BIOMETRICS 28, 859-867
September 1972

BAYESIAN COMPARISON OF TWO ORDERED MULTINOMIAL POPULATIONS

HERBERT I. WEISBERG

New York University, Graduate School of Business Administration, 100 Trinity Place, New York, N. Y.

SUMMARY

A Bayesian approach to the problem of comparing two multinomial distributions with k categories and a natural ordering of the categories is developed. Assuming independent Dirichlet priors, we obtain a convenient algorithm for computing the posterior probability that one distribution is stochastically larger than the other. The important special case $k = 2$ reduces to computing the posterior probability that one binomial proportion exceeds another. We consider the more general problem of finding the posterior distribution of the ratio of two such proportions.

1. INTRODUCTION

Suppose we have a multinomial population with k categories and a natural ordering on the categories. For example, medical patients being treated by a particular method may be rated in terms of their response as being much improved, somewhat improved, the same, somewhat worse, or much worse. Often there will be a natural numerical scale associated with the categories. When this is not the case, the categories can be coded in some convenient fashion.

If we wish to determine whether there is a difference between two such populations (for example patients treated by two different methods), the usual classical approaches are use of the chi-square test and use of a rank test with allowance for ties. In this paper we follow a Bayesian approach to this problem along lines suggested in Altham [1969] and Novick and Grizzle [1965]. The main contribution of this paper is not the approach, but the development of methods by which it can be conveniently implemented, so that it may be of use to the practicing statistician.

Let p_{ij} be the probability associated with category j for population i ($i = 1, 2$: $j = 1, \cdots k$). Let $F_1 < F_2$ read 'population 1 is stochastically smaller than population 2'. Regardless of what coding of the categories is used (consistent with their natural ordering), this is equivalent to $\sum_{j=1}^{m} p_{1j} \geq \sum_{j=1}^{m} p_{2j}$, $m = 1, 2, \cdots, k - 1$. Following the approach of Novick and Grizzle we assume independent conjugate Dirichlet prior distributions and obtain the posterior probability of $\{F_1 < F_2\}$ based on samples from the two populations.

When $k = 2$, the method reduces to finding the probability that one binomial proportion is smaller than another. We consider this case in detail and generalize to obtain the distribution functions of the ratio of two proportions.

The user of our method is of course free to choose that member of the Dirichlet family which most accurately reflects his prior information. In proposing it as a general alternative to classical methods, however, we have in mind a situation where little or no prior information is available. That is, we think of the prior as an indifference distribution, reflecting as little prior knowledge as possible. Novick and Grizzle [1965] suggest the flat $D(1, 1, \cdots, 1; 1)$ prior in this case (see equation (2.1)). For further discussion of this matter the reader is referred to their paper and accompanying references.

2. STOCHASTIC COMPARISON

In this section we derive an algorithm which can be used to obtain $P\,(F_1 < F_2)$. Following the notation of Wilks [1962] we define the $(k-1)$-variate Dirichlet distribution $D\,(r_1, r_2, \cdots, r_{k-1}, r_k)$ by the density

$$f(x_1, \cdots, x_{k-1}) = \frac{\Gamma(r_1 + r_2 + \cdots + r_k)}{\prod_{i=1}^{k} \Gamma(r_i)} \prod_{i=1}^{k-1} x_i^{r_i-1}\left(1 - \sum_1^{k-1} x_i\right)^{r_k-1} \qquad (2.1)$$

for

$$x_i \geq 0, \quad i = 1, \cdots, k, \qquad \sum_1^{k-1} x_i \leq 1$$

$$= 0 \quad \text{otherwise}$$

For a Bayesian analysis of multinomial data it is convenient to assume a conjugate prior of the Dirichlet family. Let distribution i have prior density $D(n'_{i1}, n'_{i2}, \cdots, n'_{ik})$ for $i = 1, 2$, where these two distributions are assumed independent. Let n_{ij} be the observed frequency in category j for distribution i. Then the posterior density of $(p_{i1}, p_{i2}, \cdots, p_{i(k-1)})$ is $D(n''_{i1}, n''_{i2}, \cdots, n''_{ik})$, where $n''_{ij} = n'_{ij} + n_{ij}$.

It is not difficult to show (see, e.g., Wilks [1962] p. 177) that $D(n''_{i1}, n''_{i2}, \cdots n''_{ik})$ is the distribution of $(U_{(\nu_{i1})}, U_{(\nu_{i2})} - U_{(\nu_{i1})}, \cdots, U_{(\nu_{i(k-1)})} - U_{(\nu_{i(k-2)})})$, where $U_{(i)}$ is the ith order statistic from a sample of size ν_{ik-1} from the uniform distribution on [0, 1], and

$$\nu_{im} = \sum_{j=1}^{m} n''_{ij}, \qquad m = 1, 2, \cdots, k.$$

Thus if $(U_{(1)}, U_{(2)}, \cdots, U_{(\nu_{1k}-1)})$ and $(V_{(1)}, V_{(2)}, \cdots, V_{(\nu_{2k}-1)})$ are the order statistics from two independent samples, we have

$$P(F_1 < F_2) = P(U_{(\nu_{1m})} \geq V_{(\nu_{2m})}), \qquad m = 1, 2, \cdots, k-1. \qquad (2.2)$$

This approach to comparing two ordered multinomials and the reduction to a combinatorial problem implied by (2.2) have been discussed by Altham [1969], who gives (in our notation)

$$P(F_1 < F_2) = \sum p(s_1, s_2, \cdots, s_k) \tag{2.3}$$

where

$$p(s_1, \cdots, s_k) = \frac{\binom{n''_{11} + n''_{21} - 1}{s_1}\binom{n''_{12} + n''_{22}}{s_2} \cdots \binom{n''_{1(k-1)} + n''_{2(k-1)}}{s_{k-1}}\binom{n''_{1k} + n''_{2k} - 1}{s_k}}{\binom{\nu_{1k} + \nu_{2k} - 2}{\nu_{1k} - 1}}$$

and the summation is over s_i for which $\sum_1 s_i \leq \nu_{1m}$, $m = 1, \cdots, k$.

This expression is difficult to compute and is impossible to compute by hand for even small sample sizes. We shall develop an algorithm which allows hand calculation for small values of the n''_{ij} and is very convenient for machine computation. We require a brief digression.

Suppose we have samples of m and n independent observations from two continuous distributions F and G. Denote a rank ordering (or ranking) of these observations by $z = (z_1, z_2, \cdots, z_{m+n})$ where $z_i = 1$ if the ith smallest observation in the combined sample is from F and $z_i = 0$ if it is from G. We can represent z as a path on a grid from (0, 0) to (m, n) such that the ith step is one unit to the right for $z_i = 1$ and one unit up for $z_i = 0$. For two paths z and $z' = (z'_1, z'_2, \cdots, z'_{m+n})$ we say that z dominates z' ($z \geq z'$) if and only if

$$\sum_{i=1}^{j} z_i \leq \sum_{i=1}^{j} z'_i, \qquad j = 1, 2, \cdots, m + n. \tag{2.4}$$

Thus, $z = 101100$ dominates 101100, 110100, 111000. Note that $z \geq z'$ means the path corresponding to z' is never above that corresponding to z.

Let $N(z)$ be the number of paths, or rankings dominated by a given path z. If $F = G$, then all $\binom{n+m}{n}$ rankings are equally likely, so that $N(z) \Big/ \binom{n+m}{n}$ is the probability that a ranking obtained from two independent samples of size m and n from the distribution is dominated by z.

Let $z_{(r)}$ be the path represented by the first r components of z. Let $N(q, z_{(r)})$ be the number of paths dominated by $z_{(r)}$ and having their last 1 in position q. Let u_r and v_r be respectively the number of 1's and 0's in $z_{(r)}$. Then the following result can be used to calculate $N(z)$ for any z.

$$\begin{aligned} N(q, z_{(r+1)}) &= N(q-1, z_{(r+1)}) + N(q-1, z_{(r)}) \quad &\text{if } z_{r+1} = 1 \\ &= N(q, z_{(r)}) &\text{if } z_{r+1} = 0 \end{aligned} \tag{2.5}$$

The proof is presented in Appendix A. Note that if $z_1 = 1$, we have

$$N(1, z_{(1)}) = 1$$
$$N(j, z_{(1)}) = 0 \qquad j \neq 1$$

If $z_1 = 0$, we have

$$N(0, z_{(1)}) = 1$$
$$N(j, z_{(1)}) = 0 \qquad j \neq 1$$

We can thus use (2.5) recursively to obtain

$$N(q, z) \quad \text{for} \quad q = 0, 1, \cdots, m + n \qquad \text{and}$$
$$N(z) = \sum_{q=m}^{m+n} N(q, z).$$

Returning to our problem, it can be seen that $\{U_{(\nu_{1m})} \geq V_{(\nu_{2m})}, m = 1, 2, \cdots, k - 1\}$ if and only if the ranking defined by the two samples is dominated by

$$z^* = \overbrace{00 \cdots 0}^{n''_{11} - 1} \; \overbrace{11 \cdots 1}^{n''_{21}} \; \overbrace{00 \cdots 0}^{n''_{12}} \; \overbrace{11 \cdots 1}^{n''_{22}} \cdots \overbrace{0 \cdots 0}^{n''_{1k}} \; \overbrace{1 \cdots 1}^{n''_{2k} - 1} \tag{2.7}$$

Thus we have

$$P(F_1 < F_2) = \frac{N(z^*)}{\binom{\nu_{1k} + \nu_{2k} - 2}{\nu_{1k} - 1}} \tag{2.8}$$

To illustrate how the calculations proceed, suppose $k = 3$ and

$$n''_{11} = 4 \qquad n''_{21} = 2$$
$$n''_{12} = 4 \qquad n''_{22} = 3$$
$$n''_{13} = 2 \qquad n''_{23} = 5$$

So that,

$$z^* = 000110000111001111.$$

The calculations are carried out in Table 1 and yield

$$P\{F_1 < F_2\} = \frac{39{,}680}{\binom{18}{9}} = .8161$$

Note that only values of r such that $z_r = 1$ need be displayed since there are no new calculations when $z_r = 0$. Each entry is formed by adding the entry one line above it in the previous column to the entry one line above in the same column. The process is continued until $q = r$. Then the next column is computed in the same way. The calculations for this problem, with $\nu_{13} = \nu_{23} = 10$, were performed by hand in a few minutes. Substantially larger values of ν_{1k}, ν_{2k} can be handled conveniently with the aid of a desk calculator.

As a larger example, let us consider some data presented by Novick

TABLE 1
$N(q, z_{(r)})$

q \ r	4	5	10	11	12	15	16	17	18
1	1	0	0	0	0	0	0	0	0
2	1	1	0	0	0	0	0	0	0
3	1	2	1	0	0	0	0	0	0
4	1	3	3	1	0	0	0	0	0
5		4	6	4	1	0	0	0	0
6			10	10	5	1	0	0	0
7			10	20	15	6	1	0	0
8			10	30	35	21	7	1	0
9			10	40	65	56	28	8	1
10			10	50	105	121	84	36	9
11				60	155	226	205	120	45
12					215	381	431	325	165
13						596	812	756	490
14						596	1408	1568	1246
15						596	2004	2976	2814
16							2600	4980	5790
17								7580	10770
18									18350
									39680

and Grizzle [1965]. The results of using two different operations for duodenal ulcer were as shown in Table 2.

TABLE 2
(FROM NOVICK AND GRIZZLE [1965])

	Death	Fair to poor	Good to excellent
Operation 1	7 (n_{11})	17 (n_{12})	76 (n_{13})
Operation 2	1 (n_{21})	10 (n_{22})	89 (n_{23})

Assuming a uniform $D(1, 1; 1)$ prior for each distribution means that $n''_{ij} = n_{ij} + 1$, so that we obtain

$$n''_{11} = 8 \qquad n''_{21} = 2$$
$$n''_{12} = 18 \qquad n''_{22} = 11$$
$$n''_{13} = 77 \qquad n''_{23} = 90$$

Carrying out the calculations by computer yields

$$P\{F_1 < F_2\} = 0.9747$$

It should be noted that although Altham [1969] does not discuss computational aspects in the use of (2.3) to solve such a problem, the 'largest' problem she considers involves $\nu_{1k} = \nu_{2k} = 28$.

3. COMPARING TWO BINOMIALS

In the important special case $k = 2$, the approach discussed in this paper reduces to finding $P(p_{21} \leq p_{11})$. Altham [1969] has shown that if p_{i1} has the beta (n''_{i1}, n''_{i2}) distribution, then

$$P(p_{21} \leq p_{11}) = \sum_{s=\max(n_{11}-n_{22},0)}^{n_{11}''-1} \frac{\binom{n''_{11}+n''_{21}-1}{s}\binom{n''_{12}+n''_{22}-1}{Y_{22}-1-s}}{\binom{\nu_{12}+\nu_{22}-2}{\nu_{12}-1}} \tag{3.1}$$

The required probability can thus be obtained from tables of the hypergeometric distribution. The most extensive of these are those of Lieberman and Owen [1961] which are adequate for values of $\nu_{12} + \nu_{22}$ up to 102. For larger sample sizes we can use the algorithm of section 2, since for $k = 2$, we have from (2.7) that

$$P(p_{21} \leq p_{11}) = \frac{N(z^{**})}{\binom{\nu_{12}+\nu_{22}-2}{\nu_{12}-1}} \tag{3.2}$$

where

$$z^{**} = \overbrace{00\cdots0}^{n''_{11}-1}\ \overbrace{11\cdots1}^{n''_{21}}\ \overbrace{0\cdots0}^{n''_{12}}\ \overbrace{1\cdots1}^{n''_{22}-1}.$$

The only computational limitation on this method is the size of the denominator in (3.2), and this is a function of the accuracy available on the computer used. Using double precision on an IBM 360/50, we can conveniently deal with values of $\nu_{12} + \nu_{22}$ up to about 250. This can be extended by more careful programming if necessary.

Novick and Grizzle [1965] used an approximate method to obtain bounds on these probabilities. In Table 3 we display the tightest bounds they obtained for three sets of data and the exact probabilities computed from (3.2).

TABLE 3

n''_{11}, n''_{12}	n''_{21}, n''_{22}	Approximate bounds		$P\{p_{21} \leq p_{11}\}$
8,94	2,100	.9811	.9836	.9825
8,94	4,98	.8895	.8967	.8932
4,98	2,100	.8088	.8221	.8156

In comparing two binomials we may wish to consider $P\{p_{21}/p_{11} \leq c\}$ for any value of c. To do this we need a way to calculate the distribution function of two independent beta random variables. The following theorem, whose proof is given in Appendix B, provides such a method.

Theorem: If X and Y have independent beta $(r, n - r + 1)$ and beta $(s, m - s + 1)$ distributions respectively and $0 < c \leq 1$, then

$$P\left(\frac{X}{Y} \leq c\right) = \sum_{\mu=r}^{n} \sum_{k=0}^{n-\mu} (-1)^{n-\mu-k} c^{n-k} \binom{n}{\mu} \binom{n-\mu}{k} \frac{\binom{m+n+s-k-1}{s-1}}{\binom{m+n+s-k-1}{m}} \tag{3.3}$$

Note that for $d > 1$, we have

$$P\left(\frac{X}{Y} \leq d\right) = P\left(\frac{Y}{X} \geq \frac{1}{d}\right) = 1 - P\left(\frac{Y}{X} \leq \frac{1}{d}\right) \tag{3.4}$$

which can be computed from (3.3). As an example, suppose that two treatments are each administered to 40 patients and the results rated as successes or failures. If we assume uniform priors on the p_{i1}'s and observe 12 successes in group 1 and 20 in group 2, we have

$$n''_{21} = 21 \qquad n''_{21} = 13$$

$$n''_{12} = 21 \qquad n''_{22} = 29.$$

Using (3.3) we obtain the results shown in Table 4.

TABLE 4

c	$P\{\frac{p_{21}}{p_{11}} \leq c\}$
.8	.0042
1.0	.0354
1.2	.1313
1.4	.2925
1.6	.4766
1.8	.6406
2.0	.7656
2.5	.9281
3.0	.9787
3.5	.9935
4.0	.9979

COMPARAISON BAYESIENNE DE DEUX POPULATIONS MULTINOMIALES ORDONNEES

RESUME

On développe une approche bayesienne du problème de la comparaison de deux distributions multinomiales à k catégories avec une notion d'ordre parmi les catégories. En supposant a priori des distributions de Dirichlet indépendantes, nous obtenons un algorithme satisfaisant pour calculer la probabilité a posteriori qu'une distribution est stochastiquement plus grande que l'autre. Le cas particulier important où $k = 2$ se réduit au calcul de la probabilité a posteriori qu'une proportion binomiale est plus grande que l'autre. Nous considérons le problème plus général qui consiste à trouver la distribution a posteriori du rapport de deux telles proportions.

REFERENCES

Altham, P. M. E. [1969]. Exact Bayesian analysis of a 2 × 2 contingency table, and Fisher's "Exact" significance test. *Jour. Roy. Stat. Soc. B 31*, 261–9.

Lieberman, G. J. and Owen, D. B. [1961]. *Tables of the Hypergeometric Probability Distribution.* Stanford University Press, Stanford.

Novick, M. R. and Grizzle, J. E. [1965]. A Bayesian approach to the analysis of data from clinical trials. *J. Amer. Statist. Ass. 60*, 81–95.

Weisberg, H. I. [1970]. Upper and lower probability inferences from ordered multinomial data. Ph.D. thesis, Harvard University, Cambridge, Mass.

Wilks, S. S. [1962]. *Mathematical Statistics.* Wiley, New York.

APPENDIX A

In this appendix we present the proof of (2.5).

Proof: Suppose first that $z_{r+1} = 0$. Then any path from (0, 0) to $(u_r, v_r + 1)$ never above $z_{(r+1)}$ and having its last 1 in position q must consist of a path from (0, 0) to (u_r, v_r) never above $z_{(r)}$ and having its last 1 in position q with a 0 added at the end. This proves the second line.

Suppose now that $z_{r+1} = 1$. Any allowable path from (0, 0) to $(u_r + 1, v_r)$ with right-most 1 in position q must consist of a path from (0, 0) to $(u_r, q - v_r - 1)$ followed by a 1 and $u_r + v_r - q + 1$ 0's. Note, however, that the number of paths from (0, 0) to $(u_r, q - v_r - 1)$ which are never above $z_{(r)}$ is just the number of paths from (0, 0) to (u_r, v_r) with last 1 in position $q - 1$ or less which are never above $z_{(r)}$. Thus we have

$$N(q, z_{(r+1)}) = \sum_{j=u_r}^{q-1} N(j, z_{(r)}) \quad \text{if} \quad z_{r+1} = 1. \tag{A1}$$

Now note that

$$\begin{aligned}\sum_{j=u_r}^{q-1} N(j, z_{(r)}) &= \sum_{j=u_r}^{q-2} N(j, z_{(r)}) + N(q-1, z_{(r)}) \qquad \text{(A2)}\\ &= N(q-1, z_{(r+1)}) + N(q-1, z_{(r)}),\end{aligned}$$

the last step following from (A1). This proves the first line of (2.5).

APPENDIX B

In this appendix we present the proof of the theorem in section 3, which we restate here.

Theorem: If X and Y have independent beta $(r, n - r + 1)$ and beta $(s, m - s + 1)$ distributions respectively and $0 < c \leq 1$, then

$$P\left(\frac{X}{Y} \leq c\right) = \sum_{\mu=r}^{n} \sum_{k=0}^{n-\mu} (-1)^{n-\mu-k} c^{n-k} \binom{n}{\mu} \binom{n-\mu}{k} \frac{\binom{m+n+s-k-1}{s-1}}{\binom{m+n+s-k-1}{m}}$$

Proof:

$$P\left(\frac{X}{Y} \le c\right) = \frac{1}{\beta(s, m - s + 1)} \frac{1}{\beta(r, n - r + 1)}$$

$$\cdot \int_0^1 \int_0^{cy} x^{r-1}(1 - x)^{n-r} y^{s-1}(1 - y)^{m-s}\, dx\, dy$$

$$= \frac{1}{\beta(s, m - s + 1)} \int_0^1 \sum_{\mu=r}^{n} \binom{n}{\mu} (cy)^{\mu}(1 - cy)^{n-\mu} y^{s-1}(1 - y)^{m-s}\, dy$$

$$= \frac{1}{\beta(s, m - s + 1)} \sum_{\mu=r}^{n} \binom{n}{\mu} c^{\mu}$$

$$\cdot \int_0^1 y^{\mu+s-1}(1 - y)^{m-s} \sum_{k=0}^{n-\mu} \binom{n - \mu}{k} (-1)^{n-\mu-k}(cy)^{n-\mu-k}\, dy$$

$$= \frac{1}{\beta(s, m - s + 1)} \sum_{\mu=r}^{n} \binom{n}{\mu} \sum_{k=0}^{n-\mu} (-1)^{n-\mu-k} c^{n-k} \binom{n - \mu}{k}$$

$$\cdot \int_0^1 y^{n+s-k-1}(1 - y)^{m-s}\, dy$$

$$= \frac{1}{\beta(s, m - s + 1)} \sum_{\mu=r}^{n} \binom{n}{\mu} (-1)^{n-\mu-k} \binom{n - \mu}{k} c^{n-k}$$

$$\cdot \beta(n + s - k, m - s + 1)$$

$$= \sum_{\mu=r}^{n} \sum_{k=0}^{n-\mu} (-1)^{n-\mu-k} c^{n-k} \binom{n}{\mu} \binom{n - \mu}{k} \frac{\binom{m + n + s - k - 1}{s - 1}}{\binom{m + n + s - k - 1}{m}}$$

Received May 1971, *Revised November* 1971

Key Words: Bayesian inference; Multinomial distribution; Ordered categories; Computing algorithm; Stochastic ordering; Ratio of binomial proportions.