Propriety of Intrinsic Priors in Invariant Testing Situations

Sarat C. Dass

University of Michigan

October 5, 2006

Abstract

The Theory of Intrinsic Priors, developed by Berger and Pericchi (1996a,b), is a general method of constructing objective priors for testing and model selection when proper priors are considered for the simpler null hypotheses. When this prior distribution is improper, as is typically the case for Objective Bayesian testing, they suggest approximating the (improper) prior by a sequence of proper priors on compacts. This "limiting procedure" was formalized by Moreno, Bertolino and Racugno (1998) who showed that the limiting Bayes factor is unique. Still, the natural question of whether some component of the intrinsic prior on parameters in the full model is proper in the limit (Sansó, 1997), remained to be answered. We develop a method here to partially answer this question for testing problems involving group invariance structures. We give a sufficient condition, which is easily verified, which implies that the conditional intrinsic prior under the full model is proper. This hitherto had to be verified by direct calculation. This paper also complements Berger, Pericchi and Varshavsky (1998), who develop methods, under group invariance situations, for non-nested models. For nested models, we give a class of initial non-informative priors and identify component parameters in the full model for which the default analysis results in a proper prior for these parameters. The proper prior is also identified as being the intrinsic prior arising from the use of the Arithmetic Intrinsic Bayes Factor (AIBF) methodology for the default analysis.

Key words and phrases. Intrinsic Bayes Factors; Intrinsic priors; Group invariance; Nested models; Non-informative priors.

1   Introduction

For a model selection problem, one can specify priors on model spaces and compute Bayes factors for competing models as the ratio of weighted likelihoods. The evaluation of Bayes factors, therefore, depends

significantly on prior specifications, making the choice of priors used on these model spaces important. In the preliminary stages of model selection, a Bayesian might use certain kinds of non-informative priors, say reference or Jeffreys' priors, in the event that proper subjective priors have not been carefully elicitated. These non-informative priors are typically improper. A problem which arises here with the use of improper priors is that the corresponding Bayes factors are unique only up to multiplicative constants. Hence, development of default Bayesian methodology is difficult in this respect. Efforts have been made to conventionally specify the constants for the non-informative priors, for example, Smith and Spiegelhalter (1980) , and Spiegelhalter and Smith (1982).

In order to overcome the problem described above, Berger and Pericchi (1996a,1996b) introduce the Intrinsic Bayes Factor (IBF) approach to hypothesis testing and model selection, using initial non-informative priors on model spaces and computing a quantity called the intrinsic Bayes factor. They argued, with many examples, that the IBF approach is attractive with appropriate choices of initial non-informative priors on the model spaces. Interestingly, the IBF approach relates closely to Bayes factors derived from certain priors, thus called intrinsic priors. Berger and Pericchi (1996b) show that in testing examples, the resulting intrinsic priors have many attractive features and are, indeed, proper priors. This is the strongest argument in favor of the IBF approach since proper Bayes factors (Bayes factors derived from proper priors) are easily interpreted.

Berger and Pericchi (1997) argued that there are potential benefits of using Intrinsic Priors to compute actual Bayes Factors. This approach eliminates the need for computations with real training samples and also eliminates concerns about its stability. However, what they prove is that if the initial prior on the null model is proper, then the induced intrinsic prior in the alternative is also proper. For Objective Bayesian testing, the assumed prior on the null is typically improper, and so they studied the propriety of intrinsic priors directly on a case by case basis. For testing using improper priors for the null hypotheses, Berger and Pericchi (1996a,b) suggest approximating the (improper) prior with proper priors supported on a sequence of increasing compact sets under the null. This "limiting procedure" was formalized by Moreno, Bertolino and Racugno (1998), who show that the limiting Bayes Factor is independent of the choice of the increasing compact sets. However, the fundamental question which was first raised by Sansó (1997) remained to be answered, namely, for which initial non-informative priors, the resulting intrinsic prior for certain parameters, conditional on the others in the full model, is proper. If this conditional prior is not proper, then the intrinsic prior does not obey the principle enunciated in Berger and Pericchi (1996a,b), that is, sensible Bayesian procedures should correspond to reasonable priors. Thus, given that the intrinsic prior on the full model is also improper, the question is also whether there is a decomposition of the improper prior on the full model into components that are proper. Sanso, Pericchi and Moreno (1996) study this problem for the location model. It is clear that the propriety of component priors, if they exist, depend crucially on the choice of the initial non-informative priors on the reduced and full

models. This raises another question as to whether there are classes of non-informative priors for which the resulting intrinsic priors on component parameters in the full model are, indeed, proper.

The main result in the present paper is to (partially) answer this fundamental question in the Theory of Intrinsic Priors as discussed above. For nested testing situations involving group invariance structures, we give a sufficient condition in terms of a class of initial improper priors and identify component parameters in the full model for which the default analysis results in a proper prior for these parameters, which are typically the parameters of interest. The (conditionally) proper prior is also identified as being the intrinsic prior arising from the use of the AIBF methodology for the default analysis when the intrinsic prior on the reduced model is taken to be the initial non-informative prior as in Sanso (1997).

The organization of the rest of the paper is as follows. Section 2 gives a statement of the problem that this paper aims to solve. Some background on intrinsic priors is presented in Section 3. Section 4 gives the group theoretic preliminaries and the structures on model spaces that will be relevant for the derivation of the proper intrinsic priors. The main result of propriety of the intrinsic prior is shown in Section 5. Section 6 discusses several examples relevant to the previous theory. Section 7 gives proofs of theorems and relevant definitions in the paper.

2   Statement of the Problem

Let $x_1, x_2, \ldots, x_n$ be iid f. Consider the following hypotheses testing scenario of

$$H_0 : f = f_0(\cdot|\theta_0) \quad \text{vs.} \quad H_1 : f = f_1(\cdot|\theta_1, \xi), \tag{1}$$

where $\theta_0 \in \Theta_0, \theta_1 \in \Theta_1$ and $\xi \in \Omega$ are unknown parameters.

Definition 1 *We will say that H_0 is **nested** in H_1 if for every $\theta_0 \in \Theta_0$,*

$$f_0(x|\theta_0) = f_1(x|\theta_1, \xi_1) \tag{2}$$

for some $\theta_1 \in \Theta_1$ and $\xi_1 \in \Omega$.

We will only consider nested testing situations from now on. Typically, the parameters $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta_1$ will denote "similar" parameters under the reduced and full models, respectively. For a brief discussion of intrinsic priors and intrinsic equations, the reader is referred to Section 3. For initial non-informative priors $\pi_0^N(\theta_0)$ and $\pi_1^N(\theta_1, \xi)$ on H_0 and H_1, respectively, Sanso (1997) shows that the limit intrinsic Bayes factor (LIBF) of Moreno, Bertolino and Racugno (1997) actually corresponds to the Bayes factor that is obtained when the intrinsic prior on the reduced model, $\pi_0^I(\theta_0)$, is the initial non-informative prior, $\pi_0^N(\theta_0)$. In fact, the LIBF of H_1 to H_0, B_{10}, has the expression

$$B_{10}(x) = \frac{\int_{\Theta_1} \int_{\Omega|\Theta_1} f_1(x|\theta_1, \xi) \pi_1^N(\theta_1, \xi) E_{(\theta_1, \xi)} \left[\frac{m_0^N(x(l))}{m_1^N(x(l))} \right] d\xi \, d\theta_1}{\int_{\Theta_0} f_0(x|\theta_0) \pi_0^N(\theta_0) \, d\theta_0}, \tag{3}$$

where $x(l)$ denotes a theoretical minimal training sample (see Section 4 for a definition of a minimal training sample), and m_0^N and m_1^N are the marginals under H_0 and H_1 with respect to priors π_0^N and π_1^N, respectively. Given that the intrinsic prior on the full model, $\pi_1^I(\theta_1, \xi)$, is also improper, Sanso (1997) asks the question whether the conditional intrinsic prior of ξ given θ_1, $\pi_1^I(\xi|\theta_1)$, is proper. More generally, the question is for which initial non-informative priors, π_0^N and π_1^N, does there exist a decomposition of the parameter space of the full model into components where one component gets a (conditionally) proper intrinsic prior while the other gets a non-informative prior. Sanso, Pericchi and Moreno (1996) study this problem for the location model. In the case when group invariance structures are present for both H_0 and H_1, we give a class of initial non-informative priors and identify component parameters in the full model for which the resulting conditional intrinsic prior is proper.

2.1 Example of a conditional prior that is not proper

We present an example of nested hypotheses testing where the resulting conditional intrinsic prior is not proper for a particular choice of the initial non-informative prior. This example is taken from Berger and Pericchi (1996a). See Berger and Pericchi (1996a) for details and generalizations of testing nested linear models. Consider the following testing of

$$H_0 \,:\, y = x_0\beta_0 + \epsilon_0, \quad \epsilon_0 \sim N(0, \sigma_0^2 I_n)$$

versus

$$H_1 \,:\, y = x_0\beta_1 + x_1\beta_{11} + \epsilon_1, \quad \epsilon_1 \sim N(0, \sigma_1^2 I_n),$$

where y is a $n \times 1$ vector of observations, x_i's are $n \times 1$ vector of covariates for $i = 0, 1$, σ_0^2 and σ_1^2 are unknown positive numbers and β_0, β_1 and β_{11} are unknown real numbers. Here, $\Theta_0 = (\beta_0, \sigma_0^2)$, $\Theta_1 = (\beta_1, \sigma_1^2)$ and $\Omega = (\beta_{11})$. With the choice of initial non-informative priors on H_0 and H_1 taken to be the Jeffreys' priors, namely,

$$\pi_0^N(\beta_0, \sigma_0) = \frac{1}{\sigma_0^2}\, d\sigma_0\, d\beta_0$$

and

$$\pi_1^N(\beta_{11}, \beta_1, \sigma_1) = \frac{1}{\sigma_1^3}\, d\beta_{11}\, d\beta_1\, d\sigma_1,$$

the conditional intrinsic prior, $\pi_1^I(\beta_{11}|\beta_1, \sigma_1)$, integrates to $\frac{2}{\pi} \neq 1$. Thus, not all initial choices of non-informative priors on H_0 and H_1 give rise to conditional intrinsic priors that are proper. So, the problem of finding initial non-informative priors for which intrinsic priors on component parameters are actually proper is both relevant and important.

3   Intrinsic Priors and Intrinsic Equations

This section gives a brief review of intrinsic priors and their derivation. The reader can consult Berger and Pericchi (1996a,1996b) for more details. From asymptotic considerations, Berger and Pericchi (1996b) show that the intrinsic Bayes methodology is equivalent to a Bayesian analysis using the intrinsic priors, π_1^I and π_0^I, on H_1 and H_0, respectively, where π_0^I and π_1^I are solutions of the intrinsic equations

$$\frac{\pi_1^I(\theta_1,\xi)\pi_0^N(P_0(\theta_1,\xi))}{\pi_1^N(\theta_1,\xi)\pi_0^I(P_0(\theta_1,\xi))} = B_1^*(\theta_1,\xi) \tag{4}$$

and

$$\frac{\pi_1^I(P_1(\theta_0))\pi_0^N(\theta_0)}{\pi_1^N(P_1(\theta_0))\pi_0^I(\theta_0)} = B_0^*(\theta_0), \tag{5}$$

for initial non-informative priors, π_0^N and π_1^N. In equations (4) and (5), P_0 and P_1 are the Kullback-Liebler projections onto H_0 and H_1, respectively and, $B_1^*(\theta_1,\xi)$ and $B_0^*(\theta_0)$ are constants closely related to the choice of "averaging" for the intrinsic Bayes factor. Dmochowski (1995) characterizes the general solutions of (4) and (5). In the nested case, it follows that (5) is contained in (4), that is, we have only one intrinsic equation to determine π_0^I and π_1^I, giving us the freedom to choose one arbitrarily. Based on the argument given for the LIBF in Section 2, we will choose π_0^I to be the initial non-informative prior on the reduced model, π_0^N, as in Sanso (1997). The following section presents the general framework of testing problems in the presence of group invariance structures.

4   Testing under Group Invariance Structures

For the hypotheses testing scenario of (1), that is,

$$H_0 : f = f_0(\cdot|\theta_0) \quad \text{vs.} \quad H_1 : f = f_1(\cdot|\theta_1,\xi), \tag{6}$$

we have the following group invariance assumptions. We assume that a group G acts on the observation space, $\mathcal{X}$, and that each family f_0 and f_1 is **G-invariant**. For subsequent group theoretic terms in bold type, refer to Section 7. The action of G on $\mathcal{X}$ induces groups $\bar{G}_0$ and $\bar{G}_1$ acting on Θ_0 and (Θ_1,Ω), respectively. Furthermore, assume for any $\bar{g}_1 \in \bar{G}_1$ and $(\theta_1,\xi) \in (\Theta_1,\Omega)$, $\bar{g}_1 \circ (\theta_1,\xi) = (\bar{g}_1 \circ \theta_1,\xi)$, i.e., $\xi \in \Omega$ is invariant under the action of $\bar{G}_1$. $\bar{G}_0$ and $\bar{G}_1$ are assumed to be **transitive** on their respective domains. If, in addition, $\bar{G}_0$ and $\bar{G}_1$ have trivial **isotropy subgroups**, it follows that there is an **isomorphism** that maps Θ_0 to Θ_1. The family of densities in H_0 may therefore be parameterized by $\theta_1 \in \Theta_1$ instead of $\theta_0 \in \Theta_0$, and the action of the group $\bar{G}_0$ on Θ_0 may be replaced by the action of $\bar{G}_1$ on Θ_1. Subsequently under this reparameterization, we may assume that the family of densities in (1) have the same parameter space Θ, with a common group $\bar{G}$ (arising from the action of G) acting on them.

For a set $A \subseteq \bar{G}$ and elements $\bar{g}, \bar{h} \in \bar{G}$, the set $A \cdot \bar{g}$ denotes the right translate of A and the set $\bar{g} \cdot A$ denotes the left translate of A.

Definition 2 *A measure μ on $\bar{G}$ is said to be **relatively invariant** with **left multiplier** α_l and **right multiplier** α_r if $\mu(A \cdot \bar{g}) = \alpha_r(\bar{g}) \cdot \mu(A)$ and $\mu(\bar{g} \cdot A) = \alpha_l(\bar{g}) \cdot \mu(A)$.*

Note that $\alpha_r = 1$ and $\alpha_l = \Delta_l$ (the **left-hand moduli** of $\bar{G}$) corresponds to the right-Haar measure on $\bar{G}$. We will denote the right-Haar measure by ν. Similarly, $\alpha_r = \Delta_r$ (the **right-hand moduli** of $\bar{G}$) and $\alpha_l = 1$ corresponds to the left-Haar measure on $\bar{G}$. We denote the left-Haar measure on $\bar{G}$ by μ_L.

Define a function $\phi : \bar{G} \longrightarrow \Theta$ by $\phi(\bar{g}) = \bar{g} \circ e$ where e is the identity element of Θ. Since $\bar{G}$ is assumed to be transitive, the function ϕ is onto. Subsequently, a prior μ on $\bar{G}$ induces a prior $\mu_\phi \equiv \mu \phi^{-1}$ on Θ. We will say that μ_ϕ is relatively invariant if μ is relatively invariant. As a special case, ν_ϕ will denote the prior induced by the right-Haar measure ν on $\bar{G}$.

The goal is to obtain intrinsic priors for ξ and show that it is proper.

In presence of group invariance structures, it is reasonable to restrict the choice of priors on H_0 and H_1 to a class for which posterior inferences will be invariant. Stone (1970) showed that priors on H_0 and H_1 should then be relatively invariant. Relatively invariant priors are typically improper. Thus, for the testing of (1), an initial non-informative prior, π_0^N, on H_0 can be

$$\pi_0^N(\theta_0) \propto \text{ a relatively invariant prior for } \theta_0. \tag{7}$$

A further argument based on the essential uniqueness of the left-Haar prior shows that, a non-informative prior, π_1^N, on H_1, should satisfy the **product form**

$$\pi_1^N(\theta_1, \xi) \propto \text{ a relatively invariant prior for } \theta_1 \times \text{ arbitrary measure for } \xi. \tag{8}$$

Since θ_0 and θ_1 denote "similar" parameters for H_0 and H_1, respectively, we choose the same relatively invariant prior for them. Thus, the initial "non-informative" priors on H_0 and H_1, π_0^N and π_1^N, will be chosen according to (7) and (8), that is,

$$\pi_0^N(\theta_0) = \mu_\phi(\theta_0) \tag{9}$$

$$\text{and} \quad \pi_1^N(\theta_1, \xi) = \mu_\phi(\theta_1) \cdot \pi(\xi), \tag{10}$$

where μ_ϕ denotes a relatively invariant prior and π is arbitrary. A right-Haar version of the above non-informative priors is obtained by replacing μ_ϕ with ν_ϕ in (9) and (10). The usual default priors used in testing when subjective information is not available are the reference prior (Bernardo (1979), Berger and Bernardo (1992)) and the Jeffreys' prior (Jeffreys (1961)). It will be shown later that both the reference and Jeffreys' prior correspond to right-Haar priors on θ_0 and θ_1, and π's that are typically improper. Hence, for the ensuing discussion, π will typically denote an improper prior, defined up to a multiplicative constant.

Since Θ_0 and Θ_1 are isomorphic to each other as discussed earlier, there is a common parameter space, Θ, and a common group, $\bar{G}$ (arising from the action of G on $\mathcal{X}$), acting on Θ for both H_0 and H_1.

For the initial non-informative priors chosen according to (9) and (10), the Bayes factor based on a sample, $\underline{x}_k$, of size k is given by

$$B_{01}^{\mu,\pi}(\underline{x}_k) = \frac{\displaystyle\int_{\Theta} f_0(\underline{x}_k|\theta)\, d\mu_\phi(\theta)}{\displaystyle\int_{\Omega}\int_{\Theta} f_1(\underline{x}_k|\theta,\xi)\, d\mu_\phi(\theta)\, \pi(\xi)\, d\xi}, \qquad (11)$$

for π possibly improper. Note that (11) is defined only up to a multiplicative constant due to the inherent arbitrariness in the choice of improper π. Corresponding to the right-Haar prior ν_ϕ, we can obtain an expression similar to (11), namely,

$$B_{01}^{\nu,\pi}(\underline{x}_k) = \frac{\displaystyle\int_{\Theta} f_0(\underline{x}_k|\theta)\, d\nu_\phi(\theta)}{\displaystyle\int_{\Omega}\int_{\Theta} f_1(\underline{x}_k|\theta,\xi)\, d\nu_\phi(\theta)\, \pi(\xi)\, d\xi}. \qquad (12)$$

It is clear that the integrals in the numerator and denominator of (11) and (12) must be finite for the corresponding Bayes factors to be well-defined. Let m_0 denote the smallest k for which this is so (assuming this smallest k is defined with probability one). Berger and Pericchi (1996a,1996b) define a **minimal training sample** to be a hypothetical sample, $\underline{x}_{m_0}$, of size m_0, for which the Bayes factors in (11) and (12) are well-defined. Thus, for a sample, $\underline{x}_n$, of size n, we denote the total number of distinct minimal training samples obtainable from $\underline{x}_n$ by L, and a generic minimal training sample by $x(l)$ for $l = 1, 2, \ldots, L$.

The following theorem applies to the default Bayes factors derived in (11).

Theorem 1 *For any $k \geq m_0$, the Bayes factor in (11), $B_{01}^{\mu,\pi}(\underline{x}_k)$, has a distribution which does not depend on θ_0 under H_0. Under H_1, the distribution of $B_{01}^{\mu,\pi}(\underline{x}_k)$ depends on (θ_1,ξ) only through ξ.*

Proof: See Section 7.

Recall that ξ is the invariant component under the group action in the full model, H_1. The above theorem implies that the expectation of $B_{01}^{\mu,\pi}(x(k))$ under $f_1(\cdot|\theta,\xi)$ in H_1, is a function of ξ only. Thus, for the arithmetic intrinsic prior (arising from arithmetic averaging in (4)), where the constant $B_1^*(\theta_1,\xi)$ is given by

$$B_1^*(\theta_1,\xi) = \lim_{L\to\infty} E_{(\theta_1,\xi)}\left[\frac{1}{L}\sum_{l=1}^{L} B_{01}^{\mu_\phi,\pi}(X(l))\right],$$

we have

$$B_1^*(\theta_1,\xi) = \lim_{L\to\infty}\frac{1}{L}\sum_{l=1}^{L} E_{(\theta_1,\xi)} B_{01}^{\mu_\phi,\pi}(X(l)) \qquad (13)$$

$$= \lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} E_\xi B_{01}^{\mu_\phi, \pi}(X(l)), \quad \text{(by Theorem 1)}$$

which is just a function of ξ provided the limit exists. In situations where the $\{x(l)\}$ are exchangeable, the summation over all possible training samples and the limit as $L \to \infty$ are unnecessary. Then, we have

$$
\begin{aligned}
B_1^*(\theta_1, \xi) &= E_{(\theta_1, \xi)} B_{01}^{\mu_\phi, \pi}(X(l)) \\
&= E_\xi B_{01}^{\mu_\phi, \pi}(X(l)) \quad \text{(by Theorem 1)} \\
&= B_1^*(\xi), \text{ say.}
\end{aligned}
\tag{14}
$$

We now state a theorem for $B_{01}^{\nu, \pi}(\underline{x}_k)$.

Theorem 2 *For a prior π (proper or improper) on Ω, and for $B_{01}^{\nu, \pi}(\underline{x}_k)$, we have that*

$$\int_\Omega E_\xi(B_{01}^{\nu, \pi}(\underline{X}_k)) \, \pi(\xi) \, d\xi = 1. \tag{15}$$

Proof: See Section 7.

The integrand in (15) is uniquely determined even though the choice of the improper prior π was arbitrary up to a multiplicative constant. Equation (15) also shows that $\pi^*(\xi) = E_\xi \{B_{01}^{\nu, \pi}(\underline{X}_k)\} \pi(\xi)$ is a proper prior on Ω regardless of whether π was initially proper or improper. We will use Theorem 2 applied to $k = m_0$ to show that the intrinsic priors for ξ will always be proper.

5 Propriety of the component intrinsic prior on ξ

We factor $\pi_1^I(\theta_1, \xi)$ as

$$\pi_1^I(\theta_1, \xi) = \pi_1^I(\theta_1) \cdot \pi_1^I(\xi), \tag{16}$$

that is, the intrinsic prior on H_1 also satisfies the product form (8). Choose the initial non-informative priors, π_0^N and π_1^N, to be the **right-Haar** versions of (9) and (10) and choose an arbitrary (proper or improper) π. We also choose

$$\pi_0^I(\theta_0) = \pi_0^N(\theta_0) \tag{17}$$

as in Sanso (1997). The intrinsic prior on ξ, $\pi_1^I(\xi)$, and on θ_1, $\pi_1^I(\theta_1)$, is then given by

$$\pi_1^I(\xi) = B_1^*(\xi) \cdot \pi(\xi), \tag{18}$$

and

$$\pi_1^I(\theta_1) = \pi_1^N(\theta_1). \tag{19}$$

Equations (18) and (19) above are a consequence of solving (4) for $\pi_1^I(\theta_1, \xi)$ taken as in (16). We state a theorem for the intrinsic prior on ξ.

8

Theorem 3 *If the initial non-informative priors on H_0 and H_1 are taken as*

$$\pi_0^N(\theta_0) = \nu_\phi(\theta_0) \tag{20}$$

$$and \quad \pi_1^N(\theta_1, \xi) = \nu_\phi(\theta_1) \cdot \pi(\xi), \tag{21}$$

where ν_ϕ is the right-Haar prior and π is an arbitrary prior for ξ, the intrinsic prior for ξ, $\pi_1^I(\xi)$, as given in (18), is proper.

Proof: See Section 7.

Any default choice of an improper prior $\pi(\xi)$ on Ω will result in a proper $\pi_1^I(\xi)$, as long as right-Haar priors are used as initial priors on "similar" nuisance parameters for both models. Therefore, a testing problem may acquire more than one proper intrinsic prior with the use of different initial default priors on Ω. In other words, we have identified a class of initial non-informative priors and a component of the full model, namely ξ, for which the intrinsic prior for that component is proper. The parameter ξ is the invariant parameter under the group action in the full model, H_1. In actual testing situations, we may have to reparameterize H_1 to identify ξ. We have also shown that intrinsic priors on component parameters in H_1 can be proper even when the intrinsic prior on H_0 is taken to be improper.

6   Examples

We present some examples in this section as applications of the preceding theory. All cases considered are nested testing situations.

6.1   Normal mean testing problem

Suppose $X_1, X_2, \ldots, X_n$ are iid $N(\mu, \sigma_1^2)$ under H_1 with unknown $\sigma_1^2 > 0$ and $\mu \in R$. Under H_0, the $X_i's$ are iid $N(0, \sigma_0^2)$, with unknown $\sigma_0^2 > 0$. We proceed to identify the class of non-informative priors generally given by (20) and (21), and the parameter ξ for this testing example.

  We denote the observation space by $\mathcal{X} = \{(x_1, x_2, \ldots, x_n) : x_i \in R, \ i = 1, 2, \ldots, n.\}$. The group acting on $\mathcal{X}$ which leaves both H_0 and H_1 invariant is the multiplicative group $(x_1, x_2, \ldots, x_n) \longrightarrow (cx_1, cx_2, \ldots, cx_n)$ for any $c > 0$. Using the notation in Section 4, we have that $\Theta_0 = \{\sigma_0 : \sigma_0 > 0\}$ and $\Theta_1 = \{\sigma_1 : \sigma_1 > 0\}$. Furthermore, $\bar{G}_0 = \bar{G}_1 = \{c : c > 0\} = \bar{G}$, say, where the action of $\bar{G}$ on the space Θ_i is $\sigma_i \longrightarrow c\sigma_i$ for $i = 0, 1$. If we reparameterize (μ, σ_1) in H_1 to $(\sigma_1, \mu/\sigma_1)$, it is easily seen that $\xi \equiv \mu/\sigma_1$ is invariant under the group action. The right-Haar priors on Θ_i are $\frac{1}{\sigma_i} d\sigma_i$ for $i = 0, 1$, respectively.

Lemma 1 *If π_0^N and π_1^N are chosen such that*

$$\pi_0^N(\sigma_0) = (1/\sigma_0)\, d\sigma_0 \tag{22}$$

and

$$\pi_1^N(\sigma_1, \xi) = (1/\sigma_1)\, d\sigma_1\, \pi(\xi)\, d\xi, \tag{23}$$

where π is arbitrary, the resulting intrinsic prior for $\xi \equiv \mu/\sigma_1$ is proper.

Proof: This is a direct consequence of Theorem 3.

Berger and Pericchi (1996b) take default priors on H_0 and H_1 to be $\pi_0^N(\sigma_0) = 1/\sigma_0$ and $\pi_1^N(\mu, \sigma_1) = 1/\sigma_1^2$, respectively. The non-informative prior π_0^N is the usual **reference prior** on H_0 and π_1^N is the **Jeffreys prior** on H_1. The reference prior on H_0 corresponds to the right-Haar prior on H_0, while the Jeffreys prior used for H_1 corresponds to the choice of $\pi(\xi) \equiv 1$ in (23). It follows from Lemma 1 that the intrinsic prior on ξ, $\pi_1^I(\xi)$, is proper where

$$
\begin{aligned}
\pi_1^I(\xi) &= B_1^*(\xi)\pi(\xi) \\
&= E_\xi\left[\frac{(X_i - X_j)^2}{2\sqrt{\pi}(X_i^2 + X_j^2)}\right] \cdot 1
\end{aligned}
$$

in this example. Berger and Pericchi (1996b) showed the propriety of $\pi_1^I(\xi)$ by direct calculation.

6.2 Testing exponential versus Weibull distributions

Two competing models frequently considered for failure data are the exponential and Weibull distributions. These nested models are, respectively,

$$H_0 : f_0(x|\theta_0) = \theta_0^{-1}\exp\{-x/\theta_0\} \tag{24}$$

vs.

$$H_1 : f_1(x|\gamma, \beta) = \beta x^{(\beta-1)}\gamma^{-\beta}\exp\{-(x/\gamma)^\beta\}. \tag{25}$$

The spaces Θ_0 and Θ_1 are, respectively, the spaces of θ_0 for H_0 and γ for H_1.

We will again identify a class of initial non-informative priors and the component parameter ξ for which the conditional intrinsic prior is proper. Denote the observation space by $\mathcal{X} = \{(x_1, x_2, \ldots, x_n) : x_i > 0 \,\forall\, i = 1, 2, \ldots, n\}$. As before, the group acting on $\mathcal{X}$ is the multiplicative group $(x_1, x_2, \ldots, x_n) \longrightarrow (cx_1, cx_2, \ldots, cx_n)$. The group action on Θ_0 and Θ_1 are the same namely, $\theta_0 \longrightarrow c\theta_0$ and $\gamma \longrightarrow c\gamma$. Note that β is invariant under the action of the group, so, using our notation, $\xi \equiv \beta$. The right-Haar priors on Θ_0 and Θ_1 are $\frac{1}{\theta_0}d\theta_0$ and $\frac{1}{\gamma}d\gamma$, respectively. The following lemma is again a direct consequence of Theorem 3.

Lemma 2 *For any π_0^N and π_1^N satisfying*

$$\pi_0^N(\theta_0) \quad = \quad (1/\theta_0)\, d\theta_0 \tag{26}$$

and

$$\pi_1^N(\gamma, \xi) \quad = \quad (1/\gamma)d\gamma \cdot \pi(\beta)\, d\beta, \tag{27}$$

where $\pi(\beta)$ is arbitrary, the resulting intrinsic prior for $\xi \equiv \beta$ is proper.

For H_0, the standard non-informative prior is $\pi_0^N(\theta_0) = 1/\theta_0$. Two default priors for H_1 are the Jeffreys prior given by $\pi_1^J(\gamma, \beta) = (1/\gamma)\, d\gamma d\beta$, and the reference prior, $\pi_1^R(\gamma, \beta) = (1/\gamma\beta)\, d\gamma d\beta$. For the reference prior, the requirements of Lemma 2 are satisfied by choosing $\pi(\beta) \equiv 1/\beta$. The intrinsic prior for β corresponding to the reference prior is given by

$$\pi^I(\beta) = E_\beta \left[\frac{2 X_i X_j |\log(X_i/X_j)|}{(X_i + X_j)^2} \right] \cdot \frac{1}{\beta}. \tag{28}$$

Using Lemma 2,

$$\int_{\beta > 0} E_\beta \left[\frac{2 X_i X_j |\log(X_i/X_j)|}{(X_i + X_j)^2} \right] \cdot \frac{1}{\beta} \cdot d\beta = 1, \tag{29}$$

thus, proving the propriety of the intrinsic prior in (28).

For the Jeffreys prior, the requirements of Lemma 2 are satisfied by choosing $\pi(\beta) \equiv 1$. These intrinsic priors were already shown to be proper by direct calculations in Berger and Pericchi (1996b), but the general theory here eliminates the need for such direct calculations on a case by case basis.

6.3 IBFs in Linear Models

The linear model set-up that we consider is the following:

$$H_0 : y = X_0 \beta_0 + \sigma_0 \cdot \epsilon \quad \text{vs.} \quad H_1 : y = X_0 \beta_1 + X_1 \beta_{11} + \sigma_1 \cdot \epsilon \tag{30}$$

where y is a $n \times 1$ vector of observations, X_0 and X_1 are design matrices of full column rank and of order $n \times k_0$ and $n \times k_1$, respectively, and $\epsilon \sim N_n(0, I_n)$. We consider two separate cases: (i) $\sigma_0 = \sigma_1 = \sigma$ (known) and (ii) σ_0 and σ_1 both unknown. For both cases, we identify the appropriate class of initial non-informative priors and the component parameter ξ for which the intrinsic prior is proper.

Case 1: $\sigma_0 = \sigma_1 = \sigma$ (known)

The group G that leaves H_0 and H_1 invariant is the additive group whose action is given by $y \longrightarrow y + X_0 b$ for $b \in R^{k_0}$. Here, $\Theta_0 = \{\beta_0 : \beta_0 \in R^{k_0}\}$ and $\Theta_1 = \{\beta_1 : \beta_1 \in R^{k_0}\}$, where the induced action of G on the parameter spaces is given by $\beta_0 \longrightarrow \beta_0 + b$ and $\beta_1 \longrightarrow \beta_1 + b$. The right-Haar priors on Θ_0 and Θ_1 are both proportional to 1. The parameter β_{11} is invariant under the group action, so $\xi \equiv \beta_{11}$. We have the following lemma

Lemma 3 *If π_0^N and π_1^N satisfy*

$$\pi_0^N(\beta_0) \quad = \quad 1 \cdot d\beta_0 \tag{31}$$

and

$$\pi_1^N(\beta_1, \beta_{11}) \quad = \quad 1 \cdot d\beta_1 \, \pi(\beta_{11}) d\beta_{11} \tag{32}$$

where $\pi(\beta_{11})$ is arbitrary, the resulting intrinsic prior for $\xi \equiv \beta_{11}$ is proper.

The proof of Lemma 3 again follows from Theorem 3.

An important distinction arises here compared to the previous hypothesis testing situations. The minimal training samples are not exchangeable and so one cannot simplify the expression in (4) to obtain a single expectation. Nevertheless, Theorem 3 still applies since group invariance properties are preserved for each minimal training sample separately. This is essentially how Theorem 3 is proved in the Appendix. The full generality of Theorem 3 for non-exchangeable training samples is utilized to obtain Lemma 3.

Berger and Pericchi (1996a) choose initial non-informative priors on H_0 and H_1 to be the Jeffreys' priors and show by direct calculations that the intrinsic prior on β_{11} is proper. In our case, the default Jeffreys prior is obtained as a special case by taking $\pi(\beta_1) \equiv 1$ in Lemma 3.

Case 2: σ_0 and σ_1 are unknown

The group of transformations on y, in this situation, which leaves both H_0 and H_1 invariant is $y \longrightarrow cy + X_0 b$, where $c > 0$ and $b \in R^{k_0}$. The action of G on the space of observations induces group actions on the parameter spaces given by $(\sigma_0, \beta_0) \longrightarrow (c\sigma_0, c\beta_0 + b)$ and $(\sigma_1, \beta_1, \beta_{11}) \longrightarrow (c\sigma_1, c\beta_1 + b, c\beta_{11})$. The right-Haar priors on H_0 and H_1 are

$$\nu(\sigma_i, \beta_i) = \frac{1}{\sigma_i} \, d\sigma_i \, d\beta_i \tag{33}$$

for $i = 0, 1$, respectively. An invariant under the action of the group for H_1 is β_{11}/σ_1. So, in the notation of Section 4, $\xi \equiv \beta_{11}/\sigma_1$.

Lemma 4 *For π_0^N and π_1^N satisfying*

$$\pi_0^N(\sigma_0, \beta_0) \quad = \quad \frac{1}{\sigma_0} \, d\sigma_0 \, d\beta_0 \quad and \tag{34}$$

$$\pi_1^N(\sigma_1, \beta_1, \xi) \quad = \quad \frac{1}{\sigma_1} \, d\sigma_1 d\beta_1 \, \pi(\xi) d\xi$$

$$\quad = \quad \frac{1}{\sigma_1} \, d\sigma_1 d\beta_1 \, \pi(\beta_{11}/\sigma_1) \cdot \frac{1}{\sigma_1^{k_1}} d\beta_{11}, \tag{35}$$

where π is arbitrary, the resulting intrinsic prior for $\xi \equiv \beta_{11}/\sigma_1$ is proper.

The proof of Lemma 4 follows from Theorem 3 once again.

The choice of $\pi \equiv 1$ corresponds to the **modified Jeffreys prior**, suggested by Berger and Pericchi (1996a) as a choice for a default prior for the testing in (30) for unknown variance. Berger and Pericchi

(1996a) also show that the intrinsic prior arising from the reference prior is proper. The propriety of the intrinsic prior when the reference prior is used could not be shown by our method.

Again, the minimal training samples, $x(l)$, are not exchangeable but, by the same reasoning as in Case 1, Theorem 3 applies.

7   Group-Theoretic Definitions and Proofs

The following definitions are taken from group invariance theory; see, e.g., Eaton (1989) and Wijsman (1990).

Definition 3 *Let G be a group of measurable one-to-one transformations of $\mathcal{X}$ onto itself; $g : x \longrightarrow g \circ x$ such that the family $\{P_\theta : \theta \in \Theta\}$ is closed with respect to this transformation, i.e., for $X \sim P_\theta, g \circ X \sim P_{\theta'}$ where $\theta' \in \Theta$ is given by $\theta' \equiv \bar{g} \circ \theta$. In this case we say that the family $\{P_\theta\}$ is $\boldsymbol{G\text{-}invariant}$.*

The action of the group G on $\mathcal{X}$ induces another group $\bar{G}$ on the parameter space Θ. Actually it can be shown that G and $\bar{G}$ are isomorphic to each other. Assume the family $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ admits a family of densities $\{p(\cdot|\theta) : \theta \in \Theta\}$ with respect to a σ-finite dominating measure λ. Let $G, \bar{G}$ act on $\mathcal{X}, \Theta$ respectively and λ be relatively invariant under G with a multiplier $\chi(\cdot)$. If the densities satisfy

$$p(x|\theta) = p(gx|\bar{g}\theta)\chi(g), \tag{36}$$

then the family $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is G-invariant. Furthermore, $gP_\theta = P_{\bar{g}\theta}$.

Definition 4 *For a group of measurable transformations G acting on a space $\mathcal{X}$, G is said to be transitive on $\mathcal{X}$ if for any $x, x' \in \mathcal{X}$, there is a g in G such that $x' = g \circ x$.*

G in the above definition may be applied to $\bar{G}_0$ and $\bar{G}_1$ acting on the spaces Θ_0 and Θ_1 for the hypotheses testing problem in (1).

Definition 5 *The isotropy subgroup of G at x is the subgroup G_x,*

$$G_x = \{g \in G : g \circ x = x\} \tag{37}$$

The isotropy subgroup of G is said to be trivial if $G_x = e \, \forall x \in \mathcal{X}$. Define a function $\phi : \bar{G} \longrightarrow \mathcal{X}$ by $\phi(\bar{g}) = \bar{g} \circ e$ where e is the identity element of $\mathcal{X}$. Then, transitivity of $\bar{G}$ on $\mathcal{X}$ implies that ϕ is onto. If further, G has a trivial isotropy subgroup, the function ϕ is also one-one, and in that case it is an automorphism of G and $\mathcal{X}$.

Definition 6 *Let G be a group of measurable transformations of $\mathcal{X}$ onto itself. A $\boldsymbol{maximal\ invariant}$, $\tau(x)$, is a function on $\mathcal{X}$ satisfying*

- $\tau(x)$ *is invariant under* G, *i.e.* $\tau(g \circ x) = \tau(x)$ *for* $g \in G$ *and* $x \in \mathcal{X}$;

- $\tau(x)$ *takes different values on different orbits of* G, *i.e.* $\tau(x_1) = \tau(x_2)$ *implies* $x_1 = gx_2$ *for some* $g \in G$.

Definition 7 *The* **left-hand moduli** *of* G, Δ_l, *is such that for any right-Haar measure* ν, $\nu(g \cdot A) = \Delta_l(g)\nu(A) \, \forall \, g \in G$. *Similarly, the* **right-hand moduli** *of* G, Δ_r, *is such that for any left-Haar measure* μ_L, $\mu_L(A \cdot g) = \Delta_r(g)\mu_L(A) \, \forall \, g \in G$.

PROOF OF THEOREM 1

We assume all models under consideration have the structure given in (36). Consider samples, $\underline{x}_k$ and $\underline{y}_k$, of size k. Let $\tau(\underline{x}_k)$ be a maximal invariant. Then $\tau(\underline{x}_k) = \tau(\underline{y}_k) \quad \Rightarrow \quad \underline{x}_k = g\underline{y}_k \quad$ for some g.

$$
\begin{aligned}
B_{01}^{\mu,\pi}(\underline{x}_k) &= \frac{\displaystyle\int_{\Theta} f_0(\underline{x}_k|\theta)d\mu_\phi(\theta)}{\displaystyle\int_{\Omega}\int_{\Theta} f_1(\underline{x}_k|\theta,\xi)d\mu_\phi(\theta)\pi(\xi)d\xi} \\[2ex]
&= \frac{\displaystyle\int_{\Theta} f_0(g\underline{y}_k|\theta)d\mu_\phi(\theta)}{\displaystyle\int_{\Omega}\int_{\Theta} f_1(g\underline{y}_k|\theta,\xi)d\mu_\phi(\theta)\pi(\xi)d\xi} \\[2ex]
&= \frac{\displaystyle\int_{\bar{G}} f_0(g\underline{y}_k|\bar{h}\circ e)\,d\mu(\bar{h})}{\displaystyle\int_{\Omega}\left[\int_{\bar{G}} f_1(g\underline{y}_k|\bar{h}\circ e,\xi)\,d\mu(\bar{h})\right]\pi(\xi)d\xi} \\[2ex]
&= \frac{\alpha_l(\bar{g})\chi(\bar{g})^{-1}\displaystyle\int_{\bar{G}} f_0(\underline{y}_k|\bar{h}\circ e)d\mu(\bar{h})}{\alpha_l(\bar{g})\chi(\bar{g})^{-1}\displaystyle\int_{\Omega}\left[\int_{\bar{G}} f_1(\underline{y}_k|\bar{h}\circ e,\xi)d\mu(\bar{h})\right]\pi(\xi)d\xi} \\[2ex]
&= \frac{\displaystyle\int_{\Theta} f_0(\underline{y}_k|\theta)d\mu_\phi(\theta)}{\displaystyle\int_{\Omega}\int_{\Theta} f_1(\underline{y}_k|\theta,\xi)d\mu_\phi(\theta)\pi(\xi)d\xi} = B_{01}^{\mu,\pi}(\underline{y}_k)
\end{aligned}
$$

So, $B_{01}^{\mu,\pi}(\cdot)$ is a function of the maximal invariant $\tau(\cdot)$. Transitivity of $\bar{G}$ implies that the distribution of τ is independent of θ_0 and θ_1 under H_0 and H_1 respectively, and hence the proof. QED.

PROOF OF THEOREM 2

We first state a few basic results:

Theorem 4 *[Wijsman] For* $i = 0,1$, *let* P_i *be a distribution on* $\mathcal{X}$ *with density* p_i, *with respect to a* χ-*relatively invariant measure* λ. *Let* $\tau(x)$ *be a maximal invariant with distributions* P_0^τ *and* P_1^τ, *respectively.*

Then P_0^τ and P_1^τ are absolutely continuous with respect to a dominating measure μ^τ and, for any $x \in \mathcal{X}$,

$$\frac{dP_1^\tau}{dP_0^\tau}(\tau(x)) = \frac{\int p_1(gx)\chi(g)d\mu_L(g)}{\int p_0(gx)\chi(g)d\mu_L(g)}, \tag{38}$$

where μ_L is a left-invariant measure on G.

Proof: *See Wijsman (1990).*

Theorem 5 *For $i = 0,1$, let $\mathcal{P}_i$ denote the family of distributions with densities $\{p_i(\cdot|\theta) : \theta \in \Theta\}$ with respect to a χ-relatively invariant measure λ. For $i = 0,1$, assume that the class of densities $\{p_i(\cdot|\theta) : \theta \in \Theta\}$ are G-invariant, and $\bar{G}$ is transitive on Θ. Then,*

$$\frac{dP_1^\tau}{dP_0^\tau}(\tau(x)) = \frac{\int p_1(x|\theta)d\nu_\phi(\theta)}{\int p_0(x|\theta)d\nu_\phi(\theta)}. \tag{39}$$

Proof:

$$
\begin{aligned}
\frac{\int p_1(x|\theta)d\nu_\phi(\theta)}{\int p_0(x|\theta)d\nu_\phi(\theta)} &= \frac{\int p_1(x|\bar{g}\circ e)d\nu(\bar{g})}{\int p_0(x|\bar{g}\circ e)d\nu(\bar{g})} \\[2mm]
&= \frac{\int p_1(g^{-1}x|e)\chi(g^{-1})d\nu(\bar{g})}{\int p_0(g^{-1}x|e)\chi(g^{-1})d\nu(\bar{g})} \\[2mm]
&= \frac{\int p_1(gx|e)\chi(g)d\mu_L(g)}{\int p_0(gx|e)\chi(g)d\mu_L(g)} \\[2mm]
&= \frac{dP_1^\tau}{dP_0^\tau}(\tau(x)) \quad (\text{ by Theorem 4 }),
\end{aligned}
$$

where $d\nu(\bar{g}^{-1}) = d\mu_L(g)$. QED.

Apply (39) to an $f_0(x(l)|\theta)$ in M_0 and $f_1(x(l)|\theta,\xi)$ in M_1 to obtain

$$\frac{dP_\xi^\tau}{dP_0^\tau}(\tau(x(l))) = \frac{\int_\Theta f_1(x(l)|\theta,\xi)\,d\nu_\phi(\theta)}{\int_\Theta f_0(x(l)|\theta)\,d\nu_\phi(\theta)}. \tag{40}$$

Then,

$$
\int_\Omega E_\xi\left\{B_{01}^{\nu,\pi}(X(l))\right\}\pi(\xi)\,d\xi
$$

$$
= \int_\Omega \left[\int_{\tau(\mathcal{X}(l))} \frac{\int_\Theta f_0(x(l)|\theta)\,d\nu_\phi(\theta)}{\int_\Omega \int_\Theta f_1(x(l)|\theta,\xi)\,d\nu_\phi(\theta)\pi(\xi)d\xi} \cdot dP_\xi^\tau(\tau(x(l))) \right]\pi(\xi)\,d\xi
$$

$$= \int_{\Omega} \left[\int_{\tau(\mathcal{X}(l))} \frac{\int_{\Theta} f_0(x(l)|\theta)\, d\nu_\phi(\theta)}{\int_{\Omega} \int_{\Theta} f_1(x(l)|\theta,\xi)\, d\nu_\phi(\theta)\pi(\xi)d\xi} \cdot \frac{dP_\xi^\tau}{dP_0^\tau} \cdot dP_0^\tau(\tau(x(l))) \right] \pi(\xi)\, d\xi$$

$$= \int_{\Omega} \int_{\tau(\mathcal{X}(l))} \frac{\int_{\Theta} f_0(x(l)|\theta)\, d\nu_\phi(\theta)}{\int_{\Omega} \int_{\Theta} f_1(x(l)|\theta,\xi)\, d\nu_\phi(\theta)\pi(\xi)d\xi} \cdot \frac{\int_{\Theta} f_1(x(l)|\theta,\xi)\, d\nu_\phi(\theta)}{\int_{\Theta} f_0(x(l)|\theta)\, d\nu_\phi(\theta)} \cdot$$

$$dP_0^\tau(\tau(x(l)))\pi(\xi)\, d\xi \quad \text{from (40)}$$

$$= \int_{\tau(\mathcal{X}(l))} \frac{\int_{\Omega} \int_{\Theta} f_1(x(l)|\theta,\xi)\, d\nu_\phi(\theta)\, \pi(\xi)\, d\xi}{\int_{\Omega} \int_{\Theta} f_1(x(l)|\theta,\xi)\, d\nu_\phi(\theta)\, \pi(\xi)\, d\xi} \cdot dP_0^\tau(\tau(x(l)))$$

$$= \int_{\tau(\mathcal{X}(l))} dP_0^\tau(\tau(x(l)))$$

$$= 1. \quad \text{QED.}$$

PROOF OF THEOREM 3

$$\int_{\Omega} B_1^*(\xi)\, \pi(\xi)\, d\xi$$

$$= \int_{\Omega} \lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} E_\xi\{B_{01}^{\nu;\pi}(X(l))\}\, \pi(\xi)\, d\xi$$

$$= \lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} \int_{\Omega} E_\xi\{B_{01}^{\nu;\pi}(X(l))\}\, \pi(\xi)\, d\xi$$

$$= \lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} 1 \quad \text{(by Theorem 2)}$$

$$= 1. \quad \text{QED.}$$

8 Acknowledgments

References

Berger, J. and Bernardo, J. M. (1992) On the development of reference priors (disc: P49-60). In *Bayesian Statistics 4. Proceedings of the Fourth Valencia International Meeting*, pp. 35– 49.

Berger, J. and Pericchi, L. (1996a) The intrinsic Bayes factor for linear models. In *Bayesian Statistics 5*, pp. 25–44. Oxford University Press.

Berger, J. and Pericchi, L. (1996b) The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, **91,** 109–122.

Berger, J., Pericchi, L. and Varshavsky, J. (1998) Bayes factors and marginal distributions in invariant testing situations. *Sankya Series* A, **60,** 307–321.

Berger, James O. and Pericchi, L. R. (1997) On The Justification Of Default and Intrinsic Bayes Factors. In *Modelling and prediction (Hsinchu, 1994)*. Springer, New York.

Bernardo, J. M. (1979) Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society, Series B*, **41,** 113–128.

Bertolino, F. and Racugno, E. (1996) Is Intrinsic Bayes Factor Intrinsic ? *Metron*, **1-2,** 5–15.

Dawid, A. P., Stone, M. and Zidek, J. V. (1973) Marginalization paradoxes in Bayesian and structural inference (with discussion). *Journal of the Royal Statistical Society, Series B*, **35,** 189–233.

Dmochowski, J. (1995) Properties of intrinsic Bayes factors. Ph.D. Thesis. Purdue University, W. Lafayette.

Eaton, M. L. (1983) *Multivariate Statistics: A Vector Space Approach*. Wiley, New York.

Eaton, M. L. (1989) *Group Invariance Applications in Statistics*. Institute of Mathematical Statistics, Hayward, California.

Jeffreys, H. (1961) *Theory of Probability (3rd ed.)*. Claderon Press, Oxford.

Moreno, E., Bertolino, F. and Racugno, W. (1997) Default approaches to compare means of normal populations. In *Special issue of Rasegna di Metodici Statistici ed Applicazioni* (ed. W. Racugno), pp. 133–148.

Moreno, E., Bertolino, F. and Racugno, W. (1998) An intrinsic limiting procedure for model selection and hypotheses testing. *Journal of the American Statistical Association*, **93,** 1451–1460.

Sanso, B. (1997) Discussion of "Default approaches to compare means of normal populations" by Moreno E., Bertolino F. and Racugno, W. In *Special issue of Rasegna di Metodici Statistici ed Applicazioni* (ed. W. Racugno), pp. 149–151.

Sanso, B., Pericchi, L. and Moreno, E. (1996) On the robustness of the intrinsic Bayes factor for nested models (with discussion). *IMS Lectures Notes-Monograph Series*, **29,** 157–176.

Smith, A. F. M. and Spiegelhalter, D. J. (1980) Bayes factors and choice criteria for linear models. *Journal of the Royal Statistical Society, Series B, Methodological*, **42,** 213–220.

Spiegelhalter, D. J. and Smith, A. F. M. (1982) Bayes factors for linear and loglinear models with vague prior information. *Journal of the Royal Statistical Society, Series B, Methodological*, **44,** 377–387.

Stone, M. (1970) Necessary and sufficient condition for convergence in probability to invariant posterior distributions. *The Annals of Mathematical Statistics*, **41,** 1349–1353.

Wijsman, R. A. (1990) *Invariant Measures on Groups and their Uses in Statistics.* Institute of Mathematical Statistics, Hayward, California.