The Annals of Statistics
1990, Vol. 18, No. 3, 1295–1316

LOWER BOUNDS ON BAYES FACTORS FOR MULTINOMIAL DISTRIBUTIONS, WITH APPLICATION TO CHI-SQUARED TESTS OF FIT[1]

BY MOHAN DELAMPADY AND JAMES O. BERGER

University of British Columbia and Purdue University

Lower bounds on Bayes factors in favor of the null hypothesis in multinomial tests of point null hypotheses are developed. These are then applied to derive lower bounds on Bayes factors in both exact and asymptotic chi-squared testing situations. The general conclusion is that the lower bounds tend to be substantially larger than P-values, raising serious questions concerning the routine use of moderately small P-values (e.g., 0.05) to represent significant evidence against the null hypothesis.

1. Introduction.

1.1. *Overview.* Lower bounds on Bayes factors (and posterior probabilities) in favor of point null hypotheses, H_0, have been discussed in Edwards, Lindman and Savage (1963), Dickey (1977), Good (1950, 1958, 1967), Berger (1985), Berger and Sellke (1987), Casella and Berger (1987), Berger and Delampady (1987) and Delampady (1986, 1989a, b) among others. The startling feature of these results is that they establish that the Bayes factor and posterior probability of H_0 are generally substantially larger than the P-value. When such is the case, the interpretation of P-values as measures of evidence against H_0 requires great care. (Other references concerning the relationship between P-values and Bayes factors can be found in the above articles.)

One common rejoinder is that P-values are valuable when there are no alternatives explicitly specified, as is commonly the case in tests of fit. Without alternatives, calculation of Bayes factors or posterior probabilities is impossible. The ultimate goal of this paper is to address this issue for a particularly common test of fit, the chi-squared test of fit. It will be observed that alternatives implicitly do exist, which allow for the computation of lower bounds on Bayes factors in favor of H_0 and posterior probabilities of H_0. These lower bounds will be seen to be much larger than the corresponding P-values.

Lower bounds on Bayes factors are also of interest from Bayesian and likelihood viewpoints. They provide bounds on the amount of evidence for the null hypothesis, in a Bayes factor or weighted likelihood ratio sense, that

---

Received February 1987; revised May 1989.

[1]Research supported by NSF grants DMS-84-01996 and DMS-87-02620, and Natural Sciences and Engineering Research Council (Canada) grant A9250.

AMS 1980 *subject classifications.* Primary 62A15; secondary 62F15.

Key words and phrases. Conjugate densities, unimodal spherically symmetric densities, P-values, point null hypotheses, tests of fit.

depend only on the general class of priors being considered, and not on a specific prior distribution or likelihood "weight function."

In developing the results for the chi-squared test of fit, it is first necessary to deal with testing of point null hypotheses in multinomial problems. This is the subject of Sections 2 and 3; Section 2 deals with lower bounds for Bayes factors over the class of conjugate priors, and Section 3 with lower bounds over a large class of transformed symmetric priors. Section 5 discusses the chi-squared test of fit. Some comments, comparisons and conclusions are presented in Sections 4 and 6.

1.2. *Notation.* Let $\mathbf{n} = (n_1, n_2, \dots, n_t)'$ be a sample of fixed size $N = \sum_{i=1}^t n_i$ from a t-category multinomial distribution with unknown cell probabilities $\mathbf{p} = (p_1, p_2, \dots, p_t)' \in A$, the t-dimensional simplex. The probability density (mass function) of $\mathbf{n}$ is

$$f(\mathbf{n}|\mathbf{p}) = \frac{N!}{\prod_{i=1}^t n_i!} \prod_{i=1}^t p_i^{n_i}. \tag{1}$$

The problem of interest is to test the hypothesis:

$$H_0\colon \mathbf{p} = \mathbf{p}^0 \quad \text{versus} \quad H_1\colon \mathbf{p} \neq \mathbf{p}^0,$$

where $\mathbf{p}^0$ is a specified interior point of A. The classical multinomial test has P-value

$$P = P_{\mathbf{p}=\mathbf{p}^0}\big(\mathbf{y}\colon f(\mathbf{y}|\mathbf{p}^0) \le f(\mathbf{n}|\mathbf{p}^0)\big). \tag{2}$$

However, this being difficult to calculate, the most popular approach is to use the chi-squared approximation, $P(\chi^2_{t-1} \ge S_N)$, where

$$S_N = \sum_{i=1}^t \frac{(n_i - Np_i^0)^2}{Np_i^0}, \tag{3}$$

and χ^2_m represents a chi-squared random variable with m degrees of freedom.

Approaching the testing problem from the Bayesian viewpoint, assume that π is a prior distribution on A which assigns mass π_0 to $\{\mathbf{p}^0\}$ and $1 - \pi_0$ to $\{\mathbf{p} \neq \mathbf{p}^0\}$ and such that the conditional density with respect to Lebesgue measure on $\{\mathbf{p} \neq \mathbf{p}^0\}$ is $g(\mathbf{p})$. Define $g(\mathbf{p}^0) = 0$ and

$$m_g(\mathbf{n}) = \int_A f(\mathbf{n}|\mathbf{p}) g(\mathbf{p})\, d\mathbf{p},$$

which we assume to be positive for all $\mathbf{n}$. The quantities of interest are then

1. the Bayes factor of H_0 relative to H_1:

$$B^g(\mathbf{n}) = \frac{f(\mathbf{n}|\mathbf{p}^0)}{m_g(\mathbf{n})};$$

2. the posterior probability of H_0:

$$P^{\pi}(H_0|\mathbf{n}) = \left[1 + \frac{(1-\pi_0)}{\pi_0}\frac{1}{B^g(\mathbf{n})}\right]^{-1}.$$

$B^g(\mathbf{n})$ is also of interest from the likelihood viewpoint, since it is the ratio of the likelihood of H_0 to the average or weighted likelihood of H_1, the averaging being with respect to the "weight function" g.

Of interest is that lower bounds on $B^g(\mathbf{n})$ [and hence $P^{\pi}(H_0|\mathbf{n})$] can be found for important classes of densities g, and that these lower bounds tend to be surprisingly large. If G is a class of densities g under consideration, we will consider the lower bounds

$$\underline{B}_G(\mathbf{n}) = \inf_{g\in G} B^g(\mathbf{n})$$

and

$$\underline{P}_G(H_0|\mathbf{n}) = \inf_{g\in G} P^{\pi}(H_0|\mathbf{n}) = \left[1 + \frac{(1-\pi_0)}{\pi_0}\frac{1}{\underline{B}_G}\right]^{-1}.$$

We will only present results in terms of $\underline{B}_G$, since this determines $\underline{P}_G(H_0|\mathbf{n})$ once π_0, the prior probability of H_0, has been specified.

1.3. *Choice of G.* A Bayesian might restrict G to a single distribution, g_0. A robust Bayesian might restrict g to a small class of densities, say, those in a neighborhood of some g_0 [cf. Berger and Berliner (1986) and Sivaganesan and Berger (1989)]. But any such restrictions require specific subjective input. Of interest to Bayesians and non-Bayesians alike are choices of G which require only general shape specifications concerning G. Two such possibilities are

(4) $G_{\mathrm{CU}} = \{g \text{ which are conjugate to } f(\mathbf{n}|\mathbf{p}) \text{ and such that } E^g[\mathbf{p}] = \mathbf{p}^0\}$,

(5) $G_{\mathrm{US}} = \{\text{unimodal } g, \text{ symmetric about } \mathbf{p}^0\}$

(where "symmetric about $\mathbf{p}^0$" will be defined in Section 3).

The appeal of these two classes of densities is that they seem to be somewhat objective classes. They acknowledge the central role of $\mathbf{p}^0$, and seek to spread out the prior mass around $\mathbf{p}^0$ in ways that are not biased toward particular alternatives. Many other classes could be considered; a detailed study of a number of such classes in Berger and Delampady (1987) (for the binomial case) indicates that G_{CU} and G_{US} are quite representative, and also satisfactory in terms of being neither too big nor too small. (It might appear that G_{CU} is too small, typically including only a small dimensional class of distributions; that similar results are obtained for G_{US} should allay such fears.) Further justifications for the use of G_{CU} may be found in Good and Crook (1974) (and references therein) who cite work by Johnson (1932) in the special case where $p_i^0 = 1/t$, $1 \le i \le t$. Use of G_{CU} is considered in Section 2, and use of G_{US} in Section 3.

Additional discussion of the multinomial testing problem with mixtures of conjugate priors can be found in Good (1965, 1967, 1975). Edwards, Lindman and Savage (1963) discuss the possibility of finding $\underline{B}_{\mathrm{CU}}$ for the binomial problem. Extensive discussion of the binomial problem can be found in Delampady (1986) and Berger and Delampady (1987).

2. Bounds for conjugate priors in multinomial testing.

2.1. *Introduction.* For the multinomial distribution, $f(\mathbf{n}|\mathbf{p})$ in (1), the Dirichlet densities form the usual conjugate family. The density of the Dirichlet distribution with parameters $\mathbf{k} = (k_1, k_2, \dots, k_t)'$ is

$$g_{\mathbf{k}}(\mathbf{p}) = \frac{\Gamma(\sum_{i=1}^{t} k_i)}{\prod_{i=1}^{t}\Gamma(k_i)} \prod_{i=1}^{t} p_i^{k_i - 1}, \qquad k_i > 0, i = 1, 2, \dots, t; \mathbf{p} \in A.$$

The mean of $g_{\mathbf{k}}$ is the vector $(\sum_{i=1}^{t} k_i)^{-1}\mathbf{k}$, which equals $\mathbf{p}^0 = (p_1^0, p_2^0, \dots, p_t^0)'$ only if $\mathbf{k} = c\mathbf{p}^0$ for some $c > 0$. Thus, in testing H_0: $\mathbf{p} = \mathbf{p}^0$ versus H_1: $\mathbf{p} \neq \mathbf{p}^0$, the class of conjugate densities with mean equal to $\mathbf{p}^0$ is

$$G_{\mathrm{CU}} = \{g_{\mathbf{k}} : \mathbf{k} = c\mathbf{p}^0, c > 0\}. \tag{6}$$

For convenience, define

$$\underline{B}_{\mathrm{CU}}(\mathbf{n}) = \underline{B}_{G_{\mathrm{CU}}}(\mathbf{n}) = \inf_{g \in G_{\mathrm{CU}}} B^g(\mathbf{n}).$$

2.2. *Exact results.* For the conjugate priors $g_{\mathbf{k}}$,

$$\begin{aligned} m_{g_{\mathbf{k}}}(\mathbf{n}) &= \int_A f(\mathbf{n}|\mathbf{p}) g_{\mathbf{k}}(\mathbf{p})\, d\mathbf{p} \\ &= \frac{N!}{\prod_{i=1}^{t} n_i!} \frac{\Gamma(\sum_{i=1}^{t} k_i)}{\prod_{i=1}^{t}\Gamma(k_i)} \frac{\prod_{i=1}^{t}\Gamma(n_i + k_i)}{\Gamma(N + \sum_{i=1}^{t} k_i)}. \end{aligned} \tag{7}$$

The following result is an immediate consequence.

THEOREM 1. *The lower bound on the Bayes factor over* G_{CU} *is given by*

$$\underline{B}_{\mathrm{CU}}(\mathbf{n}) = \inf_{c>0} \frac{\Gamma(c + N)\prod_{i=1}^{t}\{\Gamma(cp_i^0)(p_i^0)^{n_i}\}}{\Gamma(c)\prod_{i=1}^{t}\Gamma(n_i + cp_i^0)}. \tag{8}$$

The minimization in (8) can easily be carried out numerically. This is because, as Good (1965) conjectured and Levin and Reeds (1977) proved,

$$\frac{\Gamma(c)\prod_{i=1}^{t}\Gamma(n_i + cp_i^0)}{\Gamma(c + N)\prod_{i=1}^{t}\Gamma(cp_i^0)}$$

is unimodal, and has its maximum at a finite c if

$$S_N > t - 1 + \sum_{i=1}^{t} \frac{n_i/N - p_i^0}{p_i^0} \tag{9}$$

and at $c = \infty$ otherwise [where S_N is as defined in (3)].

For selected values of t, $\mathbf{p}^0$ and $\mathbf{n}$, $\underline{B}_{\mathrm{CU}}$ is tabulated in Tables 2 and 3 along with the corresponding P-values. These tables are given in Section 4 to which we defer discussion of the results.

2.3. *Asymptotic results.* As $N \to \infty$, the behavior of $\underline{B}_{\mathrm{CU}}(\mathbf{n})$ is given in the following theorem. For use in theorem, define $\Omega_{N,K} = \{\mathbf{n}\colon 0 < S_N \le K\}$.

THEOREM 2. *For every* $K > 0$,

$$\lim_{N\to\infty} \sup_{\Omega_{N,K}} \left| \underline{B}_{\mathrm{CU}}(\mathbf{n}) - \underline{B}_{\mathrm{CU}}^{*}(S_N) \right| = 0,$$

where

$$\underline{B}_{\mathrm{CU}}^{*}(v) = \inf_{1>a>0} a^{1-t} \exp\left(-\tfrac{1}{2}(1-a^2)v\right).$$

PROOF. See the Appendix. □

$\underline{B}_{\mathrm{CU}}^{*}$ is the bound obtained from the following normal problem. Let $\mathbf{X} = (X_1, X_2, \dots, X_{t-1})' \sim N_{t-1}(\boldsymbol{\theta}, I)$, and suppose that it is desired to test H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus H_1: $\boldsymbol{\theta} \ne \boldsymbol{\theta}_0$. Let G be the class of all unimodal densities, spherically symmetric about the vector $\boldsymbol{\theta}_0$. Then, the lower bound on the Bayes factor for this problem, over the class G, is precisely $\underline{B}_{\mathrm{CU}}^{*}(\|\mathbf{X}\|^2)$, as is proved in Delampady (1986). This lower bound, calculated for a number of different dimensions, is displayed in Table 1 in Section 3; discussion is deferred to that section. For Table 1 we have chosen values of $\|\mathbf{X}\|^2$ equal to the $1-P$ quantile of a chi-squared random variable with $t-1$ degrees of freedom, P being certain common P-values; for convenience of comparison, the table is given in terms of these P-values, instead of $\|\mathbf{X}\|^2$.

3. Bounds for symmetric priors: Multinomial testing

3.1. *Introduction.* When $\mathbf{p}^0 = (t^{-1}, \dots, t^{-1})'$, it is not difficult to define symmetry for conditional prior densities g. For general $\mathbf{p}^0$, a natural way to obtain a notion of symmetry is to consider symmetry in a suitable transformation of the parameter $\mathbf{p}$, such as in $\mathbf{u}(\mathbf{p})$ defined as follows.

Let $D(\mathbf{p}^a)$ be the diagonal matrix with ith diagonal element equal to p_i^a, $i < t$, and define $\phi(\mathbf{p}) = (\sqrt{p_1}, \sqrt{p_2}, \dots, \sqrt{p_{t-1}})'$. Then the covariance matrix of the first $t-1$ free coordinates of $\mathbf{n}$ is $\Sigma = ND(\mathbf{p}^{1/2})(I_{t-1} - \phi(\mathbf{p})\phi(\mathbf{p})')D(\mathbf{p}^{1/2})$, where I_k is the $k \times k$ identity matrix. Define $B(\mathbf{p})$ by

$$B(\mathbf{p}) = \left(I_{t-1} + \frac{1}{\sqrt{p_t} + p_t} \phi(\mathbf{p})\phi(\mathbf{p})' \right) D(\mathbf{p}^{-1/2}).$$

Note that $NB(\mathbf{p})B(\mathbf{p})' = \hat{\Sigma}$. Finally, denoting the first $t-1$ coordinates of $\mathbf{p} - \mathbf{p}^0$ by $[\mathbf{p} - \mathbf{p}^0]_*$, define $\mathbf{u}(\mathbf{p})$ as

$$\mathbf{u}(\mathbf{p}) = B(\mathbf{p})([\mathbf{p} - \mathbf{p}^0]_*) = \left(\frac{p_1 - p_1^0}{\sqrt{p_1}}, \ldots, \frac{p_{t-1} - p_{t-1}^0}{\sqrt{p_{t-1}}} \right) + \left(\frac{p_t^0 - p_t}{\sqrt{p_t} + p_t} \right) \left(\sqrt{p_1}, \ldots, \sqrt{p_{t-1}} \right).$$

The reasons for considering the transformation $\mathbf{u}(\mathbf{p})$ are as follows:

1. The range of $\mathbf{u}(\mathbf{p})$ is R^{t-1}.
2. The likelihood function of $\mathbf{u}(\mathbf{p})$ is considerably more symmetric than that of $\mathbf{p}$ when $\Sigma(p_i^0 - 1/t)^2$ is large. Further, it is approximately normal with covariance matrix I_{t-1} in a neighborhood of 0 (i.e., for $\mathbf{p}$ near $\mathbf{p}^0$).
3. Since $\mathbf{u}(\mathbf{p})$ can be written in closed form, calculations are greatly simplified.

Since $\mathbf{u}$ is approximately normal about 0 with range R^{t-1}, it is natural to define a class of "symmetric, unimodal" priors in $\mathbf{u}$ by (letting * denote the transformed problem)

$$G_{\mathrm{US}}^* = \{\text{unimodal } g^*(\mathbf{u}) \text{ which are spherically symmetric about } \mathbf{0}\}.$$

Transforming back to the original parameter yields the class ("TUS" standing for "transformed unimodal symmetric")

$$G_{\mathrm{TUS}} = \left\{ g(\mathbf{p}) = g^*(\mathbf{u}(\mathbf{p})) \left| \frac{\partial \mathbf{u}(\mathbf{p})}{\partial \mathbf{p}} \right| : g^* \text{ is unimodal and symmetric about } \mathbf{0} \right\}. \tag{10}$$

The term $|\partial \mathbf{u}(\mathbf{p})/\partial \mathbf{p}|$ is merely the Jacobian of the transformation. In calculations it is most convenient to work directly with $\mathbf{u}$ and G_{US}^*, however, so calculation of the Jacobian is not needed.

Note that there were several somewhat arbitrary choices made above, in arriving at G_{TUS}. The first was the transformation [to $\mathbf{u}(\mathbf{p})$]. Other transformations to approximate normality could have been chosen, but the above choice was easy to implement and is sensible. Also, the answers are not likely to vary much for alternative transformations, as indicated in Delampady (1986) for the binomial distribution.

In contrast, the second significant choice above, that of spherical symmetry of the prior for the transformed variable, does matter. Since $\mathbf{u}$ is approximately $N_{t-1}(\mathbf{0}, I_{t-1})$, specification of spherical symmetry in the prior is natural, but very different answers can be obtained if, say, elliptical symmetry is specified instead. Indeed, one could consider choosing $g^*(\mathbf{u}) = h(\mathbf{u}'\Lambda\mathbf{u})$, for Λ other than the identity matrix. It can be shown that the choice of Λ that minimizes the Bayes factor is the singular choice such that $g^*(\mathbf{u})$ is concentrated on the line $\mathbf{u}(N^{-1}\mathbf{n})$, while the Bayes factor is maximized by any choice such that $g^*(\mathbf{u})$ is concentrated in the perpendicular plane to this line (at $\mathbf{0}$). Achieving the absolute minimum, by allowing g^* to concentrate on the "least favorable" (to H_0) line, seems unappealing, especially because there is already a substantial bias against H_0 in the calculation of $\underline{B}$ (namely, the minimization over *all* unimodal g^*). Utilization of spherical symmetry in $\mathbf{u}$ to construct the prior is also reminiscent of the classical use of invariance to perform multivariate testing. [See also Delampady (1989a).]

3.2. *Exact results.* The following theorem gives the lower bound on the Bayes factor over all conditional densities g in G_{TUS}.

THEOREM 3.

$$\underline{B}_{\mathrm{TUS}}(\mathbf{n}) = \inf_{g \in G_{\mathrm{TUS}}} B^g(\mathbf{n}) = f(\mathbf{n}|\mathbf{p}^0) \Big/ \left\{ \sup_r \frac{1}{V(r)} \int_{\|\mathbf{u}\| \le r} l(\mathbf{u})\, d\mathbf{u} \right\}, \tag{11}$$

where $V(r)$ is the volume of a sphere of radius r,

$$l(\mathbf{u}) = \frac{N!}{\prod_{i=1}^t n_i!} \prod_{i=1}^t p(\mathbf{u})_i^{n_i},$$

and $\mathbf{p}(\mathbf{u})$ is the inverse function of $\mathbf{u}(\mathbf{p})$.

PROOF. A change of variable yields

$$\begin{aligned} \sup_{g \in G_{\mathrm{TUS}}} m_g(\mathbf{n}) &= \sup_{g \in G_{\mathrm{TUS}}} \int_A \frac{N!}{\prod_{i=1}^t n_i!} \prod_{i=1}^t p_i^{n_i} g(\mathbf{p})\, d\mathbf{p} \\ &= \sup_{h \in G_{\mathrm{US}}} \int l(\mathbf{u}) h(\mathbf{u})\, d\mathbf{u}. \end{aligned} \tag{12}$$

The conclusion follows from the standard result that the class of all unimodal spherically symmetric distributions can be represented as the class of all convex mixtures of uniform distributions over balls $B(r) = \{\mathbf{u}\colon \|\mathbf{u}\| \le r\}$, so that a linear functional of h, such as the integral in (12), will be maximized over the uniform distributions on $B(r)$. □

For selected t, $\mathbf{n}$ and $p_i^0 = 1/t$, $\underline{B}_{\mathrm{TUS}}$ is tabulated in Tables 2 and 3 in Section 4, along with the corresponding P-values. We defer discussion until then, but it is useful, for calculating the integral in (11), to record that the inverse function $\mathbf{p}(\mathbf{u})$ is given by

$$p_i(\mathbf{u}) = \left\{ \frac{u_i + \left(u_i^2 + 4p_i^0 H(p_t)\right)^{1/2}}{2H(p_t)} \right\}^2, \qquad i = 1, \dots, t-1,$$

where $H(p_t) = (1 + p_t^0/\sqrt{p_t})/(1 + \sqrt{p_t})$ and $p_t = p_t(\mathbf{u})$ is the solution to $(1 - p_t) = \sum_{i=1}^{t-1} p_i(\mathbf{u})$.

3.3. *Asymptotic results.* The calculation in (11) can be difficult if t is large. Hence an asymptotic approximation for large N is given in Theorem 4. Again, $\Omega_{N,K} = \{\mathbf{n}: 0 < S_N \le K\}$.

THEOREM 4. *For every $K > 0$,*

$$\lim_{N \to \infty} \sup_{\Omega_{N,K}} \left| \underline{B}_{\mathrm{TUS}}(\mathbf{n}) - \underline{B}^*_{\mathrm{US}}(S_N) \right| = 0,$$

where

$$\underline{B}^*_{\mathrm{US}}(v) = \frac{(2\pi)^{-(t-1)/2} \exp(-v/2)}{\sup_r (1/V(r)) P(Y \le r^2)}, \tag{13}$$

Y having a noncentral chi-squared distribution with $t-1$ degrees of freedom and noncentrality parameter v.

PROOF. See the Appendix. Note that some care is needed in establishing the result, since $\underline{B}_{\mathrm{TUS}}$ involves an infimum over g; this infimum is *inside* the limit as $N \to \infty$, so asymptotics cannot just be applied directly to the individual B^g. □

Note that $\underline{B}^*_{\mathrm{US}}(\|\mathbf{X}\|^2)$ in (13) is the lower bound on B^g over the unimodal and spherically symmetric class, G_{US}, of conditional prior densities for $\boldsymbol{\theta}$ that would be obtained in the multivariate normal problem discussed at the end of Section 2.3. Table 1 presents values of $\underline{B}^*_{\mathrm{US}}$ for a range of t and $\|\mathbf{X}\|^2$ corresponding to certain common P-values. (For a given P-value, P, the corresponding value of $\|\mathbf{X}\|^2$ is the $1 - P$ quantile of the chi-squared distribution with $t - 1$ degrees of freedom.) As could be expected, the $\underline{B}^*_{\mathrm{US}}$ are smaller than the $\underline{B}^*_{\mathrm{CU}}$ (the $\underline{B}^*_{\mathrm{US}}$ corresponding to a quite large nonparametric class of priors), but they are reassuringly similar. Though the lower bounds decrease with increasing dimension, the decrease is not dramatic. The main observation to make, of course, is that the entries are substantially larger than the corresponding P-values.

4. Comparisons and conclusions. Tables 2 and 3 tabulate the exact bounds, $\underline{B}_{\mathrm{CU}}$ and $\underline{B}_{\mathrm{TUS}}$, for $t = 3$ and $t = 4$, respectively, with $p_i^0 = 1/t$ and various choices of $N, \mathbf{n}$. Here P denotes the P-value, with "Exact-P" referring to the exact P-value from (2), and "$\chi^2 - P$" referring to the approximate P-value obtained from the chi-squared approximation.

TABLE 1
Asymptotic lower bounds

	P = 0.001		**P = 0.01**		**P = 0.05**		**P = 0.10**	
Dimension = $t - 1$	$\underline{B}^*_{CU}$	$\underline{B}^*_{US}$	$\underline{B}^*_{CU}$	$\underline{B}^*_{US}$	$\underline{B}^*_{CU}$	$\underline{B}^*_{US}$	$\underline{B}^*_{CU}$	$\underline{B}^*_{US}$
1	0.0244	0.0182	0.1538	0.1227	0.4734	0.4092	0.7001	0.6437
2	0.0198	0.0143	0.1247	0.0978	0.4067	0.3481	0.6263	0.5699
3	0.0165	0.0119	0.1142	0.0902	0.3784	0.3259	0.5818	0.5396
4	0.0156	0.0114	0.1064	0.0850	0.3615	0.3141	0.5713	0.5232
5	0.0133	0.0099	0.1020	0.0824	0.3503	0.3072	0.5576	0.5131
6	0.0129	0.0097	0.0988	0.0807	0.3419	0.3023	0.5473	0.5058
7	0.0126	0.0096	0.0964	0.0797	0.3356	0.2990	0.5392	0.5004
8	0.0124	0.0095	0.0945	0.0789	0.3305	0.2963	0.5330	0.4966
9	0.0121	0.0094	0.0929	0.0782	0.3261	0.2942	0.5277	0.4932
10	0.0119	0.0094	0.0916	0.0777	0.3228	0.2927	0.5230	0.4908
15	0.0113	0.0093	0.0859	0.0752	0.3108	0.2875	0.5078	0.4826
20	0.0109	0.0093	0.0833	0.0743	0.3036	0.2844	0.4988	0.4782
30	0.0105	0.0092	0.0803	0.0735	0.2950	0.2809	0.4879	0.4725

TABLE 2
Lower bounds for conjugate and transformed symmetric densities, $t = 3$

Exact-P	$\chi^2 - P$	N	n_1	n_2	$\underline{B}_{CU}$	$\underline{B}_{TUS}$
0.001	0.00	12	10	1	0.0285	0.0077
0.008	0.00	13	10	2	0.0565	0.0236
0.017	0.01	14	10	3	0.0995	0.0517
0.024	0.02	9	7	1	0.2010	0.0999
0.033	0.03	14	9	4	0.2378	0.1405
0.060	0.04	12	8	3	0.3163	0.1895
0.062	0.06	13	8	4	0.3833	0.2064
0.056	0.07	14	8	5	0.4053	0.2877
0.080	0.09	15	7	7	0.4345	0.3101
0.166	0.10	9	6	2	0.6086	0.3899
0.100	0.12	13	7	5	0.5733	0.4366

The first fact to be noted is that $\underline{B}_{CU}$ and $\underline{B}_{TUS}$ differ more here than in the asymptotic normal situation of Table 1. (Comparisons between these tables and Table 1 are best made by comparing entries corresponding to approximately equal P-values.) However, most of the cases in Tables 2 and 3 are extreme, with likelihoods concentrated near the boundary of A, and hence these differences are probably about as large as one would expect to find. Whether one uses $\underline{B}_{CU}$ or $\underline{B}_{TUS}$ is somewhat a matter of taste: $\underline{B}_{CU}$ is probably more representative of typical Bayes factors, while $\underline{B}_{TUS}$ is perhaps more convincing to non-Bayesians since it is based on such a large class of priors. Note that the Table 1 asymptotic bounds seem fairly reasonable as approximations to $\underline{B}_{CU}$ even for these small N, but can be rather poor as approximations to $\underline{B}_{TUS}$ for small N.

TABLE 3
Lower bounds for conjugate and transformed symmetric densities, $t = 4$

Exact-P	$\chi^2 - P$	N	n_1	n_2	n_3	$\underline{B}_{\mathrm{CU}}$	$\underline{B}_{\mathrm{TUS}}$
0.001	0.00	15	11	2	1	0.0124	0.0032
0.003	0.00	12	9	1	1	0.0356	0.0116
0.005	0.00	15	10	3	1	0.0454	0.0171
0.007	0.00	15	10	2	2	0.0660	0.0254
0.008	0.01	14	9	3	1	0.0987	0.0351
0.013	0.01	15	9	4	1	0.1101	0.0465
0.057	0.02	13	8	2	2	0.2749	0.1464
0.025	0.03	15	8	5	1	0.1919	0.0909
0.053	0.03	11	7	2	1	0.3098	0.1644
0.044	0.04	15	7	6	1	0.2506	0.1563
0.045	0.05	14	7	5	1	0.3235	0.2038
0.066	0.05	13	7	4	1	0.3712	0.2269

Finally, we come to the major point, reflected here as well as in Table 1: The "objective" lower bounds on B^g are substantially larger than the P-value. For instance, when $t = 4$, $N = 14$ and $\mathbf{n} = (7, 5, 1, 1)$, the exact P-value is 0.045 ("significant at the 0.05 level"), yet $\underline{B}_{\mathrm{CU}} = 0.323$ and $\underline{B}_{\mathrm{TUS}} = 0.2038$. Thus the data support H_1: $\mathbf{p} \neq (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ *at most* 3 and 5 times, respectively, as much as it supports H_0: $\mathbf{p} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. This would appear to be at most mild evidence against H_0, yet standard practice using P-values would consider the data to be significant evidence against H_0.

5. The chi-squared test of fit. Consider a statistical experiment in which a random sample of size N is observed from a distribution F. The problem is to test the hypothesis

$$H_0\colon F = F_0 \quad \text{versus} \quad H_1\colon F \neq F_0,$$

where F_0 is a specified distribution. The standard test procedure for this problem is the chi-squared test of fit, which first finds the vector $\mathbf{n} = (n_1, \dots, n_t)'$ of frequencies of the N observations in a partition of the sample space consisting of, say, t cells, and then computes the P-value as

$$P = P\left(\chi^2_{t-1} \geq S_N\right),$$

where S_N is as in (3), with p_i^0 being the probability under F_0 of the ith cell in the partition. Reducing the observations to the vector $\mathbf{n}$ of cell frequencies implicitly implies that one is testing H_0: $\mathbf{p} = \mathbf{p}^0$ versus H_1: $\mathbf{p} \neq \mathbf{p}^0$, where $\mathbf{n}$ has a Multinomial$(N, \mathbf{p})$ distribution. Thus we can apply the results of the previous sections to obtain "objective" lower bounds on the Bayes factors.

EXAMPLE. Thirty observations were made on the arrival times of a certain process. It is desired to test the hypothesis that the distribution of the arrival

TABLE 4
Cells and data

i	Cell	n_i	Np_i^0
1	$[0, 0.40)$	16	10
2	$[0.40, 1.01)$	9	10
3	$[1.01, \infty)$	5	10

TABLE 5
P-values and lower bounds

Exact-P	$\chi^2 - P$	$\underline{B}^*_{\mathrm{CU}}$	$\underline{B}^*_{\mathrm{US}}$	$\underline{B}_{\mathrm{CU}}$	$\underline{B}_{\mathrm{TUS}}$
0.058	0.05	0.4489	0.3881	0.3750	0.2922

time, X, is exponential with mean 1, i.e., to test

$$H_0\colon F_0(x) = 1 - \exp(-x), \qquad x > 0.$$

Suppose that it is decided to use a partition with three cells, the cells being chosen so as to have equal probability under H_0. The cells, the observed cell counts and the expected cell counts under H_0 are given in Table 4.

The chi-squared test statistic is $S_N = 6.20$ with two degrees of freedom. The exact P-value [computed by (2) for the multinomial model], the P-value using the chi-squared approximation, the exact lower bounds ($\underline{B}_{\mathrm{CU}}$ and $\underline{B}_{\mathrm{TUS}}$) on the Bayes factor from Theorems 1 and 3, and the asymptotic lower bounds ($\underline{B}^*_{\mathrm{CU}}$ and $\underline{B}^*_{\mathrm{US}}$) from Theorems 2 and 4, are all given in Table 5.

The chi-squared approximation is quite reasonable here and the lower bounds over G_{CU} and G_{TUS} are quite similar. But the differences between the P-value and the lower bounds on the Bayes factors are substantial. The lower bounds on the Bayes factor indicate that the data support H_1 by *at most* a factor of 3 to 1.

6. Comments. General discussion and debate concerning the implications of the discrepancy between Bayes factors and P-values can be found in Berger and Sellke (1987), Casella and Berger (1987), Berger and Delampady (1987) and many references therein. We feel obliged to again raise, however, within the context of this paper, the important qualification that, although the lower bounds $\underline{B}_{\mathrm{CU}}$ and $\underline{B}_{\mathrm{TUS}}$ seem much more useful than P-values, they are just lower bounds. If $\underline{B} = 0.5$, then we can be quite assured that there is no strong reason to reject H_0, but if $\underline{B} = 0.05$ what should be done? After all, this implies only that the Bayes factor is somewhere between 0.05 and ∞ [which Good (1975) shows can be attained if no n_i is equal to N], depending on the choice of g. Furthermore, it has been observed [cf. Jeffreys (1961), Lindley (1957), Good (1967) and Good and Crook (1974)] that actual Bayes factors tend to increase as $\sqrt{N}$ (when the P-value is fixed), while our various $\underline{B}$ do not so depend on N. (The minimizing g^* varies as N varies, in such a way that the dependence on N is removed). To avoid this inappropriate behavior and/or obtain a precise Bayes factor, at least partial specification of a subjective g is required. Reasonable results might often be obtained by fairly crude devices, such as considering only the conjugate $g_{\mathbf{k}}$ in Section 2.1, with $\mathbf{k} = c\mathbf{p}^0$. Then only c needs to be specified to determine the Bayes factor, and this could be done from a subjective estimate of the variability of $\mathbf{p}$ conditional on H_0 being false. Furthermore, one could graph the Bayes factor as a function of c [following the ideas of Dickey (1973)], allowing a wide range of users (with different c) to interpret the data. For the multinomial distribution under consideration, we refer to Good (1975) for related graphs.

APPENDIX

In this Appendix the basic steps of the proofs are given. For details see Delampady and Berger (1989). First we shall prove three technical results before establishing a basic result in Lemma 4. Define $\hat{\mathbf{p}} = N^{-1}\mathbf{n}$. Let $h(\mathbf{p}) = \sum_{i=1}^{t} n_i \log(p_i/p_i^0)$. Assume that $S_N \le K$ for some $K > 0$ in each of the following three lemmas.

LEMMA 1. *For each $B > 0$ and $0 \le \delta < 1/3$, there exists $N_0 = N_0(B, \delta, K)$ such that whenever $N > N_0$ and $\mathbf{p}$ satisfies $\|\mathbf{p} - \mathbf{p}^0\|^2 \le BN^{-1+\delta}$,*

$$h(\mathbf{p}) = \frac{1}{2}S_N - \frac{N}{2}\sum_{i=1}^{t}\frac{(p_i - \hat{p}_i)^2}{p_i^0} + R_N^*(\mathbf{p}), \tag{14}$$

where $R_N^*(\mathbf{p})$ is bounded by $C(B, \delta, K)N^{-(1-3\delta)/2}$, $C(B, \delta, K) > 0$ being a constant.

PROOF. The Taylor expansion of the $\log(\cdot)$ function yields

$$h(\mathbf{p}) = \frac{1}{2}S_N - \frac{N}{2}\sum_{i=1}^{t}\frac{(p_i - \hat{p}_i)^2}{p_i^0} + R_N^*(\mathbf{p}),$$

where $R_N^*(\mathbf{p})$ is the remainder,

$$R_N^*(\mathbf{p}) = \frac{N}{3}\sum_{i=1}^{t}\hat{p}_i\frac{\left[(p_i - p_i^0)/p_i^0\right]^3}{[1 + x_i^*]^3} - \frac{N}{2}\sum_{i=1}^{t}\frac{(\hat{p}_i - p_i^0)(p_i - p_i^0)^2}{(p_i^0)^2},$$

$0 < |x_i^*| < |p_i - p_i^0|/p_i^0$. From $S_N \le K$ and $\|\mathbf{p} - \mathbf{p}^0\|^2 \le BN^{-1+\delta}$ it can be shown that $R_N^*(\mathbf{p})$ is bounded by $C(B, \delta, K)N^{-(1-3\delta)/2}$, $C(B, \delta, K) > 0$ being a constant. □

Fix $K > 0$ and, for each $A > 0$, $D > 0$, $0 < \delta < 1/3$ and integer $N \ge 1$, define

$$\Omega_{A,N,K} = \left\{\mathbf{p}\colon \|\mathbf{p} - \mathbf{p}^0\|^2 \ge 2(K + A)/N\right\},$$

$$\Lambda_{M,N,K,D,\delta} = \left\{\mathbf{p}\colon DN^{-1+\delta} > \|\mathbf{p} - \hat{\mathbf{p}}\|^2 \ge MKN^{-1}\right\}.$$

LEMMA 2. *There exists* $A = A(K) > 0$ *such that, for each* $(N, \mathbf{n})$ *for which* $S_N \le K$,

$$h(\mathbf{p}^0) = 0 > h(\mathbf{p}) \quad \text{if } \mathbf{p} \in \Omega_{A,N,K}. \tag{15}$$

Further, there exists $N_2(D, \delta, K)$ *such that whenever* $N \ge N_2$ *and for each* $M \ge 2$ *such that* $2(M+2)K \le DN^{\delta}$,

$$h(\mathbf{p}) < -(M - 1.5)K/2 \quad \text{if } \mathbf{p} \in \Lambda_{M,N,K,D,\delta}. \tag{16}$$

PROOF. To establish (15) note, first of all, that $h(\mathbf{p})$ is strictly concave in $\mathbf{p}$ with unique maximum at $\mathbf{p} = \hat{\mathbf{p}}$. Second, from (14), $|R_N^*(\mathbf{p})| = |h(\mathbf{p}) - S_N/2 + (N/2)\sum_{i=1}^t (p_i - \hat{p}_i)^2/p_i^0|$ can be made arbitrarily small for fixed $B > 0$ and $\|\mathbf{p} - \mathbf{p}^0\|^2 \le B/N$ and N sufficiently large. Noting that $S_N \le K$ implies $\|\hat{\mathbf{p}} - \mathbf{p}^0\|^2 \le K/N$, we get

$$\|\mathbf{p} - \mathbf{p}^0\|^2 \le 2\|\mathbf{p} - \hat{\mathbf{p}}\|^2 + 2K/N. \tag{17}$$

Therefore, choosing $B = 8K$, we obtain

$$R_1 = \left\{\mathbf{p} \colon \|\mathbf{p} - \mathbf{p}^0\|^2 \le 8K/N\right\} \supseteq R_2 = \left\{\mathbf{p} \colon \|\mathbf{p} - \hat{\mathbf{p}}\|^2 \le 3K/N\right\}$$

for each N. Fix $0 < \varepsilon < K/4$. We can find $N_1 = N_1(K, \varepsilon)$ large such that $|R_N^*(\mathbf{p})| < \varepsilon$ for all $N > N_1$ and $\mathbf{p} \in R_1$. For $\mathbf{p} \in \{\mathbf{p}\colon 2K/N < \|\mathbf{p} - \hat{\mathbf{p}}\|^2 \le 3K/N\}$, from (14) of Lemma 1,

$$h(\mathbf{p}) = \frac{S_N}{2} - \frac{N}{2}\sum_{i=1}^t \frac{(p_i - \hat{p}_i)^2}{p_i^0} + R_N^*(\mathbf{p}) < -\frac{K}{4}.$$

Since $h(\mathbf{p})$ is strictly concave $h(\mathbf{p}) < -K/4$ if $\|\mathbf{p} - \hat{\mathbf{p}}\|^2 > 3K/N$ and $N > N_1$. (17) now shows that $A = 3K$ is a possible choice for A in (15) for $N > N_1$. To establish (15) for $N \le N_1 = N_1(K)$, one may choose a still larger A. Very similar arguments establish (16). □

LEMMA 3. *There exist* $A_2^*(K) > A_1^*(K) > 0$, *which satisfy* $A_1^*(K) < 2^{-[2+(K+2)/\{(t-1)\log(2)\}]}$, *and* $N_3(K)$ *such that whenever* $N \ge N_3$ *and* $0 < c_1 < A_1^*N < A_2^*N \le c < 2A_2^*N$,

$$\begin{aligned} &\frac{\Gamma(c)}{\prod_{i=1}^t \Gamma(cp_i^0)} \frac{\prod_{i=1}^t \Gamma(n_i + cp_i^0)}{\Gamma(N+c)} \\ &\qquad > \frac{\Gamma(c_1)}{\prod_{i=1}^t \Gamma(c_1 p_i^0)} \frac{\prod_{i=1}^t \Gamma(n_i + c_1 p_i^0)}{\Gamma(N+c_1)}. \end{aligned} \tag{18}$$

PROOF. Fix $0 < a_1 < a_2$ and consider, for each N, $c \geq a_2 N$ and $c_1 < a_1 N$. To proceed further, we shall require the following Stirling's approximation:

$$\Gamma(y) = \sqrt{2\pi}\, y^{y-1/2} \exp(-y + D(y)/y), \tag{19}$$

where $D(y)$ is a bounded function as $y \to \infty$ [Feller (1957), page 52]. As N increases, we shall apply this approximation to all the Γ terms that involve n_i, N or c and, also to those terms which involve c_1 whenever $a_1 \log(N) \leq c_1$. We shall now establish (18) by showing that

$$\begin{aligned}\mathrm{LOG}(c, c_1, \mathbf{n}, N) &\equiv \log\left(\frac{\Gamma(c)}{\prod_{i=1}^t \Gamma(cp_i^0)} \frac{\prod_{i=1}^t \Gamma(n_i + cp_i^0)}{\Gamma(N+c)} \right) \\ &\quad - \log\left(\frac{\Gamma(c_1)}{\prod_{i=1}^t \Gamma(c_1 p_i^0)} \frac{\prod_{i=1}^t \Gamma(n_i + c_1 p_i^0)}{\Gamma(N+c_1)} \right) \\ &> 0.\end{aligned}$$

We need to consider two different cases.

CASE 1. Assume $a_1 \log(N) \leq c_1 < a_1 N$. Then, for N large, (19), together with Taylor expansions of the log terms involving n_i and algebra, yield

$$\begin{aligned}\mathrm{LOG}(c, c_1, \mathbf{n}, N) &= -\frac{S_N}{2} \frac{(c/N - c_1/N)}{(1 + c/N)(1 + c_1/N)} \\ &\quad + \frac{t-1}{2}\left[\log\left(\frac{c/N}{c_1/N} \right) - \log\left(\frac{1 + c/N}{1 + c_1/N} \right) \right] \\ &\quad + R^*(c, c_1, \hat{\mathbf{p}}, N),\end{aligned}$$

where $|R^*(c, c_1, \hat{\mathbf{p}}, N)| \leq C_1(K, \mathbf{p}^0)/\log(N)$, for some $C_1 > 0$ and large N. Therefore, there exists $N_3(K)$ such that for $N \geq N_3$, $|R^*| < 1$. Hence,

$$\mathrm{LOG}(c, c_1, \mathbf{n}, N) > -\frac{Ka_2}{1 + a_2} + \frac{t-1}{2}\left\{ \log\left(\frac{a_2}{a_1} \right) - \log(1 + 2a_2) \right\} - 1.$$

Choose $a_2 = 1$, $a_1 = 2^{-x}$ and $x > (K+2)/[(t-1)\log(2)] + 2$. Then

$$\mathrm{LOG}(c, c_1, \mathbf{n}, N) > \tfrac{1}{2}\{(t-1)\log(2)(x-2) - (K+2)\} > 0.$$

CASE 2. $0 < c_1 < a_1 \log(N)$: Since $\Gamma(c_1)/\prod_{i=1}^t \Gamma(c_1 p_i^0) = (\int \prod_{i=1}^t p_i^{c_1 p_i^0 - 1}\, d\mathbf{p})^{-1}$ (integration over the simplex, $\{\mathbf{p}\colon 0 < p_i < 1,\ 1 \leq i \leq t-1,\ 0 < \sum_{i=1}^{t-1} p_i < 1\}$), $(\prod_{i=1}^t \Gamma(c_1 p_i^0)/\Gamma(c_1))^{-1}$ is an increasing function of c_1. Therefore,

$$\sup_{0 < c_1 < a_1 \log(N)} \frac{\Gamma(c_1)}{\prod_{i=1}^t \Gamma(c_1 p_i^0)} = \frac{\Gamma(a_1 \log(N))}{\prod_{i=1}^t \Gamma(a_1 \log(N) p_i^0)}.$$

Hence, replacing all the c_1 that appear in $\mathrm{LOG}(c, c_1, \mathbf{n}, N)$ unaccompanied by either n_i or N by $a_1 \log(N)$, and using (19), we get

$$\begin{aligned}
\mathrm{LOG}(c, c_1, \mathbf{n}, N) > {} & \frac{t-1}{2}[\log(c) - \log(a_1 \log(N))] \\
& + (a_1 \log(N) - c_1) \sum_{i=1}^{t} p_i^0 \log(p_i^0) \\
& - \frac{S_N}{2} \frac{(c/N - c_1/N)}{(1 + c/N)(1 + c_1/N)} \\
& - \frac{t-1}{2}[\log(1 + c/N) - \log(1 + c_1/N)] - 1 \\
> {} & \log(N) \left\{ \frac{t-1}{4} - a_1 \sum_{i=1}^{t} p_i^0 \log(1/p_i^0) \right\} \\
& - \frac{K a_2}{1 + a_2} - \frac{t-1}{2} \log(1 + 2a_2) - 1,
\end{aligned}$$

for N large enough and $\log(\log(N)) < \log(N)/2$. Choose

$$a_2 = 1 \quad \text{and} \quad a_1 < \frac{t-1}{8 \sum_{i=1}^{t} p_i^0 \log(1/p_i^0)}.$$

Now choose $N_4(K)$ such that for all $N \geq N_4$, $\frac{1}{2}(K + (t-1)\log(3) + 2) < (t-1)\log(N)/8$. Then $\mathrm{LOG}(c, c_1, \mathbf{n}, N) > 0$, whenever $N \geq N_4$ and $0 < c_1 < a_1 \log(N) < a_1 N < a_2 N \leq c < 2a_2 N$. The result follows by choosing $A_2^* = 1$ and A_1^* to be the minimum of the a_1 chosen in Cases 1 and 2. □

Now we shall prove the following key lemma (this version of which was suggested by the editor and a referee) which proves that only contiguous alternatives need be considered.

LEMMA 4. *For each $K > 0$ there exist constants $B_K^* > 0$ and $Q_K^* > 0$, $Q_K^*/(1 + Q_K^*) < (t-1)/K$, such that, whenever $S_N \leq K$, the following hold*:

(i) $\sup_{g \in G_{\mathrm{TUS}}} m_g(\mathbf{n})$ *is attained at g^* for which $P^{g^*}(\|\mathbf{u}(\mathbf{p})\|^2 \leq B_K^*/N) = 1$;*
(ii) $\sup_{g \in G_{\mathrm{CU}}} m_g(\mathbf{n})$ *is attained at g_{c^*}, where $c^* \geq Q_K^* N$.*

PROOF. It is convenient to divide $m_g(\mathbf{n})$ by $f(\mathbf{n}|\mathbf{p}^0)$ in the above supremum. Now

$$\sup_{g \in G} m_g(\mathbf{n})/f(\mathbf{n}|\mathbf{p}^0) = \sup_{g \in G} \int \exp\left(\sum_{i=1}^{t} n_i \log(p_i/p_i^0) \right) g(\mathbf{p})\, d\mathbf{p}.$$

PROOF OF PART (I). We want to maximize $\int \exp(h(\mathbf{p})) g(\mathbf{p})\, d\mathbf{p}$ over $g \in G_{\mathrm{TUS}}$. From (15) of Lemma 2 and the fact that $\|\mathbf{p} - \mathbf{p}^0\|^2 < 2(K + A(K))/N$ implies $\|\mathbf{u}(\mathbf{p})\|^2 \leq B(K)/N$, for some $B(K) > 0$, it can be shown that there exists $B(K) > 0$ for which

$$h(\mathbf{p}) < 0 \quad \text{whenever } \|\mathbf{u}(\mathbf{p})\|^2 \geq B(K)/N.$$

Now, note that any g that gives mass to the region $\{\mathbf{p}: \|\mathbf{u}(\mathbf{p})\|^2 \geq B(K)/N\}$ obtains a smaller value for $\int \exp(h(\mathbf{p}))g(\mathbf{p})\,d\mathbf{p}$ than its unnormalized restriction to $\{\mathbf{p}: \|\mathbf{u}(\mathbf{p})\|^2 < B(K)/N\}$ with the remaining mass assigned to the point $\mathbf{p}^0$. Since, for every $g \in G_{\mathrm{TUS}}$, its unnormalized restriction to $\{\mathbf{p}: \|\mathbf{u}(\mathbf{p})\|^2 < B/N\}$ belongs to G_{TUS} and $\mathbf{u}(\mathbf{p}^0) = 0$ we need only consider those g that assign all their mass to this set.

PROOF OF PART (II). Note that

$$\frac{\int \exp(h(\mathbf{p}))g_c(\mathbf{p})\,d\mathbf{p}}{\int \exp(h(\mathbf{p}))g_{c_1}(\mathbf{p})\,d\mathbf{p}} = \left\{\frac{\Gamma(c)}{\prod_i \Gamma(cp_i^0)} \frac{\prod_i \Gamma(n_i + cp_i^0)}{\Gamma(N+c)}\right\} \Bigg/ \left\{\frac{\Gamma(c_1)}{\prod_i \Gamma(c_1 p_i^0)} \frac{\prod_i \Gamma(n_i + c_1 p_i^0)}{\Gamma(N+c_1)}\right\}.$$

From (18) of Lemma 3, there exist constants $A_1^*(K) < A_2^*(K)$ and $N_4(K)$ such that whenever $N \geq N_4$ and $0 < c_1 < A_1^* N < A_2^* N \leq c < 2A_2^* N$, the ratio above is larger than 1. Thus, given any $c_1 < A_1^* N$, one can choose a c between $A_2^* N$ and $2A_2^* N$ such that

$$\int \exp(h(\mathbf{p}))g_c(\mathbf{p})\,d\mathbf{p} > \int \exp(h(\mathbf{p}))g_{c_1}(\mathbf{p})\,d\mathbf{p}.$$

It follows that we can restrict attention to g_c, $c \geq A_1^* N$. Lemma 4 proves, in addition, that $A_1^* < 2^{-[2+(K+2)/\{(t-1)\log(2)\}]}$, from which it follows that $A_1^*/(1 + A_1^*) < (t-1)/K$. The result follows if we choose $Q_K^* = A_1^*$. □

PROOF OF THEOREM 2. From (8),

$$\left[\underline{B}_{G_{\mathrm{CU}}}\right]^{-1} = \sup_{a>0} \frac{\Gamma(a)}{\prod_{i=1}^t \Gamma(ap_i^0)} \frac{\prod_{i=1}^t \Gamma(n_i + ap_i^0)}{\Gamma(N+a)\prod_{i=1}^t (p_i^0)^{n_i}}.$$

Since we need only consider $a > Q_K^* N$, from (ii) of Lemma 4, we can apply Stirling's approximation (19) to all the Γ terms. Let $n_i = Np_i^0 + b_i$, $\sum_{i=1}^t b_i = 0$. Then

$$\left[\underline{B}_{G_{\mathrm{CU}}}\right]^{-1} = \sup_{a > Q_k^* N} \left(\frac{a}{N+a}\right)^{(t-1)/2} \prod_{i=1}^t \left(1 + \frac{b_i}{(N+a)p_i^0}\right)^{(N+a)p_i^0 + b_i - 1/2} \times \exp\left(D^*(a, N, \mathbf{n})\left(\frac{1}{a} + \frac{1}{N}\right)\right),$$

where, $D^*(a, N, \mathbf{n})$, from (19), is a bounded function as N increases.

STEP 1. Using the Taylor expansion of the log function, it can be shown that

$$\log \prod_{i=1}^{t} \left(1 + \frac{b_i}{(N+a)p_i^0}\right)^{(N+a)p_i^0 + b_i - 1/2} = \frac{1}{2}(1-c^2)S_N + C(N,a,\mathbf{n}),$$

where $c = \sqrt{a/(N+a)}$,

$$C(N,a,\mathbf{n}) = -\frac{1}{2(N+a)} \sum_{i=1}^{t} \frac{b_i}{p_i^0}$$
$$+ \frac{1}{4(N+a)^2} \sum_{i=1}^{t} \frac{b_i^2}{(p_i^0)^2} - \sum_{i=1}^{t} \frac{b_i^3}{2[(N+a)p_i^0]^2}$$
$$+ \sum_{i=1}^{t} \left\{ \frac{b_i^3}{3[(N+a)p_i^0]^2} + \frac{b_i^3(b_i - 1/2)}{3[(N+a)p_i^0]^3} \right\} \Big/ (1+x_i^*)^3,$$

and $0 < |x_i^*| < |b_i/\{(N+a)p_i^0\}|$.

STEP 2. For $a > 0$, $|C(N,a,\mathbf{n})| \le C(K)^* N^{-1/2}$, $0 < C(K)^* < \infty$ independent of $a, \mathbf{n}$.

PROOF OF STEP 2. It follows from $\sum_{i=1}^{t} b_i^2/(Np_i^0) = S_N \le K$ that, for each $j \ge 1$, there exists $K_j > 0$ depending only on K and $\mathbf{p}^0$ such that $\sum_{i=1}^{t} |b_i|^j \le K_j N^{j/2}$. Therefore, noting that $|x_i^*| < |b_i/\{(N+a)p_i^0\}| \le K' N^{1/2}/(N+a) \le K' N^{-1/2}$, for some $K' > 0$, the result follows in a straightforward way.

STEP 3. Now, we show that

$$\left| [\underline{B}_{G_{\mathrm{CU}}}]^{-1} \Big/ \left\{ \sup_{1 > c > \sqrt{Q_K^*/(1+Q_K^*)}} c^{t-1} \exp((1-c^2)S_N/2) \right\} - 1 \right| \le G(K) N^{-1/2},$$

where $0 < G(K) < \infty$ is a constant.

PROOF OF STEP 3. Recalling that $c^2 = a/(N+a)$ we have, from Step 1 above and (ii) of Lemma 4, for large N,

$$[\underline{B}_{G_{\mathrm{CU}}}]^{-1} = \sup_{1 > c > \sqrt{Q_K^*/(1+Q_K^*)}} \{ c^{t-1} \exp((1-c^2)S_N/2)$$
$$\times \exp[C(N, Nc^2/(1-c^2), \mathbf{n})$$
$$+ D^*(Nc^2/(1-c^2), N, \mathbf{n})/(Nc^2)]\},$$

where both C and D^* are bounded, C by $C(K)^* N^{-1/2}$, from Step 2, and D^*

by, say, $D'(K)$. Let

$$B(K,N) = \left|C(K)^* N^{-1/2} + D'(K)(1 + Q_K^*)(Q_K^*)^{-1} N^{-1}\right|.$$

Then since, for large N,

$$\sup_{\sqrt{Q_K^*/(1+Q_K^*)} < c < 1} C(N, Nc^2/(1-c^2), \mathbf{n}) + D^*(Nc^2/(1-c^2), N, \mathbf{n})/(Nc^2)$$

$$< B(K,N),$$

$$\left|\left[\underline{B}_{G_{\mathrm{CU}}}\right]^{-1} \Big/ \left\{ \sup_{1 > c > \sqrt{Q_k^*/(1+Q_K^*)}} c^{t-1} \exp\left((1-c^2) S_N/2\right)\right\} - 1\right|$$

$$\leq \max\{|\exp(-B(K,N)) - 1|, |\exp(B(K,N)) - 1|\}$$

$$\leq B(K,N)\exp(B(K,N)) \leq G(K) N^{-1/2},$$

noting that $B(K,N) = O(N^{-1/2})$.

STEP 4. We show here that

$$\left|\left[\underline{B}_{G_{\mathrm{CU}}}\right]^{-1} - \left\{\sup_{1>c>0} c^{t-1} \exp\left((1-c^2) S_N/2\right)\right\}\right| \leq G_1(K) N^{-1/2},$$

where $0 < G_1(K) < \infty$ is a constant.

PROOF OF STEP 4. Let $a_N(c) = c^{t-1} \exp((1-c^2) S_N/2)$. Note that

$$\sup_{0<c<1} a_N(c) = \begin{cases} a_N\left(\sqrt{\dfrac{(t-1)}{SN}}\right) = \left(\sqrt{\dfrac{t-1}{S_N}}\right)^{t-1} \exp\left(\left(1 - \dfrac{t-1}{S_N}\right) S_N/2\right), \\ \qquad\qquad \text{if } t-1 \leq S_N \\ a_N(1) = 1, \qquad \text{if } t-1 > S_N. \end{cases}$$

Further, $S_N \leq K$ implies that

$$\sqrt{(t-1)/S_N} \geq \sqrt{(t-1)/K} > \sqrt{Q_K^*/(1+Q_K^*)}\,.$$

Hence,

$$\sup_{1>c>0} a_N(c) = \sup_{1 > c > \sqrt{Q_K^*/(1+Q_K^*)}} a_N(c).$$

Together with Step 3 this implies that

$$\left|\frac{[\underline{B}_{G_{\mathrm{CU}}}]^{-1}}{\sup_{1>c>0} a_N(c)} - 1\right| = \left|\frac{[\underline{B}_{G_{\mathrm{CU}}}]^{-1} - \sup_{1>c>0} a_N(c)}{\sup_{1>c>0} a_N(c)}\right| \le G(K)N^{-1/2}.$$

Now note that $((t-1)/K)^{t-1} < \sup_{0<c<1} a_N(c) < \exp(K/2)$ if $S_N \ge t-1$, and $\sup_{0<c<1} a_N(c) = 1$ if $S_N \le t-1$. Therefore,

$$\left|[\underline{B}_{G_{\mathrm{CU}}}]^{-1} - \sup_{1>c>0} c^{t-1}\exp\bigl((1-c^2)S_N/2\bigr)\right| \le G(K)\max\{\exp(K/2), 1\}N^{-1/2}.$$

This proves Step 4. The theorem is proved by observing that

$$\left|\underline{B}_{G_{\mathrm{CU}}} - \Bigl(\sup_{1>c>0} c^{t-1}\exp\bigl((1-c^2)S_N/2\bigr)\Bigr)^{-1}\right| \le \underline{B}_{G_{\mathrm{CU}}} K_1^{-1}K_2 N^{-1/2},$$

where $K_1 = \min\{((t-1)/K)^{(t-1)}, 1\}$ and $K_2 = G(K)\max\{\exp(K/2), 1\}$. □

REMARK. Note that when the sequence $(N, \mathbf{n})$ satisfies $S_N < t-1$, for each N, Theorem 2 is very easy to obtain and does not require the above proof. This is simply due to the fact that in that case $\underline{B}^*_{\mathrm{CU}}(S_N) = 1$ and since, for large N, $|\sum_{i=1}^t (n_i/N - p_i^0)/p_i^0| = O(N^{-1/2})$, condition (9) of Levin and Reeds (1977) implies that $\underline{B}_{G_{\mathrm{CU}}}(\mathbf{n}) = 1$.

PROOF OF THEOREM 4. We want to maximize

$$\frac{m_g(\mathbf{n})}{f(\mathbf{n}|\mathbf{p}^0)} = \int \exp\left(\sum_i^t n_i \log\left(\frac{p_i}{p_i^0}\right)\right) g(\mathbf{p})\, d\mathbf{p}$$

over $g \in G_{\mathrm{TUS}}$. If $\|\mathbf{p} - \mathbf{p}^0\|^2 \le A_1/N$, $A_1 > 0$, from (14) of Lemma 1, we get

$$h(\mathbf{p}) = \frac{1}{2}S_N - \frac{N}{2}\sum_{i=1}^t \frac{(p_i - \hat{p}_i)^2}{p_i^0} + R_N^*(\mathbf{p}),$$

where R_N^* is bounded by $C_1(A_1, K)/\sqrt{N}$, $C_1(A_1, K) > 0$ is a constant.

Noting that $u_i(\mathbf{p}) = (p_i - p_i^0)/\sqrt{p_i} + \sqrt{p_i}(p_t - p_t^0)/(\sqrt{p_t} + p_t)$, it can be shown that

$$N\sum_{i=1}^t \frac{(p_i - \hat{p}_i)^2}{p_i^0} = N\sum_{i=1}^{t-1} [u_i(\mathbf{p}) - a_i]^2 + L_N(\mathbf{p}, \hat{\mathbf{p}}),$$

where

$$a_i = \frac{\hat{p}_i - p_i^0}{\sqrt{p_i^0}} + \frac{\sqrt{p_i^0}\,(p_t^0 - \hat{p}_t)}{\sqrt{p_t^0} + p_t^0}$$

so that $\sum_{i=1}^{t-1} a_i^2 = S_N/N$, and the remainder, $L_N(\mathbf{p}, \hat{\mathbf{p}})$, is bounded by $C_2(A_1, K)/\sqrt{N}$, $C_2(A_1, K) > 0$, is a constant.

From (i) of Lemma 4, $\mathbf{u} = \mathbf{u}(\mathbf{p}) = (u_1, u_2, \dots, u_{t-1})'$ has a unimodal symmetric density with support in $\|\mathbf{u}\|^2 \le B_K^*/N$, and the extreme points of this class are uniform distributions on $\|\mathbf{u}\| \le r/\sqrt{N}$ for $r \le \sqrt{B_K^*}$. Therefore [from (10)],

$$\sup_{g \in G_{\mathrm{TUS}}} \frac{m_g(\mathbf{n})}{f(\mathbf{n}|\mathbf{p}^0)}$$

$$= \sup_{g \in G_{\mathrm{TUS}}} \frac{\int_{\|\mathbf{u}\| \le \sqrt{B_K^*/N}} l(\mathbf{u}) g(\mathbf{u})\, d\mathbf{u}}{f(\mathbf{n}|\mathbf{p}^0)}$$

$$= \max_{r \le \sqrt{B_K^*}} \frac{N^{(t-1)/2}}{V(r)} \int_{\|\mathbf{u}\| \le r/\sqrt{N}} \exp\left(\frac{1}{2} S_N - \frac{N}{2} \sum_{i=1}^{t-1} (u_i - a_i)^2 + L_N^*(\mathbf{u}) \right) d\mathbf{u}$$

$$= \max_{r \le \sqrt{B_K^*}} \frac{1}{V(r)} \int_{\|\mathbf{v}\| \le r} \exp\left(\frac{1}{2} S_N - \frac{1}{2} \sum_{i=1}^{t-1} \left(v_i - a_i\sqrt{N}\right)^2 + L_N^*(\sqrt{N}\mathbf{v}) \right) d\mathbf{v},$$

where $V(r)$ is the volume of a sphere of radius r and $L_N^*(\mathbf{u}) = L_N(\mathbf{p}(\mathbf{u}), \hat{\mathbf{p}})$ is bounded by $D_1(B_K^*)/\sqrt{N}$, $D_1 > 0$ depending on B_K^* and $\mathbf{p}^0$ only. Since (i) of Lemma 4 holds with B_K^* replaced by any B, $B > B_K^*$, we have, for any $B > B_K^*$,

$$\sup_{g \in G_{\mathrm{TUS}}} \frac{m_g(\mathbf{n})}{f(\mathbf{n}|\mathbf{p}^0)}$$

$$= \max_{r \le \sqrt{B}} \frac{1}{V(r)} \int_{\|\mathbf{v}\| \le r} \exp\left(\frac{1}{2} S_N - \frac{1}{2} \sum_{i=1}^{t-1} \left(v_i - a_i\sqrt{N}\right)^2 + L_N^*(\sqrt{N}\mathbf{v}) \right) d\mathbf{v}.$$

Since

$$\frac{1}{V(r)} \int_{\|\mathbf{v}\| \le r} \exp\left(\frac{1}{2} S_N - \frac{1}{2} \sum_{i=1}^{t-1} \left(v_i - a_i\sqrt{N}\right)^2 + L_N^*(\sqrt{N}\mathbf{v}) \right) d\mathbf{v}$$

$$= \frac{1}{V(r)} \int_{\|\mathbf{v}\| \le r} \left[\frac{l(\mathbf{v})}{f(\mathbf{n}|\mathbf{p}^0)} \right] d\mathbf{v}$$

$$= \frac{1}{V(r)} \int_{\|\mathbf{v}\| \le r} \exp(h(\mathbf{p}(\mathbf{v})))\, d\mathbf{v},$$

from Lemma 2, for each fixed $\varepsilon > 0$, we can find $R = R(\varepsilon) > B_K^*$ such that whenever $r \ge \sqrt{R}$,

$$\frac{1}{V(r)} \int_{\|\mathbf{v}\| \le r} \exp\left(\frac{1}{2} S_N - \frac{1}{2} \sum_{i=1}^{t-1} \left(v_i - a_i\sqrt{N}\right)^2 + L_N^*(\sqrt{N}\mathbf{v}) \right) d\mathbf{v} < \varepsilon.$$

Now, note that

$$\max_{r>0} \frac{1}{V(r)} \int_{\|\mathbf{v}\| \le r} \exp\left(\frac{1}{2} S_N - \frac{1}{2} \sum_{i=1}^{t-1} \left(v_i - a_i \sqrt{N} \right)^2 \right) d\mathbf{v}$$

$$> (2\pi)^{(t-1)/2} \max_{r>0} \left\{ \frac{\exp(-rK^{1/2}) P(T \le r^2)}{V(r)} \right\}$$

$$> (2\pi)^{(t-1)/2} \frac{\exp(-K^{1/2}) P(T \le 1)}{V(1)} \equiv \text{LOB},$$

where T has a central chi-squared distribution with $t-1$ degrees of freedom. Fix any $\varepsilon < \text{LOB}/2$. Then, for each $B > R = R(\varepsilon)$,

$$\sup_{g \in G_{\text{TUS}}} \frac{m_g(\mathbf{n})}{f(\mathbf{n}|\mathbf{p}^0)}$$

$$= \max\left\{ \max_{r<\sqrt{R}} \frac{1}{V(r)} \int_{\|\mathbf{v}\| \le r} \exp\left(\frac{1}{2} S_N - \frac{1}{2} \sum_{i=1}^{t-1} \left(v_i - a_i \sqrt{N} \right)^2 + L_N^*(\sqrt{N}\mathbf{v}) \right) d\mathbf{v}, \varepsilon \right\}.$$

Noting that $\max_{\|\mathbf{v}\| \le \sqrt{R}} |L_N^*(\sqrt{N}\mathbf{v})| \le C_3(R)/\sqrt{N}$, for some $C_3 > 0$, we conclude that

$$\sup_{g \in G_{\text{TUS}}} m_g(\mathbf{n})/f(\mathbf{n}|\mathbf{p}^0)$$

$$= \exp(S_N/2)(2\pi)^{(t-1)/2} \max_{r>0} \{P(Y \le r^2)/V(r)\}(1 + O(N^{-1/2})),$$

where Y is a noncentral chi-squared random variable with $t-1$ degrees of freedom and noncentrality parameter S_N. □

Acknowledgments. We are very grateful to the editor, who greatly simplified the proofs and strengthened the theorems, and to the referees, who provided extensive sets of valuable suggestions.

REFERENCES

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, New York.

BERGER, J. O. and BERLINER, L. M. (1986). Robust Bayes and empirical Bayes analysis with ε-contaminated priors. *Ann. Statist.* **14** 461–486.

BERGER, J. O. and DELAMPADY, M. (1987). Testing precise hypotheses (with discussion). *Statist. Sci.* **2** 317–348.

BERGER, J. O. and SELLKE, T. (1987). Testing a point null hypothesis: The irreconcilability of significance levels and evidence (with discussion). *J. Amer. Statist. Assoc.* **82** 112–139.

CASELLA, G. and BERGER, R. L. (1987). Reconciling Bayesian and frequentist evidence in the one-sided testing problem (with discussion). *J. Amer. Statist. Assoc.* **82** 106–111.

DELAMPADY, M. (1986). Testing a precise hypothesis: Interpreting p-values from a robust Bayesian viewpoint. Ph.D. thesis, Dept. Statist., Purdue Univ..

DELAMPADY, M. (1989a). Lower bounds on Bayes factors for invariant testing situations. *J. Multivariate Anal.* **28** 227–246.

DELAMPADY, M. (1989b). Lower bounds on Bayes factors for interval null hypotheses. *J. Amer. Statist. Assoc.* **84** 120–124.

DELAMPADY, M. and BERGER, J. (1989). Lower bounds on Bayes factors for multinomial distributions, with application to chi-squared testing. Technical Report, Dept. Statist., Univ. British Columbia.

DICKEY, J. M. (1973). Scientific reporting. *J. Roy. Statist. Soc. Ser. B* **35** 285–305.

DICKEY, J. M. (1977). Is the tail area useful as an approximate Bayes factor. *J. Amer. Statist. Assoc.* **72** 138–142.

EDWARDS, W., LINDMAN, H. and SAVAGE, L. J. (1963). Bayesian statistical inference for psychological research. *Physchol. Rev.* **70** 193–242. Reprinted in *Robustness of Bayesian Analyses.* (J. Kadane, ed.) (1984). North-Holland, Amsterdam.

FELLER, W. (1957). *An Introduction to Probability and Its Applications* **1**, 2nd ed. Wiley, New York.

GOOD, I. J. (1950). *Probability and the Weighing of Evidence.* Charles Griffin, London.

GOOD, I. J. (1958). Significance tests in parallel and in series. *J. Amer. Statist. Assoc.* **53** 799–813.

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

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

GOOD, I. J. (1975). The Bayes factor against equiprobability of a multinomial population assuming a symmetric dirichlet prior. *Ann. Statist.* **3** 246–250.

GOOD, I. J. and CROOK, J. F. (1974). The Bayes/non-Bayes compromise and the multinomial distribution. *J. Amer. Statist. Assoc.* **69** 711–720.

JEFFREYS, H. (1961). *Theory of Probability*, 3rd ed. Clarendon, Oxford.

JOHNSON, W. E. (1932). Probability: The deductive and inductive problems. *Mind* **41** 409–423.

LEVIN, B. and REEDS, J. (1977). Compound multinomial likelihood functions are unimodal: Proof of a conjecture of I. J. Good. *Ann. Statist.* **5** 79–87.

LINDLEY, D. V. (1957). A statistical paradox. *Biometrika* **44** 187–192.

SIVAGANESAN, S. and BERGER, J. (1989). Ranges of posterior measures for priors with unimodal contaminations. *Ann. Statist.* **17** 868–889.

DEPARTMENT OF STATISTICS
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER V6T1W5
CANADA

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907