baskerville and Solomon, 1973).

THEOREM 1. Suppose w has support $A = A_0 \cup A_1$ where $A_0 = \{\delta | -\infty < \delta \le \delta_0\}$ and $A_1 = \{\delta | \delta_0 < \delta < \infty\}$. If $P\{z|\delta\}$ is a continuous function of δ for each z, then a Bayes rule t_0, with respect to w, with finite risk $R(t_0,w)$, is admissible.

Theorem 1 shows that any procedure of the form (1) of section 2.2 with A as prescribed is admissible (e.g., example of section 3.2). In the case of a simple hypothesis where $A_0 = \{\delta_0\}$ and $w(\delta_0) = p > 0$, the fact that the discontinuity of $r(t,\delta)$ at δ_0 is simple leads to a proof of the admissibility of t_0 as in theorem 1. Therefore, any procedure of form (3) in section 3.1 where the support of h is an interval, A_1, is admissible.

THEOREM 2. <u>The Bayes two-decision procedure for the problem</u> $(\mathscr{A}, \Omega, D, L)$ <u>is admissible if there is at most one rank order</u> z <u>such that</u>

$$\int_{A_1} P\{z|\delta\}dW(\delta) \Big/ \int_{A_0} P\{z|\delta\}dW(\delta) = a/b \qquad (8)$$

It should be noted that strict monotonicity of the ratio in (8) for some ordering of the z's implies the condition of theorem 2. A partial ordering used by Saxena (1965) has been shown by Saxena and Savage (1969) to produce a monotone rank order likelihood ratio $P\{z|\delta\}/P\{z'|\delta\}$ for certain types of alternatives.

More precisely, define a relation R^* by zR^*z', if $z_i = z_i'$ for $i=1,2,\cdots,N$ except for some j and j+1 where $z_j = z_{j+1}' = 0$ and $z_{j+1} = z_j' = 1$. Then define zRz' if ZR^*z' or if there exist rank orders $z^{(1)}, z^{(2)}, \cdots, z^{(K)}$ such that $zR^*z^{(1)}R^*z^{(2)}R^* \ldots R^*z^{(K)}R^*z'$ Savage and Saxena show that in the case of Lehmann alternatives the likelihood ratio is a strictly increasing function of δ for zRz'. Furthermore, it is shown that $P\{z|\delta\} > P\{z'|\delta\}$ when $\delta > 1$. Therefore, for rank orders that are R-related, if zRz' then $\int P\{z|\delta\}dH(\delta) > \int P\{z'|\delta\}dH(\delta)$.

This fact, along with theorem 2, provides a useful tool for investigation of the admissibility of Bayes procedures such as (6). If the procedure takes d_1 for rank order z' then it will take d_1 for all z such that zRz'. Similarly, if d_0 is taken for z', d_0 will be taken for those rank orders z such that z'Rz. Such a Bayes two-decision procedure is said to be monotone.

6. BEHAVIOR OF THE BAYES RISK FOR LARGE SAMPLES

The Bayes risk of the procedure t with respect to the prior CDF W is given by

$$R(t,W) = a \sum_z Q_t(z) \int_{A_0} P\{z|\delta\} dW(\delta) + b \sum_z [1-Q_t(z)] \int_{A_1} P\{z|\delta\} dW(\delta) ,$$

where $Q_t(z)$ is the indicator function of $\{z|t(z) = d_1\}$. The following theorems demonstrate the behavior of the Bayes risk of the procedures discussed in section 3. Theorem 3 is a generalization of a result of Saxena (1965) and theorem 4 treats the important case where $A_0 = \{\delta_0\}$.

THEOREM 3. <u>Let</u> $A_0 = \{\delta|-\infty < \delta \le \delta_0\}$ <u>and</u> $A_1 = \{\delta|\delta_0 < \delta < \infty\}$. <u>Suppose</u> Δ <u>has an absolutely continuous</u> CDF W, $P(\delta) = P\{X < Y|\delta\}$ <u>is a strictly increasing function of</u> δ, <u>and</u> t^* <u>is Bayes with respect to</u> W. <u>Then</u> $\lim_{m,n\to\infty} R(t^*,W) = 0$.

<u>Proof</u>. Let U(z) be the Mann-Whitney statistic, i.e., the number of pairs (X_i,Y_j) from a combined sample of size N with $X_i < Y_j$, and t_0 the procedure that takes d_1 if $z \in \{z|U(z)/mn > P(\delta_0)\}$. Since t^* is Bayes with respect to W, $R(t^*,W) \le R(t_0,W)\ \forall\ m,n$. It is therefore sufficient to show that $\lim_{m,n\to\infty} R(t_0,W) = 0$, where

$$R(t_0,W) = a \int_{A_0} P\{U(z)/mn > P(\delta_0)|\delta\} dW(\delta) + b \int_{A_1} P\{U(z)/mn \le P(\delta_0)|\delta\} dW(\delta) .$$

It is known (see Hajek and Šidák, 1967) that U(z)/mn converges in probability to $P(\delta)$. Thus, since $P(\delta)$ is strictly increasing in δ, $P(\delta) \le P(\delta_0)$ for $\delta \in A_0$ and $\lim_{m,n\to\infty} P\{U(z)/mn > P(\delta_0)|\delta\} = 0,\ \forall\ \delta < \delta_0$. Similarly, $\lim_{m,n\to\infty} P\{U(z)/mn \le P(\delta_0)|\delta\} = 0,\ \forall\ \delta \in A_1$.

Define $(m_1,n_1) < (m_2,n_2)$ if $m_1 \le m_2$, $n_1 \le n_2$ and $m_1 + n_1 < m_2 + n_2$. Then $P\{U(z)/mn > P(\delta_0)|\delta\}$ and $P\{U(z)/mn \le P(\delta_0)|\delta\}$ are sequences of measurable functions in δ bounded by 1, where the measure is that induced on the reals by W. Thus applying the Lebesgue Dominated Convergence Theorem and noting that $W(\delta)$ is absolutely continuous we have $\lim_{m,n\to\infty} R(t_0,W) = 0$.

This theorem shows that Bayes procedures for Lehmann alternatives where $A_0 = \{\gamma|0 < \gamma \le \frac{1}{2}\}$, $A_1 = \{\gamma|\frac{1}{2} < \gamma < 1\}$ have Bayes risk approaching zero as m and n increase. The next theorem shows that this is also the case when $A_0 = \{\delta_0\}$.

THEOREM 4. <u>Let</u> $A_0 = \{\delta_0\}$, <u>and</u> Δ <u>have</u> <u>prior</u> CDF W <u>given</u> <u>by</u> $W(\delta) = pM(\delta) + (1-p)H(\delta)$ <u>where</u> $M(\delta) = 0(1)$ <u>for</u> $\delta < \delta_0$ $(\delta \ge \delta_0)$ <u>and</u> H <u>an</u> <u>absolutely</u> <u>continuous</u> <u>or</u> <u>discrete</u> CDF <u>with</u> <u>support</u> <u>in</u> $A_1 = \{\delta|\delta_0 < \delta < \infty\}$. <u>Then</u> <u>if</u> $P(\delta)$ <u>is</u> <u>a</u> <u>strictly</u> <u>increasing</u> <u>function</u> <u>of</u> δ <u>and</u> t^* <u>is</u> <u>Bayes</u> <u>with</u> <u>respect</u> <u>to</u> W, $\lim_{m,n\to\infty} R(t^*,W) = 0$.

<u>Proof</u>. Using the same approach as in the proof of theorem 3, let $\epsilon > 0$ be arbitrary and t_0 be the procedure that takes d_1 if $z \in \{z|U(z)/mn > P(\delta_0+\epsilon)\}$. Then

$$R(t_0,W) = apP\{U(z)/mn > P(\delta_0+\epsilon)|\Delta = \delta_0\} + b(1-p)\int_{A_1} P\{U(z)/mn \le P(\delta_0+\epsilon)|\delta\}dH(\delta) . \quad (9)$$

Now for $\delta = \delta_0$, $U(z)/mn$ converges in probability to $P(\delta_0)$, but $P(\delta_0+\epsilon) > P(\delta_0)$, $\forall\ \epsilon > 0$ so $\lim_{m,n\to\infty} P\{U(z)/mn > P(\delta_0+\epsilon)|\Delta = \delta_0\} = 0$.

Suppose H is discrete with $\delta_1 > \delta_0$ the first point at which H has a jump discontinuity. Choosing ϵ^* such that $P(w_0+\epsilon^*) < P(\delta_1)$, we have

$$\lim_{m,n\to\infty} P\{U(z)/mn \le P(\delta_0+\epsilon^*)|\delta\} = 0; \qquad \delta \ge \delta_1,\ \epsilon \le \epsilon^* .$$

Therefore, when H is discrete $\lim_{m,n\to\infty} R(t^*,W) = 0$.

If H is continuous we can write the integral in (9) as

$$\int_{\delta_0}^{\delta_0+\epsilon} P\{U(z)/mn \le P(\delta_0+\epsilon)|\delta\}dH(\delta) + \int_{\delta_0+\epsilon}^{\infty} P\{U(z)/mn \le P(\delta_0+\epsilon)|\delta\}dH(\delta) .$$

Taking the limit and interchanging limit and integral yields $P_H\{\delta_0 < \Delta < \delta_0 + \epsilon\}$. Since ϵ is arbitrary and H is continuous we have $\lim_{m,n\to\infty} R(t^*,W) = 0$.

REFERENCES

BASKERVILLE, J. & SOLOMON, D. L. (1973). Bayes two-decision procedures for distribution-free two-sample problems and rank-order data. Paper Number BU-464-M in the Mimeo Series, Biometrics Unit, Cornell University.

BIRNBAUM, Z. W. & McCARTY, R. C. (1958). A distribution-free upper confidence bound for $Pr\{Y < X\}$, based on independent samples of X and Y. Ann. Math. Statist. 29, 558-62.

DAVID, F. N. & JOHNSON, N. L. (1954). Statistical treatment of censored data. I. Fundamental formulae. Biometrika 41, 228-40.

FERGUSON, T. S. (1967). Mathematical Statistics. New York: Academic Press.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. Ann. of Statict. 1, 209-30.

HAJEK, J. & ŠIDÁK, Z. (1967). Theory of Rank Tests. New York: Academic Press.

HOEFFDING, W. (1950). 'Optimum' non-parametric tests. Proc. Second Berkeley Symposium, 83-92.

LEHMANN, E. L. (1953). The power of rank tests. Ann. Math. Statist. 24, 23-43.

SAVAGE, I. R. (1956). Contributions to the theory of order statistics: the two-sample case. Ann. Math. Statist. 27, 590-615.

SAXENA, K. M. L. (1965). Some nonparametric Bayesian estimation problems. F. S. U. Statistics Report M100, Florida State University, Tallahassee, Florida.

SAXENA, K. M. L. & SAVAGE, I. R. (1969). Monotonicity of rank order likelihood ratio. Ann. of the Inst. of Statist. Math. 21, 265-75.