

An Optimality Property of Bayes' Test Statistics

Raghu Raj Bahadur and Peter J. Bickel^{1,*}

University of Chicago and Imperial College, London

Dedicated to Erich Lehmann on his 90th Birthday

Contents

1	Introduction	19
2	A Generalization of a Theorem of Bahadur	19
3	General Assumptions and a Useful Lemma	21
4	The Main Theorem	26
5	Optimality of Minimax Tests	29
6	References	30

Preface:

This paper dates back to the late 60's when I collaborated with Raj Bahadur, who is unfortunately no longer with us. The reason it has not appeared until now is that he felt it had to be accompanied by a number of multivariate examples. We both went on to other things; the examples were not worked out although we both knew of the existence of some of them. So why is this paper appearing here (with the approval of Steve Stigler, an executor of the Bahadur estate)? First, in addition to attesting to Erich's continued vital presence, it gives me the opportunity of paying a tribute to Bahadur, who was a friend of both of ours. Second, it is an interesting reminder of how writing styles have changed on the whole I think for the better – from rigorous abstract formulation and mathematically rigorous presentation to more motivation and a lot of hand waving. Third, and most importantly, the result is an example of what I think both Erich and I consider an important endeavor, the reconciliation of the Bayesian and frequentist points of view (in context of now rather unfamiliar asymptotics). In an important paper in the 5th Berkeley Symposium [4], Bahadur showed that the maximum likelihood ratio statistic possessed an optimality property from the view of a large deviation based frequentist comparison of tests he introduced in 1960 [1]. Our paper shows that this property is shared by Bayes test statistics for reasonable priors and conjectures that a corresponding Bayesian optimality property holds for the maximum likelihood ratio statistic. If true this can be viewed as the large deviation analogue of the well-known Bernstein von Mises' theorem – see Lehmann and Casella [10] p.489, which establishes the equivalence at the $n^{-\frac{1}{2}}$ scale of Bayesian and maximum likelihood estimates. Establishing this conjecture is left as a challenge to the reader.

¹On leave from University of California, Berkeley (1965-66).

*Prepared with partial support of N.S.F. Grant G.P. 2593.

Given the historical interest I have not changed the text save for typos and only brought references up to date.

1. Introduction

In [4] one of the authors established the optimality of the classical likelihood ratio test statistic in terms of a method of stochastic comparison previously introduced by him in [1], [2], and [3].

In the main theorem of this paper, Theorem 2 of Section 4, we show that this property is shared by Bayes test statistics (averages of likelihood ratios with respect to probability measures on the parameter space) under conditions which are slightly different from, and in some respects weaker than those given in [4]. These assumptions are given and discussed in Section 3. Section 5 contains a theorem establishing the asymptotic optimality of minimax tests under appropriate restrictions.

In Section 2 we give a strengthening Theorem 1 of [4], which established a lower bound for the slope of any family of tests in terms of the Kullback-Leibler information numbers. The proof given here drops assumption 1 of [4] and weakens assumption 2 considerably. This argument seems to give some insight into the necessity of an assumption such as our modification of assumption 2 of [4].

2. A Generalization of a Theorem of Bahadur

Even as in [4] we let X be an abstract space, \mathcal{A} a field on X , P_θ , $\theta \in \Theta$, a set of probability measures on (X, \mathcal{A}) , and Θ_0 a given subset of Θ . For any θ, θ' we define,

$$(2.1) \quad K(\theta, \theta') = - \int_X \log \frac{dP_{\theta'}}{dP_\theta}(x) dP_\theta(x) ,$$

where $\frac{dP_{\theta'}}{dP_\theta}$ is the ratio of the Radon Nikodym derivatives of $P_\theta, P_{\theta'}$ with respect to (say) $P_\theta + P_{\theta'}$ and $0/0$ is by convention equal to 1. Also let,

$$(2.2) \quad J(\theta) = \inf \{ K(\theta, \theta') : \theta' \in \Theta_0 \} .$$

It is well known that (cf. [4]), $0 \leq K(\theta, \theta') \leq \infty$, and necessarily the same is true of $J(\theta)$. Following [4], let T_n be any sequence of extended real valued measurable functions of the infinite product space $(X^\infty, \mathcal{A}^\infty)$ such that T_n is a function of the first n co-ordinates only. Denote the cumulative distribution of T_n , when θ obtains, by $F_n(t, \theta)$, i.e.,

$$(2.3) \quad F_n(t, \theta) = P_\theta [T_n(s) < t] ,$$

where P_θ now denotes the infinite product measure extension of P_θ to X^∞ . Finally, let,

$$(2.4) \quad L_n(s) = \sup \{ 1 - F_n(T_n(s), \theta) : \theta \in \Theta_0 \} .$$

We assume that L_n is measurable. This for instance holds if $F_n(T_n, \theta)$ is a separable stochastic process in θ for $\theta \in \Theta_0$.

We can now state and prove, in the above framework,

Theorem 1. *If*

$$\int_X \left[\log \left(\frac{dP_\theta}{dP_{\theta'}} \right) \right]^2 dP_\theta(x) < \infty ,$$

for every $\theta \in \Theta - \Theta_0$, $\theta' \in \Theta_0$ such that $K(\theta, \theta') < \infty$, then

$$(2.5) \quad \liminf_n \frac{1}{n} \log L_n(s) \geq -J(\theta)$$

with P_θ probability 1 for every $\theta \in \Theta - \Theta_0$.

Proof. Fix $\theta \in \Theta - \Theta_0$. Assume the Theorem has been proved for Θ_0 simple. Clearly we can suppose $J(\theta) < \infty$ and can find $\{\theta_m\}$ with $K(\theta, \theta_m) < \infty$ and $K(\theta, \theta_m) \rightarrow J(\theta)$. But then

$$(2.6) \quad \frac{1}{n} \log L_n(s) \geq \frac{1}{n} \log (1 - F_n(T_n(s), \theta_m)) .$$

By our assumption of the theorem for Θ_0 simple, we have

$$(2.7) \quad \liminf_n \frac{1}{n} \log (1 - F_n(T_n(s), \theta_m)) \geq -K(\theta, \theta_m)$$

with probability 1. Inequalities (2.6) and (2.7) then imply (2.5). If $\Theta_0 = \{\theta_0\}$,

$$P_\theta \left[\liminf_n \frac{1}{n} \log L_n(s) \geq -J(\theta) \right] = 1$$

if and only if,

$$(2.8) \quad P_\theta [1 - F_n(T_n, \theta_0) < a^n \exp -nK(\theta, \theta_0) \text{ infinitely often}] = 0$$

for every $0 \leq a \leq 1$. Fix a . Let $A_n = [1 - F_n(T_n, \theta_0) < a^n \exp\{-nK(\theta, \theta_0)\}]$. Then,

$$(2.9) \quad P_{\theta_0}(A_n) \leq a^n \exp\{-nK(\theta, \theta_0)\} .$$

By the Neyman-Pearson lemma there exists c_n , such that,

$$(2.10) \quad P_{\theta_0} \left\{ \sum_{j=1}^n \log \frac{dP_\theta}{dP_{\theta_0}}(x_j) > nc_n \right\} \leq a^n \exp\{-nK(\theta, \theta_0)\}$$

and,

$$(2.11) \quad P_{\theta_0} \left\{ \sum_{i=1}^n \log \frac{dP_\theta}{dP_{\theta_0}}(x_i) > nc_n \right\} \geq P_{\theta_0}(A_n) .$$

We require,

Lemma 1. *If (2.10) holds for all n , then there exists an $\varepsilon > 0$ such that $\liminf_n c_n \geq K(\theta, \theta_0) + \varepsilon$.*

Proof. By a Theorem of Chernoff [7],

$$(2.12) \quad \frac{1}{n} \log P_{\theta_0} \left\{ \sum_{i=1}^n \log \frac{dP_\theta}{dP_{\theta_0}}(x_i) > nz \right\} \rightarrow \inf_t \log H(t, z) ,$$

for

$$z \geq \int_X \log \frac{dP_\theta}{dP_{\theta_0}}(x) dP_{\theta_0}(x) ,$$

where

$$H(t, z) = e^{-tz} \int_X \left(\frac{dP_\theta}{dP_{\theta_0}} \right)^t(x) dP_{\theta_0}(x) .$$

By the theory of the Laplace transform, $0 \leq H(t, z) \leq \infty$, $H(t, z)$ is strictly convex in t wherever it is finite and if $\inf_t H(t, z) < 1$, the infimum is obtained for a unique $t(z)$ given by the solution of the equation

$$(2.13) \quad z = \frac{\int \log \frac{dP_\theta}{dP_{\theta_0}}(x) \left[\frac{dP_\theta}{dP_{\theta_0}}(x) \right]^t dP_{\theta_0}(x)}{\int \left[\frac{dP_\theta}{dP_{\theta_0}}(x) \right]^t dP_{\theta_0}(x)} .$$

It is easily seen that if $z_0 = K(\theta, \theta_0)$, then $t(z_0) = 1$, and,

$$(2.14) \quad \inf_t \log H(t, z_0) = -K(\theta, \theta_0) .$$

From (2.10), (2.12) and (2.14) we can immediately conclude that $\liminf_n c_n \geq K(\theta, \theta_0)$. But, in fact, by the implicit function theorem as $z \rightarrow K(\theta, \theta_0)$ we have $t(z) \rightarrow 1$, and $\log H(z, t(z)) \rightarrow K(\theta, \theta_0)$, by dominated convergence. Choose $z_1 > z_0$ such that $H(z_1, t(z_1)) > a \exp\{-K(\theta, \theta_0)\}$. Then $\varepsilon = z_1 - z_0$ will do for the Lemma.

It now follows from the basic assumption of the theorem, by a result of Erdős, Hsu and Robbins [9] that,

$$(2.15) \quad \sum_n P_{\theta_0} \left[\sum_{i=1}^n \log \frac{dP_{\theta}}{dP_{\theta_0}}(x_i) \geq nc_n \right] < \infty ,$$

and this by (2.11) and the Borel Cantelli lemma suffices for (2.8) and the theorem to hold.

Remarks. 1. Erdős has shown in [9] that our second moment assumption is necessary as well as sufficient for (2.15) to hold. Although, of course, (2.15) is not necessary for (2.8) the relative arbitrariness of the A_n apart from condition (2.9) would suggest that the Theorem may be false if some condition such as the one imposed does not hold.

2. As in [4], if we define $N(\varepsilon, s) =$ least positive m such that $L_n \geq \varepsilon$ for all $n \geq m$ and ∞ otherwise, we have under the assumptions of our theorem 1,

$$(2.16) \quad \liminf_{\varepsilon \rightarrow 0} \frac{N(\varepsilon, s)}{-\log \varepsilon} \geq \frac{1}{J(\theta)} \text{ a.s. } P_\theta .$$

3. General Assumptions and a Useful Lemma

Before giving further structural assumptions needed for Sections 4 and 5 we prove a simple general lemma already implicit in [4] stating a useful sufficient condition for a sequence $\{T_n\}$ to be optimal. We say $\{T_n\}$ is asymptotically optimal if,

$$(3.1) \quad \lim_n \frac{1}{n} \log L_n(s) = -J(\theta)$$

with P_θ probability 1 for all $\theta \in \Theta - \Theta_0$. Then (3.1) implies, (cf. [4])

$$(3.2) \quad \lim_{\varepsilon \rightarrow 0} \frac{N(\varepsilon, s)}{-\log \varepsilon} = \frac{1}{J(\theta)} \text{ a.s. } P_\theta .$$

Lemma 2. *If the conclusion of Theorem 1 holds and*

$$i) \liminf_n T_n \geq J(\theta) \text{ a.s. } P_\theta$$

$$ii) \limsup_n \log(1 - G_n(t)) \leq -t$$

where $G_n(t) = \inf \{F_n(t, \theta_0) : \theta_0 \in \Theta_0\}$ and θ ranges over $\Theta - \Theta_0$, then $\{T_n\}$ is asymptotically optimal.

Proof. It clearly suffices to show that

$$(3.3) \quad \limsup_n \frac{1}{n} \log L_n(s) \leq -J(\theta)$$

with P_θ probability 1. But since $L_n(s) = 1 - G_n(T_n)$ and $1 - G_n(t)$ is monotone decreasing, i) and ii) obviously imply (3.3).

We begin by giving nine general assumptions which are sufficient to ensure the validity of Theorem 2 of the main section.

Assumption 1. There exists a c finite measure μ on (X, \mathcal{A}) which dominates the family $\{P_\theta\}$. We denote the density of P_θ with respect to μ by $f(x, \theta)$. Then,

$$\frac{dP_\theta(x)}{dP_{\theta'}} = \frac{f(x, \theta)}{f(x, \theta')} \text{ a.e. } P_\theta + P_{\theta'} .$$

Assumption 2. Θ is a metric space. The topological Borel field on Θ is denoted by \mathcal{B} . $f(x, \theta)$ is bimeasurable in (x, θ) on $(X \times \Theta, \mathcal{A} \times \mathcal{B})$.

Assumption 3. We are given a probability measure ν on (Θ, \mathcal{B}) , $\Theta_0 \in \mathcal{B}$, and $\nu(\Theta_0) > 0$. Moreover, if $S(\theta, d)$ is the open sphere of centre θ and radius d , $\nu\{S(\theta, d) \cap [\Theta - \Theta_0]\} > 0$ for all $\theta \in \Theta - \Theta_0$ and $d > 0$.

Assumption 4. There exists a suitable metric compactification $\bar{\Theta}_0$ of Θ_0 (viz [4]). That is, we first define $\hat{S}(\theta, d)$ to be the sphere of radius d and centre θ in $\bar{\Theta}_0$ and then take,

$$(3.4) \quad g_0(x, \theta, d) = \sup \{f(x, \lambda) : \lambda \in \hat{S}(\theta, d) \cap \Theta_0\} .$$

We assume g_0 is measurable in x for d sufficiently small and define,

$$(3.5) \quad g_0(x, \theta, 0) = \lim_{d \rightarrow 0} g_0(x, \theta, d) .$$

The final assumption, (see [4]) is,

$$(3.6) \quad E_\theta \left(\frac{g(x, \theta', 0)}{f(x, \theta)} \right) \leq 1 ,$$

1 where $E_\theta(h(x))$ denotes $\int_X h(x)dP_\theta(x)$ for any integrable function h . 1

2
3 **Assumption 5.** Define for all $\theta' \in \bar{\Theta}_0$, $\theta \in \Theta - \Theta_0$, 3

$$4 \quad (3.7) \quad \bar{K}(\theta, \theta') = -E_\theta \left(\frac{\log g_0(x, \theta', 0)}{f(x, \theta)} \right). \quad 4$$

5
6
7 (3.6) and Jensen's inequality guarantee $0 \leq \bar{K} \leq \infty$. Assume, 7

$$8 \quad (3.8) \quad J(\theta) = \inf \{K(\theta, \theta') : \theta' \in \bar{\Theta}_0\}. \quad 8$$

9
10
11
12 **Assumption 6.** 12

$$13 \quad (3.9) \quad E_\theta \left(\log \frac{g_0(x, \theta', d)}{f(x, \theta)} \right) < \infty, \quad 13$$

14
15
16 for all $\theta \in \Theta - \Theta_0$, $\theta' \in \Theta_0$. As in, [4] p.22, this is equivalent to, 16

$$17 \quad (3.10) \quad \lim_{d \rightarrow 0} E_\theta \left(\frac{\log g_0(x, \theta', d)}{f(x, \theta)} \right) \leq -K(\theta, \theta'). \quad 17$$

18
19
20
21
22 **Assumption 7.** Define, 22

$$23 \quad (3.11) \quad \eta(x, \theta, d) = \inf \left\{ \log \frac{f(x, \lambda)}{f(x, \theta)} : \lambda \in S(\theta, d) \cap [\Theta - \Theta_0] \right\}. \quad 23$$

24
25
26 Assume that η is a measurable function of x for d sufficiently small and that, 26

$$27 \quad (3.12) \quad \lim_{d \rightarrow 0} E_\theta(\eta(x, \theta, d)) = 0 \quad 27$$

28
29
30 for all $\theta \in \Theta - \Theta_0$. 30

31
32 **Assumption 8.** Define, for $\theta' \in \Theta_0$, 32

$$33 \quad (3.13) \quad \gamma(x, \theta', d) = \log \inf \left\{ \frac{f(x, \lambda)}{f(x, \theta')} : \lambda \in S(\theta', d) \cap \Theta_0 \right\}, \quad 33$$

$$34 \quad (3.14) \quad \varphi(t, \theta', d) = E_{\theta'}(\exp\{-t\gamma(x, \theta', d)\}). \quad 34$$

35
36
37 For every $0 < \rho < 1$, $\beta > 0$, there exists $d(\theta', \rho, \beta)$ such that 37

$$38 \quad \inf_t e^{-t\beta} \varphi(t, \theta', d(\theta', \rho, \beta)) \leq \frac{\rho}{2}. \quad 38$$

39
40
41
42
43
44 **Assumption 9.** 44

$$45 \quad \inf \{ \nu[S(\theta', d(\theta', \rho, \beta)) \cap \Theta_0 : \theta' \in \Theta_0] \} = m(\rho, \beta) > 0. \quad 45$$

46
47
48 We shall now examine these assumptions in turn giving where necessary stronger, 48

49 but more easily checkable conditions, which we shall denote by primes. Thus, (say) 49

50 assumption 4' will imply assumption 4 and the conclusion of theorem 2 will continue 50

51 51

to hold if 4 is replaced by 4'. The more important of these useful weakenings of Theorem 2 will be isolated as a Corollary.

Assumption 1 is self-explanatory and clearly cannot be weakened appreciably.

The requirement that Θ be a metric space can clearly be dropped and replaced by the requirement that Θ be a topological space and \mathcal{B} the topological Borel field. However, the notational convenience involved in being able to define quantities in terms of spheres of a given radius rather than neighbourhood bases seems well worth the loss of generality. On the other hand, assumption 2 is obviously satisfied if we have the usual,

Assumption 2'. Θ is a subset of k dimensional Euclidean space with the usual metric topology. \mathcal{B} is the Borel σ field and $f(x, \theta)$ is bimeasurable.

Weakenings of assumption 3 do not fit readily into this program, but we mention that we can drop the requirement that ν be a probability (finite) measure if the following two conditions hold, as well as the second part of assumption 3:

- (1) There exists N such that, $\int_{\Theta} \prod_{i=1}^N f(x_i, \lambda) \nu(d\lambda) < \infty$ a.s. P_{θ} and
- (2) $\int_{\Theta_0} \prod_{i=1}^N f(x_i, \lambda) \nu(d\lambda) > 0$, $\int_{\Theta - \Theta_0} \prod_{i=1}^N f(x_i, \lambda) \nu(d\lambda) > 0$ a.s. P_{θ} for all $n \geq N$.

For details of the proof of i) of Lemma 2 for \bar{T}_n under these assumptions we refer to [6]. The basic idea of this generalization is to consider the process of observation as really starting after N with prior distribution, the posterior distribution of θ given x_1, \dots, x_N , which by (1) is a true probability distribution. In fact, we can in general make dependent ν on the observations all along if we modify the second part of assumption 3 and assumption 9 suitably. The generalization given above is of interest in the case when reasonable tests arise from improper "priors", e.g. Lebesgue measure.

The most natural replacement of assumption 4 is of course assuming that Θ_0 is already compact. In this case, we have, **Assumption 4'**. Θ_0 is compact.

We can then drop 5, but must replace 6 by its equivalent form,

Assumption 6'. $\lim_{d \rightarrow 0} E_{\theta} \log \left(\frac{g_0(x, \theta', d)}{\rho(x, \theta)} \right) \leq -K(\theta, \theta')$ for all $\theta \in \Theta - \Theta_0$, $\theta' \in \Theta_0$.

Measurability of g_0 must still of course be invoked. Assumption 6 may replace 6' if $f(x, \theta)$ is continuous in θ for almost all x . A less stringent modification in some senses which is most useful is combining 4, 5, and 6 with 2' to give:

Assumption (4,5,6)'' Assumption 2' holds and

- (a) $\lim_{d \rightarrow 0} E_{\theta} \left(\log \frac{g_0(x, \theta', d)}{f(x, \theta)} \right) \leq -K(\theta, \theta')$ for $\theta \in \Theta - \Theta_0$, $\theta' \in \Theta_0$.
- (b) $\lim_{d \rightarrow \infty} E_{\theta} \left(\log \sup \left\{ \log \frac{f(x_{\lambda}, \lambda)}{f(x, \theta)} : \lambda \in \Theta_0, \|\lambda\| \geq d \right\} \right) \leq -J(\theta)$, for all $\theta \in \Theta - \Theta_0$, where $\|\cdot\|$ is the usual Euclidean norm.

This assumption is clearly equivalent to 5 and 6 if in 4 we take $\bar{\Theta}_0$ to be the closure of Θ_0 in the one point compactification of R .

Assumption 7 is most readily replaced by,

Assumption 7'. $f(x, \theta)$ is continuous in θ for almost all $x(\mu)$ and, $E_{\theta}(\eta(x, \theta, d)) < \infty$ for some $d > 0$ for each $\theta \in \Theta - \Theta_0$.

Assumption 7' and the dominated convergence theorem readily imply 7. We need not require measurability of g_0 in this case in view of 4 or 2' since Θ_0 being a subset of a separable metric space is separable. The same is true of η if 2' holds or more generally if Θ is a separable metric space.

A useful substitute for assumption 8 is

Assumption 8'. There exists an $M < \infty$ such that for every $0 < T < \infty$ we can find a $d^x(\theta, T)$ with $\varphi(T, \theta', d^x) \leq M$. Assumption 8' easily implies 8 since, then,

$$(3.15) \quad \inf_t e^{-t\beta} \varphi(T, \theta', d^x) \leq e^{-T\beta} M < \frac{\rho}{2},$$

for T sufficiently large.

In many situations 8' is most easily verified by showing that,

$$(3.16) \quad \varphi(t, \theta', d) \rightarrow 1,$$

uniformly on compacts in t as $d \rightarrow 0$. This in turn is implied by,

Assumption 8''. $f(x, \theta)$ is continuous in θ for almost all $x(\mu)$ and for every $0 < T < \infty$, $\varphi(T, \theta', d) < \infty$ for d sufficiently small.

Assumption 8'' implies (3.16) by way of the dominated convergence theorem if we remark that $\varphi(t, \theta', d)$ is monotone increasing in t for every fixed θ', d .

Finally, we can replace assumption 9 by,

Assumption 9'. $d(\theta, \rho, \beta)$ is independent of θ' , assumption 4' holds, and $\nu[S(\theta', d) \cap \Theta_0] > 0$ for all $\theta' \in \Theta_0$.

To show that 9' implies 9 we need only prove that,

$$(3.17) \quad \inf \{ \nu[S(\theta', d) \cap \Theta_0] : \theta' \in \Theta_0 \} > 0.$$

Then, if $\theta'_n \rightarrow \theta'$ and, $S^*(\theta, d) = S(\theta, d) \cap \Theta_0$

$$(3.18) \quad \begin{aligned} \nu[S^*(\theta', d)] - \nu[S^*(\theta'_n, d)] &= \nu[\lambda : \delta(\lambda, \theta') < d, \quad \delta(\lambda, \theta'_n) \geq d, \quad \lambda \in \Theta_0] \\ &\quad - \nu[\lambda : \delta(\lambda, \theta') \geq d, \quad \delta(\lambda, \theta'_n) < d, \quad \lambda \in \Theta_0]. \end{aligned}$$

Clearly, if $n \rightarrow \infty$, the first term of the above difference tends to 0 since the set whose measure is computed tends to the empty set. Therefore, for fixed d , $\nu[S^*(\theta', d)]$ is a lower semi-continuous function of θ' on Θ_0 and 2', and the third part of 9', imply (3.17).

This completes our roster of simplifying assumptions. Clearly if further restrictions are put on $f(x, \theta)$ the verification of most can be very easy. A strong form of Lemma 3, $\tau_n \rightarrow J(\theta)$, is given under very weak assumptions in [5] if $f(x, \theta)$ is of exponential type, (Theorem 4.1). If the conditions 1-7 of this paper are made two sided (viz. [5] Theorem 4.2) one can obtain this strengthening of Lemma 3 in general. In fact, the conditions detailed in theorem 4.2 of [5] are somewhat restrictive versions of our conditions 2'-7'.

The assumptions required by this paper but not by [4] are 7, 8, and 9. On the other hand, assumption 6 of that paper and the fact that Θ can be suitably compactified (rather than just Θ_0) is not required by us. Since simple hypotheses tend to be somewhat more common than simple alternatives this seems a gain. Otherwise the non structural assumptions of this paper and [4] coincide.

4. The Main Theorem

Given ν in assumption 3 we now define,

$$(4.1) \quad \bar{T}_n = \frac{1}{n} \log \frac{\int_{\Theta - \Theta_0} \prod_{i=1}^n f(x_i, \lambda) \nu(d\lambda)}{\int_{\Theta_0} \prod_{i=1}^n f(x_i, \lambda) \nu(d\lambda)} .$$

By assumption 2, \bar{T}_n is well defined ($\frac{\infty}{\infty} = 1$, $\frac{0}{0} = 1$). In fact, \bar{T}_n is a version of the test statistic a Bayesian with prior ν would use to test $H : \theta \in \Theta_0$, rejecting for large values of \bar{T}_n . We can now state the principal theorem of the paper.

Theorem 2. *If assumptions 1–9 and the conclusion of Theorem 1 holds, then $\{\bar{T}_n\}$ is asymptotically optimal.*

Proof. The proof proceeds by way of some lemmas. We have first,

Lemma 3. *Under assumptions 1–7, $\{\bar{T}_n\}$ satisfies condition i) of Lemma 2.*

Proof. Suppose that $\theta \in \Theta - \Theta_0$ holds. Define,

$$(4.2) \quad U_n(s, \theta) = \int_{\Theta - \Theta_0} \prod_{i=1}^n \frac{f(x_i, \lambda)}{f(x_i, \theta)} \nu(d\lambda)$$

$$(4.3) \quad V_n(s, \theta) = \int_{\Theta_0} \prod_{i=1}^n \frac{f(x_i, \lambda)}{f(x_i, \theta)} \nu(d\lambda) .$$

Then,

$$(4.4) \quad \bar{T}_n = \frac{1}{n} \log \frac{U_n(s, \theta)}{V_n(s, \theta)} .$$

We show first,

$$(4.5) \quad \limsup_n \frac{1}{n} \log V_n(s, \theta) \leq -J(\theta)$$

a.s. P_θ . Note that,

$$(4.6) \quad \frac{1}{n} \log V_n(s, \theta) \leq \frac{1}{n} \log \left(\sup \left\{ \frac{1}{n} \sum_{i=1}^n \log \frac{f(x_i, \lambda)}{f(x_i, \theta)} : \lambda \in \Theta_0 \right\} \right) .$$

And the right hand side equals $R_0(\Theta_0, \theta)$ in the notation of [4] which converges to $-J(\theta)$ a.s. P_θ by Lemma 4 of that paper. An examination of the proof of this Lemma will show that only our assumptions 1–6 are used.

To establish the Lemma we need now only show,

$$(4.7) \quad \liminf_n \frac{1}{n} \log U_n(s, \theta) \geq 1 \text{ a.s. } P_\theta .$$

By assumption 7 we can find $d_1(\theta, \varepsilon)$ such that,

$$(4.8) \quad E_\theta(\eta(X_i, \theta, d_1)) \geq -\varepsilon .$$

1 But,

$$\begin{aligned}
 & \frac{1}{n} \log U_n \geq \log \nu[S(\theta, d_1) \cap [\Theta - \Theta_0]] \\
 & \quad + \inf \left\{ \frac{1}{n} \sum_{i=1}^n \frac{f(X_i, \lambda)}{f(X_i, \theta)} : \lambda \in S(\theta, d_1) \cap [\Theta - \Theta_0] \right\} \\
 & \geq \frac{1}{n} \sum_{i=1}^n \eta(x_i, \theta, d_1) + \log \nu[S(\theta, d_1) \cap [\Theta - \Theta_0]] .
 \end{aligned}$$

10 Letting $n \rightarrow \infty$ and then $\varepsilon \rightarrow 0$, (4.9), assumption 3, and the strong law of large numbers imply (4.7). The Lemma follows.

13 We complete the proof of the Theorem by way of two further Lemmas.

15 **Lemma 4.** Under the first part of assumption 3, for all $n, t, \theta' \in \Theta_0$,

$$(4.10) \quad P_{\theta'} \left[\frac{1}{n} \log U_n(s, \theta') \geq t \right] \leq e^{-nt} .$$

19 *Proof.*

$$\begin{aligned}
 P_{\theta'} \left[\frac{1}{n} \log U_n(s, \theta') \geq t \right] &= \int_W \prod_{i=1}^n f(x_i, \theta') \mu(dx_1) \dots \mu(dx_n) \\
 &\text{where } W = \left[s : \prod_{i=1}^n f(x_i, \theta') \leq e^{-nt} \int_{\Theta - \Theta_0} \prod_{i=1}^n f(x_i, \lambda) \nu(d\lambda) \right] .
 \end{aligned}$$

27 Thus our probability is bounded above by

$$(4.11) \quad e^{-nt} \int_{X^n} \int_{\Theta - \Theta_0} \prod_{i=1}^n f(x_i, \lambda) \nu(d\lambda) \mu(dx_1) \dots \mu(dx_n) .$$

31 But the right hand side of (4.11), by Fubini's theorem, is

$$e^{-nt} \int_{\Theta - \Theta_0} \int_{X^n} \prod_{i=1}^n f(x_i, \lambda) \mu(dx_1) \dots \mu(dx_n) \nu(d\lambda) \leq e^{-nt} .$$

37 **Lemma 5.** Under assumptions 8 and 9 if $\theta' \in \Theta_0$, for every $0 < \rho < 1$, $\beta > 0$, there exist $N(\rho, \beta)$ such that for $n \geq N(\rho, \beta)$

$$(4.12) \quad P_{\theta'} \left[\frac{1}{n} \log V_n(s, \theta') \geq -\beta \right] \leq \rho^n .$$

43 *Proof.* Choose $d(\theta', \rho, \beta)$ as in assumptions 8 and 9. Then since

$$\begin{aligned}
 & \frac{1}{n} \log V_n(S, \theta') \geq \log \nu[S(\theta', d(\theta', \rho, \beta) \cap \Theta_0)] \\
 & \quad + \inf \left\{ \frac{1}{n} \sum_{i=1}^n \log \frac{f(x_i, \lambda)}{f(x_i, \theta')} : \lambda \in S(\theta', d(\theta', \rho, \beta) \cap \Theta_0) \right\} \\
 & \geq \log \nu[S(\theta', d(\theta', \rho, \beta)) \cap \Theta_0] + \frac{1}{n} \sum_{i=1}^n \gamma(x_i, \theta', d(\theta', \rho, \beta)) ,
 \end{aligned}$$

we have,

(4.14)

$$P_{\theta'} \left[\frac{1}{n} \log V_n(s, \theta') \leq -\beta \right] \leq P_{\theta'} \left[\sum_{i=1}^n \gamma(x_i, \theta', d(\theta', \rho, \beta)) \leq -n\beta - \log m(\rho, \beta) \right] .$$

By Lemma 1 of [4]

$$(4.15) \quad P_{\theta'} \left[\frac{1}{n} \sum_{i=1}^n \gamma(x_i, \theta', d(\theta', \rho, \beta)) \leq -n\beta - \log m(\rho, \beta) \right] \leq \inf_t \left\{ \exp \left[-\beta t - \log m(\rho, \beta) \frac{t}{n} \right] \varphi(t, \theta', d(\theta', \rho, \beta)) \right\}^n$$

for $n \geq -\frac{\log m(\rho, \beta)}{\beta}$ which is finite by assumption 9. If $\inf_t e^{-\beta t} \varphi(t, \theta', d(\theta', \rho, \beta))$ is attained for $t = t_0$ which is strictly positive by lemma 1 of [4], we can choose,

$$N(\rho, \beta) \leq \max \left(-\frac{\log m(\rho, \beta)}{\beta}, -\frac{t_0 \log m(\rho, \beta)}{\log 2} \right) < \infty .$$

Now,

$$(4.16) \quad P_{\theta'} \left[\frac{1}{n} \log \bar{T}_n \geq t \right] \leq P_{\theta'} \left[\frac{1}{n} \log U_n(s, \theta') \geq -\beta, \frac{1}{n} \log V_n(s, \theta') \geq -\beta \right] + P_{\theta'} \left[\frac{1}{n} \log V_n(s, \theta') \leq -\beta \right] .$$

By Lemmas 4 and 5 the right hand side of (4.16) is bounded above by $e^{-n(t-\beta)} + \rho^n$ for $n \geq N(\rho, \beta)$. Hence,

$$(4.17) \quad \limsup_n \frac{1}{n} \log \sup \{ P_{\theta'} [\bar{T}_n \geq t] : \theta \in \Theta_0 \} \leq \limsup_n \frac{1}{n} \log (e^{-n(t-\beta)} + \rho^n) = \max[-(t-\beta), \log \rho] .$$

Letting $\rho \rightarrow 0$ first and then $\beta \rightarrow 0$, we find that ii) of Lemma 2 is satisfied by \bar{T}_n and the Theorem is proved. Gathering the most useful of the ‘‘prime’’ assumptions together we state,

Corollary 1. *If assumptions 1, 2', 3, (456)'', 7', 8'', and 9 hold, then the conclusion of Theorem 2 is valid.*

The most immediate field of application of Corollary 1 is when $f(x, \theta)$ is the density of a Koopman-Darmois (exponential) family.

$$(4.18) \quad f(x, \theta) = e^{\theta \cdot t(x)}$$

where $\theta = (\theta_1, \dots, \theta_k)$, $t(x) = (t_1(x), \dots, t_k(x))$ and the θ_j, t_j are real. Assumptions 1, 2', (456)'', 7' and 8'' are then automatically satisfied and we need only impose conditions 3, and 9 on ν . If Θ_0 is compact, 9' is automatic and 3 is all that is needed.

5. Optimality of Minimax Tests

The main result of this section is an immediate consequence of the following Lemma. We retain the notation of the previous section, defining only

$$(5.1) \quad \bar{F}_n(t, \theta) = P_\theta[\bar{T}_n < t] .$$

$$(5.2) \quad \bar{G}_n(t) = \inf \{ \bar{F}_n(t, \theta') : \theta' \in \Theta_0 \} .$$

Lemma 6. *Suppose that there exists a measurable subset S of Θ_0 such that,*

iii) $1 - \bar{F}_n(t, \theta')$ is a constant on S (for fixed n, t) as a function of θ' .

iv) $\sup \{ 1 - \bar{F}_n(t, \theta') : \theta' \in S \} = 1 - \bar{G}_n(t)$.

v) $\nu[\Theta_0 - S] = 0$.

Then,

$$(5.3) \quad \limsup_n \frac{1}{n} \log (1 - \bar{G}_n(t)) \leq -t .$$

Proof.

$$(5.4) \quad (1 - \bar{G}_n(t)) = \sup \{ 1 - \bar{F}_n(t, \theta') : \theta' \in S \} = \int_{\Theta_0} P_{\theta'}[\bar{T}_n \geq t] \nu(d\theta')$$

by iii), iv), v). Now, let

$$C_n = \left[\int_{\Theta_0} \prod_{i=1}^n f(x_i, \lambda) \nu(d\lambda) \leq e^{-nt} \int_{\Theta - \Theta_0} \prod_{i=1}^n f(x_i, \lambda) \nu(d\lambda) \right] .$$

Then,

$$(5.5) \quad \begin{aligned} \int_{\Theta_0} P_{\theta'}[\bar{T}_n < t] \nu(d\theta') &= \int_{\Theta_0} \int_{C_n} \prod_{i=1}^n f(x_i, \lambda) \mu(dx_1) \dots \mu(dx_n) \nu(d\lambda) \\ &= \int_{C_n} \int_{\Theta_0} \prod_{i=1}^n f(x_i, \lambda) \mu(dx_1) \dots \mu(dx_n) \nu(d\lambda) \\ &\leq e^{-nt} \int_{C_n} \int_{\Theta - \Theta_0} \prod_{i=1}^n f(x_i, \lambda) \nu(d\lambda) \mu(dx_1) \dots \mu(dx_n) \leq e^{-nt} . \end{aligned}$$

We formulate,

Assumption 10. For every n, t

$$a) \nu \left[\theta \in \Theta - \Theta_0 : \sup \{ \bar{F}_n(t, \lambda) : \lambda \in \Theta - \Theta_0 \} > \bar{F}_n(t, \theta) \right] = 0 .$$

$$b) \nu \left[\theta' : \bar{G}_n(t) < \bar{F}_n(t, \theta') : \theta' \in \Theta_0 \right] = 0 .$$

We can now state,

Theorem 3. *If assumption 10 holds the test which rejects if $\bar{T}_n \geq t$ is minimax for $H : \theta \in \Theta_0$ vs $K : \theta \in \Theta - \Theta_0$ at level $1 - \bar{G}_n(t)$. If assumptions 1-7 and assumption 10b) hold, then the sequence of test statistics $\{\bar{T}_n\}$ is asymptotically optimal.*

Proof. The first part of the Theorem is classical (c.f. Lehmann [8] p.327). The second part is an immediate consequence of Theorem 2 and Lemma 5 since assumption 10b) is equivalent to iii), iv), v).

This Theorem is of interest in connecting the classical finite sample optimality results with stochastic comparison. The most immediate application of this result is in the one-parameter exponential family where minimax tests of (say) $H : \theta \leq \theta_0$ vs $K : \theta = \theta_1 > \theta$ are. Bayes tests with respect to two point distributions satisfy assumption 10 (cf. [8]). Unfortunately proving optimality directly is trivial in this case. More interesting candidates are in the normal situation the t -statistic and the S^2 statistic used in testing $H : \mu < \mu_0$ and $H : \sigma \leq \sigma_0$ when μ , and σ are respectively unknown. Although we are here presented with a situation which does not quite fall under Theorem 3, (ν satisfying 10a,b depends on n , (cf. [8], p.94), one can easily check the conclusion of Lemma 2 directly and then apply Lemma 5 to obtain optimality.

Finally, it may be interesting to see that from a quasi Bayesian point of view, if stochastic comparison is defined in terms of the observed expected level of significance, (where the expectation is taken under the prior) then Lemma 5 and assumptions 1–7 and an analogue of Theorem 1 guarantee the optimality of \bar{T}_n in this sense. Formally for any sequence $\{T_n\}$ we would then consider not L_n but,

$$L_n^*(s) = (1 - G_n^*(T_n))$$

where $G_n^*(t) = \int_{\Theta_0} F_n(t, \theta') \nu(d\theta')$.

The analogue of the conclusion of Theorem 1 needed would be that a.s. P_θ ,

$$\liminf_n \frac{1}{n} \log L_n^* \geq -J(\theta) .$$

References

- [1] BAHADUR, R. R. (1960). Stochastic comparison of tests. *Ann. Math. Stat.*, **31** 276–295.
- [2] BAHADUR, R. R. (1960). Asymptotic efficiency of tests and estimates. *Sankhya*, **22** 229–252.
- [3] BAHADUR, R. R. (1960). Simultaneous comparison of the optimum and sign tests of a normal mean. In *Contributions to Probability and Statistics*. Stanford Univ. Press, 79–88.
- [4] BAHADUR, R. R. (1965). An optimal property of the likelihood ratio statistic. In *Proceedings of the 5th Berkeley Symposium*, **1**, 13–26.
- [5] BICKEL, P. J. and YAHAV, J. A. (1965). Asymptotically pointwise optimal procedures in sequential analysis. In *Proceedings of the 5th Berkeley Symposium*, **1**, 401–415.
- [6] BICKEL, P. J. and YAHAV, J. A. (1969). Asymptotic theory of Bayes solution. *Prob. Theory and Related Fields*, **11**, 257–276.
- [7] CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of hypothesis based on the sum of observations. *Ann. Math. Stat.*, **23**, 493–507.
- [8] ERDÖS, P. (1949). On a theorem of Hsu and Robbins. *Ann. Math. Stat.*, **20**, 286–291.
- [9] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. J. Wiley and Sons.