Lower Bounds on Bayes Factors for Invariant Testing Situations

by

Mohan Delampady
Purdue University

Technical Report #86-36

Department of Statistics
Purdue University

August 1986

Lower Bounds on Bayes Factors for Invariant Testing Situations *

Mohan Delampady
Purdue University and
University of British Columbia

August 26, 1986

Abstract

Lower bounds on Bayes factors are studied in situations where invariance under suitable groups of transformations exist. Expressions for these lower bounds are derived under fairly general conditions. The lower bounds are usually much larger than the corresponding P–values.

1 Introduction

1.1 Set Up

A random quantity X, assuming values in a space $\mathcal{X}$, having density (or mass function) $f(x|\theta)$ with respect to a σ–finite measure m, is observed. The unknown parameter of interest is $\theta \in \Theta \subseteq R^n$.

The problem of interest is to test

$$H_0 : \theta \in \Theta_0 \text{ against } H_1 : \theta \in \Theta_1.$$

The following assumptions are made:
(i) There is a group G acting on $\mathcal{X}$ which induces a group $\overline{G}$ on Θ.
(ii) G and $\overline{G}$ are isomorphic and are locally compact Hausdorff topological groups.

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

(iii) Either G is compact; or $\mathcal{X}$ and Θ are isomorphic and Θ is σ–compact.
(iv) The measure m is relatively invariant under the action of G with multiplier k; i.e, k is a continuous homomorphism from G to $(0, \infty)$ and $m(gA) = m(A)k(g)$ for all measurable subsets of $\mathcal{X}$.
(v) $f(gx|\overline{g}\theta) = f(x|\theta)k(g)$.
(vi) $\overline{g}\Theta_0 = \Theta_0$, $\overline{g}\Theta_1 = \Theta_1$, $\overline{g}\Theta = \Theta$.
(vii) $\int_G f(x|\overline{g}\theta)d\mu(g) < \infty \quad a.s \quad (m)$, where, μ is the Haar measure on G if G is compact and, μ is any right invariant Haar measure on G if G is non–compact.

Remark: The measure μ of (vii) above exists from (ii). Further, if G is compact, then the only continuous homomorphism from G to $(0, \infty)$ is $k \equiv 1$. If G is not compact then we have assumed in 3. above that both Θ and $\mathcal{X}$ have the same multiplier under G for the measures with respect to which densities are calculated.

The action of G and $\overline{G}$ induce maximal invariants $t(X)$ on $\mathcal{X}$ and $\eta(\theta)$ on Θ. The following example illustrates these points:

Example 1: $\underline{X} \sim N_n(\underline{\theta}, I)$ and we want to test $H_0 : \underline{\theta} = \underline{0}$ against $H_1 : \underline{\theta} \neq \underline{0}$. Then $\mathcal{X} = R^n = \Theta$. The problem is invariant under the group, G, of all orthogonal transformations; i.e., if O is an orthogonal matrix of order n, $g_O\underline{X} = O\underline{X}$. Also, $O\underline{X} \sim N(O\underline{\theta}, I)$ so that $\overline{g_O}\underline{\theta} = O\underline{\theta}$. Here G is a compact topological group. Note that

$$f(x|\theta) = \frac{1}{(2\pi)^{n/2}} \exp(-\frac{1}{2}(x-\theta)'(x-\theta))$$

and

$$\begin{aligned} f(g_O x|\overline{g_O}\theta) &= \frac{1}{(2\pi)^{n/2}} \exp(-\frac{1}{2}(Ox - O\theta)'(Ox - O\theta)) \\ &= \frac{1}{(2\pi)^{n/2}} \exp(-\frac{1}{2}(x-\theta)'(x-\theta)) \\ &= f(x|\theta), \end{aligned}$$

so that $k(g) \equiv 1$. Also, $\overline{g_O}\underline{0} = \underline{0}$ and $\overline{g_O}\theta = \theta$. □

Assume that a classical significance test is based on a test statistic $T(X)$, a function of the maximal invariant under the group G of transformations. Large values of this statistic provide evidence against the null hypothesis. The P–value, or the observed significance level, of data x is defined to be

$$\alpha = \sup_{\theta \in \Theta_0} P_\theta \left(T(X) \geq T(x) \right).$$

Approaching the above testing problem from the Bayesian assume that π is a prior distribution for θ with $\pi_0 = P^\pi(\Theta_0)$ and g_0, g_1 are the densities of π conditional on the sets Θ_0 and Θ_1 respectively. From the likelihood viewpoint g_0 and g_1 may be viewed as weight functions for the likelihood function on the respective parameter sets. Then the quantities of interest are
(i) the Bayes factor of H_0 to H_1:

$$B^\pi(x) = \frac{\int_{\Theta_0} f(x|\theta) g_0(\theta) d\theta}{\int_{\Theta_1} f(x|\theta) g_1(\theta) d\theta};$$

(ii) the posterior probability of H_{0a} given x:

$$P^\pi(H_0|x) = \left[1 + \frac{(1-\pi_0)}{\pi_0} \frac{1}{B^\pi(x)}\right]^{-1},$$

$B(x)$, which is also called the "weighted likelihood ratio" of H_0 to H_1, measures the impact of the data. By considering the Bayes factor the prior probabilities can be ignored. A likelihoodist is also interested in $B^\pi(x)$, since it is the ratio of the averaged or weighted likelihoods of H_0 and H_1, the averaging being with respect to the "weight functions" g_0 and g_1 respectively. If indeed both the hypotheses are simple (i.e, $H_0 : p = p_0$, $H_1 : p = p_1$) then

$$B(x) = \frac{f(x|p_0)}{f(x|p_1)},$$

which is the likelihood ratio, used widely as a standard test statistic.

1.2 Lower Bounds on Bayes Factors

Specification of π or g_0 and g_1 is natural and important to a Bayesian or likelihoodist, 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 distributions π, and that these lower bounds tend to be surprisingly large. If I is a class of priors π under consideration, we will consider the lower bounds

$$\underline{B}_I(x) = \inf_{\pi \in I} B^\pi(x), \tag{1}$$

$$\underline{P}_I(H_0|x) = \inf_{\pi \in I} P^\pi(H_0|x), \tag{2}$$

$$= \left[1 + \frac{(1-\pi_0)}{\pi_0} \cdot \frac{1}{\underline{B}_I}\right]^{-1}. \tag{3}$$

It has been shown by a number of authors that there is a vast discrepancy between the P–value and the lower bounds on Bayes factors or posterior probabilities over "objective" classes in the situations of testing a point null hypotheses. Since testing general hypotheses when the problems are invariant forms a very important part of multivariate statistics the comparison of P–values with the above mentioned conditional measures of evidence seems to be of great interest in this general set up. This study is important from a Bayesian (and likelihood) viewpoint because there is no "objective" Bayesian (or likelihood) analysis for these problems as it is not possible to objectively specify the prior π (or the weight functions g_0 and g_1). This is also important from a classical viewpoint because it shows the need for a careful interpretation of P–values.

This need is especially serious as some practitioners tend to interpret P–values quantitatively in terms of the probability of H_0 being true. The danger of this interpretation is indicated in the sections that follow, where an attempt is made to show that P–values tend to be an order of magnitude smaller than, say, posterior probabilities that H_0 is true. However, posterior probabilities of H_0 are, typically, very dependent on the prior chosen, making comparisons of this kind unclear. Therefore as an alternative to the "objective Bayesian" analysis with a specific prior, a robust Bayesian approach is suggested here. In this approach a reasonable solution seems to be using lower bounds on posterior probabilities of H_0 (or likelihood ratios) over an "objective" class of priors (or weight functions). It is very striking to note that in all situations considered here these lower bounds are substantially larger than the corresponding P–values, making it clear that careful interpretation of P–values is necessary.

A lower bound, such as $\underline{B}_I$, is particularly useful when I 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 I, one could argue that $\underline{B}_I$ is not actually a valid lower bound. If I 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 π has the effect of finding that $\pi \in I$ which is <u>most</u> <u>favorable</u> 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 I in as many ways as are considered reasonable. (Surprising results can be obtained even if I is allowed to contain <u>all</u> distributions; indeed, Edwards, Lindman and Savage (1963) show that $\underline{B}_I(x)$ is often still much bigger than P–values.)

At this point it is very important to note that the P–value is computed after reducing the problem using the invariance under I. This suggests a Bayesian assumption of invariance on priors, for any reasonable comparison of P–values and lower bounds on the posterior probabilities. Further, leaving out the assumption of invariance on priors would make the comparisons very unclear. Therefore, our discussion in this paper will mainly concentrate on the class I of priors that satisfy the following condition (viii).
(viii) $\pi(\overline{g}A) = \pi(A) = \pi(A\overline{g})$ if G is compact, and
$\pi(A\overline{g}) = \pi(A)$ if G is non–compact.

The class I consists of all spherically symmetric densities in the example discussed earlier.

1.3 History

Lower bounds on Bayes factors and posterior probabilities of point null hypotheses have been studied extensively by Edwards, Lindman and Savage(1963), Dickey (1977), Good(1950, 1958, 1976), Berger (1985), Casella and Berger (1985), Berger and Sellke (1986), Berger and Delampady (1986a) and Delampady and Berger (1986) among others. The discussion in the case of an interval null hypothesis is found in Delampady (1986b). See Berger and Delampady (1986) for further references to works dealing with P–values. Jeffreys (1961) has works related to Bayesian analysis in invariant testing situations.

1.4 Overview

In the next section some results on the densities of maximal invariants will be reviewed. These results involve representing the integral of any function with respect to a relatively invariant measure on $\mathcal{X}$, as an iterated integral by factoring $\mathcal{X}$ as $\mathcal{X}/G \times G$. This is easily done for compact topological groups, G. Results are found in Eaton(1983). However it is necessary to make a number of assumptions to get this representation in the non–compact case. The main references here are Wijsman(1967, 1985), Koehn(1970), Bondar(1976), Andersson(1982) and Farrell(1985). These results will be used in the later sections. In the next two sections results for lower bounds on the weighted likelihood ratios will be obtained. In section 3, situations where G is a compact topological group will be considered. Section 5.4 will generalize the results of section 3 to locally compact groups.

When G is compact, as mentioned in (viii) above, the weight functions or priors are G–invariant; i.e., $\pi(gA) = \pi(A)$ and $\pi(Ag) = \pi(A)$. In the non–compact case right invariant measures π^r (i.e., $\pi^r(Ag) = \pi^r(A)$) are used. These measures are relatively left invariant with multiplier Δ^{-1} where Δ is the right hand modulus of G; i.e., if μ^r and μ^l are the right and left invariant Haar measures on G then, they satisfy, $d\mu^r(g) = \Delta(g^{-1})d\mu^l(g)$ and $\pi^r(gA) = \Delta(g^{-1})\pi^r(A)$. Here justifications for the use of right invariant (improper) weight functions will also be given. Conclusions will be given in section 5.

2 Some Results on Maximal Invariants

2.1 Compact Topological Groups

For compact topological groups the following factorization theorem is immediate. This factorization gives rise to an expression for the density of maximal invariants.

Theorem 2.1.1. *Suppose that G is a compact topological group acting measurably on $\mathcal{X}$, μ is the invariant probability measure on G, and $f(x|\theta)$ is the density of X with respect to a G–invariant, σ–finite measure m on $\mathcal{X}$. Further, suppose that t: $\mathcal{X} \longrightarrow \mathcal{T}$ is a measurable mapping inducing a measure n on $\mathcal{T}$ given by $n(c) = m(t^{-1}C)$ and $\int_{\mathcal{X}} h\left(t(x)\right) dm(x) = \int_{\mathcal{T}} h(t)dn(t)$ for all integrable functions h on $\mathcal{T}$. Then, if $q(t|\eta)$ is the density of T with respect to n and h is any integrable function,*

$$q\left(t(x)|\eta(\theta)\right) = \int_G f(gx|\theta)d\mu(g), \tag{4}$$

$$\int_{\mathcal{X}} h(x)dm(x) = \int_{\mathcal{T}} \left(\int_G h(gt)d\mu(g)\right) dn(t). \tag{5}$$

Proof: Eaton(1983) Propositions 7.15 and 7.16 on pages 271–273 and 280–281. □

2.2 Locally Compact Topological Groups

Here Theorem 2.1.1 will be generalized. However this result needs a number of assumptions on all the spaces involved.

Theorem 2.2.1 (Bondar 1976). *Assume that*
(1) G is a separable completely metrizable locally compact topological group.

(2) $\mathcal{X}$ is a separable completely metrizable locally compact topological space.
(3) G acts continuously on $\mathcal{X}$ (i.e., the map $(g,x) \longrightarrow gx$ is continuous).
(4) There exists a Borel cross–section Z for the orbits of G in $\mathcal{X}$ (A Borel cross–section is a Borel subset of $\mathcal{X}$ which intersects each orbit Gx precisely once).
Then
(A) If G acts freely on $\mathcal{X}$ (i.e., if g is not the identity then $gx \neq x$ for all $x \in \mathcal{X}$) there exists a Borel cross–section Z which is a bimeasurable image of $\mathcal{X}/G$ and for any f, integrable with respect to a relatively invariant measure m with modulus(multiplier) k, the following is true:

$$\int_{\mathcal{X}} f(x)dm(x) = \int_Z \left(\int_G f(gz)k(g)d\mu^l(g) \right) dn(z), \tag{6}$$

for some measure n on Z, where μ^l is the left Haar measure on G.
(B) If G is a Lie group of non–null dimension with all stability subgroups compact and conjugate to each other and m as in (A), then there exists an m–null invariant set N and Borel cross–section Z which is a bimeasurable image of $(\mathcal{X} - N)/G$ such that the conclusion of (A) holds. □

Another version of the same result is given below under somewhat different assumptions.

Theorem 2.2.2 (Andersson 1982). *Assume that*
(1) G is a locally compact, σ–compact Hausdorff topological group.
(2) $\mathcal{X}$ is a locally compact, σ–compact Hausdorff topological space.
(3) G acts properly on $\mathcal{X}$ (i.e., G acts continuously on $\mathcal{X}$ and further the map $(g,x) \longrightarrow (gx,x)$ satisfies the condition that inverse image of a compact set is compact).
Then, for f and m as in Theorem 2.2.1, the same conclusion as in that theorem holds. □

Remark: If G is a Hausdorff topological group operating properly on $\mathcal{X}$, if $\mathcal{X}$ is locally compact, so are G and $\mathcal{X}/G$ (Bourbaki 1963, Proposition 11).
Remark: If G is a Hausdorff topological group acting continuously on a topological space $\mathcal{X}$, then free action of G is proper iff the graph C of the equivalence relation defined by G is closed in $\mathcal{X} \times \mathcal{X}$ and the canonical map $\phi : C \longrightarrow G$ is continuous (Bourbaki 1963, Proposition 6).

The following are some related results from Andersson(1982):
1. Every continuous action of a compact group is proper.

2. Action of the group of non–singular matrices $((A, \Sigma) \longrightarrow (A\Sigma A', \Sigma))$ on the set of positive definite matrices is proper.
3. Action of the translation group on an affine space is free and proper.
4. If G acts properly on $\mathcal{X}$, H is a closed subgroup of G, $\mathcal{X} \subseteq \mathcal{Y}$ is closed and $H\mathcal{Y} = \mathcal{Y}$, then the restriction of the proper action $G \times \mathcal{X} \longrightarrow \mathcal{X}$ to $H \times \mathcal{Y} \longrightarrow \mathcal{Y}$ is proper.

Farrell(1985) and Wijsman(1967, 1985) have similar factorization results under similar conditions on the spaces G, Θ and $\mathcal{X}$.

2.3 Likelihood Ratios for Maximal Invariants

The factorization theorems of the previous sections yield the following results on ratios of densities of maximal invariants.

Theorem 2.3.1. *Under the assumptions of Theorem 2.1.1,*

$$\frac{q\left(t(x)|\eta(\theta_1)\right)}{q\left(t(x)|\eta(\theta_2)\right)} = \frac{\int_G f(gx|\theta_1)d\mu(g)}{\int_G f(gx|\theta_2)d\mu(g)}. \tag{7}$$

Proof: Immediate from (1), the first part of Theorem 2.1.1. □

Theorem 2.3.2. *Under the assumptions which obtain the conclusions of Theorem 2.2.1,*

$$\frac{q\left(t(x)|\eta(\theta_1)\right)}{q\left(t(x)|\eta(\theta_2)\right)} = \frac{\int_G f(gx|\theta_1)k(g)d\mu^l(g)}{\int_G f(gx|\theta_2)k(g)d\mu^l(g)}. \tag{8}$$

Proof: Follows immediately from Theorem 2.2.1, Bondar(1976), and Theorem 2.2.2, Andersson(1982). □

3 Lower Bounds on Bayes Factors when G is Compact

With the basic results on factorization of integrals and likelihood ratios of maximal invariants it is now possible to obtain results related to the lower bounds on weighted likelihood ratios over the class of invariant weight functions.

Theorem 3.1. *Under the set up described in Theorem 2.3.1,*

$$\inf_{\pi\in I} B^\pi(x) = \frac{\inf_{\eta_1\in\Theta_0/G} q\left(t(x)|\eta_1\right)}{\sup_{\eta_2\in\Theta_1/G} q\left(t(x)|\eta_2\right)}.$$

Proof:

$$\begin{aligned}
\inf_{\pi\in I} B^\pi(x) &= \inf_{\pi\in I} \frac{\int_{\Theta_0} f(x|\theta)d\pi(\theta)}{\int_{\Theta_1} f(x|\theta)d\pi(\theta)} \\
&= \inf_n \frac{\int_{\Theta_0/G}\left(\int_G f(x|g\theta)d\mu(g)\right)dn\left(\eta(\theta)\right)}{\int_{\Theta_1/G}\left(\int_G f(x|g\theta)d\mu(g)\right)dn\left(\eta(\theta)\right)} \\
&= \inf_n \frac{\int_{\Theta_0/G}\left(\int_G f(gx|\theta)d\mu(g)\right)dn\left(\eta(\theta)\right)}{\int_{\Theta_1/G}\left(\int_G f(gx|\theta)d\mu(g)\right)dn\left(\eta(\theta)\right)} \\
&= \inf_n \frac{\int_{\Theta_0/G} q\left(t(x)|\eta_1\right)dn(\eta_1)}{\int_{\Theta_1/G} q\left(t(x)|\eta_2\right)dn(\eta_2)} \\
&= \frac{\inf_{\eta_1\in\Theta_0/G} q\left(t(x)|\eta_1\right)}{\sup_{\eta_2\in\Theta_1/G} q\left(t(x)|\eta_2\right)},
\end{aligned}$$

using Theorem 2.3.1. □

Corollary. *If $\Theta_0/G = \{0\}$, then under the same conditions as in Theorem 3.1,*

$$\inf_{\pi\in I} B^\pi(x) = \frac{q\left(t(x)|0\right)}{\sup_{\eta\in\Theta/G} q\left(t(x)|\eta\right)}.$$

□

Example 1 continued: For the Normal example described in the introductory section we get

$$\inf_{\pi\in I} B^\pi(x) = \frac{q\left(t(x)|0\right)}{q\left(t(x)|\hat{\eta}\right)},$$

where $q(t|\eta)$ is the density of a non–central χ^2 random variable with n degrees of freedom and non-centrality parameter η, and $\hat{\eta}$ is the maximum likelihood estimate of η from data $t(x)$. For selected values of $t(x)$ and n the lower bounds are tabulated against their P–values in Table 1.1.

Table 1: Invariant Test for Normal Means

α	n	$t(x)$	$\underline{P}(H_0 \mid t(x))$
.01	2	9.2205	.0685
	3	11.3246	.0697
	4	13.2797	.0693
	8	20.0820	.0696
	11	24.7545	.0687
	15	30.6054	.0686
	20	37.5914	.0684
.05	2	5.9948	.2259
	3	7.8167	.2256
	4	9.4916	.2250
	9	16.9252	.2229
	13	22.3667	.2218
	15	24.9997	.2213
	21	32.6776	.2200

Notice that the lower bounds on the posterior probabilities of the null hypothesis are anywhere from 4 to 7 times as large as the corresponding P–values, indicating that there is a vast discrepancy between P–values and posterior probabilities.

4 Lower Bounds on Bayes Factors when G is Locally Compact

4.1 Convergence of Posterior Distributions

In an invariant testing situation, the assumption that the weight functions are invariant is reasonable. However, if the group G is not compact (but only locally compact) these invariant priors or weight functions are not probability measures. The use of these improper priors is justified if they can be approximated appropriately by finite measures, in the sense that the posterior distributions obtained by these approximating measures converge to that of the invariant prior under consideration. In what follows this convergence is discussed.

Definition : A sequence π_n of finite measures on Θ is said to be asymptotically invariant with limit π if there exists a G–invariant measure π on Θ (i.e., $\pi(gA) = \pi(A)$ for all measurable subset A of Θ and $g \in G$) and

$$\lim_{n\to\infty} \pi_n(A) = \pi(A)$$

for all measurable subset A of Θ.

Theorem 4.1.1. *If $f(x|\theta)$ is a density function on $\mathcal{X}$ for each θ such that*
(i) $f(x|\theta)$ is integrable with respect to a regular G–invariant measure π on Θ,
(ii) π_n is asymptotically invariant with limit π,
(iii) π_n is absolutely continuous with respect to π, and
(iv) $\frac{d\pi_n}{d\pi}$ is a bounded function,
then

$$\lim_{n\to\infty} \int_A f(x|\theta)d\pi_n(\theta) = \int_A f(x|\theta)d\pi(\theta) \qquad \text{a.s } (m)$$

for all measurable subsets A of Θ. (This says that the posterior probability of any event under π_n converges to that under π.)

Proof: Clearly

$$\int_A f(x|\theta)d\pi_n(\theta) = \int_A f(x|\theta)\frac{d\pi_n}{d\pi}(\theta)d\pi(\theta),$$

$$\frac{d\pi_n}{d\pi}(\theta) \longrightarrow 1 \text{ as } n \to \infty \qquad a.s(\pi),$$

and

$$|\frac{d\pi_n}{d\pi}(\theta)| < M,$$

for some $M > 0$. Therefore

$$f(x|\theta)\frac{d\pi_n}{d\pi}(\theta)I_A \longrightarrow f(x|\theta)I_A \qquad a.s(\pi),$$

$$|f(x|\theta)\frac{d\pi_n}{d\pi}(\theta)| \leq Mf(x|\theta),$$

and f is integrable with respect to π. Therefore, by the Lebesgue dominated convergence theorem,

$$\lim_{n\to\infty} \int_A f(x|\theta)d\pi_n(\theta) = \int_A f(x|\theta)d\pi(\theta) \qquad \text{a.s } (m).$$

$\square$

Remark 1: Since Θ is assumed to be σ–compact, there is a sequence $\{\Theta_n\}$ of compact subsets increasing to Θ. Then a convenient sequence, π_n, of finite measures, which approximates π, is

$$\frac{d\pi_n}{d\pi}(\theta) = I_{\Theta_n}(\theta).$$

Remark 2: The condition that $\frac{d\pi_n}{d\pi}$ is bounded may be weakened by assuming only that

$$\lim_{n\to\infty} \int_{\frac{d\pi_n}{d\pi}(\theta)>M} f(x|\theta)d\pi_n(\theta) = 0$$

as $n \to \infty$ for some $M > 0$.

Remark 3: In fact, the condition that the sequence of functions $\frac{d\pi_n}{d\pi}$ is uniformly integrable with respect to the measure m defined by $dm(u) = f(x|u)d\pi(u)$ is necessary and sufficient in view of Vitali's theorem (Rudin(1974), page 143–144).

4.2 Lower Bounds on Bayes Factors

As in section 4.1 we would like to look at the lower bounds on Bayes factors and posterior probabilities, over the class of G–invariant weight functions or priors. Note, however, that in hypothesis testing the prior distributions need to be of total mass one. For compact G this poses no difficulty, but when G is non–compact, G–invariant priors are not finite. However, Theorem 4.1.1 allows us to approximate all G–invariant priors by finite, asymptotically invariant distributions, resolving the dilemma. Therefore, with this approximation in mind, the following discussion will only consider weight functions or priors which are G–invariant.

Theorem 4.2.1. *Assume the conditions of Theorem 2.3.2. In addition assume that the density f is with respect to Lebesgue measure on $\mathcal{X}$, and that the class I contains priors π which are right invariant with respect to G and satisfy*

$$\begin{aligned} d\pi(\theta) &= h(\theta)d\theta \\ &= h_1(z)dn(z)d\mu^r(g_\theta) \\ &= h_1(z)dn(z)\Delta_G(g_\theta)d\mu^l(g_\theta), \end{aligned}$$

where $\theta = (z, g_\theta)$ (from the factorization $\Theta = \Theta/G \times G$). Assume also that h_1 defines a probability measure, on the space Z of maximal invariants, such that

$$\int_{\Theta_0/G} h_1(z)dn(z) = \int_{\Theta_1/G} h_1(z)dn(z).$$

Then

$$\inf_{\pi\in I} B^{\pi}(x) = \frac{\inf_{\eta_1\in\Theta_0/G} q\left(t(x)|\eta_1\right)}{\sup_{\eta_2\in\Theta_1/G} q\left(t(x)|\eta_2\right)}.$$

Proof: First note that, for $i = 0, 1$,

$$\begin{aligned}
\int_{\Theta_i} f(x|\theta)d\pi(\theta) &= \int_{\Theta_i} f(x|\theta)h(\theta)d\theta \\
&= \int_{Z_i} h_1(z)\left(\int_G f(x|gz)\Delta_G(g^{-1})d\mu^l(g)\right)dn(z) \\
&= \int_{Z_i} h_1(z)\left(\int_G f(x|gz)d\mu^r(g)\right)dn(z),
\end{aligned}$$

using (3) and $h(gz) = h_1(z)\frac{1}{k(g)}$. Also

$$\begin{aligned}
\int_G f(x|gz)d\mu^r(g) &= \int_G f(g^{-1}x|z)\frac{1}{k(g)}d\mu^r(g) \\
&= \int_G f(gx|z)\frac{1}{k(g^{-1})}d\mu^r(g^{-1}) \\
&= \int_G f(gx|z)k(g)d\mu^l(g),
\end{aligned}$$

since $k(g^{-1}) = \frac{1}{k(g)}$, $f(x|gz) = f(g^{-1}x|z)\frac{1}{k(g)}$ and $d\mu^r(g^{-1}) = d\mu^l(g)$. Also let $dm(z) = h_1(z)dn(z)$. Then

$$\begin{aligned}
\inf_{\pi\in I} B^{\pi}(x) &= \inf_{\pi\in I}\frac{\int_{\Theta_0} f(x|\theta)d\pi(\theta)}{\int_{\Theta_1} f(x|\theta)d\pi(\theta)} \\
&= \inf_{m}\frac{\int_{Z_1}\left(\int_G f(gx|z)k(g)d\mu^l(g)\right)dm(z)}{\int_{Z_2}\left(\int_G f(gx|z)k(g)d\mu^l(g)\right)dm(z)} \\
&= \inf_{z_1\in Z_1}\sup_{z_2\in Z_2}\frac{\int_G f(gx|z_1)k(g)d\mu^l(g)}{\int_G f(gx|z_2)k(g)d\mu^l(g)} \\
&= \frac{\inf_{\eta_1\in\Theta_0/G} q\left(t(x)|\eta_1\right)}{\sup_{\eta_2\in\Theta_1/G} q\left(t(x)|\eta_2\right)},
\end{aligned}$$

using (5). □

Corollary. *If $\Theta_0/G = \{0\}$ then, under the same conditions as in Theorem 3.2.1,*

$$\inf_{\pi\in I} B^{\pi}(x) = \frac{q\left(t(x)|0\right)}{\sup_{\eta\in\Theta/G} q\left(t(x)|\eta\right)}.$$

Ex 2: Assume that $X_1, X_2, \cdots, X_n$ is a random sample from the $N\left(\theta,\sigma^2\right)$ distribution; both θ and σ are unknown. The problem is to test the hypothesis $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$. A sufficient statistic for (θ,σ) is $\underline{x} = (\overline{X}, S)$, $\overline{X} = \frac{1}{n}\sum_1^n X_i$ and $S = \left[\frac{1}{n}\sum_1^n (X_i - \overline{X})^2\right]^{1/2}$. Then

$$f(x|\theta,\sigma) = K\sigma^{-n}s^{n-2}\exp(-\frac{n}{2\sigma^2}\left[(\overline{X}-\theta)^2 + s^2\right])$$

where K is a constant. Also

$$\mathcal{X} = \{(\overline{x}, s) : \overline{x} \in R^1, s > 0\}, \text{and} \Theta = \{(\theta,\sigma) : \theta \in R^1, \sigma > 0\}.$$

The problem is invariant under the group

$$G = \{g_c = c : c > 0\} \text{and} g_c(x) = c(\overline{x}, s) = (c\overline{x}, cs).$$

We have $\mathcal{X}$ and Θ isomorphic, $k(g_c) = c^2$, $d\mu^l(g_c) = \frac{dc}{c}$, $f(x|\theta,\sigma) = \frac{1}{c^2}f(x|\theta,\sigma)$, $t(x) = \frac{\overline{x}}{s}$ and $\eta(\theta,\sigma) = \frac{\theta}{\sigma}$. Then, defining

$$I \;=\; \{\pi : d\pi(\theta,\sigma) = h_1(\eta)d\eta\frac{d\sigma}{\sigma}, h_1 \text{ is any density for } \eta\}, \tag{9}$$

$$\inf_{\pi\in I} B^{\pi}(x) = \frac{q\left(t(x)|0\right)}{q\left(t(x)|\hat{\eta}\right)},$$

where $q(t|\eta)$ is the density of a non–central t random variable with $n-1$ degrees of freedom and non–centrality parameter η, and $\hat{\eta}$ is the maximum likelihood estimate of η. The fact that all the necessary conditions are satisfied is shown in Andersson(1982) and Wijsman(1967). For selected values of $t(x)$ and n the lower bounds are tabulated along with the P–values in Table 2.

For small values of n the lower bounds in Table 2 are comparable to the corresponding P–values and as n gets large the difference between these lower bounds and the P–values get larger.

Table 2: Test for Normal Means, Variance unknown

	n	2	8	12	16	32
	α=.01	.0116	.0135	.0208	.0282	.0317
$\underline{P}(H_0 \mid t(x))$	α=.05	.0482	.0860	.1107	.1151	.1213
	α=.10	.0858	.1745	.1876	.1919	.1986

It can be seen from Table 2 that, for large n, the lower bounds are very close to the Normal "all prior" bounds. In our notation, an explanation for this phenomenon is that

$$\begin{aligned}\frac{q(t|0)}{q(t|\hat{\eta})} &= \sup_{\eta} \frac{1}{2^{(n-1)/2}\Gamma(\frac{n+1}{2})} \exp(-\frac{1}{2}\eta^2) \\ &\quad \times \int_0^{\infty} v^n \exp(-\frac{1}{2}v^2 + v\frac{t\eta}{\sqrt{n+t^2}})\, dv\end{aligned}$$

converges to, $\exp(\frac{1}{2}t^2)$, the Normal "all prior" lower bound on the Bayes factor.

5 Conclusions

We have succeeded in emphasizing that there is a clear need for a careful interpretation of P–values. We have also suggested a robust Bayesian interpretation of P–values which looks at the lower bounds on posterior probabilities and weighted likelihood ratios. We recommend looking at such lower bounds as better "objective" measures of evidence than P–values.

In all invariant testing situations where the group is compact, the lower bounds over the class of invariant weights on the weighted likelihood ratios of the original problem reduces to a lower bound on the likelihood ratios of the maximal invariants. This same phenomenon is observed for locally compact non–compact groups, under some regularity conditions, when the class of weight functions is chosen appropriately. This simplifies the problem considerably both in terms of numerical work and ease of understanding the problem. Usually the class of invariant weight functions are quite reasonable and the lower bounds over this class are substantial.

Another look at Example 1 explains some of these points. Observe that the invariant priors here are spherically symmetric. It is clear that the lower bounds here must be smaller than those for the unimodal spherically symmetric weight functions (see Berger and Delampady (1986)). However, even though they are somewhat smaller for small dimensions, for large dimensions they are very much comparable. This possibly indicates that the use of invariant weight functions is not inappropriate for problems in high dimensions. This is significant especially because in high dimensions the tool of invariance is very useful.

However, the class I of weight functions may be too large in some situations, leading to excessively small lower bounds as in example 2. Notice that this class allows arbitrary functional forms for the induced priors on the maximal invariant parameter which probably leads to an excessive bias in favour of the alternative hypothesis. Quite often it may be more appropriate to restrict the class I so that the induced priors satisfy other criteria, such as symmetry.

It must be again emphasized here that the lower bounds on the posterior probabilities in Example 1 are anywhere from 4 to 7 times as large as the corresponding P–values. Our recommendation is thus replacing P–values with the lower bounds on the Bayes factor over "objective" classes of priors. The point is that typical users of P–values can not be expected to interpret P–values quantitatively since this depends on the particular situation. 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 <u>all</u> reasonable priors makes such an argument more difficult. Other arguments that have been raised are given and discussed in Berger and Sellke (1987), Casella and Berger (1987), and Berger and Delampady (1986).

Acknowledgements. The author is grateful to his doctoral thesis advisor, Professor James Berger, for all his valuable suggestions, attention and help in this work.

References

[1] Andersson, S. (1982), "Distributions of Maximal Invariants Using Quotient Measures", *The Annals of Statistics*, **10**, 955–961.

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

[3] 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**.

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

[5] Bondar, J. V. (1976), "Borel Cross-Sections and Maximal Invariants", *The Annals of Statistics*, **4**, 866–877.

[6] Bourbaki, N. (1966), *Elements of Mathematics, General Topology, Part 1*, Hermann, Paris; and Addison-Wesley, Reading, Massachusetts.

[7] Casella, G. and Berger, R. L. (1987), "Reconciling Bayesian and Frequentist Evidence in the One-Sided Testing Problem", *Journal of the American Statistical Association*, **82**.

[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 Hypotheses", Tech. Report 86-35, Department of Statistics, Purdue University.

[10] Delampady, M. and Berger, J. (1986), "Lower Bounds on Bayes Factors for the multinomial problem and Chi-Square Test of Fit", Tech. Report 86-37, Department of Statistics, Purdue University.

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

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

[13] 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.)

[14] Eaton, M. L. (1983), *Multivariate Statistics*, New York: Wiley.

[15] Farrel, R. H. (1985), *Multivariate Calculations*, New York: Springer-Verlag.

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

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

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

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

[20] Koehn, U. (1970), "Global Cross Sections and Densities of Maximal Invariants", *The Annals of Mathematical Statistics*, **41**, 2045–2056.

[21] Rudin, W. (1974), *Real and Complex Analysis*, New Delhi: Tata McGraw-Hill.

[22] Wijsman, R. A. (1967), "Cross-sections of Orbits and their application to Densities of Maximal Invariants", *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*, Berkeley: University of California Press, 1, 389–400.

[23] Wijsman, R. A. (1985), "Proper Action in Steps, with application to Density Ratios of Maximal Invariants", *The Annals of Statistics*, **13**, 395–402.