LOWER BOUNDS ON BAYES FACTORS FOR MULTINOMIAL AND CHI-SQUARED TESTS OF FIT*

by

Mohan Delampady
Purdue University and
University of British Columbia

James Berger
Purdue University

Technical Report#86-37

Department of Statistics
Purdue University

January 1987

* Research supported by the National Science Foundation, Grant DMS-8401996.

Lower Bounds on Bayes Factors for Multinomial and Chi-Squared Tests of Fit*

Mohan Delampady
Purdue University and
University of British Columbia

James Berger
Purdue University

January 30, 1987

Abstract

Lower bounds on Bayes factors and posterior probabilities in multinomial tests of point null hypotheses are developed. These are then applied to derive lower bounds on Bayes factors and posterior probabilities for χ^2–test of fit, in both exact and asymptotic situations. The general conclusion is that the lower bounds tend to be at least an order of magnitude larger than P–values, causing serious questions to be raised concerning the routine use of P–values for these problems.

*Research supported by the National Science Foundation, Grant DMS-8401996.

1 Introduction

1.1 Overview

Lower bounds on Bayes factors and posterior probabilities of point null hypotheses, H_0, have been discussed in Edwards, Lindman and Savage (1963), Dickey (1977), Good (1950, 1958, 1976), Berger and Sellke (1987), Casella and Berger (1987), Berger and Delampady (1986), and Delampady (1986a, 1986b, 1986c) among others. The startling feature of these results is that they establish that the Bayes factor and posterior probability of H_0 are generally at least an order of magnitude larger than the P–value. When such is the case, the usefulness of P–values as measures of evidence against H_0 is highly questionable.

One common rejoinder is that P–values are valuable when there are no alternatives 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 χ^2–test of fit. It will be argued that alternatives implicitly do exist, which allow for the computation of lower bounds on Bayes factors and posterior probabilities. The overall conclusion is that P–values for χ^2–tests of fit are highly misleading.

The lower bounds on Bayes factors and posterior probabilities are also of direct interest to Bayesians and likelihoodists. They provide a bound on the amount of evidence against the null hypothesis, in a Bayes factor or weighted likelihood ratio sense, irrespective of the prior distribution or likelihood "weight function".

In developing the results for the χ^2–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 over the class of conjugate priors, and Section 3 with lower bounds over a large class of transformed symmetric priors. Section 4 discusses the χ^2–test of fit, and Section 5 presents some comments and conclusions.

1.2 Notation

We let $f(x|\theta)$, $\theta \in \Theta \subseteq R^t$, denote the density of the random quantity X. It will be desired to test

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0,$$

where θ_0 is a specified element of Θ. Assume that a classical significance test is based on some test statistic $T(X)$, large values of which provide evidence against the null hypothesis. The P–value, or the observed significance level, of data x is defined to be

$$\alpha \quad = \quad P_{\theta_0}\left(T(X) \geq T(x)\right). \tag{1}$$

The example that will be of most interest in this paper is the multinomial situation. When specifically discussing this, we will use $\underset{\sim}{n} = (n_1, n_2, \cdots, n_t)$ instead of X to denote the random multinomial cell frequencies, and $\underset{\sim}{p} = (p_1, p_2, \cdots, p_t)$ instead of θ to denote the unknown cell probabilities. Thus $\underset{\sim}{n}$ will be $Multinomial(N, \underset{\sim}{p})$, with $N = \sum_{i=1}^{t} n_i$ the fixed sample size; $0 \leq n_i \leq N$ for $i = 1, 2, \ldots, t$;

$$\underset{\sim}{p} \in A \quad = \quad \{\underset{\sim}{p} : 0 \leq p_i \leq 1 \text{ for } 1 \leq i \leq t; \sum_{i=1}^{t} p_i = 1\}; \tag{2}$$

and probability density (mass function)

$$f(\underset{\sim}{n}|\underset{\sim}{p}) \quad = \quad \frac{N!}{\prod_{i=1}^{t} n_i!} \prod_{i=1}^{t} p_i^{n_i}. \tag{3}$$

The problem of interest is then to test the hypothesis:

$$H_0 : \underset{\sim}{p} = \underset{\sim}{p}^0 \quad \text{against} \quad H_1 : \underset{\sim}{p} \neq \underset{\sim}{p}^0, \tag{4}$$

where $\underset{\sim}{p}^0$ is a specified vector in A. The classical multinomial test has a P–value of

$$\alpha \;=\; P_{\underset{\sim}{p}=\underset{\sim}{p}^0}\left(\underset{\sim}{y} : f(\underset{\sim}{y}|\underset{\sim}{p}^0) \leq f(\underset{\sim}{n}|\underset{\sim}{p}^0)\right). \tag{5}$$

However, this being a hard computation, the most popular approach is to use the χ^2 approximation, $P(\chi^2_{t-1} \geq O)$, where $O = \sum_{i=1}^{t} \frac{(n_i - Np_i^0)^2}{Np_i^0}$. This approximation will be considered in Section 4.

Approaching the general testing problem from the Bayesian viewpoint, assume that π is a prior distribution on Θ with prior probabilities $\pi_0 = P^\pi(H_0)$ and $1 - \pi_0 = P^\pi(H_1)$, and conditional density given H_1 is true, $g(\theta)$ (on $\{\theta \neq \theta_0\}$; we define $g(\theta_0) = 0$). The conditional marginal density of X with respect to g is

$$m_g(x) \;=\; \int_\Theta f(x|\theta)g(\theta)\, d\theta > 0, \tag{6}$$

which we assume to be positive. The quantities of interest are then

(i) The Bayes factor of H_0 to H_1:

$$B^\pi(x) \;=\; \frac{f(x|\theta_0)}{m_g(x)}; \tag{7}$$

(ii) The posterior probability of H_0:

$$P^\pi(H_0|x) \;=\; \left[1 + \frac{(1-\pi_0)}{\pi_0} \cdot \frac{1}{B^\pi(x)}\right]^{-1}. \tag{8}$$

The Bayes factor is of interest to a Bayesian because of the well known relationship that

$$\frac{P^\pi(H_0|x)}{P^\pi(H_1|x)} \;=\; \frac{\pi_0}{(1-\pi_0)} \cdot B^\pi(x); \tag{9}$$

thus $B^\pi(x)$ is the factor by which the data changes prior odds of H_0 to H_1 to posterior odds. By considering $B^\pi(x)$, one thus considers the impact of the data; the prior probabilities of the hypotheses can be ignored. A likelihoodist is also interested in $B^\pi(x)$, 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.

Specification of g is natural and important to a Bayesian, but is resisted by others. Of interest is that lower bounds on $B^\pi(x)$ (and hence $P^\pi(H_0|x)$) 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

$$\underline{B}_G(x) = \inf_{g \in G} B^\pi(x), \tag{10}$$

and

$$\begin{aligned} \underline{P}_G(H_0|x) &= \inf_{g \in G} P^\pi(H_0|x) \\ &= \left[1 + \frac{(1-\pi_0)}{\pi_0} \cdot \frac{1}{\underline{B}_G}\right]^{-1}. \end{aligned} \tag{11}$$

For the most part, we will only present results in terms of $\underline{B}_G$, since $\underline{P}_G(H_0|x)$ requires the additional specification of π_0.

1.3 Choice of G

A lower bound, such as $\underline{B}_G(x)$, is particularly useful when G is large enough to include all densities which are plausible, but is not so large as to include unreasonable densities. If reasonable densities are omitted from G, one could argue that

$\underline{B}_G$ is not actually a valid lower bound. If G contains unreasonable distributions, on the other hand, then the lower bounds may be driven too low to be useful. Note, in particular, that minimizing $B^\pi(x)$ over g has the effect of finding that $g \in G$ which is most favorable to H_1.

All these lower bounds thus contain a potential substantial bias towards H_1, and it is obviously desirable to minimize this bias; this can best be done by restricting G in as many ways as are considered reasonable. (Surprising results can be obtained even if G is allowed to contain all distributions; indeed, Edwards, Lindman and Savage (1963) show that $\underline{B}_G(x)$ is often still much bigger than P–values.)

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 (1986)). 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

$$G_C = \{g \text{ which are conjugate to } f(x|\theta) \text{ and such that } E^g[\theta] = \theta_0\}, \quad (12)$$

$$G_{US} = \{\text{unimodal } g, \text{ symmetric about } \theta_0\}. \quad (13)$$

The appeal of these two classes of densities is that they seem to be "objective" classes. They acknowledge the central role of θ_0, and seek to spread mass about θ_0 in ways that are not biased towards particular alternatives. Lower bounds derived from such G could be termed "objective lower bounds," and are thus of interest when subjective input is unavailable or being avoided. Many other similarly "objective" classes could be considered; a detailed study of a number of such classes in Berger and Delampady (1986) (for the binomial case) indicates that G_C and G_{US} are quite representative, and also satisfactory in terms of being neither too big nor too small. (It might appear that G_C is too small, typically including only a small dimensional class of distributions; use also of G_{US} should allay any such fears.) Use of G_C is considered in Section 2, and use of G_{US} in Section 3.

2 Bounds for Conjugate Priors in Multinomial Testing

2.1 Introduction

For the Multinomial distribution, $f(\underset{\sim}{n}|\underset{\sim}{p})$ in (3), the Dirichlet densities form the usual conjugate family. The density of the Dirichlet distribution with parameters $\underset{\sim}{k} = (k_1, k_2, \cdots, k_t)$, is

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

The mean of $g_{\underset{\sim}{k}}$, is the vector $(\sum_{j=1}^{t} k_j)^{-1}\underset{\sim}{k}$, which equals $\underset{\sim}{p}^0 = (p_1^0, p_2^0, \cdots, p_t^0)$ only if $\underset{\sim}{k} = c\underset{\sim}{p}^0$ for some $c > 0$. Thus, in testing $H_0 : \underset{\sim}{p} = \underset{\sim}{p}^0$ versus $H_1 : \underset{\sim}{p} \neq \underset{\sim}{p}^0$, the class of conjugate densities with mean equal to $\underset{\sim}{p}^0$ is

$$G_C = \{g_{\underset{\sim}{k}} : \underset{\sim}{k} = c\underset{\sim}{p}^0\ c > 0\}. \tag{14}$$

As discussed in Subsection 1.3, we will seek

$$\underline{B}_C(\underset{\sim}{n}) = \underline{B}_{G_C}(\underset{\sim}{n}) = \inf_{g \in G_C} B^{\pi}(\underset{\sim}{n}) \tag{15}$$

and

$$\underline{P}_C(\underset{\sim}{n}) = \inf_{g \in G_C} P^{\pi}(H_0|x) \quad = \quad \left[1 + \frac{(1-\pi_0)}{\pi_0} \cdot \frac{1}{\underline{B}_C(\underset{\sim}{n})}\right]^{-1}. \tag{16}$$

It should be observed that Good(1967) extensively discusses the multinomial testing situations with conjugate priors, and Edwards, Lindman and Savage (1963) discuss the possibility of finding $\underline{B}_C$ for the binomial problem. Extensive discussion of the binomial problem can be found in Delampady (1986a) and Berger and Delampady (1986).

2.2 Exact Results

For the conjugate priors $g_{\underset{\sim}{k}}$, it is possible to exactly calculate the marginal density as

$$\begin{aligned} m_{g_{\underset{\sim}{k}}}(\underset{\sim}{n}) &= \int_A f(\underset{\sim}{n}|\underset{\sim}{p}) g_{\underset{\sim}{k}}(\underset{\sim}{p})\, d\underset{\sim}{p} \\ &= \frac{\Gamma(\sum_{i=1}^t k_i)}{\prod_{i=1}^t \Gamma(k_i)} \cdot \frac{\prod_{i=1}^t \Gamma(n_i + k_i)}{\Gamma(N + \sum_{i=1}^t k_i)}. \end{aligned} \tag{17}$$

The following result is an immediate consequence.

Theorem 1. *The lower bound on the Bayes factor over G_C is given by*

$$\underline{B}_C(\underset{\sim}{n}) = \inf_{c>0} \frac{\left[\prod_{i=1}^t \Gamma(cp_i^0)\right] \Gamma(c+N) \prod_{i=1}^t (p_i^0)^{n_i}}{\Gamma(c) \prod_{i=1}^t \Gamma(n_i + cp_i^0)}. \tag{18}$$

The minimization in (18) can easily be carried out numerically. For selected values of t, $\underset{\sim}{p}^0$ and $\underset{\sim}{n}$, $\underline{B}_C(\underset{\sim}{n})$ is tabulated in Table 2 along with the corresponding P–values. Table 2 is given at the end of Section 3 until which we defer discussion of the results.

2.3 Asymptotic Results

As $N \longrightarrow \infty$, the behavior of $\underline{B}_C(\underset{\sim}{n})$ is given in the following theorem.

Theorem 2. *As $N \longrightarrow \infty$, suppose that the observation vector $\underset{\sim}{n}$ satisfies*

$$\sum_{i=1}^{t} \frac{(n_i - Np_i^0)^2}{Np_i^0} = K + O(\frac{1}{N}),$$

where $K > t - 1$ is a fixed constant. (The sum will, of course, converge to some fixed constant with probability one.) Then,

$$\lim_{N \to \infty} \underline{B}_C(\underset{\sim}{n}) \quad = \quad \underline{B}_C^*(K), \tag{18}$$

where

$$\underline{B}_C^*(K) = \inf_{a>0} a^{1-t} \exp(-\frac{1}{2}(1 - a^2)K).$$

Proof: See the Appendix. □

$\underline{B}_C^*(K)$ is the bound obtained from the following normal problem. Let $\underset{\sim}{X} = (X_1, X_2, \cdots, X_{t-1}) \sim N_{t-1}(\underset{\sim}{\theta}, I)$, and suppose that it is desired to test $H_0 : \underset{\sim}{\theta} = \underset{\sim}{\theta}_0$ versus $H_1 : \underset{\sim}{\theta} \neq \underset{\sim}{\theta}_0$. Let G be the class of all multivariate normal densities, spherically symmetric about the vector $\underset{\sim}{\theta}_0$. Then, the lower bound on the Bayes factor for this problem, over the class G, is precisely $\underline{B}_C^*(K)$, where $K = ||\underset{\sim}{X} - \underset{\sim}{\theta}_0||^2$, as is proved in Delampady (1986a). This lower bound, calculated for a number of different dimensions, is displayed in Table 1 in Section 3; discussion is deferred to the same section.

3 Bounds for Symmetric Priors: Multinomial Testing

When $\underset{\sim}{p}^0 = (t^{-1}, \ldots, t^{-1})'$, it is not difficult to define a notion of symmetry for conditional prior densities g, leading to a class such as G_{US} in (13). For general $\underset{\sim}{p}^0$, however, usual notions of symmetry are quite inappropriate. To see this, suppose that $\underset{\sim}{p}^0$ is near the edge of A (the parameter space). Then any "symmetric" prior would be concentrated quite near this edge. Calculations with such priors were indeed carried out, for the most part resulting in lower bounds on $B^{\pi}(\underset{\sim}{n})$ that were much larger than our other bounds (and hence even more of a contrast with P–values).

A natural way to obtain a notion of symmetry is to consider symmetry in a suitable transformation of the parameter $\underset{\sim}{p}$. One such transformation is suggested by the Normal approximation to the Multinomial likelihood function. Thus, if $H_0 : \underset{\sim}{p} = \underset{\sim}{p}^0$ is to be tested, it may be reasonable to specify symmetry in the variable $\underset{\sim}{u}(p)$ of $\underset{\sim}{p}$ defined as follows. Let $D(\underset{\sim}{p}^a)$ be the diagonal matrix with i th diagonal element equal to p_i^a and $\phi(\underset{\sim}{p}) = \left(\sqrt{p}_1, \sqrt{p}_2, \cdots, \sqrt{p}_{t-1}\right)$. Then the covariance matrix of the first $t-1$ free coordinates of $\underset{\sim}{n}$ (recall $\underset{\sim}{n}$ is *Multinomial* $(N, \underset{\sim}{p})$) is $D(\underset{\sim}{p}^{1/2})\,(I_{t-1} - \underset{\sim}{\phi}\underset{\sim}{\phi}')\,D(\underset{\sim}{p}^{1/2})$, where I_k is the identity matrix of dimension $k \times k$. Let

$$B(\underset{\sim}{p})B(\underset{\sim}{p})' \;=\; D(\underset{\sim}{p}^{-1/2})(I - \underset{\sim}{\phi}\underset{\sim}{\phi}')^{-1}D(\underset{\sim}{p}^{-1/2}) \qquad (19)$$

be a decomposition. Then $\underset{\sim}{u}(\underset{\sim}{p})$ is defined as

$$\underset{\sim}{u}(\underset{\sim}{p}) \;=\; B\,(\underset{\sim}{p})\,(\underset{\sim}{p} - \underset{\sim}{p}^0)\,. \qquad (20)$$

The reasons for considering the transformation $\underset{\sim}{u}(\underset{\sim}{p})$ are the following:

(i) The range of $\underset{\sim}{u}(\underset{\sim}{p})$ is R^{t-1}.

(ii) The likelihood function of $\underset{\sim}{u}(\underset{\sim}{p})$ is approximately normal with covariance matrix I_{t-1} in a neighbourhood of $\underset{\sim}{p}^0$ as is shown in Result 4.1 in the Appendix.

(iii) It is possible, as follows, to explicitly give $\underset{\sim}{u}(\underset{\sim}{p})$ in any dimension. By choosing

$$B(\underset{\sim}{p}) = \left(I + \frac{1}{\sqrt{p_t} + p_t}\phi(\underset{\sim}{p})\phi(\underset{\sim}{p})'\right) D(\underset{\sim}{p}^{-\frac{1}{2}}),$$

(see Result 4.1 in the Appendix) $\underset{\sim}{u}(\underset{\sim}{p})$ can be represented as

$$\begin{aligned} \underset{\sim}{u}(\underset{\sim}{p}) &= \left(\frac{p_1 - p_1^0}{\sqrt{p_1}}, \cdots, \frac{p_{t-1} - p_{t-1}^0}{\sqrt{p_{t-1}}}\right) \\ &\quad + \left(\frac{p_t^0 - p_t}{\sqrt{p_t} + p_t}\right)\left(\sqrt{p_1}, \cdots, \sqrt{p_{t-1}}\right). \end{aligned}$$

For a normal likelihood, the natural class of conditional prior densities in (13) can be used, yielding (with $*$ denoting the transformed problem)

$$G^*_{US} = \{\text{unimodal } g^*(\underset{\sim}{u}) \text{ which are symmetric about } \underset{\sim}{0}\}. \tag{21}$$

Transforming back to the original parameter yields the class ("TUS" standing for "Transformed Unimodal Symmetric")

$$\begin{aligned} G_{TUS} = \{&g(\underset{\sim}{p}) = g^*\left(\underset{\sim}{u}(\underset{\sim}{p})\right)\left|\frac{\partial \underset{\sim}{u}(\underset{\sim}{p})}{\partial \underset{\sim}{p}}\right| : g^* \text{ is} \\ &\text{unimodal and symmetric about } \underset{\sim}{0}\}. \end{aligned} \tag{22}$$

The term $\left|\frac{\partial \underset{\sim}{u}(\underset{\sim}{p})}{\partial \underset{\sim}{p}}\right|$ is merely the Jacobian of the transformation. In all calculations it will be most convenient to work directly with $\underset{\sim}{u}$ and G^*_{US}, however, so calculation of the Jacobian will not be needed.

3.1 Exact Results

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

Theorem 3.

$$\begin{aligned} \underline{B}_{TUS}(\underset{\sim}{n}) &= \inf_{g \in G_{TUS}} B^{\pi}(\underset{\sim}{n}) \\ &= \frac{f(\underset{\sim}{n}|\underset{\sim}{p}^0)}{\sup_r \frac{1}{V(r)} \int_{\|\underset{\sim}{u}\| \le r} l(\underset{\sim}{u})\, d\underset{\sim}{u}}, \end{aligned} \tag{23}$$

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

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

and $\underset{\sim}{p}(\underset{\sim}{u})$ *is the inverse function of* $\underset{\sim}{u}(\underset{\sim}{p})$.

Proof:

$$\begin{aligned} & \sup_{g \in G_{TUS}} m_g(\underset{\sim}{n}) \\ &= \sup_{g \in G_{TUS}} \int_A \frac{N!}{\prod_{i=1}^{t} n_i!} \prod_{i=1}^{t} p_i^{n_i} g(\underset{\sim}{p})\, d\underset{\sim}{p} \\ &= \sup_{h \in G_{US}} \int_A \frac{N!}{\prod_{i=1}^{t} n_i!} \prod_{i=1}^{t} p_i^{n_i} h\left(\underset{\sim}{u}(\underset{\sim}{p})\right) \left|\frac{\partial \underset{\sim}{u}(\underset{\sim}{p})}{\partial \underset{\sim}{p}}\right| d\underset{\sim}{p} \\ &= \sup_{h \in G_{US}} \int \frac{N!}{\prod_{i=1}^{t} n_i!} \prod_{i=1}^{t} p(u)_i^{n_i} h(\underset{\sim}{u})\, d\underset{\sim}{u} \\ &= \sup_r \frac{1}{V(r)} \int_{\|\underset{\sim}{u}\| \le r} l(\underset{\sim}{u})\, d\underset{\sim}{u}, \end{aligned}$$

using the result that the extreme points of the class of all unimodal spherically symmetric distributions are uniform distributions on spheres symmetric about the origin. The last equality proves the result. □

In Delampady (1986a) it is shown that the maximizing r in (24) is finite. Other versions of transformed symmetry were also considered therein, and yielded similar or larger lower bounds.

For selected t, $\underset{\sim}{n}$ and $p_i^0 = 1/t$, $\underline{B}_{TUS}$ is tabulated in Table 2 (at the end of this section), along with the corresponding P–values. We defer general discussion until then, but it is useful, for calculating the integral in (24), to record that the inverse function $\underset{\sim}{p}(\underset{\sim}{\mu})$ is given by

$$p_i(\underset{\sim}{\mu}) = \left[\frac{\mu_i + (\mu_i^2 + 4p_i^0 H(p_t))^{\frac{1}{2}}}{2H(p_t)}\right]^2, i = 1, \ldots, t-1,$$

where

$$H(p_t) = (1 + p_t^0/\sqrt{p_t})/(1 + \sqrt{p_t})$$

and $p_t = p_t(\underset{\sim}{\mu})$ is the solution to

$$(1 - p_t) = \sum_{i=1}^{t-1} p_i(\underset{\sim}{\mu}).$$

3.2 Asymptotic Results

The calculation in (24) can be very difficult if t is large. Hence an asymptotic approximation for large N is particularly important here, and is given by the following development. For any multinomial observation, $\underset{\sim}{n}$, let $\underset{\sim}{Z}(\underset{\sim}{n}) = \sqrt{N}B(\underset{\sim}{p}^0)\left(\frac{1}{N}\underset{\sim}{n} - \underset{\sim}{p}^0\right)$, where $B(\underset{\sim}{p}^0)$ is as in (21).

Theorem 4. *Consider testing $H_0 : \underset{\sim}{p} = \underset{\sim}{p}^0$ against $H_1 : \underset{\sim}{p} \neq \underset{\sim}{p}^0$ in the multinomial situation. Suppose the observation vector $\underset{\sim}{n}$ satisfies $\|\underset{\sim}{Z}(\underset{\sim}{n})\|^2 = K + O(\frac{1}{N})$, where* *the fixed constant $K > t-1$. Then, we have,*

$$\lim_{N\to\infty} \underline{B}_{TUS}(\underset{\sim}{n}) = \underline{B}^*_{US}(K),$$

where

$$\underline{B}^*_{US}(K) = \frac{\frac{1}{(2\pi)^{(t-1)/2}} \exp(-\frac{1}{2}K)}{\sup_r \frac{1}{V(r)} P\left(Y \le r^2\right)}, \tag{25}$$

Y having a non-central chi-squared distribution with $t-1$ degrees of freedom and non-centrality parameter K.

Proof: See the Appendix. □

Note that $\underline{B}^*_{US}(K)$ in (25) is the lower bound on B^π over the class G_{US} in (13) of conditional prior densities for $\underset{\sim}{\theta}$ that would be obtained in the multivariate normal problem discussed at the end of Subsection 2.3. Table 1 presents values of $\underline{B}^*_{US}$ for a range of t and K corresponding to certain common P–values.

A comment concerning the complexity of the proof of Theorem 4 is in order. Standard asymptotic theory would yield that

$$\underline{B}^*_{US}(K) = \inf_{g\in G_{US}} \lim_{N\longrightarrow\infty} B^\pi(\underset{\sim}{n}).$$

The difficulty in the proof of Theorem 4 was caused by first taking the infimum, and then taking the limit. This more difficult version is needed to ensure that $\underline{B}^*_{US}(K)$ really is approximating the actual global minimum of $B^\pi(\underset{\sim}{n})$.

Table 1: Asymptotic Lower Bounds

	α =.001		α =.01		α =.05		α =.10	
Dimension=$t-1$	$\underline{B}_C^*$	$\underline{B}_{US}^*$	$\underline{B}_C^*$	$\underline{B}_{US}^*$	$\underline{B}_C^*$	$\underline{B}_{US}^*$	$\underline{B}_C^*$	$\underline{B}_{US}^*$
1	.0244	.0182	.1538	.1227	.4734	.4092	.7001	.6437
2	.0198	.0143	.1247	.0978	.4067	.3481	.6263	.5699
3	.0165	.0119	.1142	.0902	.3784	.3259	.5818	.5396
4	.0156	.0114	.1064	.0850	.3615	.3141	.5713	.5232
5	.0133	.0099	.1020	.0824	.3503	.3072	.5576	.5131
6	.0129	.0097	.0988	.0807	.3419	.3023	.5473	.5058
7	.0126	.0096	.0964	.0797	.3356	.2990	.5392	.5004
8	.0124	.0095	.0945	.0789	.3305	.2963	.5330	.4966
9	.0121	.0094	.0929	.0782	.3261	.2942	.5277	.4932
10	.0119	.0094	.0916	.0777	.3228	.2927	.5230	.4908
15	.0113	.0093	.0859	.0752	.3108	.2875	.5078	.4826
20	.0109	.0093	.0833	.0743	.3036	.2844	.4988	.4782
30	.0105	.0092	.0803	.0735	.2950	.2809	.4879	.4725

4 Comparisons and Conclusions

Tables 2 and 3 tabulate the exact bounds, $\underline{B}_C$ and $\underline{B}_{TUS}$, respectively, for $t = 3$ and $t = 4$ with $p_i^0 = 1/t$, and various choices of N, $\underset{\sim}{n}$. Here α denotes the P–value, with "exact" referring to the exact P–value from (5), and "χ^2" referring to the approximate P–value obtained from the chi-squared approximation.

Table 2: Lower Bounds for Conjugate and Transformed Symmetric Densities

t=3						
α		N	n_1	n_2	$\underline{B}_C$	$\underline{B}_{TUS}$
χ^2	exact					
0.00	.001	12	10	1	.0285	.0077
0.00	.008	13	10	2	.0565	.0236
0.01	.017	14	10	3	.0995	.0517
0.02	.024	9	7	1	.2010	.0999
0.03	.033	14	9	4	.2378	.1405
0.04	.060	12	8	3	.3163	.1895
0.06	.062	13	8	4	.3833	.2064
0.07	.056	14	8	5	.4053	.2877
0.09	.080	15	7	7	.4345	.3101
0.10	.166	9	6	2	.6086	.3899
0.12	.100	13	7	5	.5733	.4366

Table 3: Lower Bounds for Conjugate and Transformed Symmetric Densities

t=4							
α		N	n_1	n_2	n_3	$\underline{B}_C$	$\underline{B}_{TUS}$
0.00	.001	15	11	2	1	.0124	.0032
0.00	.003	12	9	1	1	.0356	.0116
0.00	.005	15	10	3	1	.0454	.0171
0.00	.007	15	10	2	2	.0660	.0254
0.01	.008	14	9	3	1	.0987	.0351
0.01	.013	15	9	4	1	.1101	.0465
0.02	.057	13	8	2	2	.2749	.1464
0.03	.025	15	8	5	1	.1919	.0909
0.03	.053	11	7	2	1	.3098	.1644
0.04	.044	15	7	6	1	.2506	.1563
0.05	.045	14	7	5	1	.3235	.2038
0.05	.066	13	7	4	1	.3712	.2269

The first fact to be noted is that $\underline{B}_C$ and $\underline{B}_{TUS}$ differ more here than in the asymptotic normal situation of Table 1. 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 $\underline{B}_C$ or $\underline{B}_{TUS}$ is somewhat a matter of taste: $\underline{B}_C$ is probably more representative of typical Bayes factors, while $\underline{B}_{TUS}$ is more saleable as a true objective lower bound. Note also that the Table 1 asymptotic bounds seem fairly reasonable as approximations to $\underline{B}_C$ even for these small N, but can be rather poor as approximations to $\underline{B}_{TUS}$ for small N.

Finally, we come to the major point, reflected here as well as in Table 1: the "objective" lower bounds on B^π are substantially larger than the P–value. For instance, when $t = 4$, $N = 14$, and $\underset{\sim}{n} = (7,5,1,1)$, the exact P-value is .045 ("significant at the .05 level"), yet $\underline{B}_C = .323$ and $\underline{B}_{TUS} = .2038$. Thus (using the likelihood interpretation discussed in Section 1) the data supports $H_1 : \theta \neq (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ at most 3 and 5 times, respectively, as much as it supports $H_0 : \theta = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. This would appear to at most be 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-Square Test of Fit

5.1 Set Up and Chi-Square Test

Consider a statistical experiment in which N independent and identically distributed random quantities $X_1, X_2, \cdots, X_N$ are observed from a distribution F. The problem is to test the hypothesis

$$H_0 : F = F_0 \text{ against } H_1 : F \neq F_0,$$

where F_0 is a specified distribution. The standard test procedure for this problem is the χ^2 test of fit, given as follows. First, a partition $\{a_i\}_{i=0}^{t}$ of the real line is considered. Then the frequencies of the N observations in this partition are found. Let $\underset{\sim}{n} = (n_1, \cdots, n_t)'$ denote these frequencies; thus n_i = the number of X_i's in

$(a_{i-1}, a_i]$. Let

$$p_i = F(a_i) - F(a_{i-1}) = P_F(a_{i-1} < X \leq a_i),$$

$$p_i^0 = F_0(a_i) - F_0(a_{i-1}) = P_{F_0}(a_{i-1} < X \leq a_i),$$

and

$$\underset{\sim}{p} = (p_1, \cdots, p_t)',$$

$$\underset{\sim}{p}^0 = \left(p_1^0, \cdots, p_t^0\right)'.$$

Then the χ^2 test procedure is to calculate the test statistic,

$$o \;=\; \sum_{i=1}^{t} \frac{(n_i - Np_i^0)^2}{Np_i^0}, \tag{26}$$

and compute the P–value assuming a χ^2_{t-1} distribution for O, as

$$\alpha = P(\chi^2_{t-1} \geq o).$$

5.2 Likelihood and Bayesian Lower Bounds

Reducing the observations to the vector $\underset{\sim}{n}$ of cell frequencies implicitly implies that one is testing

$$H_0 : \underset{\sim}{p} = \underset{\sim}{p}^0 \text{ versus } H_1 : \underset{\sim}{p} \neq \underset{\sim}{p}^0,$$

where $\underset{\sim}{n}$ has a $Multinomial(N, \underset{\sim}{p})$ distribution given $\underset{\sim}{p}$. Thus we can apply the results of the previous sections to obtain "objective" lower bounds on the Bayes factor and posterior probability of H_0. The exact lower bounds, for the class of conjugate and transformed symmetric unimodal densities, respectively, were given in Theorems 1 and 3 and Tables 2 and 3, while asymptotic approximations to the lower bounds for large N were given in Theorems 2 and 4 and Table 1.

Example: 30 observations were made on the arrival times of a certain process. It is desired to test the hypothesis that the distribution of the arrival time, X, is exponential with mean 1, i.e. to test

$$H_0 : F_0(x) = 1 - \exp(-x), \quad x > 0.$$

Suppose that it is decided to use a 3 cell partition, with the cells 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.

Table 4: Cells and Data

cell	boundaries	n_i	Np_i^0
1	[0, 0.40)	16	10
2	[0.40, 1.01)	9	10
3	[1.01, ∞)	5	10

Using (26), the χ^2 test statistic can be seen to be $O = 6.20$ with 2 degrees of freedom. The exact P–value (computed by (5) for the multinomial model), the P–value using the chi–square approximation, the exact lower bounds ($\underline{B}_C$ and $\underline{B}_{TUS}$) on the Bayes factor from Theorems 1 and 3, and the asymptotic lower bounds ($\underline{B}_C^*$ and $\underline{B}_{US}^*$) from Theorems 2 and 4, are all given in Table 5.

Table 5: P-values and Lower Bounds

Exact-α	χ^2-α	$\underline{B}_C^*$	$\underline{B}_{US}^*$	$\underline{B}_C$	$\underline{B}_{TUS}$
.058	.05	.4489	.3881	.3750	.2922

The chi-square approximation is quite reasonable here and the lower bounds over G_C and G_{TUS} are quite similar. But the difference between the P–value and the lower bound on the Bayes factor is enormous. The lower bounds on the Bayes factor indicate that the data support H_1 by at most a factor of 3 to 1; the common interpretation of a P–value of .05 is that the evidence against H_0 is much stronger than 3 to 1.

6 Comments and Conclusions

We have seen that lower bounds on Bayes factors are typically 4 or more times larger than P–values in multinomial testing of a point hypothesis or chi-square tests of fit. Our recommendation is thus to abandon use of P–values, at the very least replacing them with the lower bounds on the Bayes factor. The point is that typical users of P–values can not be expected to understand that a P–value of .05 really means at most 3 to 1 evidence in favor of H_1 or that a P–value of .01 really means at most 9 to 1 evidence in favor of H_1. This is especially so because the relationship between P–values and Bayes factors is highly dependent on the problem, sample size, type of hypothesis, and stopping rule (see Berger and Sellke (1987) and Berger and Delampady (1986)). Note that there is nothing "wrong" with a P–value; it is after all just a specific well-defined function of the data. The problem lies in attempting to interpret the meaning of a P–value. In some problems a P–value will correspond to Bayes factors against H_0; in others, such as those discussed here, it will be an order of magnitude smaller than all sensible Bayes factors.

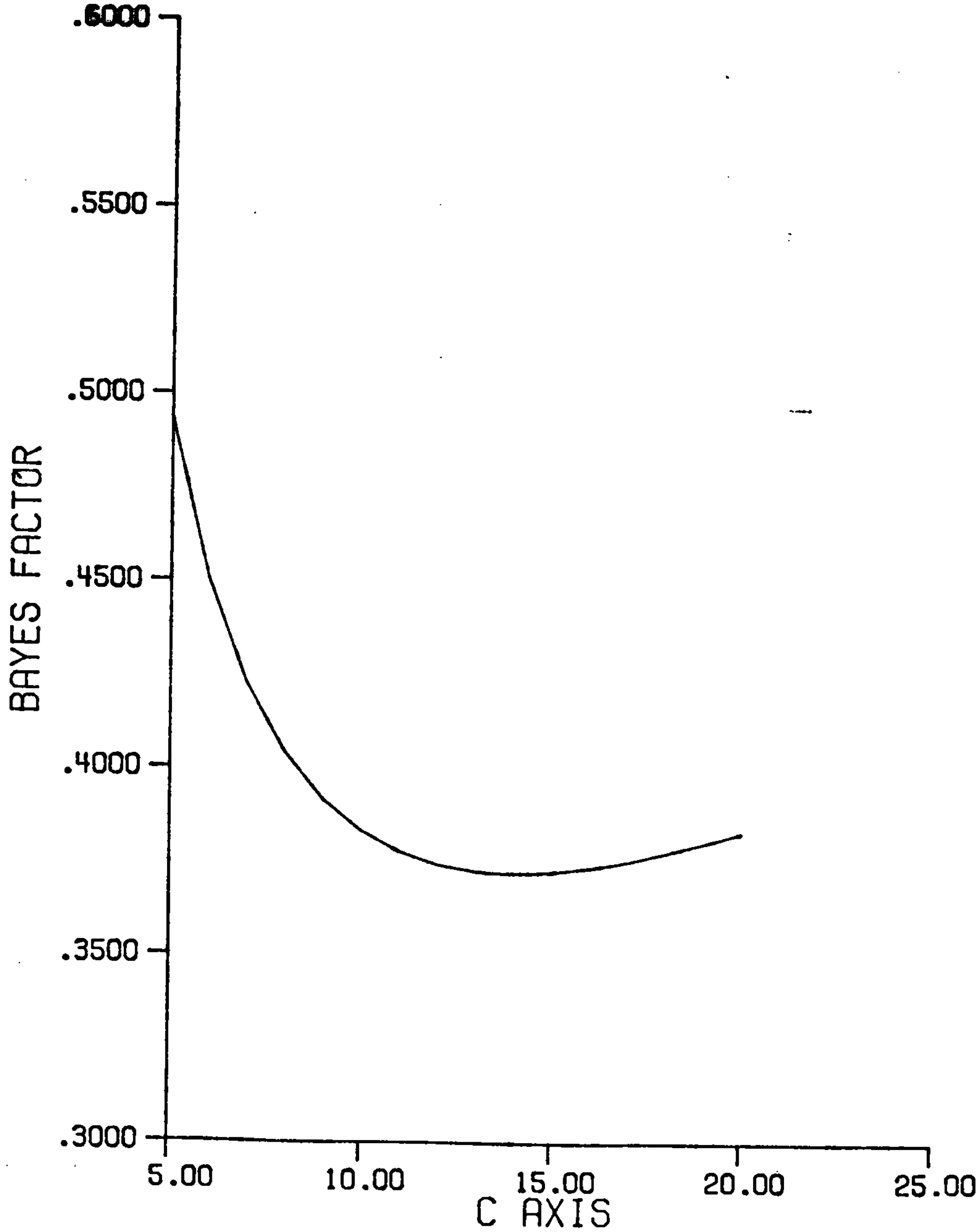
Many arguments can be raised concerning the development here. One can always argue against the Bayesian formulation, but the fact that we are working with lower bounds on the Bayes factor over all reasonable priors makes such an argument more difficult. (One can quibble with the classes G_C and G_{TUS} that we chose, but we do not think that the answers will be qualitatively different for other "objective" classes.)

One can also argue with the entire enterprise of testing a point null. Two replies are that: (i) Testing a point null is very often done, so why not do it right? and (ii) Testing of a point null can be shown to frequently be a good approximation to testing of a precise hypothesis, such as $H_0 : |\theta - \theta_0| \leq \epsilon$, in the sense that the Bayes factor is essentially unchanged if H_0 is replaced by $H_0^* : \theta = \theta_0$ (see Berger and Delampady (1986)). Other arguments that have been raised are given and discussed in Berger and Sellke (1987), Casella and Berger (1987), and Berger and Delampady (1986).

An important qualification is that, although the lower bounds $\underline{B}_C$ and $\underline{B}_{TUS}$ are much more useful than P–values, they are just lower bounds. If $\underline{B} = .5$, then we can be quite assured that there is no strong reason to reject H_0, but if $\underline{B} = .05$ what should be done? After all, this implies only that the Bayes factor is somewhere between .05 and ∞ (which can be shown to be the upper bound), depending on the choice of g. The answer, of course, is that one cannot avoid at least crude specification of subjective g, if a precise Bayes factor is to be determined. Reasonable results might often be obtained by fairly crude devices, such as considering only the conjugate $g_{\underset{\sim}{k}}$ in Subsection 2.1, with $\underset{\sim}{k} = c\underset{\sim}{p}^0$. 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 $\underset{\sim}{p}$ 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.

Example (continued). From (7) and (17), we can graph $B^{\pi}(\underset{\sim}{n})$ for the conjugate priors $g_{c\underset{\sim}{p}^0}$, as a function of c. This is done in Figure 1. Note that the lower bound over all c (i.e. $\underline{B}_C(\underset{\sim}{n})$) is attained at $c = 14.008$.

Figure 1: Graph of Bayes Factors



References

[1] Berger, J. (1985), *Statistical Decision Theory and Bayesian Analysis*, New York, Springer-Verlag.

[2] Berger, J. and Berliner, L. M. (1986), "Robust Bayes and Empirical Bayes Analysis with ϵ–Contaminated Priors", *Annals of Statistics*, **14**, 461–486.

[3] Berger, J. and Delampady, M. (1986), "Bayesian Testing of Precise Hypotheses", Tech. Report 86–38, Department of Statistics, Purdue University.

[4] Berger, J. and Sellke, T. (1987), "Testing a Point Null Hypothesis: The Irreconcilability of Significance Levels and Evidence", *Journal of the American Statistical Association*, **82**.

[5] Berkson, J. (1942), "Tests of Significance Considered as Evidence", *Journal of American Statistical Association*, **37**, 325–335.

[6] Casella, G. and Berger, R. L. (1985), "Reconciling Bayesian and Frequentist Evidence in the One-Sided Testing Problem", *Journal of the American Statistical Association*, **82**.

[7] DeGroot, M. H. (1973), "Doing What Comes Naturally: Interpreting a Tail Area as a Posterior Probability or as a Likelihood Ratio", *Journal of the American Statistical Association*, **68**, 966–969.

[8] Delampady, M. (1986a), "Testing a Precise Hypothesis: Interpreting P–values from a Robust Bayesian Viewpoint", Ph. D Thesis, Purdue University.

[9] Delampady, M. (1986b), "Lower Bounds on Bayes Factors for Interval Null Hypotheses", Tech. Report 86–35, Department of Statistics, Purdue University.

[10] Delampady, M. (1986c), "Lower Bounds on Bayes Factors for Invariant Testing Situations", Tech. Report 86–36, Department of Statistics, Purdue University.

[11] Dempster, A. P. (1973), "The Direct Use of Likelihood for Significance Testing", *Proceedings of the Conference on Foundational Questions in Statistical Inference*, O. Barndorff-Nielsen et. al. (Eds.), University of Aarhus.

[12] Dickey, J. M. (1971), "The Weighted Likelihood Ratio, Linear Hypotheses on Normal Location Parameters", *Annals of Mathematical Statistics*, **42**, 204–223.

[13] Dickey, J. M. (1973), "Scientific Reporting", *Journal of the Royal Statistical Society B*, **35**, 285–305.

[14] Dickey, J. M. (1974), "Bayesian Alternatives to the F-test and Least Squares Estimate in the Linear Model", *Studies in Bayesian Econometrics and Statistics*, S. E. Fienberg and A. Zellner (Eds.), Amsterdam: North-Holland.

[15] Dickey, J. M. (1977), "Is the Tail Area Useful as an Approximate Bayes Factor", *Journal of the American Statistical Association*, **72**, 138–142.

[16] Dickey, J. M. and Lientz, B. P. (1970), "The Weighted Likelihood Ratio, Sharp Hypotheses About Chances, The Order of a Markov Chain", *Annals*

of Mathematical Statistics, **41**, 214–226.

[17] Edwards, W., Lindman, H., and Savage, L. J. (1963), "Bayesian Statistical Inference for Psychological Research", *Psychological Review*, **70**, 193–242. (Reprinted in *Robustness of Bayesian Analyses*, J. Kadane (Eds.), Amsterdam: North-Holland.)

[18] Feller, W. (1968), *An Introduction to Probability Theory and Its Applications*, Vol. 1, (3rd Edition), New York: Wiley.

[19] Good, I. J. (1950), *Probability and the Weighing of Evidence*, London: Charles Griffin.

[20] Good, I. J. (1958), "Significance Tests in Parallel and in Series", *Journal of the American Statistical Association*, **53**, 799–813.

[21] Good, I. J. (1965), *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*, Cambridge: M. I. T. Press.

[22] Good, I. J. (1967), "A Bayesian Significance Test for the Multinomial Distribution", *Journal of the Royal Statistical Society B*, **29**, 399–431.

[23] Good, I. J. (1983), *Good Thinking: The Foundations of Probability and Its Applications*, Minneapolis: University of Minnesota Press.

[24] Hill, B. (1982), "Comment on 'Lindley's Paradox"', *Journal of the American Statistical Association*, **77**, 344–347.

[25] Jeffreys, H. (1961), *Theory of Probability* (3rd Edition), Oxford: Oxford University Press.

[26] Lempers, F. B. (1971), *Posterior Probabilities of Alternative Linear Models*, Rotterdam: University of Rotterdam Press.

[27] Lindley, D. V. (1977), "A Problem in Forensic Science", *Biometrika*, **64**, 207–213.

[28] Pratt, J. W. (1965), "Bayesian Interpretation of Standard Inference Statements (with Discussion)", *Journal of the Royal Statistical Society B*, **27**, 169–203.

[29] Shafer, G. (1982), "Lindley's Paradox", *Journal of the American Statistical Association*, **77**, 325–351.

[30] Sivaganesan, S. and Berger, J. (1986), "Ranges of Posterior Measures for Priors with Unimodal Contaminations", Tech. Report 86-41 , Department of Statistics, Purdue University.

[31] Smith, A. F. M. and Spiegelhalter, D. J. (1980), "Bayes Factors and Choice Criteria for Linear Models", *Journal of the Royal Statistical Society B*, **42**, 213–220.

[32] Zellner, A. (1984), "Posterior Odds Ratios for Regression Hypotheses: General Considerations and Some Specific Results", *Basic Issues in Econometrics*, Chicago: University of Chicago Press.

[33] Zellner, A. and Siow, A. (1980), "Posterior Odds Ratios for Selected Regression Hypotheses", *Bayesian Statistics*, J. M. Bernardo et. al. (Eds.), Velencia: University Press.

Appendix

Proof of Theorem 2. From (18),

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

First of all, it can be shown that Stirling's approximation is valid for all the Γ terms. Now let $n_i = Np_i^0 + b_i$, $\sum_{i=1}^t b_i = 0$. Then, using Stirling's approximation,

$$\begin{aligned}
\left[\underline{B}_{G_C}\right]^{-1} &= \sup_{a>0}\left(\frac{a}{N+a}\right)^{\frac{(t-1)}{2}} \prod_{i=1}^t \left[\frac{(N+a)p_i^0 + b_i}{(N+a)p_i^0}\right]^{(N+a)p_i^0+b_i-1/2} l_N^{(1)}(a) \\
&= \sup_{a>0}\left(\frac{a}{N+a}\right)^{\frac{(t-1)}{2}} \prod_{i=1}^t \left(1 + \frac{b_i}{(N+a)p_i^0}\right)^{(N+a)p_i^0+b_i-1/2} l_N^{(1)}(a),
\end{aligned}$$

where, $N\left(l_N^{(1)}(a) - 1\right)$ is bounded. Further,

$$\begin{aligned}
&\log \prod_{i=1}^t \left(1 + \frac{b_i}{(N+a)p_i^0}\right)^{(N+a)p_i^0+b_i-1/2} \\
&= \sum_{i=1}^t \left[b_i + \frac{b_i(b_i - \frac{1}{2})}{(N+a)p_i^0} - \frac{b_i^2}{2(N+a)p_i^0} - \frac{b_i^2(b_i - 1/2)}{2[(N+a)p_i^0]^2} + \frac{b_i^3}{3[(N+a)p_i^0]^2} + \cdots\right] \\
&= \sum_{i=1}^t \frac{b_i^2}{2(N+a)p_i^0} - \sum_{i=1}^t \frac{b_i^3}{6\left[(N+a)p_i^0\right]^2} + O\left(\frac{1}{\sqrt{N}}\right) \\
&= \left(\frac{N}{2(N+a)} \sum_{i=1}^t \frac{b_i^2}{Np_i^0}\right) l_N^{(2)}(a) \\
&= \left(\frac{1}{2}(1-c)^2 K\right) l_N^{(3)}(a),
\end{aligned}$$

where $c = \left(\frac{a}{N+a}\right)^{1/2}$, both $\sqrt{N}\left(l_N^{(2)}(a) - 1\right)$ and $\sqrt{N}\left(l_N^{(3)}(a) - 1\right)$ are bounded functions. The rest of the argument follows by noting that $l_N^{(1)} l_N^{(3)} \longrightarrow 1$ uniformly. (See Theorem 3.2.2 (Delampady(1986a))). □

Proof of Theorem 4. We first need the following lemma which shows that, for large N, only contiguous alternatives to H_0 need be considered. Recall that

$l(\underset{\sim}{u}) = \frac{N!}{\prod_{i=1}^{t} n_i!} \prod_{i=1}^{t} p(\underset{\sim}{u})_i^{n_i}$, from Theorem 3. Let

$$\begin{aligned} m_g(\underset{\sim}{n}) &= \int_A \frac{N!}{\prod_{i=1}^{t} n_i!} \prod_{i=1}^{t} p_i^{n_i} g(\underset{\sim}{p})\, d\underset{\sim}{p}, \text{ for } g \in G_1, \text{ and} \\ H(r) &= \frac{1}{V(r)} \int_{\|\underset{\sim}{u}\| \le r} l(\underset{\sim}{u}) d\underset{\sim}{u}. \end{aligned}$$

Again from Theorem 3,

$$\sup_{g \in G_1} m_g(\underset{\sim}{n}) = \sup_r H(r).$$

To proceed further we need the following lemma.

Lemma 4.1. *For large N, the r which maximizes $H(r)$ is $O(\frac{1}{\sqrt{N}})$.*

Proof: Letting $\underset{\sim}{u} = r\underset{\sim}{v}$, we get

$$H(r) = \int_{\|\underset{\sim}{v}\| \le 1} l(r\underset{\sim}{v}) d\underset{\sim}{v}.$$

Thus we see that $H(r) \longrightarrow l(\underset{\sim}{0})$ as $r \to 0$ and $H(r) \longrightarrow 0$ as $r \to \infty$. Since $H(r)$ is continuously differentiable and

$$\frac{dH(r)}{dr} = \frac{1}{V(r)} \frac{d}{dr} \int_{\|\underset{\sim}{u}\| \le r} l(\underset{\sim}{u}) d\underset{\sim}{u} - \frac{(t-1)}{rV(r)} \int_{\|\underset{\sim}{u}\| \le r} l(\underset{\sim}{u}) d\underset{\sim}{u},$$

we get all the extreme points of H by solving $\frac{dH(r)}{dr} = 0$. If 0 is the global maximum, the lemma is proved. Therefore, let $r_0 > 0$ be the global maximum of $H(r)$. Then, r_0 satisfies the following:

$$r_0 = \frac{(t-1) \int_{\|\underset{\sim}{u}\| \le r_0} l(\underset{\sim}{u}) d\underset{\sim}{u}}{\left(\frac{d}{dr} \int_{\|\underset{\sim}{u}\| \le r} l(\underset{\sim}{u}) d\underset{\sim}{u} \right) |_{r=r_0}}.$$

Also

$$\begin{aligned} \left(\frac{d}{dr} \int_{\|\underset{\sim}{u}\| \le r} l(\underset{\sim}{u}) d\underset{\sim}{u} \right) |_{r=r_0} &= \frac{(t-1)}{r_0} \int_{\|\underset{\sim}{u}\| \le r_0} l(\underset{\sim}{u}) d\underset{\sim}{u} \\ &\ge \frac{(t-1)}{r_0} V(r_0) l(\underset{\sim}{0}), \end{aligned}$$

since

$$\frac{1}{V(r_0)}\int_{\|\underset{\sim}{u}\|\le r_0} l(\underset{\sim}{u})d\underset{\sim}{u} \ge l(\underset{\sim}{0}).$$

Therefore

$$r_0 \le \frac{(t-1)\int_{\|\underset{\sim}{u}\|\le r_0} l(\underset{\sim}{u})d\underset{\sim}{u}}{\frac{(t-1)}{r_0}V(r_0)l(\underset{\sim}{0})},$$

and hence

$$V(r_0) \le \frac{\int l(\underset{\sim}{u})d\underset{\sim}{u}}{l(\underset{\sim}{0})}.$$

But

$$\begin{aligned}\int l(\underset{\sim}{u})d\underset{\sim}{u} &= \int \frac{N!}{\prod_{i=1}^{t} n_i!}\prod_{i=1}^{t} p_i^{n_i} d\underset{\sim}{p} \\ &= O(\frac{1}{N^{t-1}})\left(1+O(\frac{1}{\sqrt{N}})\right)\end{aligned}$$

and

$$l(\underset{\sim}{0}) = O\left(\frac{1}{N^{\frac{t-1}{2}}}\right)$$

from Result D given below. Therefore

$$V(r_0) = O\left(\frac{1}{N^{\frac{t-1}{2}}}\right).$$

Since

$$V(r_0) = C(t-1)r_0^{t-1},$$

where $C(t-1)$ is a constant independent of N,

$$r_0 = O(\frac{1}{\sqrt{N}}).$$

Therefore the global maximum is $O(\frac{1}{\sqrt{N}})$. This proves the Lemma. □

Result D. *If* $\frac{1}{\sqrt{N}}|\sum_{i=1}^{t}\left(\frac{n_i-Np_i}{\sqrt{Np_i}}\right)^3\frac{1}{\sqrt{p_i}}| \to 0$ *as* $N \to \infty$, *then*

$$\begin{aligned}\frac{N!}{\prod_{i=1}^{t} n_i!}\prod_{i=1}^{t} p_i^{n_i} &= \frac{1}{(2\pi)^{(t-1)/2}(\prod_{i=1}^{t} p_i)^{\frac{t-1}{2}} N^{\frac{t-1}{2}}}\exp(-\frac{N}{2}\sum_{i=1}^{t}(p_i-\frac{n_i}{N})^2\frac{1}{p_i}) \\ &\times\left[1+\frac{A(\underset{\sim}{p})}{N}+\frac{C(\underset{\sim}{p})}{\sqrt{N}}|\sum_{i=1}^{t}(\frac{n_i-Np_i}{\sqrt{Np_i}})^3\frac{1}{\sqrt{p_i}}|\right]\end{aligned}$$

for bounded functions A *and* C.

Proof: Exactly as in Feller(1968) for the Binomial case or Jeffreys (1961). □

Continuing with the proof of the theorem, let $\underset{\sim}{a}(p) = \sqrt{N}\underset{\sim}{u}(p)$,
$\underset{\sim}{Z} = \sqrt{N}B(\underset{\sim}{p}^0)(\frac{1}{N}\underset{\sim}{n} - \underset{\sim}{p}^0)$ and $\underset{\sim}{Y} = \sqrt{N}B(\underset{\sim}{p})(\frac{1}{N}\underset{\sim}{n} - \underset{\sim}{p}^0)$. Then we have the following result.

Result 4.1. *For* $\|\underset{\sim}{p} - \underset{\sim}{p}^0\| = O(\frac{1}{\sqrt{N}})$,

$$\frac{N!}{\prod_{i=1}^{t} n_i!}\prod_{i=1}^{t} p_i^{n_i} = \frac{1}{(2\pi N)^{(t-1)/2}(\prod_{i=1}^{t} p_i^0)^{1/2}}\exp(-\frac{1}{2}\|\underset{\sim}{a}(p)-\underset{\sim}{Z}\|^2)\left[1+\frac{F(p)}{\sqrt{N}}\right]$$

where F *is a bounded function.*

Proof: Using Result D,

$$\begin{aligned}\frac{N!}{\prod_{i=1}^{t} n_i!}\prod_{i=1}^{t} p_i^{n_i} &= \frac{1}{(2\pi)^{(t-1)/2}}\exp(-\frac{1}{2}\|\underset{\sim}{Y}\|^2)\frac{1}{(\prod_{i=1}^{t} p_i)^{1/2} N^{\frac{t-1}{2}}} \\ &\times\left[1+\frac{A(\underset{\sim}{p})}{N}+\frac{C(\underset{\sim}{p})}{\sqrt{N}}|\sum_{i=1}^{t}(\frac{n_i-Np_i}{\sqrt{Np_i}})^3\frac{1}{\sqrt{p_i}}|\right] \\ &= \frac{1}{(2\pi)^{(t-1)/2}}\exp(-\frac{1}{2}\|\underset{\sim}{a}(p)-\underset{\sim}{Z}\|^2)\frac{1}{(\prod_{i=1}^{t} p_i)^{1/2} N^{\frac{t-1}{2}}} \\ &\times\left[1+\frac{A(\underset{\sim}{p})}{N}+\frac{C(\underset{\sim}{p})}{\sqrt{N}}|\sum_{i=1}^{t}(\frac{n_i-Np_i}{\sqrt{Np_i}})^3\frac{1}{\sqrt{p_i}}|\right] \\ &\times\prod_{i=1}^{t}\left(\frac{p_i^0}{p_i}\right)^{1/2}\exp(\frac{1}{2}\left[\|\underset{\sim}{a}(p)-\underset{\sim}{Z}\|^2-\|\underset{\sim}{Y}\|^2\right]),\end{aligned}$$

where A and C are some bounded functions. Also,

$$\begin{aligned}
\|\underset{\sim}{a}(p)-\underset{\sim}{Z}\|^2-\|\underset{\sim}{Y}\|^2 &= (\underset{\sim}{a}(p)-\underset{\sim}{Z})'(\underset{\sim}{a}(p)-\underset{\sim}{Z})-\underset{\sim}{Y}'\underset{\sim}{Y} \\
&= N(\frac{1}{N}\underset{\sim}{n}-\underset{\sim}{p}^0)'\left[B(\underset{\sim}{p}^0)'B(\underset{\sim}{p}^0)-B(\underset{\sim}{p})'B(\underset{\sim}{p})\right](\frac{1}{N}\underset{\sim}{n}-\underset{\sim}{p}^0) \\
&\quad -2\sqrt{N}\underset{\sim}{a}(p)'\left[B(\underset{\sim}{p}^0)-B(\underset{\sim}{p})\right](\frac{1}{N}\underset{\sim}{n}-\underset{\sim}{p}^0) \\
&= \underset{\sim}{Z}'\left(I-B(\underset{\sim}{p}^0)'^{-1}B(\underset{\sim}{p})'B(\underset{\sim}{p})B(\underset{\sim}{p}^0)^{-1}\right)\underset{\sim}{Z} \\
&\quad -2\underset{\sim}{a}(p)'\left[I-B(\underset{\sim}{p})B(\underset{\sim}{p}^0)^{-1}\right]\underset{\sim}{Z}.
\end{aligned}$$

Therefore, it is enough to show that

$$B(\underset{\sim}{p})B(\underset{\sim}{p}^0)^{-1} = \left(1+O(\frac{1}{\sqrt{N}})\right)I, \tag{27}$$

for $\|\underset{\sim}{p}-\underset{\sim}{p}^0\|=O(\frac{1}{\sqrt{N}})$, since $\|\underset{\sim}{Z}\|^2$ is bounded. We shall prove (27) now.

Let $D(\underset{\sim}{p}^k)$ be the diagonal matrix with i th diagonal element equal to p_i^k. Then we can choose $B(\underset{\sim}{p})$ to be $\left(I+\frac{1}{\sqrt{p_t}+p_t}\phi(\underset{\sim}{p})\phi(\underset{\sim}{p})'\right)D(\underset{\sim}{p}^{-1/2})$ because

$$B(\underset{\sim}{p})'B(\underset{\sim}{p})=D\left(\underset{\sim}{p}^{-1/2}\right)\left(I+\frac{1}{\sqrt{p_t}+p_t}\phi(\underset{\sim}{p})\phi(\underset{\sim}{p})'\right)^2 D\left(\underset{\sim}{p}^{-1/2}\right)$$

and

$$\left(I+\frac{1}{\sqrt{p_t}+p_t}\phi(\underset{\sim}{p})\phi(\underset{\sim}{p})'\right)^2 = \left(I+\left(\frac{1}{\sqrt{p_t}+p_t}(2+\frac{1-\sqrt{p_t}}{\sqrt{p_t}})\right)\phi(\underset{\sim}{p})\phi(\underset{\sim}{p})'\right).$$

Therefore

$$\begin{aligned}
&\left(I+\frac{1}{\sqrt{p_t}+p_t}\phi(\underset{\sim}{p})\phi(\underset{\sim}{p})'\right)^2\left(I-\phi(\underset{\sim}{p})\phi(\underset{\sim}{p})'\right) \\
&= \left(I+\left(\frac{1}{\sqrt{p_t}+p_t}(2+\frac{1-\sqrt{p_t}}{\sqrt{p_t}})\right)\phi(\underset{\sim}{p})\phi(\underset{\sim}{p})'\right)\left(I-\phi(\underset{\sim}{p})\phi(\underset{\sim}{p})'\right) \\
&= \left(I+\left(\frac{1}{\sqrt{p_t}+p_t}(2+\frac{1-\sqrt{p_t}}{\sqrt{p_t}})p_t-1\right)\phi(\underset{\sim}{p})\phi(\underset{\sim}{p})'\right) \\
&= I,
\end{aligned}$$

which shows that

$$(B(\underset{\sim}{p})'B(\underset{\sim}{p}))^{-1} = D(\underset{\sim}{p}^{1/2})\left(I - \phi(\underset{\sim}{p})\phi(\underset{\sim}{p})'\right)D(\underset{\sim}{p}^{1/2}).$$

Furthermore,

$$B(\underset{\sim}{p})^{-1} = D(\underset{\sim}{p}^{1/2})\left(I - \left(\frac{1}{1+\sqrt{p_t}}\right)\phi(\underset{\sim}{p})\phi(\underset{\sim}{p})'\right).$$

Therefore defining $p_i^1 = (p_i^0/p_i)$,

$$\begin{aligned}
B(\underset{\sim}{p})B(\underset{\sim}{p}^0)^{-1} &= \left(I + \frac{1}{\sqrt{p_t}+p_t}\phi(\underset{\sim}{p})\phi(\underset{\sim}{p})'\right)D(\underset{\sim}{p}^{-1/2}) \\
&\quad \times D((\underset{\sim}{p}^0)^{1/2})\left(I - \left(\frac{1}{1+\sqrt{p_t^0}}\right)\phi(\underset{\sim}{p}^0)\phi(\underset{\sim}{p}^0)'\right) \\
&= D((\underset{\sim}{p}^1)^{1/2}) - D((\underset{\sim}{p}^1)^{1/2})\frac{1}{1+\sqrt{p_t^0}}\phi(\underset{\sim}{p}^0)\phi(\underset{\sim}{p}^0)' \\
&\quad + \frac{1}{\sqrt{p_t}+p_t}\phi(\underset{\sim}{p})\phi(\underset{\sim}{p})'D((\underset{\sim}{p}^1)^{1/2}) \\
&\quad - \frac{1}{(\sqrt{p_t}+p_t)(1+\sqrt{p_t^0})}\phi(\underset{\sim}{p})\phi(\underset{\sim}{p})'D(\underset{\sim}{p}^1)\phi(\underset{\sim}{p}^0)\phi(\underset{\sim}{p}^0)' \\
&= D((\underset{\sim}{p}^1)^{1/2}) + \left(\frac{\sqrt{p_t^0}}{\sqrt{p_t}+p_t}I - \frac{1}{\sqrt{p_t^0}+1}D(\underset{\sim}{p}^1)\right)\phi(\underset{\sim}{p})\phi(\underset{\sim}{p}^0)'.
\end{aligned}$$

Now

$$\begin{aligned}
\sqrt{p_i/p_i^0} &= \sqrt{\frac{p_i^0 + (p_i - p_i^0)}{p_i^0}} \\
&= \sqrt{1 + O(\frac{1}{\sqrt{N}})}
\end{aligned}$$

because $|p_i - p_i^0| = O(\frac{1}{\sqrt{N}})$. Therefore

$$D((\underset{\sim}{p}^1)^{1/2}) = \left(1 + O(\frac{1}{\sqrt{N}})\right)I$$

and

$$D(\underset{\sim}{p}^1) = \left(1 + O(\frac{1}{\sqrt{N}})\right) I.$$

Therefore

$$\begin{aligned}
&\left(\frac{\sqrt{p_t^0}}{\sqrt{p_t} + p_t} I - \frac{1}{\sqrt{p_t^0} + 1} D(\underset{\sim}{p}^1)\right) \phi(\underset{\sim}{p})\phi(\underset{\sim}{p}^0)' \\
= &\left(\frac{\sqrt{p_t^0}}{\sqrt{p_t} + p_t} I - \frac{1}{\sqrt{p_t^0} + 1}\left(1 + O(\frac{1}{\sqrt{N}})\right) I\right) \phi(\underset{\sim}{p})\phi(\underset{\sim}{p}^0)' \\
= &\left((\frac{\sqrt{p_t^0}}{\sqrt{p_t} + p_t} - \frac{1}{\sqrt{p_t^0} + 1}) I + O(\frac{1}{\sqrt{N}}) I\right) \phi(\underset{\sim}{p})\phi(\underset{\sim}{p}^0)' \\
= &\frac{\sqrt{p_t^0} + p_t^0 - \sqrt{p_t} - p_t}{(1 + \sqrt{p_t^0})(\sqrt{p_t} + p_t)} \phi(\underset{\sim}{p})\phi(\underset{\sim}{p}^0)' + O(\frac{1}{\sqrt{N}})\phi(\underset{\sim}{p})\phi(\underset{\sim}{p}^0)' \\
= &O(\frac{1}{\sqrt{N}})\phi(\underset{\sim}{p})\phi(\underset{\sim}{p}^0)'
\end{aligned}$$

Since $\|\phi(\underset{\sim}{p})\| < 1$ and $\|\phi(\underset{\sim}{p}^0)\| < 1$, (27) is proved. □

Using the above lemmas, it is now possible to conclude the proof of the theorem. Let $\underset{\sim}{v} = \sqrt{N}\underset{\sim}{u}$. Then

$$\begin{aligned}
&\frac{\sup_r \frac{1}{V(r)} \int_{\|\underset{\sim}{u}\| \le r} l(\underset{\sim}{u}) d\underset{\sim}{u}}{\frac{N!}{\prod_{i=1}^t n_i!} \prod_{i=1}^t p_i^{0\,n_i}} \\
= &\frac{\sup_r \frac{1}{V(r)} \int_{\|\underset{\sim}{v}\| \le \sqrt{N} r} \exp(-\frac{1}{2}\|\underset{\sim}{v} - \underset{\sim}{Z}\|^2) \left(1 + \frac{F(\underset{\sim}{v})}{\sqrt{N}}\right) \frac{1}{\sqrt{N}} d\underset{\sim}{v}}{\exp(-\frac{1}{2}K) \left(1 + \frac{A}{n} + \frac{B}{\sqrt{N}} K^{3/2}\right)} \\
= &\frac{\sup_r \frac{1}{V(r)\sqrt{N}} \int_{\|\underset{\sim}{v}\| \le \sqrt{N} r} \exp(-\frac{1}{2}\|\underset{\sim}{v} - \underset{\sim}{Z}\|^2) \left(1 + \frac{F(\underset{\sim}{v})}{\sqrt{N}}\right) d\underset{\sim}{v}}{\exp(-\frac{1}{2}K) \left(1 + \frac{A}{n} + \frac{B}{\sqrt{N}} K^{3/2}\right)}
\end{aligned}$$

$$= \frac{\sup_s \frac{1}{V(s)} \int_{\|\underset{\sim}{w}\| \le s} \exp(-\frac{1}{2}\|\underset{\sim}{w} - \underset{\sim}{Z}\|^2) \left(1 + \frac{F(\underset{\sim}{w})}{\sqrt{N}}\right) d\underset{\sim}{w}}{\exp(-\frac{1}{2}K)\left(1 + \frac{A}{n} + \frac{B}{\sqrt{N}} K^{3/2}\right)}.$$

The denominator of the last ratio converges to $\exp(-\frac{1}{2}K)$ as $N \to \infty$. It is, therefore, now enough to show that

$$\sup_s \frac{1}{V(s)} \int_{\|\underset{\sim}{w}\| \le s} \exp(-\frac{1}{2}\|\underset{\sim}{w} - \underset{\sim}{Z}\|^2) \left(1 + \frac{F(\underset{\sim}{w})}{\sqrt{N}}\right) d\underset{\sim}{w}$$

$$\longrightarrow \sup_s \frac{1}{V(s)} \int_{\|\underset{\sim}{w}\| \le s} \exp(-\frac{1}{2}\|\underset{\sim}{w} - \underset{\sim}{S}\|^2) d\underset{\sim}{w},$$

where $\underset{\sim}{S}$ is any vector such that $\|\underset{\sim}{S}\|^2 = K$. Let

$$h_N(s) = \frac{1}{V(s)} \int_{\|\underset{\sim}{w}\| \le s} \exp(-\frac{1}{2}\|\underset{\sim}{w} - \underset{\sim}{Z}\|^2) \left(1 + \frac{F(\underset{\sim}{w})}{\sqrt{N}}\right) d\underset{\sim}{w},$$

$$h_N^1(s) = \frac{1}{V(s)} \int_{\|\underset{\sim}{w}\| \le s} \exp(-\frac{1}{2}\|\underset{\sim}{w} - \underset{\sim}{Z}\|^2)\, d\underset{\sim}{w},$$

and

$$h(s) = \frac{1}{V(s)} \int_{\|\underset{\sim}{w}\| \le s} \exp(-\frac{1}{2}\|\underset{\sim}{w} - \underset{\sim}{S}\|^2) d\underset{\sim}{w}.$$

It can be shown that (see Result 7 of Appendix D in Delampady(1986a)) the global maximum of $h(s)$ is attained for a value of s which is less than $\sqrt{K}$. Moreover, in this range of values of s, $h_N \longrightarrow h$ uniformly. This is because

$$\begin{aligned}
&\max_s |h_N(s) - h(s)| \\
&\le \max_s |h_N(s) - h_N^1(s)| + \max_s |h_N^1(s) - h(s)| \\
&= \frac{1}{\sqrt{N}} \max_s \left| \frac{1}{V(s)} \int_{\|\underset{\sim}{w}\| \le s} F(\underset{\sim}{w}) \exp(-\frac{1}{2}\|\underset{\sim}{w} - \underset{\sim}{Z}\|^2) d\underset{\sim}{w} \right| \\
&\quad + \max_s |h_N^1(s) - h(s)| \\
&\le \frac{1}{\sqrt{N}} \max_{\underset{\sim}{w}} |F(\underset{\sim}{w})| \max_s \frac{1}{V(s)} \int_{\|\underset{\sim}{w}\| \le s} \exp(-\frac{1}{2}\|\underset{\sim}{w} - \underset{\sim}{Z}\|^2) d\underset{\sim}{w}
\end{aligned}$$

$$
\begin{aligned}
& + \max_s |h_N^1(s) - h(s)| \\
= \quad & \frac{1}{\sqrt{N}} \max_{\underset{\sim}{w}} |F(\underset{\sim}{w})| \max_s h_N^1(s) + \max_s |h_N^1(s) - h(s)| \\
\longrightarrow \quad & 0 \text{ as } N \to \infty,
\end{aligned}
$$

the last step following from the fact that h_N^1 is bounded and $h_N^1 \longrightarrow h$ uniformly. ($h_N^1(s)$ is a function of $\|\underset{\sim}{Z}\|^2$ which differs from $h(s)$ by at most $O(\frac{1}{\sqrt{N}})$ uniformly.) Finally, $\lim_{N\to\infty} \sup_s h_N(s) = \sup_s h(s)$. This follows because uniform convergence is equivalent to convergence in supremum norm and hence, if two functions are ϵ apart in supremum norm, their maxima cannot be more than ϵ apart. This concludes the proof. □