

# A Note on the Consistency of Bayes Factors for Testing Point Null versus Nonparametric Alternatives

Sarat C. Dass

Michigan State University

Jaeyong Lee

Pennsylvania State University

July 30, 2002

## **Abstract**

When testing a point null hypothesis versus an alternative that is vaguely specified, a Bayesian test usually proceeds by putting a non-parametric prior on the alternative and then computing a Bayes factor based on the observations. This paper addresses the question of consistency, that is, whether the Bayes factor is correctly indicative of the null or the alternative as the sample size increases. We establish several consistency results in the affirmative under fairly general conditions. Consistency of Bayes factors for testing a point null versus a *parametric* alternative has long been known. The results here can also be viewed as the non-parametric extension of the parametric counterpart.

*MSC:* 62G20; 62C10

*Key words:* Bayes Factor, Consistency, Dirichlet process, Polya Tree, Infinite Dimensional Exponential Family.

# 1 Introduction

Non-parametric Bayesian methods have been popular and successful in many estimation problems but their relevance in hypotheses testing situations have become of interest only recently. In particular, the testing of a parametric null versus a non-parametric alternative has received considerable attention from Bayesians, e.g., Berger and Guglielmi (1998), Verdinelli and Wasserman (1998), Carota and Parmigiani (1996), and Florens, Richard and Rolin (1996). Berger and Guglielmi (1998) consider the problem of goodness of fit in the framework of testing a parametric null versus a non-parametric alternative and derive measures of goodness of fit closely related to the Bayes factor. By looking at goodness of fit as a Bayesian test of hypotheses, one can take advantage of many of its attractive features. Bayesian hypothesis testing is not based on asymptotic results, and thus, can be used equally effectively on small or moderate sample sizes. Bayesian hypotheses testing uses Bayes factors to decide between accepting or rejecting the null hypothesis. Thus, as the sample size increases, one can ask if the Bayes factor is correctly indicative of  $H_0$  or  $H_1$  given that the sampling density belongs to one of the two hypotheses. This is the question of consistency.

Even though Bayesian answers in hypothesis testing problems are not operationally based on asymptotics, consistency of the resulting Bayes factor is an important issue that needs to be addressed. In the case of estimation using non-parametric priors, Diaconis and Freedman (1986) show that some posteriors based on  $n$  samples need not be consistent, that is, the posterior may not put mass tending to one for sufficiently small neighborhoods of the true parameter value. Thus, inference based on such inconsistent posteriors can be highly misleading.

Analogously, in hypotheses testing, it is important to know if the Bayes procedure based on the Bayes factor actually leads to sensible answers as the sample size increases. Consistency holds for Bayes factors when parametric families are involved in the testing scenario. Even when the sampling distribution does not belong to either  $H_0$  or  $H_1$  in the paramet-

ric case, the Bayes factor eventually chooses the hypothesis that is closest to the sampling density in a Kullback Liebler sense. Exact rates of convergence are also well known.

In the case of infinite dimensional parameter spaces, relatively little is known about consistency and rates of convergence of Bayes factors in the case of general non-parametric priors. We establish consistency for the Bayes factor when the null hypothesis is true for any arbitrary non-parametric prior. In the case when the alternative hypothesis is true, we show that the set of all sampling densities under which consistency holds has measure one with respect to the non-parametric prior, regardless of the prior chosen. Our goal is to establish consistency in terms of conditions satisfied by a sampling density in the support of an arbitrary non-parametric prior, and not only on a case by case basis. We only consider non-parametric priors on the space of all probability density functions for reasons explained in Section 3.

The remainder of this paper is organized as follows. Section 2 gives the motivation and definition of consistency pertaining to Bayes factors. Section 3 discusses some well known examples of non-parametric priors on the space of all densities. Section 4 gives the proofs of theorems in Section 2. We end this paper with a discussion of testing a composite parametric null versus a non-parametric alternative in Section 5.

## 2 Consistency of Bayes Factors

The following notations will be used throughout the paper. Let  $\mathcal{X}$  be a complete separable metric space (or Polish space),  $\mu$  be a  $\sigma$ -finite measure on  $\mathcal{X}$  and  $\mathcal{F}$  be the space of all probability densities with respect to  $\mu$  with support  $\mathcal{X}$ . Also, denote by  $X_1, X_2, \dots$ , random variables taking values in  $\mathcal{X}$ , which are independent and identically distributed (iid) with a density  $f \in \mathcal{F}$ . Consider the following hypothesis testing scenario

$$H_0 : f = f_0 \text{ versus } H_1 : f \neq f_0. \quad (1)$$

Equation (1) is the most general form of testing a point null versus a non-parametric alternative. A Bayesian testing procedure would proceed by first specifying prior probabilities,  $\pi_0$  and  $\pi_1$ , of the null hypothesis and the alternative, respectively, and a non-parametric prior  $\pi$  on the space of the alternative,  $H_1$ . We postpone the discussion of what an appropriate prior should be until Section 3 but for now, assume that a non-parametric prior is given. The Bayes factor for the testing of (1), based on a sample,  $\underline{x}_n$ , of size  $n$ , is the ratio of the marginal under  $H_0$  to the marginal under  $H_1$ , and is given by the expression

$$B(\underline{x}_n) = \frac{\prod_{i=1}^n f_0(x_i)}{\int_{H_1} \prod_{i=1}^n f(x_i) \pi(df)}. \quad (2)$$

The Bayes factor in (2) can also be interpreted as the ratio of posterior odds to the prior odds of  $H_0$  to  $H_1$ . To see this, define an *overall* prior on  $H_0 \cup H_1$  as

$$\pi^*(f) = \pi_0 \cdot I_{f_0}(f) + \pi_1 \cdot I_{H_1}(f) \cdot \pi(f), \quad (3)$$

where  $I_A(\cdot)$  stands for the indicator function of the set  $A$ , i.e.,  $I_A(f) = 0$  if  $f \notin A$  and  $I_A(f) = 1$  if  $f \in A$ . We use the following notation for generic priors and posteriors, namely, if  $\nu(\cdot)$  is a prior on  $H_0 \cup H_1$ , then we will denote the posterior derived from  $\nu$  based on a sample,  $\underline{x}_n$ , of size  $n$ , by  $\nu(\cdot | \underline{x}_n)$ . Thus, for the prior  $\pi^*$ , the posterior and prior odds ratio is related to the Bayes factor by

$$\frac{\pi^*(H_0 | \underline{x}_n)}{\pi^*(H_1 | \underline{x}_n)} = \frac{\pi_0}{\pi_1} \cdot B(\underline{x}_n). \quad (4)$$

For all subsequent discussions, we take the default choice for  $\pi_0$  and  $\pi_1$ , namely,  $\pi_0 = \pi_1 = 1/2$ . In this case, the posterior odds ratio is exactly equal to the Bayes factor. Thus, given the observations  $x_1, x_2, \dots, x_n$ , large values of  $B$  would indicate that there is strong evidence for  $H_0$  based on the data whereas small values of  $B$  would indicate otherwise. As the sample size increases indefinitely, we would expect to get perfect information about the sampling density, say  $f$ , and the Bayes factor should also correctly and overwhelmingly be able to decide between  $H_0$  and  $H_1$ . This motivates the following definition for the consistency of the Bayes factor.

Let  $\mathcal{X}^n$  and  $\mathcal{X}^\infty$  be the products of  $n$  and infinite copies of  $\mathcal{X}$ . Also, let  $P_f^n$  and  $P_f^\infty$  be the  $n$  and infinite products of the probability measure  $P_f$ , which has density  $f$ , on  $\mathcal{X}^n$  and  $\mathcal{X}^\infty$ , respectively.

**Definition 1** *The Bayes factor,  $B(\underline{x}_n)$ , for the testing of (1) is said to be consistent if*

$$\lim_{n \rightarrow \infty} B(\underline{x}_n) = \infty, \quad a.s. P_{f_0}^\infty,$$

and for any  $f \neq f_0$ ,

$$\lim_{n \rightarrow \infty} B(\underline{x}_n) = 0, \quad a.s. P_f^\infty.$$

Before we give the theorems establishing consistency of Bayes factors, we need a few more definitions. The Kullback-Leibler divergence,  $K(f, g)$ , provided it exists, between two densities  $f$  and  $g$  in  $\mathcal{F}$  is defined as

$$K(f, g) = \int f(x) \log \frac{f(x)}{g(x)} \mu(dx). \quad (5)$$

Also let

$$K_\epsilon(f) = \{g \in \mathcal{F} : K(f, g) < \epsilon\}, \quad \text{for } \epsilon > 0. \quad (6)$$

We say  $f$  is in the Kullback-Leibler support of  $\pi$ , if

$$\pi(K_\epsilon(f)) > 0, \quad \text{for all } \epsilon > 0.$$

With the above definitions, we can now state three theorems establishing consistency of the Bayes factor.

**Theorem 1** *Under  $f_0 \in H_0$ ,*

$$\lim_{n \rightarrow \infty} B(\underline{x}_n) = \infty, \quad a.s. P_{f_0}^\infty$$

Note that one cannot use Schwartz's criteria for consistency for the prior  $\pi^*$  to obtain Theorem 1. This is because Schwartz's criteria (Schwartz (1965)) gives consistency only for weak

neighborhoods of  $f_0$  whereas we want consistency at  $f = f_0$ . Our goal here is to establish strong consistency (almost sure convergence) of the Bayes factor for a general prior  $\pi$  on  $H_1$ . Our proof of Theorem 1 follows from a rather simple observation that  $\pi^*$  puts positive mass on  $H_0$  and the fact that the posterior odds ratio is proportional to the Bayes factor (see (4)). Thus, the argument of Doob (1949) is applied without much change. For the completeness, the proof is given in Section 4. An involved proof of a weaker result appeared in Verdinelli and Wasserman (1998) in the special case when  $\pi$  is taken to belong to the class of infinite dimensional exponential family priors. However, it should be mentioned that their proof can be modified to give rates and power for local alternatives whereas the proofs presented here cannot.

**Theorem 2** *Let  $\Theta = \{f \in H_1 : B(\underline{x}_n) \longrightarrow 0 \text{ a.s. } P_f^\infty\}$ . Then,  $\pi(\Theta) = 1$ .*

Theorem 2 states that the Bayes factor is, indeed, consistent for a large set of densities in  $H_1$ , namely, a set which has  $\pi$ -probability 1. However, Theorem 2 does not say much about any one *particular* sampling density,  $f$  in  $H_1$ . To obtain consistency for a particular sampling density,  $f$ , we have to further assume that  $f$  belongs to the Kullback-Leibler support of the prior,  $\pi$ . This is the result of

**Theorem 3** *Suppose  $f \in H_1$  is such that  $f$  is in the Kullback-Leibler support of the prior  $\pi$ . Then, under  $f$ ,*

$$\lim_{n \rightarrow \infty} B(\underline{x}_n) = 0 \text{ a.s. } P_f^\infty.$$

We give proofs of the above theorems in Section 4. Verdinelli and Wasserman (1998) gave the same result with the infinite dimensional exponential family model. Also, their proof can be modified to give a more general result and our proof of Theorem 3 is in the same spirit. In the following section, we give examples of non-parametric priors, where the support condition of Theorem 3 has been established for estimation problems. Note that this condition is also sufficient to establish consistency of Bayes factors in hypotheses testing situations by the result of Theorem 3.

### 3 Examples

#### 3.1 Posterior Consistency of Dirichlet Normal Mixtures

Let  $\mathcal{X}$  be the space of observables and  $\alpha$  be a finite measure on  $\mathcal{X}$ . Define  $u_1, u_2, \dots$  to be iid random variables from  $Beta(1, \alpha(\mathcal{X}))$  and  $Y_1, Y_2, \dots$  to be independent random variables, independent of  $u_1, u_2, \dots$ , each distributed according to the probability measure  $\alpha_0(\cdot) = \alpha(\cdot)/\alpha(\mathcal{X})$ . Define  $p_1 = u_1$ , and  $p_i = u_i \prod_{j=1}^{i-1} (1 - u_j)$  (so that  $p_i \geq 0$  for all  $i \geq 1$  and  $\sum_{i=1}^{\infty} p_i = 1$ , see Sethuraman (1994)).

Denote  $\frac{1}{\sigma} \phi(\frac{\cdot - \mu}{\sigma})$  to be the normal density function with mean  $\mu$  and standard deviation  $\sigma$ . Consider the random density function,  $g(x)$ , given by

$$g(x) = \sum_{i=1}^{\infty} p_i \frac{1}{\sigma} \phi\left(\frac{x - Y_i}{\sigma}\right), \quad (7)$$

and  $\sigma \sim \nu$ , where  $\nu$  is a probability measure on positive real line. The Dirichlet normal mixture distribution is the distribution on probability measures  $P$  arising from the random densities  $g(x)$  in (7). There can be various other choices of mixtures based on different choices of the kernel function  $\phi$  and  $\nu$ . The modelling and computational aspects of Dirichlet mixtures were studied, for instance, by MacEachern and Müller (1998).

Ghosal, Ghosh and Ramamoorthi (1999b) studied the issue of posterior consistency in the context of density estimation using Dirichlet mixtures. They gave conditions under which the sampling density,  $f$ , belongs to the Kullback-Leibler support of the Dirichlet mixture prior. For example, if

$$f(x) = \int \frac{1}{h} \phi\left(\frac{x - y}{h}\right) dP(y),$$

where  $h > 0$  is in the support of  $\nu$  and  $P$  has compact support, then  $f$  belongs to the Kullback-Leibler support of the Dirichlet mixture prior (see their Theorems 3). Ghosal, Ghosh and Ramamoorthi (1999b) also give a more general theorem for posterior consistency based on the properties of the underlying measure of Dirichlet process priors. We refer the reader to their Theorem 5 for a complete statement and proof.

The conditions for posterior consistency stated in the theorems in Ghosal, Ghosh and Ramamoorthi (1999b) are also sufficient for the consistency of Bayes factors for the hypotheses testing situation by Theorem 3.

### 3.2 Posterior Consistency of Polyá Tree Priors

We quote a theorem from Ghosal, Ghosh and Ramamoorthi (1999a) that establishes the Kullback-Liebler support of the sampling density. Here,  $\{B_{\epsilon_1, \epsilon_2, \dots, \epsilon_k}\}$  represents a usual hierarchical partition of the real line associated with the construction of a Polyá Tree Prior, and the conditional probabilities  $P(B_{\epsilon_1, \epsilon_2, \dots, \epsilon_k} | B_{\epsilon_1, \epsilon_2, \dots, \epsilon_{k-1}})$  are distributed according to  $Beta(\alpha_{\epsilon_1, \epsilon_2, \dots, \epsilon_k}, 1 - \alpha_{\epsilon_1, \epsilon_2, \dots, \epsilon_k})$  for some constants  $0 < \alpha_{\epsilon_1, \epsilon_2, \dots, \epsilon_k} < 1$ .

The following theorem gives conditions for a density to be in the Kullback-Leibler support of a Polyá tree prior.

**Theorem 4** (Ghosal, Ghosh and Ramamoorthi (1999a)) *Suppose that  $\lambda$  is a continuous probability measure with  $\lambda(B_{\epsilon_1, \epsilon_2, \dots, \epsilon_k}) = 2^{-k}$  for all  $k$  and further  $\alpha_{\epsilon_1, \epsilon_2, \dots, \epsilon_k} = a_k$ . If  $\sum_k a_k^{-1/2} < \infty$ , then any density  $f$  with respect to  $\lambda$  with  $\int f \log f d\lambda < \infty$  belongs to the Kullback-Leibler support of the Polyá tree.*

Thus, when a Polyá tree prior is chosen as the prior for the non-parametric alternative in (1), Theorem 4 provides conditions ensuring consistency of the resulting Bayes factor.

### 3.3 Infinite Dimensional Exponential Family Priors

Verdinelli and Wasserman (1998) discuss the use of the infinite dimensional exponential family priors for testing goodness of fit. They cast the testing problem (1) into the testing problem of

$$H_0 : F_0(X_1), F_0(X_2), \dots, F_0(X_n) \stackrel{\text{iid}}{\sim} Uniform(0, 1)$$

versus

$$H_1 : F_0(X_1), F_0(X_2), \dots, F_0(X_n) \stackrel{\text{iid}}{\not\sim} \text{Uniform}(0, 1),$$

where  $F_0$  is the cumulative distribution function (cdf) of the density  $f_0$  in (1). The infinite dimensional exponential family is constructed for distributions with support on the unit interval  $[0, 1]$ . They use a sequence of Legendre polynomials,  $\{\xi_j(\cdot), j = 1, 2, \dots\}$ , defined by

$$\xi_j(x) = \frac{1}{2^j j!} \frac{d^j}{dx^j} (x^2 - 1)^j,$$

and use the Legendre polynomials together with other coefficients,  $\phi = (\phi_1, \phi_2, \dots)$ , to define infinite exponential densities of the form

$$g(u|\phi) = \exp\left(\sum_{j=1}^{\infty} \phi_j \xi_j(u) - c(\phi)\right),$$

where  $c(\phi) = \log \int_0^1 \exp(\sum_j \phi_j \xi_j(u)) du$  is the normalizing constant. In order to get random densities, Verdinelli and Wasserman (1998) put priors on the coefficients,  $\phi$ , given by

$$\phi_j \sim \text{independent } N(0, \tau^2/c_j^2), \quad (8)$$

where  $\tau$  and  $c_j$ 's are constants. Using a theorem from Barron (1988), they show that

**Theorem 5** *If  $K(f_0, f) < \infty$  and  $\pi$  is the infinite dimensional exponential family prior with  $c_j = j^k$  in (8) where  $k > 8$  and  $\tau > 0$ , then  $f$  is in the Kullback-Leibler support of  $\pi$ .*

Thus, consistency of the Bayes factor at  $f \in H_1$  follows easily from Theorem 5 (as was also shown in Verdinelli and Wasserman (1998)).

## 4 Proof of Theorems

In this section, we give the proofs of Theorems 1, 2 and 3. The reader is referred to Section 2 for the notation used here. We use the weak topology on  $\mathcal{F}$  and the usual topology on  $\mathcal{X}^\infty$ . The product topology on the space  $\mathcal{F} \times \mathcal{X}^\infty$  is generated in the usual way.

Let  $\pi$  be a probability measure on  $\mathcal{F}$ . Given an  $f$  sampled from  $\pi$ , the observations  $X_1, X_2, \dots$  are independent and identically distributed according to  $P_f$ , the probability measure corresponding to  $f$ . The probability measure  $P_f$  is uniquely determined by  $f$  and vice versa upto an equivalence class resulting from the equivalence relation,  $\sim$ , defined by

$$f \sim g \quad \text{if and only if} \quad f = g \text{ a.e. } \mu \quad (9)$$

The notation  $f$  will now stand for the equivalence class that it generates. For the prior probability  $\pi$  on  $\mathcal{F}$ , write  $Q_\pi$  for the probability measure on  $\mathcal{F} \times \mathcal{X}^\infty$  defined by

$$Q_\pi(A \times B) = \int_A P_f^\infty(B) \pi(df), \quad (10)$$

where  $A$  is Borel in  $\mathcal{F}$  and  $B$  is Borel in  $\mathcal{X}^\infty$ .

We will need the following result for our proofs.

**Lemma 1** *Let  $\pi^*$  be a prior on  $H_0 \cup H_1$ . The posterior,  $\pi^*(\cdot | \underline{x}_n) \longrightarrow \delta_f(\cdot)$  weakly as  $n \rightarrow \infty$   $Q_{\pi^*}$  - a.s., where  $\delta_f(\cdot)$  is the degenerate probability at  $f$ .*

In other words, if we define

$$\Theta_0 = \{(f, x_1, x_2, \dots) : \pi^*(\cdot | \underline{x}_n) \longrightarrow \delta_f(\cdot) \text{ weakly} \},$$

then Lemma 1 states that  $Q_{\pi^*}(\Theta_0) = 1$ . Lemma 1 is an easy consequence of Doob's theorem. We refer the reader to Diaconis and Freedman (1986) (Corollary A.2 on page 16), and Ghosh and Ramamoorthi (2002) for further details and proofs. Ghosh and Ramamoorthi (2002) also discuss general measurability conditions needed for the application of Doob's theorem in the non-parametric set-up.

**Proof of Theorem 1** The proof follows easily by noting that

$$\pi^*(\{f_0\} | \underline{x}_n) = \frac{B}{B+1}$$

and the fact that  $\pi^*(\{f_0\}) = 1/2 > 0$  (since the posterior is always consistent at points with positive mass, from Lemma 1).  $\square$

To prove Theorem 2, we need the following lemma.

**Lemma 2** *Define*

$$\Theta^* = \{(x_1, x_2, \dots) : B(\underline{x}_n) \longrightarrow 0\}$$

and

$$\Theta_f = \{(x_1, x_2, \dots) : \pi^*(\cdot|\underline{x}_n) \longrightarrow \delta_f(\cdot) \text{ weakly}\}.$$

If the posterior is consistent at  $f \in H_1$ , then  $\Theta^* \supseteq \Theta_f$ .

**Proof.** Choose a  $(x_1, x_2, \dots)$  in  $\Theta_f$ , and a sufficiently small weak neighborhood of  $f$ ,  $N$ , not intersecting  $H_0$ . Since

$$B(\underline{x}_n) = \frac{\pi^*(H_0|\underline{x}_n)}{\pi^*(H_1|\underline{x}_n)}$$

and  $N \subseteq H_1$ , we have  $\pi^*(H_1|\underline{x}_n) \longrightarrow 1$  and  $\pi^*(H_0|\underline{x}_n) \longrightarrow 0$ . It follows that  $B(\underline{x}_n) \longrightarrow 0$ .

□

**Proof of Theorem 2.** By Result 1, we have  $Q_{\pi^*}(\Theta_0) = 1$ . Since

$$Q_{\pi^*}(\Theta_0) = \frac{1}{2} \cdot P_{f_0}^\infty(\Theta_{f_0}) + \frac{1}{2} \cdot \int_{H_1} P_f^\infty(\Theta_f) \pi(df),$$

we have that  $P_{f_0}^\infty(\Theta_{f_0}) = 1$  and  $P_f^\infty(\Theta_f) = 1$ ,  $\pi - a.s.$ . By Lemma 2,  $P_f^\infty(\Theta^*) = 1$ ,  $\pi - a.s.$

□

To prove Theorem 3, we need the following lemma which is obtained while proving Lemma 8.2 in Verdinelli and Wasserman (1998) (see also Ghosh and Ramamoorthi (2002)).

**Lemma 3** *Let  $f \in H_1$ . Suppose  $X_1, X_2, \dots$  are iid from  $f$  and  $f$  is in the Kullback-Leibler support of  $\pi$ . Then, for all  $\epsilon > 0$ ,*

$$\liminf_{n \rightarrow \infty} e^{n\epsilon} \int \frac{\prod_{i=1}^n g(X_i)}{\prod_{i=1}^n f(X_i)} \pi(dg) = \infty. \quad a.s. P_f^\infty.$$

**Proof of Theorem 3.** The Bayes factor for testing (1) can be written as

$$\begin{aligned} B &= \frac{\prod_{i=1}^n f_0(X_i)/f(X_i)}{\int \prod_{i=1}^n g(X_i)/f(X_i) \pi(dg)} \\ &= \frac{\exp(-n \frac{1}{n} \sum_{i=1}^n \log f(X_i)/f_0(X_i))}{\int \prod_{i=1}^n g(X_i)/f(X_i) \pi(dg)}. \end{aligned} \tag{11}$$

Let  $\epsilon = K(f, f_0)/2 > 0$ . We will show the numerator in (11) multiplied by  $e^{n\epsilon}$  goes to 0 and the denominator multiplied by  $e^{n\epsilon}$  goes to  $\infty$ . First, since  $\lim_{n \rightarrow \infty} (\epsilon - \frac{1}{n} \sum_{i=1}^n \log f(X_i)/f_0(X_i)) \leq -\epsilon/2$  a.s.  $P_f^\infty$ ,

$$\limsup_{n \rightarrow \infty} \exp\left\{n\left(\epsilon - \frac{1}{n} \sum_{i=1}^n \log \frac{f(X_i)}{f_0(X_i)}\right)\right\} = 0 \quad \text{a.s. } P_f^\infty.$$

Second, by Lemma 3,

$$\liminf_{n \rightarrow \infty} e^{n\epsilon} \int \prod_{i=1}^n g(X_i)/f(X_i) \pi(dg) = \infty \quad \text{a.s. } P_f^\infty.$$

Combining these two, we get the conclusion

$$B \rightarrow 0 \quad \text{a.s. } P_f^\infty. \quad \square$$

## 5 Discussion

In this paper, we only considered the problem of testing a point null versus non-parametric alternative and showed that under very weak conditions, the resulting Bayes factor was consistent. Of course, what is more interesting is to see if the consistency results hold for the more general composite testing of

$$H_0 : \quad f \text{ belongs to the } N(\mu, \sigma^2) \text{ family}$$

versus

$$H_1 : \quad f \text{ does not belong to the } N(\mu, \sigma^2) \text{ family,}$$

for example. The consistency of the Bayes factors for composite hypotheses testing situations such as the above is still an open question.

## Acknowledgments

The authors would like to thank Professors J. K. Ghosh and R. V. Ramamoorthi for their advice and comments on the paper. The authors would also like to thank the two anonymous referees for their helpful comments which have significantly enhanced the presentation of the paper.

## References

- Barron, A. R. (1988). The exponential convergence of posterior probabilities with implications for Bayes estimators of density functions. *Technical Report, Department of Statistics, University of Illinois, Champaign, IL.*
- Berger, J. O. and Guglielmi, A. (2001). Bayesian and conditional frequentist testing of a parametric model versus nonparametric alternatives. *Jour. Amer. Statist. Assoc.*, **96**, 174-184.
- Carota, C. and Parmigiani, G. (1996). On Bayes factors for nonparametric alternatives. *In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics 5, London, Oxford University Press, 507-511.*
- Doob, J.L. (1949). Application of the theory of martingales. *Colloque Internal du CNRS*, 22-28.
- Diaconis, P. and Freedman, D. (1986). On the consistency of Bayes estimates. *Ann. Statist.*, **14**, 1-67.
- Florens, J. P., Richard, J. F. and Rolin, J. M. (1996). Bayesian encompassing specifications tests of a parametric model against a nonparametric alternative. *Technical Report 96.08, Institut de Statistique, Université Catholique de Louvain, Belgium.*

- Ghosal, S., Ghosh, J. K. and Ramamoorthi, R. V. (1999a). Consistency issues in Bayesian nonparametrics. *In: Ghosh, S. (Eds.), Asymptotic Nonparametrics and Time Series: A Tribute to Madan Lal Puri, Marcel Dekker, Inc.*, 639-667.
- Ghosal, S., Ghosh, J. K. and Ramamoorthi, R. V. (1999b). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.*, **27**, 143-158.
- Ghosh, J. K. and Ramamoorthi, R. V. (2002). Notes on Bayesian Nonparametrics. Unpublished.
- MacEachern, S. N. and Müller, P. (1998). Estimating mixture of Dirichlet process models. *J. Comp. Graph. Statist.*, **7**, 223-239.
- Schwartz, L. (1965). On Bayes procedures. *Zeit. Wahrsch. Verw. Gebiete*, **4**, 10-26.
- Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, **4**, 639-650.
- Verdinelli, I. and Wasserman, L. (1998). Bayesian goodness of fit testing using infinite dimensional exponential families. *Ann. Statist.*, **20**, 1215-1241.