The Annals of Statistics
1975, Vol. 3, No. 1, 246–250

THE BAYES FACTOR AGAINST EQUIPROBABILITY OF A MULTINOMIAL POPULATION ASSUMING A SYMMETRIC DIRICHLET PRIOR[1]

BY I. J. GOOD

Virginia Polytechnic Institute and State University

A sample $(n_1, n_2, \cdots, n_t)$ is taken from a t-category multinomial population. The hypothesis of equiprobability, that the t physical probabilities associated with the cells are all equal to $1/t$, is called the null hypothesis. Conditional on the non-null hypothesis, a symmetric Dirichlet prior of parameter k is assumed $(k \geqq 0)$ and the Bayes factor against the null hypothesis, with this assumption, is denoted by $F(k)$. A conjecture made in 1965 is almost proved, namely that $F(k)$ has a unique local maximum and that this occurs for a finite value of k if and only if Pearson's X^2 exceeds its number of degrees of freedom. The result is required for the calculation of max $F(k)$, which provides a non-Bayesian significance criterion whose simple asymptotic distribution is good even in the extreme tail, and even for sample sizes less than t. This criterion arose from an attitude involving a Bayes/non-Bayes compromise.

The purpose of this paper is to prove a conjecture made by Good (1965, page 37) related to a Bayesian "significance test" for "equiprobability" of a multinomial population, but which has application to a useful non-Bayesian criterion. To avoid repetition, I assume that the reader has access to Good (1967) for some of the background and terminology. The proof requires a side condition that is probably unnecessary for the result, judging by some numerical results programmed by J. F. Crook.

Let $(n_1, n_2, \cdots, n_t) = \mathbf{n}$ denote the t cell frequencies in a sample from a multinomial distribution having t categories, where $n_1 + n_2 + \cdots + n_t = N$, the sample size. Let p_i denote the physical probability corresponding to cell i $(i = 1, 2, \cdots, t)$, where $\sum p_i = 1$. Let the *null hypothesis* H or *equiprobable case* be defined by $p_i = 1/t$ $(i = 1, 2, \cdots, t)$.

If a symmetric Dirichlet prior, of density proportional to $\prod p_i^{k-1}$ $(k > 0)$, is assumed, then the Bayes factor against the null hypothesis of equiprobability is

$$F(k) = t^N \Gamma(tk)[\prod \Gamma(n_i + k)][\Gamma(k)]^{-t}/\Gamma(N + tk) \tag{1}$$

$$= \left[\prod_{i=1}^{t} \prod_{j=1}^{n_i-1} \left(1 + \frac{j}{k}\right)\right] \prod_{j=1}^{N-1} \left(1 + \frac{j}{tk}\right)^{-1} \tag{2}$$

where $\prod_{j=1}^{n_i-1} 1 + j/k$ is to be interpreted as 1 when $n_i = 0$ or 1.

Received December 1973; revised February 1974.

[1] This work was supported in part by H.E.W., N.I.H. Grant # 1 R01 GM18870.

AMS 1970 *subject classifications*. Primary 62E15; Secondary 62G10.

Key words and phrases. Multinomial significance test, Type II likelihood ratio, hierarchical probability judgments, Bayes/non-Bayes compromise, Bayes factor, weight of evidence, symmetric Dirichlet distribution.

Let $F_{\max} = F(k_{\max})$ denote the least upper bound of $F(k)$ for $k > 0$. ($k_{\max}$ might be 0 or ∞.) Then $F_{\max}$ is a "Type II likelihood ratio" and can be used as a non-Bayesian criterion. In fact, if

$$G = (2 \log_e F_{\max})^{\frac{1}{2}}\,, \tag{3}$$

then, for large N, if $G > 0$,

(4) $P(G > x)/c_t \approx$ double tail corresponding to a standard normal deviate x, where c_t is the probability that χ^2 with $t-1$ degrees of freedom exceeds its expectation,

$$c_t = P(\chi^2_{[t-1]} > t-1)\,. \tag{5}$$

Calculations show that (4) is a very good approximation, even if $N < t$, down to amazingly small tail-areas such as 10^{-16} (Good and Crook, 1972), and is therefore a useful non-Bayesian criterion, especially for cryptanalysts.

For calculating $k_{\max}$ we need the following theorems. We first define *Condition* C. *There is not more than one i for which*

$$N/t < n_i < (N + t - 1)/t\,.$$

The condition is always satisfied if N is a multiple of t because there is then no integer in the interval mentioned. In any case there is at most one such integer and Condition C states that there are not two values of i for which n_i is equal to this integer.

THEOREM 1. *Suppose that* $N \neq 1$ *and* $t \neq 1$, *and assume Condition* C. *Then* $F(k)$ *takes its maximum at* $k = \infty$ *if* $X^2 \leqq t-1$ *and for a finite value of* k *if* $X^2 > t-1$, *where*

$$X^2 = \frac{t}{N} \sum \left(n_i - \frac{N}{t}\right)^2.$$

Thus $F(k) > 1$ *if and only if* $X^2 > t-1$, *the number of degrees of freedom.*

It is convenient first to dispose of the easy case $n_i = N$ for some i. In this case, $F(0) = t^{N-1}$ and $F(k)$ is easily seen, from (2), to be a strictly decreasing function of k that tends to 1 when $k \to \infty$; apart from the entirely trivial cases $N = 1$ or $t = 1$ when $F(k) = 1$ for all k. When $n_i \neq N$ for any i, we can readily see that $F(0) = 0$ (which shows that the value $k = 0$ is untenable at least for "significance testing") and that $F(k) \to 1$ as $k \to \infty$, and of course $F(k) \geqq 0$ for all $k \geqq 0$ and is continuous. Therefore, in order to prove Theorem 1, it is enough to prove the following two theorems.

THEOREM 2. *If* $n_i \neq N$ *for any i (which rules out the trivial cases), then, when k is sufficiently large (depending on t, N, and* $\mathbf{n}$*), the derivative* $F'(k) < 0$ *if* $X^2 > t-1$, *and* $F'(k) > 0$ *if* $X^2 \leqq t-1$.

THEOREM 3. *If* $N \neq 1$ *and* $t \neq 1$, *and Condition* C *is satisfied, then* $F(k)$ *has at most one local stationary value, and when it exists it must be a maximum point.*

All possible forms of the graph of $F(k)$ can be inferred from the three theorems. (See Fig. 1.)

PROOF OF THEOREM 2. From (1) we have

$$W'(k) = t\psi(tk) + \sum_{i=1}^{t} \psi(n_i + k) - t\psi(N + tk) - t\psi(k) ,$$

where $W(k) = \log_e F(k) =$ weight of evidence against H, and asymptotically for large z

$$\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)} = \log z - \frac{1}{2z} - \frac{1}{12z^2} + \frac{1}{120z^4} - \cdots .$$

We can accordingly expand $W'(k)$ in powers of $1/k$, and we obtain for large k

$$W'(k) = a_0 + a_1 k^{-1} + a_2 k^{-2} + a_3 k^{-3} + \cdots ,$$

where

$$a_0 = a_1 = 0,\ a_2 = \frac{N}{2t}(t - 1 - X^2) \tag{6}$$

and, if $X^2 = t - 1$,

$$6a_3 = 2(\sum n_i^3 - N^3 t^{-2} - N) + 3(N^2 t^{-2} - N^2 t^{-1} + N t^{-1}) - N t^{-2} . \tag{7}$$

From (6), Theorem 2 follows for the cases $X^2 \neq t - 1$, as already mentioned in Good (1965, page 37).

When $X^2 = t - 1$ we have $\sum n_i^2 = N - Nt^{-1} + N^2 t^{-1}$; and, by the Cauchy–Schwarz inequality, $\sum n_i \sum n_i^3 \geqq (\sum n_i^2)^2$. Therefore, from (7), after some elementary algebra, we see that

$$6a_3 \geqq N(N - 1)(t - 1)t^{-2}$$

and this completes the proof of Theorem 2.

PROOF OF THEOREM 3. The theorem is equivalent to the statement that the equation $F'(x) = 0$ has at most one positive root. But, by (2), we have

$$\frac{F'(x)}{F(x)} = \sum_{i=1}^{t} \sum_{j=0}^{n_i - 1} \frac{1}{x + 1} - \sum_{j=0}^{N-1} \frac{1}{x + j/t} \qquad (N > 0) , \tag{8}$$

where summations with minus 1 terms are taken as zero. The right side can be written

$$\int_0^\infty e^{-\lambda x} \phi(\lambda)\, d\lambda$$

where

$$\phi(\lambda) = \sum_i \sum_{j=0}^{n_i - 1} e^{-\lambda j} - \sum_{j=0}^{N-1} e^{-\lambda j/t} .$$

Hence

$$\phi(\lambda)(1 - e^{-\lambda}) = \sum_i (1 - e^{-\lambda n_i}) - (1 - e^{-\lambda N/t})(1 + e^{-\lambda/t} + \cdots + e^{-(t-1)\lambda/t}) .$$

By Pólya and Szegö (1945, page 50) the theorem will follow if we can prove that $\phi(\lambda)$ has at most one change of sign in the interval $0 < \lambda < \infty$. Write $e^{-\lambda/t} = u$ so that we have

$$\Phi(u) = \phi(\lambda)(1 - e^{-\lambda}) = t - 1 - (u + u^2 + \cdots + u^{t-1}) - \sum u^{tn_i} + (u^N + u^{N+1} + \cdots + u^{N+t-1}) .$$

By Descartes' Rule of signs (see, for example, Pólya and Szegö, 1945, page 43), combined with Condition C, it follows that $\Phi(u)$ has at most three roots in $0 < u < \infty$. But $\Phi(1) = \Phi'(1) = 0$ so two of the roots are coincident at $u = 1$. Therefore $\Phi(u)$ has at most one root in $0 < u < 1$ and $\phi(\lambda)$ has at most one in $0 < \lambda < \infty$, as we needed to show.

It can be shown that $\Phi''(1) = tN(t - 1 - X^2)$ which is negative if and only if $X^2 > t - 1$. This again shows that $F'(x) = 0$ has a finite solution if and only if $X^2 > t - 1$.

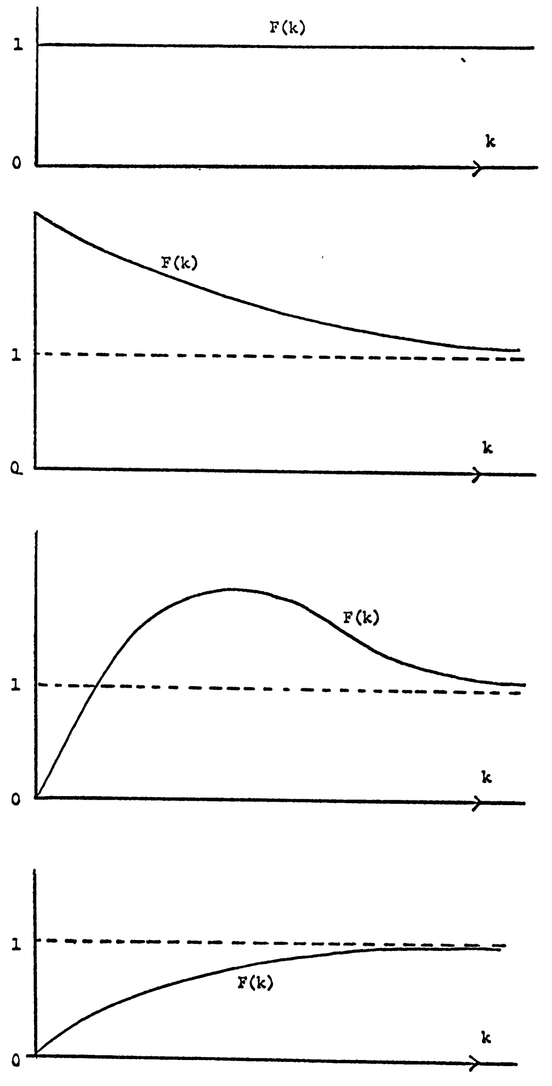

FIG. 1. Graphs of $F(k)$ $(0 \leqq k < \infty)$ under all possible circumstances. (a) $N = 1$ or $t = 1$; (b) $t > 1$, $N > 1$, $n_i = N$ for some i; (c) $n_i \neq N$ for any i; $X^2 > t - 1$; (d) $n_i \neq N$ for any i; $X^2 \leqq t - 1$.

THEOREM 4. *Under Condition* C, *each successive derivative of the weight of evidence against H,*

$$\left(\frac{d}{dk}\right)^{\nu} \log F(k) \qquad\qquad (\nu = 0, 1, 2, 3, \cdots)$$

vanishes for at most one positive value of k.

The derivative can be obtained from the Laplace integral that occurs in the proof of Theorem 3 by replacing $\phi(\lambda)$ by $\lambda^{\nu}\phi(\lambda)$. But this function again vanishes for at most one positive value of λ and the result follows as before.

The case $\nu = 3$ shows that $F'(k)/F(k)$ is convex (from below) to the left of $k_{\max}$. Hence Newton's method for calculating $k_{\max}$, applied to $F'(k)/F(k)$, succeeds when the initial value of k is less than $k_{\max}$.

I am indebted to Bruce Levin and James Reeds for pointing out an error in my original proof of Theorem 3. In addition they appear to have found another proof together with some extensions.

REFERENCES

GOOD, I. J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods.* MIT Press.

GOOD, I. J. (1967). A Bayesian significance test for multinomial distributions. *J. Roy. Statist. Soc. Ser. B* **29** 399–431.

GOOD, I. J. and CROOK, J. F. (1972). The Bayes/non-Bayes compromise and the multinomial distribution. Submitted to *J. Amer. Statist. Assoc.*

PÓLYA, G. and SZEGÖ, G. (1945). *Aufgaben und Lehrsatze aus der Analysis.* Dover, New York. (Originally published in 1925.)

WHITTAKER, E. T. and WATSON, G. N. (1935). *A Course of Modern Analysis*, 4th ed. University Press, Cambridge.

DEPARTMENT OF STATISTICS
VIRGINIA POLYTECHNIC INSTITUTE
AND STATE UNIVERSITY
BLACKSBURG, VIRGINIA 24061