Exact Bayesian Analysis of a 2×2 Contingency Table, and Fisher's "Exact" Significance Test

By Patricia M. E. Altham

M.R.C. Applied Psychology Unit, and Newnham College, Cambridge

[Received June 1968. Revised October 1968]

Summary

A relationship is derived between the posterior probability of negative association of rows and columns of a 2×2 contingency table and Fisher's "exact" probability, as given in existing tables for testing the hypothesis of no association of rows and columns. The result for the 2×2 table is generalized to provide the posterior probability that one discrete-valued random variable is stochastically larger than another.

1. Introduction

Let θ_{ij}, $1\leqslant i, j\leqslant 2$ be the parameters of a 2×2 table, with $\theta_{ij}\geqslant 0$, $1\leqslant i, j\leqslant 2$ and $\sum\sum\theta_{ij} = 1$. With the convention of capital letters for random variables and small letters for the value they take, let the prior density of $(\Theta_{11}, \Theta_{12}, \Theta_{21})$ be Dirichlet with parameters (ν'_{ij}), so that the density function is proportional to $\prod \theta_{ij}^{\nu'_{ij}-1}$: in this case (Θ_{ij}) is said to have prior $D(\nu'_{ij})$.

Suppose that a random sample of size $n_{..}$ yields observations (n_{ij}), then (Θ_{ij}) has posterior density $D(\nu_{ij})$, where $\nu_{ij} = n_{ij}+\nu'_{ij}$, $1\leqslant i, j\leqslant 2$. Assume the usual dot notation for summation over a suffix, and that the (ν_{ij}) are all strictly positive integers.

In choosing a measure of association of rows and columns of the table, it is assumed here that such a measure can be written both as a function of the pair $(\theta_{11}/\theta_{1.}, \theta_{21}/\theta_{2.})$ and as a function of the pair $(\theta_{11}/\theta_{.1}, \theta_{12}/\theta_{.2})$. Edwards (1963) has proved that in this case the measure of association must necessarily be a function of $(\theta_{11}\,\theta_{22}/\theta_{12}\,\theta_{21})$, the cross-ratio. The cross-ratio itself is therefore taken as the measure of association. In general the exact posterior distribution of $(\Theta_{11}\,\Theta_{22}/\Theta_{12}\,\Theta_{21})$ is not easy to deal with, but the posterior probability that $\Theta_{11}\,\Theta_{22}/\Theta_{12}\,\Theta_{21} < 1$, in other words that the association between rows and columns is negative, can readily be computed as a finite sum of hypergeometric probabilities.

2. Derivation of the Posterior Probability of Negative Association

Let $X = \Theta_{11}/\Theta_{1.}$, and $Y = \Theta_{21}/\Theta_{2.}$, then X and Y have independent posterior distributions with densities proportional to $x^{\nu_{11}-1}(1-x)^{\nu_{12}-1}$ and $y^{\nu_{21}-1}(1-y)^{\nu_{22}-1}$ respectively, and

$$
\begin{aligned}
P(\Theta_{11}\,\Theta_{22}/\Theta_{12}\,\Theta_{21} < 1\,|\,\mathbf{n}) &= P(X < Y) \\
&= \int_{y=0}^{1} \frac{\Gamma(\nu_{2.})}{\Gamma(\nu_{21})\,\Gamma(\nu_{22})}\, y^{\nu_{21}-1}(1-y)^{\nu_{22}-1} \\
&\quad \times \int_{x=0}^{y} \frac{\Gamma(\nu_{1.})}{\Gamma(\nu_{11})\,\Gamma(\nu_{12})}\, x^{\nu_{11}-1}(1-x)^{\nu_{12}-1}\, dx\, dy.
\end{aligned}
$$

Using the identity between the lower tail of the beta-distribution and the upper tail of the binomial distribution, given, for example, in Raiffa and Schlaifer (1961, Section 7.1), this expression may be rewritten as

$$\int_{y=0}^{1} \frac{\Gamma(\nu_{2.})}{\Gamma(\nu_{21})\,\Gamma(\nu_{22})} y^{\nu_{21}-1}(1-y)^{\nu_{22}-1} \sum_{r=\nu_{11}}^{\nu_{1.}-1} \binom{\nu_{1.}-1}{r} y^r(1-y)^{\nu_{1.}-1-r}\,dy.$$

Interchanging the summation and integral signs, and evaluating the resulting beta-integral, gives

$$P(\Theta_{11}\,\Theta_{22}/\Theta_{12}\,\Theta_{21}<1\,|\,\mathbf{n}) = \sum_{r=\nu_{11}}^{\nu_{1.}-1} \frac{\Gamma(\nu_{2.})}{\Gamma(\nu_{21})\,\Gamma(\nu_{22})} \binom{\nu_{1.}-1}{r} \frac{\Gamma(\nu_{21}+r)\,\Gamma(\nu_{..}-1-\nu_{21}-r)}{\Gamma(\nu_{..}-1)}. \quad (1)$$

This is the upper tail of the beta-binomial, or hypergeometric waiting-time distribution, whose relation to the hypergeometric distribution is demonstrated by Raiffa and Schlaifer (1961, Section 7.11). Rewriting (1) in terms of the hypergeometric distribution gives

$$P(\Theta_{11}\,\Theta_{22}/\Theta_{12}\,\Theta_{21}<1\,|\,\mathbf{n}) = \sum_{s=max(\nu_{21}-\nu_{12},0)}^{\nu_{21}-1} \binom{\nu_{.1}-1}{s}\binom{\nu_{.2}-1}{\nu_{2.}-1-s}\bigg/\binom{\nu_{..}-2}{\nu_{1.}-1}. \quad (2)$$

3. Discussion of the Result for the 2×2 Table

The right-hand side of equation (2) is just Fisher's "exact" probability for the table

ν_{11}	$\nu_{12}-1$
$\nu_{21}-1$	ν_{22}

that the classical statistician would compute when testing the hypothesis $\theta_{11}\,\theta_{22}=\theta_{12}\,\theta_{21}$ against the hypothesis $\theta_{11}\,\theta_{21}<\theta_{12}\,\theta_{21}$. If $\nu'_{11}=\nu'_{22}=0$ and $\nu'_{12}=\nu'_{21}=1$, which seems to correspond to a strong prior belief in negative association of rows and columns, then (2) is just the same as the classical "exact" probability computed for data (n_{ij}). If $\nu'_{ij}=p$, $1\leqslant i, j\leqslant 2$, where $0\leqslant p\leqslant 1$, then Fisher's test is conservative with respect to the Bayes procedure, but the difference between the two significance levels is less than the probability of the data on the null hypothesis. Of course, for large (n_{ij}), this difference is negligible.

Denote the Fisher's significance level by α, so that

$$\alpha = P(\text{observations or more extreme cases}\,|\,\theta_{11}\,\theta_{22}=\theta_{12}\,\theta_{21}).$$

Here "more extreme cases" are all configurations with the same marginal totals $(n_{i.}), (n_{.j})$ but showing stronger positive association.

Let

$$\beta = P(\text{more extreme cases}\,|\,\theta_{11}\,\theta_{22}=\theta_{12}\,\theta_{21})$$

and

$$\pi = P(\Theta_{11}\,\Theta_{22}<\Theta_{12}\,\Theta_{21}\,|\,\mathbf{n}) = \pi(\nu_{11},\nu_{12},\nu_{21},\nu_{22}),$$

say. The inequalities $\beta<\pi<\alpha$ are proved as follows.

Clearly $\pi(.,.,.,.)$ is strictly decreasing in its first and fourth arguments, and strictly increasing in its second and third arguments. The fact that this particular prior $D(\nu_{ij})$ does not necessarily give (ν_{ij}) all integers, presents no problem here.

Although the integral π cannot be expressed as a hypergeometric sum, equation (2) shows that the sums α and β can be expressed as integrals, which are then easy to compare with π.

Indeed $\alpha = \pi(n_{11}, n_{12}+1, n_{21}+1, n_{22})$ and $\beta = \pi(n_{11}+1, n_{12}, n_{21}, n_{22}+1)$. Now $\nu_{ij} = n_{ij}+p$, for $1 \leqslant i, j \leqslant 2$, and so

$$\begin{aligned} n_{11}+1 &\geqslant \nu_{11} \geqslant n_{11}, \\ n_{12} &\leqslant \nu_{12} \leqslant n_{12}+1, \\ n_{21} &\leqslant \nu_{21} \leqslant n_{21}+1, \\ n_{22}+1 &\geqslant \nu_{22} \geqslant n_{22}. \end{aligned}$$

Hence $\pi(n_{11}+1, n_{12}, n_{21}, n_{22}+1) < \pi(\nu_{11}, \nu_{12}, \nu_{21}, \nu_{22}) < \pi(n_{11}, n_{12}+1, n_{21}+1, n_{22})$, and so $\beta < \pi < \alpha$. Fisher's significance level exceeds the Bayes significance level by an amount

$$\alpha-\pi, \text{ and } \alpha-\pi < \alpha-\beta = \binom{n_{1.}}{n_{11}}\binom{n_{2.}}{n_{21}}\Big/\binom{n_{..}}{n_{.1}}.$$

The inequalities $\beta < \pi < \alpha$ are not true for general p. For example, if $p = 2$, and $n_{11} = n_{22} = 5$, $n_{12} = n_{21} = 0$ then $\alpha = 0{\cdot}004$ and $\pi = 0{\cdot}005$ (to three decimal places).

The special cases $p = 0$, $0{\cdot}5$ and 1 are of interest as they correspond respectively to the "non-committal" prior distribution, the prior favoured by Jeffreys on invariance grounds, and the uniform prior distribution.

If $\nu_{2.} \leqslant \nu_{1.} \leqslant 41$, $\nu_{11}\nu_{22} \geqslant (\nu_{12}-1)(\nu_{21}-1)$ and $\pi < 0{\cdot}05$, then π may be found in the tables produced by Finney *et al.* (1963). The hypergeometric distribution function has been tabulated by Lieberman and Owen (1961). I have compiled tables of π to three decimal places, for $20 \geqslant \nu_{1.} \geqslant \nu_{2.} \geqslant 2$ and $\nu_{11}\nu_{22} \geqslant \nu_{12}\nu_{21}$, which I hope are laid out in a form convenient for Bayesian statisticians, and I should be pleased to send a copy to anyone who is interested. These tables show that π is generally small for $\nu_{11}\nu_{22}/\nu_{12}\nu_{21}$ large, which is not surprising, but there exist tables (ν_{ij}) for which $\nu_{11}\nu_{22}/\nu_{12}\nu_{21} > 1$ and $\pi > 0{\cdot}5$; for example $\nu_{11} = 13$, $\nu_{12} = 3$, $\nu_{21} = 4$, $\nu_{22} = 1$. Here $\pi = 0{\cdot}530$.

Bayesians may also use the hypergeometric sum (2) to compute π if the data are collected with, say, the row totals $n_{1.}$ and $n_{2.}$ fixed, but not of course when both row and column totals are fixed. In this case difficulties arise in making the choice of prior density; see Lindley (1964, Section 5). For this reason it may not be strictly appropriate to compare β, π and α as above; the distribution of n_{11}, given $\theta_{11}\theta_{22} = \theta_{12}\theta_{21}$, is of course hypergeometric only when both the row and column totals are fixed, and in this case π is not a hypergeometric sum. However, the comparison between β, π and α seems important all the same, since Fisher's test is often used when the data were not collected with both margins fixed.

There does not appear to be an attractive equation corresponding to (2) that links the posterior probability of negative second-order interaction in a $2\times2\times2$ table (θ_{ijk}), i.e. $P\{\Theta_{111}\Theta_{122}\Theta_{212}\Theta_{221} < \Theta_{222}\Theta_{211}\Theta_{121}\Theta_{112} \mid (n_{ijk})\}$ with the exact significance test of the hypothesis: $\theta_{111}\theta_{122}\theta_{212}\theta_{221} = \theta_{222}\theta_{211}\theta_{121}\theta_{112}$ that was proposed by Bartlett (1935).

4. Normal Approximations

The relationship between the posterior distribution of the cross-ratio and the hypergeometric distribution makes two asymptotic normal approximations available for π. Bloch and Watson (1967) found a normal approximation to the distribution of contrasts of the logarithms of Dirichlet variables.

Writing $\pi = P(\log\Theta_{11} - \log\Theta_{12} - \log\Theta_{21} + \log\Theta_{22} < 0\,|\,\mathbf{n})$, their result in this particular case gives

$$\pi \simeq \pi_1 = \Phi\left(-\frac{\mu_1}{\sigma_1}\right),$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-z^2/2}\,dz, \quad \mu_1 = \log\left(\frac{\lambda_{11}\lambda_{22}}{\lambda_{12}\lambda_{21}}\right), \quad \sigma_1^2 = \sum\sum\frac{1}{\lambda_{ij}},$$

where

$$\lambda_{ij} = \nu_{ij} - 0{\cdot}5 + \frac{1}{24(\nu_{ij}-1)}, \quad 1 \leqslant i,j \leqslant 2.$$

The normal approximation to the hypergeometric distribution, using Yates's half-correction, gives

$$\pi \simeq \pi_2 = 1 - \Phi\left(\frac{\nu_{11} - 0{\cdot}5 - \mu_2}{\sigma_2}\right),$$

$$\mu_2 = \frac{(\nu_{1.}-1)(\nu_{.1}-1)}{(\nu_{..}-2)} \quad \text{and} \quad \sigma_2^2 = \frac{(\nu_{1.}-1)(\nu_{2.}-1)(\nu_{.1}-1)(\nu_{.2}-1)}{(\nu_{..}-2)(\nu_{..}-1)^2}.$$

Numerical comparisons of these two approximations to π were computed for $\nu_{ij} \geqslant 2$, $\nu_{11}\nu_{22} \geqslant \nu_{12}\nu_{21}$; firstly for $12 \geqslant \nu_{1.} \geqslant \nu_{2.} \geqslant 5$, secondly for $25 = \nu_{1.} \geqslant \nu_{2.} \geqslant 4$, and finally for some values of (ν_{ij}) for which $\nu_{1.} = 42, 46$. The last set was chosen for comparison with some of the results already published by Finney *et al.* (1963, p. 6). Some of the values obtained are given in Table 1. For (ν_{ij}) all large, and π not too small, ($\pi > 0{\cdot}1$, say) π_1 seems a slightly better approximation to π than does π_2. However, if π is small, π_2 is generally a closer estimate that π_1. For example, for $25 = \nu_{1.} \geqslant \nu_{2.} \geqslant 4$ and $\nu_{ij} \geqslant 2$, $\nu_{11}\nu_{22} \geqslant \nu_{12}\nu_{21}$, the number of values of (ν_{ij}) for which $\pi < 0{\cdot}05$ is 1,463. Of these, 1,308 values give $|\pi_2 - \pi| < |\pi_1 - \pi|$. It is well-known that π_2 tends to over-estimate π when π is small, and the approximation π_1 suffers from this fault too, usually to a greater extent. Analytical comparisons of π, π_1 and π_2 have proved intractable so far.

5. Exact Bayesian Comparison of two Discrete-valued Scores

Let X and Y be two scores, taking values $z_1, z_2, \ldots, z_s$, where $z_1 < z_2 < \ldots < z_s$. Let

$$P(X = z_i\,|\,\boldsymbol{\theta}) = \theta_i, \quad 1 \leqslant i \leqslant s, \quad \sum\theta_i = 1, \quad \theta_i \geqslant 0, \quad 1 \leqslant i \leqslant s$$

and

$$P(Y = z_i\,|\,\boldsymbol{\phi}) = \phi_i, \quad 1 \leqslant i \leqslant s, \quad \sum\phi_i = 1, \quad \phi_i \geqslant 0, \quad 1 \leqslant i \leqslant s.$$

Let $(\Theta_1, \ldots, \Theta_{s-1})$ and $(\Phi_1, \ldots, \Phi_{s-1})$ have independent Dirichlet prior densities, with parameters $(\nu'_1, \ldots, \nu'_s)$ and $(\mu'_1, \ldots, \mu'_s)$ respectively. Suppose that independent random samples of sizes $n_.$ and $m_.$ on X and Y reveal that X takes the value z_i exactly n_i times, and Y takes the value z_i exactly m_i times. Then $(\Theta_1, \ldots, \Theta_{s-1})$ and $(\Phi_1, \ldots, \Phi_{s-1})$ have independent posterior densities, with parameters $(\nu_1, \ldots, \nu_s)$ and $(\mu_1, \ldots, \mu_s)$ respectively, where $\nu_i = \nu'_i + n_i$ and $\mu_i = \mu'_i + m_i$, $1 \leqslant i \leqslant s$. Assume $(\nu_i), (\mu_i)$ are all strictly positive integers.

Write "X stoch. > Y" for the event "X is stochastically greater than Y". Then

$$P(X \text{ stoch.} > Y \mid \mathbf{n},\mathbf{m}) = P(\Theta_1 + \ldots + \Theta_i < \Phi_1 + \ldots + \Phi_i,\ 1 \leqslant i \leqslant s-1 \mid \mathbf{n},\mathbf{m}).$$

TABLE 1

Comparison of normal approximations to π

	$\nu_{1.}$	$\nu_{2.}$	ν_{11}	ν_{21}	π	π_1	π_2
(1)	12	12	6	2	0·032	0·043	0·037
(2)	12	12	6	3	0·091	0·097	0·097
(3)	12	12	6	4	0·193	0·195	0·198
(4)	12	12	6	5	0·335	0·335	0·338
(5)	12	12	6	6	0·500	0·500	0·500
(6)	12	6	8	2	0·077	0·087	0·083
(7)	12	6	8	3	0·242	0·245	0·250
(8)	12	6	8	4	0·516	0·522	0·528
(9)	25	25	12	2	0·000	0·003	0·001
(10)	25	25	12	4	0·006	0·009	0·007
(11)	25	25	12	6	0·034	0·038	0·037
(12)	25	25	12	7	0·068	0·070	0·070
(13)	25	25	12	8	0·119	0·120	0·121
(14)	25	25	12	10	0·281	0·281	0·282
(15)	25	25	12	11	0·386	0·386	0·387
(16)	25	25	12	12	0·500	0·500	0·500
(17)	25	12	12	2	0·022	0·035	0·027
(18)	25	12	12	3	0·077	0·085	0·082
(19)	25	12	12	4	0·187	0·188	0·189
(20)	25	12	12	5	0·352	0·350	0·352
(21)	42	42	30	22	0·034	0·036	0·035
(22)	42	42	30	19	0·007	0·008	0·007
(23)	42	42	30	18	0·004	0·004	0·004
(24)	42	42	34	11	0·005	0·006	0·005
(25)	42	6	32	2	0·018	0·025	0·015
(26)	46	6	32	2	0·040	0·051	0·038
(27)	42	42	21	20	0·413	0·413	0·413
(28)	42	42	21	18	0·253	0·253	0·255
(29)	42	42	21	16	0·133	0·134	0·134
(30)	42	42	21	10	0·362	0·361	0·361
(31)	42	42	21	9	0·239	0·239	0·240

Now a Dirichlet distribution can be constructed from the spacings between the order statistics of a random sample from the uniform distribution on the unit interval (Wilks, 1962, Section 8.7). Consider two independent random samples of sizes $(\nu_{.}-1)$ and $(\mu_{.}-1)$ from the uniform distribution on (0, 1): call their representation by points in the interval (0, 1) "red points" and "black points" respectively. Let $\Theta_1+\ldots+\Theta_i$ be the distance from 0 to the $(\nu_1+\ldots+\nu_i)$th red point, and $\Phi_1+\ldots+\Phi_i$ be the distance from 0 to the $(\mu_1+\ldots+\mu_i)$th black point, for $1 \leqslant i \leqslant s-1$.

The following three events are equivalent:

(a) $\Theta_1+\ldots+\Theta_i<\Phi_1+\ldots+\Phi_i, \quad 1\leqslant i\leqslant s-1;$ (3)

(b) the $(\nu_1+\ldots+\nu_i)$th red point is encountered before the $(\mu_1+\ldots+\mu_i)$th black point, proceeding from 0, $1\leqslant i\leqslant s-1$; (4)

(c) the number of black points encountered in the first

$(\mu_1+\ldots+\mu_i+\nu_1+\ldots+\nu_i-1)$ points is $\leqslant(\mu_1+\ldots+\mu_i-1)$, for $1\leqslant i\leqslant s-1$. (5)

The equivalence between (3) and (5) provides a direct proof of the connection between the posterior probability of negative association and Fisher's exact probability for the 2×2 contingency table.

Let N_1 be the number of black points selected in the first $(\mu_1+\nu_1-1)$ points, proceeding from 0, N_2 the number of black points selected in the next $(\mu_2+\nu_2)$ points, N_3 the number of black points selected in the next $(\mu_3+\nu_3)$ points . . . and N_s the number of black points in the final $(\mu_s+\nu_s-1)$ points. Then $N_1+\ldots+N_s=\mu_.-1$, and from the equivalence between (3) and (5),

$$P(X\,\text{stoch.}>Y\,|\,\mathbf{n},\mathbf{m})=\sum p(n_1,\ldots,n_s), \tag{6}$$

where

$$p(n_1,\ldots,n_s)=\frac{\binom{\mu_1+\nu_1-1}{n_1}\binom{\mu_2+\nu_2}{n_2}\cdots\binom{\mu_{s-1}+\nu_{s-1}}{n_{s-1}}\binom{\mu_s+\nu_s-1}{n_s}}{\binom{\mu_.+\nu_.-2}{\nu_.-1}}$$

for $n_i\geqslant 0$, $1\leqslant i\leqslant s$, $n_1\leqslant\mu_1+\nu_1-1$, $n_i\leqslant\mu_i+\nu_i$ for $2\leqslant i\leqslant s-1$, and $n_s\leqslant\mu_s+\nu_s-1$ and $n_1+\ldots+n_s=\mu_.-1$. The summation sign $\sum$ in (6) extends over (n_i) for which $n_1+\ldots+n_i\leqslant\mu_1+\ldots+\mu_i-1$, $1\leqslant i\leqslant s-1$.

6. Remarks on the Application of Expression (6)

The expression (6) is simple to calculate by computer, and some numerical examples are given in Table 2.

TABLE 2

Posterior probability that one score is stochastically greater than another

s	$(\nu_1,\ldots,\nu_s)$	$(\mu_1,\ldots,\mu_s)$	$P(X$ stoch. $>Y)$	$P(Y$ stoch. $>X)$
5	(1, 2, 14, 4, 7)	(4, 10, 8, 4, 2)	0·873	0·000
5	(1, 1, 11, 2, 3)	(3, 10, 5, 3, 1)	0·639	0·000
4	(1, 7, 4, 5)	(8, 10, 2, 2)	0·948	0·000
4	(3, 4, 10, 11)	(7, 12, 4, 5)	0·902	0·000
4	(3, 4, 10, 11)	(2, 6, 9, 11)	0·170	0·234
4	(2, 6, 9, 11)	(7, 12, 4, 5)	0·943	0·000

Clearly, if $0<\alpha<1$, and X_α, Y_α are the α-quantiles of the distribution functions of the scores X, Y respectively, then $P(X\,\text{stoch.}>Y\,|\,\mathbf{n},\mathbf{m})<P(X_\alpha>Y_\alpha\,|\,\mathbf{n},\mathbf{m})$. X_α and Y_α

have independent posterior distributions. McLaren (1967) has shown that for $1 \leqslant i \leqslant s$ $P(X_\alpha \leqslant z_i \mid \mathbf{n}) = P(\Theta_1 + \ldots + \Theta_i > \alpha \mid \mathbf{n}) = \sum \binom{\nu_. - 1}{r} \alpha^r (1-\alpha)^{\nu_.-1-r}$ where $\sum$ extends over $0 \leqslant r \leqslant \nu_1 + \nu_2 + \ldots + \nu_i - 1$. Hence $P(X_\alpha > Y_\alpha \mid \mathbf{n},\mathbf{m})$ can readily be computed, and may provide a useful upper bound to $P(X \text{ stoch.} > Y \mid \mathbf{n},\mathbf{m})$.

In general, for $s > 2$, $P(X \text{ stoch.} > Y \mid \mathbf{n},\mathbf{m}) + P(Y \text{ stoch.} > X \mid \mathbf{n},\mathbf{m}) < 1$. However, the program to compute $P(X \text{ stoch.} > Y \mid \mathbf{n},\mathbf{m})$ may be checked by equations (7) and (8) below.

Define the events $E_1, E_2, \ldots, E_s$ by

$$E_1 = \Theta_1 < \Phi_1, \quad \Theta_1 + \Theta_2 < \Phi_1 + \Phi_2, \ldots, \Theta_1 + \ldots + \Theta_{s-1} < \Phi_1 + \ldots + \Phi_{s-1}.$$
$$E_2 = \Theta_2 < \Phi_2, \quad \Theta_2 + \Theta_3 < \Phi_2 + \Phi_3, \ldots, \Theta_2 + \ldots + \Theta_s < \Phi_2 + \ldots + \Phi_s.$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$
$$E_s = \Theta_s < \Phi_s, \quad \Theta_s + \Theta_1 < \Phi_s + \Phi_1, \ldots, \Theta_s + \Theta_1 + \ldots + \Theta_{s-2} < \Phi_s + \Phi_1 + \ldots + \Phi_{s-2}.$$

Then

$$E_1 = X \text{ stoch.} > Y.$$

It is proved in the Appendix that

$$\sum_{j=1}^{s} P(E_j) = 1. \tag{7}$$

From (7), if $\nu_i = \nu$, $\mu_i = \mu$, $1 \leqslant i \leqslant s$, by symmetry

$$P(E_j) = \frac{1}{s}, \quad 1 \leqslant j \leqslant s. \tag{8}$$

The proof of (7) uses only the facts that $\Theta_i, \Phi_i \geqslant 0$, $1 \leqslant i \leqslant s$ and $\sum \Theta_i = \sum \Phi_i$, and the continuity of the Dirichlet distribution.

The Bayesian method presented here does involve heavier computation than that required by most classical non-parametric tests of whether one random variable is stochastically larger than another. Computation for these classical tests is of course generally simplified by the existence of special tables, for example, those constructed by Mann and Whitney for their U-test. However, such tests were initially constructed for continuous data. In fact they are often used for discrete data, and then a certain amount of complication is caused when there are ties in the rankings. This Bayesian method has the advantage that it is designed for discrete data.

Acknowledgements

I should like to thank my research supervisor Mr A. D. McLaren for much helpful advice and encouragement, and the Medical Research Council for financial support. I am also grateful to the Cambridge University Mathematical Laboratory for computing time on TITAN. Finally, I wish to thank the referee for constructive suggestions to improve the original form of this paper.

References

Bartlett, M. S. (1935). Contingency table interactions. *J. R. Statist. Soc., Suppl.*, **2**, 248–252.

Bloch, D. A. and Watson, G. S. (1967). A Bayesian study of the multinomial distribution. *Ann. Math. Statist.*, **38**, 1423–1435.

EDWARDS, A. W. F. (1963). The measure of association in a 2×2 table. *J. R. Statist. Soc.* A, **126**, 109–114.
FINNEY, D. J., LATSCHA, R., BENNETT, B. M. and HSU, P., with an introduction by PEARSON, E. S. (1963). *Tables for Testing Significance in a* 2×2 *Contingency Table*. Cambridge: University Press.
LIEBERMAN, G. J. and OWEN, D. B. (1961). *Tables of the Hypergeometric Probability Distribution*. Stanford: University Press.
LINDLEY, D. V. (1964). The Bayesian analysis of contingency tables. *Ann. Math. Statist.*, **35**, 1622–1643.
MCLAREN, A. D. (1967). Appendix to CHOWN, S., BELBIN, E. and DOWNS, S. (1967). Programmed instruction as a method of teaching paired associates to older learners. *J. Gerontology*, **22**, 212–219.
MANN, H. B. and WHITNEY, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.*, **18**, 50–60.
RAIFFA, H. and SCHLAIFER, R. (1961). *Applied Statistical Decision Theory*. Boston: Graduate School of Business Administration, Harvard University.
WILKS, S. S. (1962). *Mathematical Statistics*. New York: Wiley.

APPENDIX

To prove (7), it is sufficient to show that the E_j's are pairwise disjoint and mutually exhaustive events (neglecting sets of measure zero).

Consider the points $(1,2,\ldots,s)$ arranged clockwise on the circumference of a circle. Denote by $S(j,k)$ the sum $\sum(\Theta_l-\Phi_l)$, where l runs from j to $k-1$ inclusive, so that for example,

$$\begin{aligned} S(1,2) &= \Theta_1-\Phi_1,\\ S(1,3) &= \Theta_1+\Theta_2-\Phi_1-\Phi_2,\\ S(s-1,1) &= \Theta_{s-1}+\Theta_s-\Phi_{s-1}-\Phi_s,\\ S(s,2) &= \Theta_s+\Theta_1-\Phi_s-\Phi_1. \end{aligned}$$

Then since

$$\sum\Theta_i = \sum\Phi_i, \quad S(j,k) = -S(k,j) \quad \text{for} \quad 1 \leqslant j,\, k \leqslant s \tag{9}$$

and it is easily verified that

$$S(j,k)+S(k,l) = S(j,l) \quad \text{for} \quad 1 \leqslant j,\, k,\, l \leqslant s. \tag{10}$$

This notation for the partial sums of $(\Theta_l-\Phi_l)$ means that E_j may be written $\bigcap\{S(j,l)<0\}$ where l runs over all elements of the set $(1,2,\ldots,s)$ except j.

Hence, for any distinct j and k, $E_j \subset \{S(j,k)<0\}$, and $E_k \subset \{S(k,j)<0\}$, and so $E_j \cap E_k = \varnothing$ by (9).

It remains to show that $\bigcup E_j = \Omega$, and this is achieved by *reductio ad absurdum*.

Suppose there exists an element ω of Ω for which $\omega \notin \bigcup E_j$, so that for $1 \leqslant j \leqslant s$, $\omega \notin E_j$. Let j_1 be any element of the set $(1,2,\ldots,s)$. Then $\omega \notin E_{j_1}$ implies that there exists $j_2 \neq j_1$, such that $S(j_1,j_2) \geqslant 0$ and so, from (9), $S(j_2,j_1) \leqslant 0$. Now $\omega \notin E_{j_2}$, and so there exists $j_3 \neq j_1, j_2$ such that $S(j_2,j_3) \geqslant 0$ and so, from (9) $S(j_3,j_2) \leqslant 0$. $S(j_3,j_2), S(j_2,j_1) \leqslant 0$ imply, from (10) that $S(j_3,j_1) \leqslant 0$. $\omega \notin E_{j_3}$ implies that there exists $j_4 \neq j_1,j_2,j_3$ for which $S(j_3,j_4) \geqslant 0$ so that $S(j_4,j_3) \leqslant 0$. Using (10) again, it is apparent that $S(j_4,j_i) \leqslant 0$ for $1 \leqslant i \leqslant 3$. Repeating this argument gives a set $(j_1,\ldots,j_{s-1})$ of distinct numbers from $(1,2,\ldots,s)$, for which $S(j_{s-1},j_i) \leqslant 0$ for $1 \leqslant i \leqslant s-2$. Now $\omega \notin E_{j_{s-1}}$ implies that there exists $j_s \neq j_1,j_2,\ldots,j_{s-1}$ for which $S(j_{s-1},j_s) \geqslant 0$, so that $S(j_s,j_{s-1}) \leqslant 0$. But $S(j_{s-1},j_i) \leqslant 0$ for $1 \leqslant i \leqslant s-2$, so that $S(j_s,j_i) \leqslant 0$, $1 \leqslant i \leqslant s-1$.

The set $(j_1, \ldots, j_s)$ is a permutation of $(1, 2, \ldots, s)$, so that $\omega \in E_{j_s}$ which gives the required contradiction. Hence $\Omega = \bigcup E_j$, and so $P(\Omega) = 1 = \sum P(E_j)$.

Equations (7) and (8) have applications to the theory of order statistics. For example, suppose a, μ and ν are positive integers, and $n = a\nu - 1$, $m = a\mu - 1$. Let F and G be continuous distribution functions, and let $(X_1, \ldots, X_n), (Y_1, \ldots, Y_m)$ be independent random samples from F and G respectively. If $(X_{(1)}, \ldots, X_{(n)})$ and $(Y_{(1)}, \ldots, Y_{(m)})$ are the ordered random samples, so that $X_{(1)} \leqslant \ldots \leqslant X_{(n)}$ and $Y_{(1)} \leqslant Y_{(2)} \leqslant \ldots \leqslant Y_{(m)}$, then $\{F(X_{(1)}), \ldots, F(X_{(n)})\}$ and $\{G(Y_{(1)}), \ldots, G(Y_{(m)})\}$ constitute independent ordered random samples from the uniform distribution on $(0, 1)$. The independently distributed Dirichlet random vectors $\mathbf{\Theta}$ and $\mathbf{\Phi}$ may be defined by

$$\begin{aligned}
&\Theta_1 + \ldots + \Theta_i = F(X_{(i\nu)}), \quad 1 \leqslant i \leqslant a-1,\\
&\Theta_1 + \ldots + \Theta_a = 1,\\
&\Phi_1 + \ldots + \Phi_i = G(Y_{(i\mu)}), \quad 1 \leqslant i \leqslant a-1,\\
&\Phi_1 + \ldots + \Phi_a = 1.
\end{aligned}$$

The vectors $\mathbf{\Theta}, \mathbf{\Phi}$ have distributions $D(\boldsymbol{\nu})$, $D(\boldsymbol{\mu})$ respectively, where $\nu_1 = \ldots = \nu_a = \nu$; and $\mu_1 = \ldots = \mu_a = \mu$. Hence, from (8),

$$P\{F(X_{(i\nu)}) < G(Y_{(i\mu)}),\ 1 \leqslant i \leqslant a-1\} = \frac{1}{a}.$$

In particular, if $F = G$, this equation becomes

$$P(X_{(i\nu)} < Y_{(i\mu)},\ 1 \leqslant i \leqslant a-1) = \frac{1}{a}.$$