Bayes Two Decision Procedures for Rank Order Data and Lehmann Alternatives

by

Jon Baskerville

BU-400-M February, 1972

Abstract

The two sample problem for rank order data is put in a decision theoretic context in which the states of nature are Lehmann alternatives. Rules which are Bayes, with respect to prior distributions on the indexing parameter of the states of nature, are derived for the resulting two decision problem. Consideration is given to the admissibility and applicability of such procedures, and methods for subjective and empirical assignment of prior probabilities are discussed.

Bayes Two Decision Procedures for Rank Order Data and Lehmann Alternatives

by

Jon Baskerville

BU-400-M February, 1972

1. Introduction and summary. The classical rank tests for the two sample problem have been shown to be optimal (locally most powerful) against very restricted classes of alternatives (see [1]). In this paper the two sample problem is formulated as a two decision problem, as in [5], and Bayes procedures are derived for a larger class of alternatives. The class of alternatives considered is that introduced by Lehmann [3]. Its members have a meaningful interpretation, and the probability of observing a given rank order from a pair of distributions in this class depends only on the relationship of the distributions producing this rank order, not on their parametric form.

In section 2 the general problem and procedure are outlined. Sections 3 and 4 specialize these results for two types of prior distributions and examples are given. In section 5 a sufficient condition for the admissibility of the general procedure is presented. Section 6 discusses possibilities for application including subjective and empirical methods for determining prior distributions. In section 7 the Bayes procedure for some simple priors is compared with the Wilcoxon statistic, for a complete enumeration of rank orders with $m = 2,3$, and $n = 3$.

2. The problem and procedure. Given independent random samples $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ from cumulative distribution functions F and G respectively, let $S = \left(S_1, \ldots, S_n\right)$ be the vector of ordered ranks of the Y's in the combined sample

of n+m = N observations, and $s = (s_1,\ldots,s_n)$ a realization of S. Let $\Omega_0 = \{(F,G)|F(x) = G(x),\ \forall x\}$, $\Omega_1 = \{(F,G)|G(x) = F^{\delta}(x),\ \forall x,\ \text{some } \delta > 1\}$, and $\Omega = \Omega_0 \cup \Omega_1$. The object is then to choose between decisions $d_0 : (F,G) \in \Omega_0$ and $d_1 : (F,G) \in \Omega_1$.

It was shown by Lehmann [3] that

$$(2.1)\quad P\{S = s|(F,G) \in \Omega\} = \binom{N}{n}^{-1} \delta^n \prod_{j=1}^{n} \frac{\Gamma(S_j+j\delta-j)\Gamma(S_{j+1})}{\Gamma(S_j)\Gamma(S_{j+1}+j\delta-j)}$$

where we put $S_{n+1} = N+1$.

Thus for $(F,G) \in \Omega$ the probability of observing any rank order is independent of F and the set of states of nature Ω can be indexed by $\delta \geq 1$. We may then consider δ as a realization of a random variable Δ with cumulative distribution function W, of the mixed type. More precisely, $W(\delta) = pM(\delta) + (1-p)H(\delta)$, where $0 \leq p \leq 1$, $M(\delta) = 0(1)$ for $\delta < 1$ $(\delta \geq 1)$, and H is an absolutely continuous or discrete cdf with support $(1,\infty)$. We let h denote the continuous probability density of H in the former case and the discrete density of H in the latter. Thus the density of Δ, say w, can be written as

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

If we employ the loss function,

$$L(d_i,\delta) = \begin{cases} a \text{ if } i = 1,\ \delta = 1 \\ b \text{ if } i = 0,\ \delta > 1 \\ 0 \text{ otherwise} \end{cases}$$

where a and b are positive, then the conditional risks of taking decisions d_i, given $S = s$ are

$$R(d_o, s) = \frac{b(1-p)\int P\{S = s|\delta\}dH(\delta)}{P\{S = s\}}$$

and $R(d_1, s) = ap\binom{N}{n}^{-1} / P\{S = s\}$. Therefore a Bayes procedure chooses d_1 if

(2.2) $$\int P\{S = s|\delta\}dH(\delta) > ap/b(1-p)\binom{N}{n}$$

and d_o otherwise.

It will be convenient in the successive sections to adopt the following notation as used by Savage [4]. Let $Z_i = 1$ if $S_j = i$ for some $j = 1,2,\ldots,n$ and $Z_i = 0$ otherwise, $i = 1,2,\ldots,N$. Put $Z = (Z_1,\ldots,Z_N)$ and let $z = (z_1,\ldots,z_N)$ 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 [4] shows that 2.1 can be written as

(2.3) $$P\{Z = z|\delta\} = m!n!\delta^n / \prod_{i=1}^{N} (u_i + \delta v_i)$$

so that by 2.2 d_1 is taken if

(2.4) $$B(z) = \int\left[\delta^N / \prod_{i=1}^{N} (u_i + \delta v_i)\right]dH(\delta) > ap/b(1-p)N!$$

3. H absolutely continuous. The problem of expressing 2.4 in a convenient computational form when h is a density of the continuous type is rather complex. The form of the denominator suggests that the integral might be expressed as the sum

of simpler integrals by the method of partial fractions. If we note that since $v_i = 0$ for $i < s_1$, we have $u_i = i$ for $i < s_1$, so that

$$(3.1)\qquad \frac{\delta^n}{\prod_{i=1}^{N} (u_i+\delta v_i)} = \frac{1}{(s_1-1)!\prod_{i=s_1}^{N} v_i} \times \frac{\delta^n}{\prod_{i=s_1}^{N} (\delta+u_i/v_i)}$$

where $v_i > 0$ for $i \geq s_1$, and the degree of the polynomial, in δ, of the denominator is $N-s_1+1$ which is an integer in $[n,N]$. The method employed to decompose the last factor of 3.1 depends on the degree of the denominator and whether or not the denominator contains repeated factors, that is if for the given rank order u_i/v_i, $i \geq s_1$, takes one or a number of values with a frequency greater than one. These considerations suggest that it is not worthwhile to give a general form of the decomposition, but rather that it should be done for the observed rank order at hand.

Example 3.1. When the degree of the denominator in the last factor of 3.1 is greater than n, which is the case unless z consists of m zeros followed by n ones, and contains no repeated factors 3.1 becomes (see [2]);

$$(3.2)\qquad \frac{1}{(s_1-1)!\prod_{i=s_1}^{N} v_i} \sum_{i=0}^{N-s_1} \frac{A_i}{\delta+u_{s_1+i}/v_{s_1+i}} \quad \text{with } A_i = \frac{\left(-u_{s_1+i}/v_{s_1+i}\right)^n}{\prod_{j=s_1}^{N} (u_j/v_j-u_{s_1+i}/v_{s_1+i})}$$

where we do not allow $j = s_1+i$ in the product. Thus for rank orders satisfying the given conditions we have,

$$(3.3)\qquad B(z) = \frac{1}{(s_1-1)!\prod_{i=1}^{N} v_i} \sum_{i=0}^{N-s_1} A_i \int_1^{\infty} \frac{h(\delta)\,d\delta}{\delta+u_{s_1+i}/v_{s_1+i}}$$

Example 3.2. Suppose w puts probability p on $\delta = 1$ and the rest uniformly on $(1,R)$ so that $h(\delta) = 1/(R-1)$; $1 < \delta < R$. Then for z with the above properties,

$$B_w(z) = \frac{1}{(s_1-1)!(R-1)\prod_{i=s_1}^{N} v_i} \sum_{i=0}^{N-s_1} A_i\left[\log\left(u_{s_1+i}+Rv_{s_1+i}\right)-\log(s_1+i)\right]$$

4. H discrete. When h is a density of the discrete type so is w, and calculations similar to those leading to 2.2 yield a Bayes rule that chooses d_1 when

$$(4.1)\qquad B_w(Z) = \sum_{\delta=2}^{\infty} \frac{\delta^n w(\delta)}{\prod_{i=1}^{N}(u_i+\delta v_i)} > \frac{ap}{bN!}$$

When the support of w is finite, which will be the case in many situations as indicated in section 6, computation of B_w presents no great problems. If the support of w is infinite the method of partial fractions as discussed in the previous section may be useful.

5. Admissibility. In this section, we consider the question of admissibility of Bayes procedures of the form 2.2, for the decision problem specified by $(\Omega,\mathcal{D},L,n,N)$, where $\mathcal{D} = \{d_0,d_1\}$. It is well known (see [1]) that a Bayes procedure which is unique up to equivalence of risk functions is admissible. Also the existence of

a Bayes rule with respect to a certain prior distribution guarantees the existence of a non-randomized Bayes rule with respect to that prior. The non-randomized Bayes procedures considered here are unique unless $R(d_0, z) = R(d_1, z)$ for some rank order z, where $R(d_i, z)$ is the conditional risk associated with decision d_i given $Z = z$. Therefore a sufficient condition for the admissibility of the Bayes two decision procedure with respect to W is

$$\int P(Z = z|\delta)dH(\delta) \neq \frac{ap}{b(1-p)\binom{N}{n}}$$

for all $\binom{N}{n}$ rank orders.

It is possible to get a weaker condition by noting that if the procedure fails to be unique then its inadmissibility depends on the existence of a "better" Bayes rule. That is if D is a Bayes procedure mapping rank orders into $\mathcal{D}$, and D is inadmissible, and there is a procedure D^* such that the risk

$$r(D^*, \delta) \leq r(D, \delta) \ \forall\delta \text{ and } r(D^*, \delta) < r(D, \delta) \text{ for some } \delta$$

then $Er(D^*, \delta) \leq Er(D, \delta)$ and D^* must be Bayes. This fact and the structure of the problem under consideration yield the following result.

Theorem 5.1. The Bayes two decision procedure (2.2) for the problem $(\Omega, \mathcal{D}, L, n, N)$ is admissible if there is at most one rank order z such that

$$\int P(Z = z|\delta)dH(\delta) = \frac{ap}{b(1-p)\binom{N}{n}} \quad . \tag{5.1}$$

Proof. In light of the above remarks it is sufficient to show that the procedure is admissible when there is exactly one rank order, say z_0, which satisfies (5.1).

Let $Q_D(z)$ be the indicator function of $\{z \mid D(z) = d_1\}$. The risk function for the procedure D is given by

$$r(D,\delta) = \begin{cases} a\binom{N}{n}^{-1} \sum_z Q_D(z) & \text{if } \delta = 1 \\ \\ b \sum_z \left[1-Q_D(z)\right] P\{Z=z \mid \delta\} & \text{if } \delta > 1 \end{cases}$$

Since $R(d_0,z_0) = R(d_1,z_0)$ there are two Bayes rules, D_0 and D_1, such that $D_0(z_0) = d_0$ and $D_1(z_0) = d_1$ with $D_0(z) = D_1(z)$ for $z \neq z_0$. Then since $Q_{D_0}(z_0) = 0$, and $Q_{D_1}(z_0) = 1$ and $Q_{D_0}(z) = Q_{D_1}(z)$ for $z \neq z_0$ we have

$$r(D_0,1) - r(D_1,1) = a\binom{N}{n}^{-1} \sum_z \left[Q_{D_0}(z) - Q_{D_1}(z)\right] = - a\binom{N}{n}^{-1} < 0$$

and for $\delta > 1$

$$r(D_0,\delta) - r(D_1,\delta) = b \sum_z \left[Q_{D_1}(z) - Q_{D_0}(z)\right] P\{Z = z \mid \delta\}$$

$$= bP\{Z = z_0 \mid \delta\} > 0$$

Therefore both D_0 and D_1 are admissible and the procedure 2.2 which corresponds to D_0 is admissible.

It should be noted that strict monotonicity of the integral in 5.1 for some ordering of the z's implies the condition of theorem 5.1. A partial ordering used by Saxena [5] has been shown by Saxena and Savage [6] to produce a monotone rank order likelihood ratio in the case of Lehmann Alternatives. More precisely let $zR^{*}z'$ if $z_i = z'_i$ for $i = 1,2,\ldots,N$ 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 $z^{(1)},\ldots,z^{(K)}$ such that $zR^{*}z^{(1)}R^{*}z^{(2)}R^{*}\ldots R^{*}z^{(K)}R^{*}z'$. Saxena and Savage show that when zRz', $P\{Z = z|\delta\}/P\{Z = z'|\delta\}$ is an increasing function of δ and furthermore $P\{Z = z|\delta\} > P\{Z = z'|\delta\}$ when $\delta > 1$. Therefore, for rank orders that are R-related, if zRz' then

$$\int P\{Z = z|\delta\}dH(\delta) > \int P\{Z = z'|\delta\}dH(\delta) \quad .$$

This fact along with theorem 5.1 provides a useful tool for testing the admissibility of a Bayes procedure, for if $\int P\{Z = z'|\delta\}dH(\delta) > ap/b(1-p)\binom{N}{n}$ then so is $\int P\{Z = z|\delta\}dH(\delta)$ for all z such that zRz'. Similarly, if $\int P\{Z = z'|\delta\}dH(\delta)$ $< ap/b(1-p)\binom{N}{n}$ then so is $\int P\{Z = z|\delta\}dH(\delta)$ for all z such that $z'Rz$. Such a Bayes two decision procedure is said to be monotone.

6. Some Possibilities for Application. The class Ω_1 is admittedly not a natural set of alternatives for most practical problems, and has mainly been of interest because of the mathematical niceties it provides. However, the fact that the procedures mentioned above are optimal, in the Bayes sense, with respect to this rather large class, whereas the optimality of the classical rank tests is for more limited classes of alternatives, suggests that it is worthwhile to explore some possibilities for application.

It should be of interest, in considering problems where the class Ω_1 provides a reasonable set of alternatives, to note that $G = F^\delta$, δ a positive integer greater than 1, implies that the Y's are distributed as $\max(X_1,\ldots,X_\delta)$. Thus if the Y observations are a result of a process that in some sense chooses the largest of a group of X observations, Ω_1 provides a reasonable set of alternatives. Because we are considering δ as the realization of a random variable we need not specify a single δ, but rather a probability distribution for the variable. The remainder of this section deals with methods of determining such a distribution.

The assessment of subjective probabilities when Δ is discrete may be aided by the following device. If $G = F^\delta$ then $P\{X < Y|\delta\} = \delta/\delta+1$, so that a prior distribution for Δ might be constructed from the experimenters feelings about $P\{X < Y|\delta\}$ for various values of δ.

δ	$P\{X < Y\|\delta\}$	Experimenters prior feelings
1	1/2	w(1)
2	2/3	w(2)
3	3/4	w(3)
⋮	⋮	⋮

The Bayes procedure would then choose d_1 if $B_w(z) > \frac{aw(1)}{bN!}$ where a/b is the relative importance of classical type I to type II error.

If data from previous samples from F and G is available, the following method might be used for empirical selection of prior probabilities. Suppose r pairs of samples of sizes m_i and n_i; $i = 1,2,\ldots,r$ from F and G are available. It is desired to estimate what proportion, $w(\delta)$, of the samples from G come from F^δ. This suggests that we estimate δ for each of the r samples and take

$$w(\delta) = \frac{\#(\hat{\delta} \simeq \delta)}{r} \quad . \tag{6.1}$$

We next consider two estimators of δ, based on ranks, which might be used for this purpose.

The first $\hat{\delta}_{AH}$ is an Ad-Hoc estimator which is intuitively appealing and is based on an estimate of $P\{X < Y|\delta\}$ for a given sample. Let U be the Mann-Whitney statistic, namely the number of pairs (X_i, Y_j), from a combined sample of size m + n, with $X_i < Y_j$. We then take

$$\frac{U}{mn} = \frac{T - 1/2\ n(n+1)}{mn} \tag{6.2}$$

as an estimator of $P\{X < Y|\delta\}$, where T is the sum of the ranks of the Y's in the ordered pooled sample. Since under Ω, $P\{X < Y|\delta\} = \delta/\delta+1$, if we set (6.2) equal to $\delta/\delta+1$ and solve for δ we get

$$\hat{\delta}_{AH} = \frac{T - 1/2\ n(n+1)}{mn + 1/2\ n(n+1) - T} \tag{6.3}$$

as an estimator of δ. It should be noted that $\hat{\delta}_{AH}$ is an increasing function of T, and when all n observations from G exceed all m observations from F, $T = mn + 1/2\ n(n+1)$ so that $\hat{\delta}_{AH}$ is undefined. Such an observation, of course, provides the strongest possible evidence for $(F,G) \in \Omega_1$ and should be weighted accordingly.

A second estimator of δ is based on the "principle" of maximum likelihood which suggests that we determine the proportion of samples, $w(\delta)$, from G that have the greatest likelihood of having come from F^{δ}. Since the likelihood of a rank order is given by (2.3), treating δ as continuous and differentiating yields the maximum likelihood estimator $\hat{\delta}_{ML}$ as a solution of

$$(6.4) \qquad L(\hat{\delta}) = \sum_{i=1}^{n} \frac{\hat{\delta} V_i}{u_i + \hat{\delta} V_i} = n \quad .$$

Since each non-zero term of the sum in 6.4 is continuous, strictly increasing in $\hat{\delta}$ and positive for $\hat{\delta} > 0$; $L(\delta)$ is a strictly increasing continuous function of δ for $\delta > 0$. Also $\lim_{\delta \to 0^+} L(\delta) = 0$ and $\lim_{\delta \to \infty} L(\delta) = \#$ of non-zero terms in the sum, which is greater than n unless all the observations from G exceed all the observations from F. In the latter case the number of non-zero terms is n(i.e., $\lim_{\delta \to \infty} L(\delta) = n$). Therefore $\hat{\delta}_{ML}$ yields a unique estimate of δ except in this extreme case where $\hat{\delta}_{ML}$, as $\hat{\delta}_{AH}$, is undefined.

The properties of these estimators have not been investigated, but it should be noted that since Z is not a vector of IID random variables, the regular theory of maximum likelihood does not apply to $\hat{\delta}_{ML}$.

7. Some Numerical Comparisons.

In the following tables the values of the Bayes statistic $B_w(z)$ of (4.1) are given for all rank orders (z) in the cases $m = 2,3$; $n = 3$ and for various priors (w). The vectors are prior probability vectors $(w(1), w(2), \ldots, w(k))$.

m = 2, n = 3

z	rank sum(S)	(.5,.5)	(.5,.3,.2)	(.5,.0,.5)	(.5,.0,.0,.5)	(.5,.0,.0,.0,.5)
11100	6	.00149	.00120	.00076	.00046	.00031
11010	7	.00179	.00146	.00097	.00061	.00042
11001	8	.00208	.00174	.00122	.00079	.00056
10110	8	.00238	.00201	.00146	.00098	.00070
10101	9	.00278	.00240	.00183	.00127	.00093
10011	10	.00347	.00311	.00256	.00191	.00146
01110	9	.00476	.00461	.00438	.00391	.00348
01101	10	.00556	.00552	.00548	.00508	.00464
01011	11	.00694	.00723	.00767	.00762	.00729
00111	12	.01041	.01238	.01534	.01905	.02188

m = 3, n = 3

z	rank sum(S)	(.5,.5)	(.5,.3,.2)	(.5,.0,.5)	(.5,.0,.0,.5)	(.5,.0,.0,.0,.5)
111000	6	.00017	.00012	.00006	.00003	.00002
110100	7	.00020	.00015	.00008	.00004	.00002
110010	8	.00023	.00018	.00010	.00005	.00003
110001	9	.00026	.00021	.00012	.00007	.00004
101100	8	.00026	.00021	.00012	.00007	.00004
101010	9	.00031	.00025	.00015	.00008	.00005
101001	10	.00035	.00029	.00019	.00011	.00007
100110	10	.00038	.00031	.00021	.00013	.00008
100101	11	.00044	.00037	.00026	.00016	.00011
100011	12	.00053	.00046	.00035	.00023	.00016
011100	9	.00053	.00046	.00037	.00026	.00019
011010	10	.00062	.00055	.00046	.00034	.00026
011001	11	.00071	.00065	.00056	.00043	.00034
010110	11	.00077	.00072	.00064	.00051	.00041
010101	12	.00088	.00084	.00078	.00065	.00053
010011	13	.00105	.00105	.00104	.00092	.00079
001110	12	.00116	.00121	.00128	.00127	.00122
001101	13	.00132	.00142	.00156	.00162	.00159
001011	14	.00159	.00179	.00208	.00231	.00238
000111	15	.00211	.00266	.00347	.00462	.00556

The rank orders corresponding to the values of $B_w(z)$ below the lines in the table 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 $(F,G) \epsilon \Omega_1$ have larger values of $B_w(z)$ as prior probability shifts to larger δ, the set of rank orders leading to d_1 decreases in size.

At the level of precision given in the table it appears that $B_w(z)$ is increasing for the given ordering of z's. However, in examining more precision some switchovers are apparent.

References

[1] Ferguson, T. S. (1967). Mathematical Statistics. Academic Press, New York.

[2] Hodgman, C. D., Editor (1963). Chemical Rubber Company Standard Mathematical Tables. Twelfth Edition. Chemical

[3] Lehmann, E. L. (1953). "The Power of Rank Tests", Ann. Math. Stat. 24, pp. 24-42.

[4] Savage, I. R. (1956). "Contributions to the Theory of Order Statistics: The Two Sample Case", Ann. Math. Stat. 27, pp. 590-615.

[5] Saxena, K. M. L. (1965). Some Nonparametric Bayesian Estimation Problems. F.S.U. Statistics Report M100, Florida State University, Tallahassee, Florida.

[6] Saxena, K. M. L. and Savage, I. R. (1969). "Monotonicity of Rank Order Likelihood Ratio", Ann. of the Inst. of Stat. Math. 21, pp. 265-275.