A REVIEW OF CONSISTENCY AND CONVERGENCE OF POSTERIOR DISTRIBUTION

By Subhashis Ghosal

Indian Statistical Institute, Calcutta

Abstract

In this article, we review two important issues, namely consistency and convergence of posterior distribution, that arise in Bayesian inference with large samples. Both parametric and non-parametric cases are considered.

In this article we address the issues of consistency and convergence of posterior distribution. This review is aimed for non-specialists. Technical conditions (like measurability) and mathematical expressions are avoided as far as possible. No proof of the results mentioned are given here. Unsatisfied readers are encouraged to look at the references mentioned. The list of the references is certainly not exhaustive, but other references may be found from the mentioned ones. A general reference on the topic of Bayesian asymptotics is Ghosh and Ramamoorthi [36].

In Bayesian analysis, one starts with a prior knowledge (sometimes imprecise) expressed as a distribution on the parameter space and updates the knowledge according to the posterior distribution given the data. It is therefore of utmost importance to know whether the updated knowledge becomes more and more accurate and precise as data are collected indefinitely. This requirement is called the consistency of the posterior distribution. Although it is an asymptotic property, consistency is one of the benchmarks since the violation of consistency is clearly undesirable and one may have serious doubts against inferences based on an inconsistent posterior distribution.

The above formulation of consistency is from the point of view of a classical Bayesian who believes in an *unknown true model*. Subjective Bayesians do not believe in such true models and thinks only in terms of the predictive distribution of a future observation. Nevertheless, consistency is important to subjective Bayesians also. Blackwell and Dubins [7] and Diaconis and Freedman [16] show that consistency is equivalent to *intersubjective agreement*, which means that two Bayesian will ultimately have very close predictive distributions.

To define consistency, we consider a sequence of statistical experiments indexed by a parameter θ taking values in a parameter space Θ. The parameter space need not be a subset of a Euclidean space, so that non-parametric problems are also included. The observation at the nth stage is denoted by $X^{(n)}$. Often it consists of n i.i.d. observations $(X_1, \ldots, X_n)$, but generally it need not be so. The law of $X^{(n)}$ is a probability $P_\theta^{(n)}$ controlled by the parameter θ. By a prior distribution, we mean a probability distribution μ on Θ. The set of all priors will be denoted by $\mathcal{M}$. A prior μ gives rise to a joint distribution of $(\theta, X^{(n)})$. The posterior distribution μ_n is by definition the conditional distribution of θ given $X^{(n)}$. Generally, the posterior is not unique and there are different possible versions. When the family $P_\theta^{(n)}$ is dominated, the version is essentially unique and is given by Bayes theorem. For simplicity, we shall restrict our attention only to this case.

DEFINITION. The posterior distribution μ_n is said to be consistent at θ_0 if for every neighbourhood U of θ_0, $\mu_n(U) \to 1$ almost surely under the law determined by θ_0.

Below, we consider only the i.i.d. situation unless explicitly mentioned otherwise.

When observations take only finitely many possible values (multinomial), it is long known (and can be proved easily) that the posterior is consistent at any point θ_0 which belongs to the support of μ, i.e., if all the neighbourhoods of θ_0 get positive μ-probability. However, the situation changes if there are infinitely many possible outcomes. The following is a classical counter-example due to Freedman [22].

EXAMPLE. Consider the infinite multinomial problem of estimating an unknown probability mass function θ on the set of positive integers. Let θ_0 stand for the geometric distribution with parameter $\frac{1}{4}$. Then one can construct a prior μ which gives positive mass to every neighbourhood of θ_0 but the posterior concentrates in the neighbourhoods of a geometric distribution with parameter $\frac{3}{4}$.

The above problem is the simplest non-parametric problem. Freedman's example cautions us against mechanical uses of Bayesian methods particularly in non-parametric problems. Later we shall see that under very nominal conditions, posterior is consistent for parametric problems.

The following general result on consistency is obtained by Doob [19]. Some extensions of Doob's result may be found in [9].

THEOREM 1. *For any prior μ, the posterior is consistent at every θ except possibly on a set of μ-measure zero.*

Doob's theorem thus implies that a Bayesian will almost always have consistency. Null sets are negligibly small in measure-theoretic sense, and so a troublesome value of the parameter "will not obtain" if one is certain about the prior. Such a view is however very dogmatic. No one in practice can be so certain about the prior and troublesome values of the parameter may really obtain. Quite a different conclusion is reached when smallness in measured in a topological sense. A set A is considered to be topologically small if it can be written as a countable union $\cup_{n=1}^{\infty} F_n$, where each F_n is nowhere dense, i.e., the closure of F_n has no interior point. Such sets are called *meager* or *sets of the first category.* The following result of Freedman [23] shows that the misbehaviour of the posterior described in the example is not just a pathological case, rather it is generic in a topological sense.

THEOREM 2. *Consider the setup of the example and endow the set of prior $\mathcal{M}$ with the topology of weak convergence. Then the set of all pairs $(\theta, \mu) \in \Theta \times \mathcal{M}$ for which the posterior based on μ is consistent at θ is a meager set.*

Freedman's result thus shows that in a topological sense, most priors are troublesome, which led to some criticism of Bayesian methods. The point however is that there are many reasonable priors for which posterior consistency holds at every point of the parameter space. That there are too many bad priors is not such a serious concern because there are also enough good priors approximating any given subjective belief. Tail-free priors (Freedman [22]) and neutral-to-the-right priors (Doksum [18]) are general class of examples of good priors.

However, one should be careful about dogmatic and arbitrary specifications of the prior. It is therefore of importance to have general results providing sufficient conditions for the consistency for a given pair (θ, μ). An important general result in this direction was obtained by Schwartz [57] which is stated below.

THEOREM 3. *Let $\{f_\theta : \theta \in \Theta\}$ be a class of densities and let $X_1, X_2, \ldots$ be i.i.d. with density f_{θ_0}, where $\theta_0 \in \Theta$. Suppose for every neighbourhood U of θ_0, there is a test for $\theta = \theta_0$ against $\theta \notin U$ with power strictly greater than the size. Let μ be a prior on Θ such that for every $\varepsilon > 0$,*

$$\mu\left\{\theta : \int f_{\theta_0} \log \frac{f_{\theta_0}}{f_\theta} < \varepsilon\right\} > 0. \tag{1}$$

Then the posterior is consistent at θ_0.

Schwartz's theorem is an indispensable result in the study of Bayesian methods particularly for non-parametric and semi-parametric problems. Except for some special priors (like tail-free), Schwartz's method is practically the only general method for establishing consistency. It is worth noting that if Θ is itself a class of densities with $f_\theta = \theta$, then the condition on existence of tests in Schwartz's theorem is satisfied if Θ is endowed with the topology of weak convergence. More generally, existence of a uniformly consistent estimator implies the existence of such a test. For a recent application of Schwartz's theorem to the location problem, see [27]. Barron [2] observes the following variation of Schwartz's theorem where condition (1) is replaced by the condition

$$\lim_{n\to\infty} n^{-1} \log \left\{\mu : \int |f_\theta - f_{\theta_0}| < \frac{\varepsilon_n}{n}\right\} = 0, \tag{2}$$

where ε_n is a summable sequence. A similar idea is also used in [28] where it is shown that the use of a hierarchical form of the uniform prior on a discrete approximation of the parameter space leads to consistent posterior in many non-parametric problems including that of the density estimation. In fact, for careful choices of the discrete approximation, the posterior concentrates in the neighbourhoods of the optimal size around the true value of the parameter [29]; here the optimal rate means the best possible rate of convergence of estimators. Barron [2] also discusses about a more general form of Schwartz's theorem when observations may not be i.i.d.

For families smoothly parametrized by a real or vector valued parameter, consistency is obtained if and only if the true value of the parameter lies in the support of the parameter ([22], [57]). This is a nice characterization ensuring consistency for reasonable priors. Consistency however may not hold if any of the assumptions of smoothness or finite dimensionality is dropped. We shall later discuss about non-smooth parametric families. In case of infinite dimension, we have already seen Freedman's counter-example. Another simple yet important non-parametric problem is that of the estimation of an unknown distribution function on the real line. Traditionally, a Dirichlet process prior, introduced by Ferguson [20, 21], is used for this problem. This gives rise to a computable and consistent posterior. In this case, consistency is obtained by efficiently exploiting the special structure of the Dirichlet process. A richer class of priors maintaining the advantages of the Dirichlet but with much more flexibility is obtained by considering mixtures of Dirichlet process [1]. The posterior is consistent with Dirichlet process prior even if the data are censored, see [37].

Although Dirichlet process prior (and its modifications) can be successfully used for estimating the distribution function, many difficulties arise when one considers more complicated problems. For example, Dirichlet process cannot be used for the problem of density estimation because samples from Dirichlet process are always Discrete distributions [6]. One remedy is to consider the prior obtained by convoluting Dirichlet with a kernel as in [54]. Weak consistency for this prior can be proved using Schwartz's theorem [30]. Related computational issues are discussed in [62]. Some authors developed Bayesian methods for density estimation ([53], [51, 52]) based on Gaussian process priors which works well in practice but consistency and other desirable properties are yet to be proved.

A gross misbehaviour of the posterior distribution based on a Dirichlet process prior for estimating the location of an unknown symmetric distribution was pointed out by Diaconis and Freedman [16, 17]. Here the posterior concentrates near the two wrong values of the location parameter, thus seriously violating consistency. Such a prior is seemingly natural and in fact, is suggested by practicing Bayesians. The fact that this leads to inconsistency reminds us the need for reviewing carefully any particular prior before it is actually used, particularly in high dimensional or infinite dimensional problems. The difficulty with Dirichlet prior for this problem can be overcome by the use of suitable Polya tree priors proposed recently [45, 46, 55] and consistency can be obtained through Schwartz's theorem [27].

The author likes to take the opportunity to mention that in the high or infinite dimensional problems where Bayesian methods may suffer from certain difficulties compared to frequentist methods, if enough care is not taken. Bayesian methods have not yet been well developed for many non-parametric problems. Many more research is needed in coming years to develop sensible and practically feasible Bayesian methods which avoids pitfalls like inconsistency and achieve some optimality, at least approximately.

In finite dimension, Schwartz's theorem is not of that importance. One reason for that is that it is applicable only to proper priors, whereas one often uses improper priors in parametric problems. Moreover, for nonregular (unsmooth) families where the support of the density depends on the parameter, condition (1) may not hold. To see an example, consider the family $U(\theta-1,\theta+1)$, where θ varies on the real line. Then for any $\theta \neq \theta_0$, $\int f_{\theta_0} \log(f_{\theta_0}/f_\theta) \equiv \infty$, so no prior can satisfy Schwartz's condition (1). A general theory for finite dimensional inference problems, including both regular and nonregular cases, was developed by Ibragimov and Has'minskii [41]. Their theory is based on the following three basic conditions on the likelihood ratios described below. These conditions hold for almost all parametric problems of interest and the i.i.d. assumption is not important here at all.

To describe these conditions, fix a parameter point θ_0 and consider $Z_n(u)$ as the ratio of the likelihoods at the points $\theta_0 + \varphi_n u$ and θ_0, where φ_n is the appropriate normalizing factor. For example, $\varphi_n = n^{-1/2}$ in the regular (smooth) cases.

Conditions (IH).

(IH1). For some constants $M, m, \alpha > 0$,

$$E|Z_n^{1/2}(u_1) - Z_n^{1/2}(u_2)|^2 \leq M(1 + R^m)|u_1 - u_2|^\alpha$$

for every u_1, u_2 with $|u_1|, |u_2| \leq R$ and $R > 0$.

(IH2). $EZ_n^{1/2}(u) \leq \exp[-g_n(\|u\|)]$, where $g_n : (0, \infty) \to (0, \infty)$ is an increasing function satisfying $\lim_{\substack{n\to\infty \\ y\to\infty}} y^N \exp[-g_n(y)] = 0 \ \forall N \geq 0$.

(IH3). Finite dimensional distributions of the stochastic process $Z_n(\cdot)$ converge to the corresponding finite dimensionals of another process $Z(\cdot)$.

Roughly speaking, (IH1) implies the mean-square continuity of the process $Z_n^{1/2}(\cdot)$ whereas (IH2) controls the tail behaviour. Condition (IH3) is the most important one deciding the limiting distribution. Ibragimov and Has'minskii [41] obtained the limiting distribution of Bayes estimates and the maximum likelihood estimate (MLE) in terms of the process $Z(\cdot)$. Thus in particular, Bayes estimates are consistent in the frequentist sense. In fact, much more is true—Bayes estimates are always (asymptotically) efficient while MLE need not be always so. Thus even as merely frequentist estimators, Bayes estimates are at least as good as any other competitor.

The above general setup is also very convenient for studying asymptotic properties of the posterior distribution and we can extract many useful information about the posterior, see [35, 31]. First, posterior is consistent under (IH1) and (IH2) if the prior density is positive and continuous at θ_0 and grows at most like a polynomial. In fact, with large probability, most of the mass of the posterior is concentrated in a neighbourhood of size φ_n around θ_0. As a result of a theorem in [35], in finite dimension, consistency implies that the posterior based on any two reasonable priors are almost the same. Thus in large samples, Bayesian methods are fairly insensitive to the choice of the prior.

We now discuss another important issue called convergence, and more generally, approximation of the posterior distribution. For this, we entirely restrict our attention to finite dimension. Generally, actual expression of the posterior is quite complicated. For this reason, classical Bayesian analysis was mostly based on conjugate priors, see [10]. A mechanical use of conjugate priors may suffer from many difficulties as mentioned in [3]. Nowadays, with the help of modern computing techniques (like Markov chain Monte-Carlo) and fast computing devices, many of the computations are possible which was out of question a few years ago. Still, such methods require extensive computer time, and more importantly, a lot of expertise. It is therefore of importance to have good analytic approximations which are much simpler to compute. Often, a good approximation is as good as the exact expression itself. For example, in classical statistics, the normal approximation obtained from the central limit theorem plays a dominating role. Below we discuss some approximations which are applicable unless the sample size is small.

Consider the following example taken from [3]. Let us have five random observations (4.0, 5.5, 7.5, 4.5, 3.0) from a population which is modelled as the Cauchy distribution with an unknown location parameter. The improper uniform distribution is used as the prior. The analytic form of the posterior here is quite complicated, but the normal distribution with mean 4.61 and variance 0.376 gives a good approximation to the posterior.

The fact observed in this example is not just accidental. That the posterior distribution is well approximated by a normal distribution with mean at the MLE (or Bayes estimate) and dispersion matrix equal to the inverse of the Fisher information, is a long known fact first discovered by Laplace [44]. It was independently rediscovered by Bernstein [4] and von Mises [60], and often is termed as the Bernstein-von Mises phenomenon. Because of its formal similarity with the CLT, it is also called the Bayesian CLT. However, the statement and the proof were not quite precise at the time of Laplace or Bernstein-von Mises. The first formally correct statement and proof appeared in Le Cam [47, 48]. Since then, many authors worked in this area obtaining various extensions and modifications; see the references [5], [11], [61], [49], [15], [40], [12], [13] and [14]. Bayesian CLT has also been obtained when observations are not independent, see [8], [39] and [58]. A very general but more abstract form of the Bayesian CLT was obtained in [50]. Approximations better than the normal one can be obtained by expanding the posterior ([42, 43], [38], [63]) or by using Laplace's method of approximation of integrals [59]. The result of Johnson [43] may be viewed as a Bayesian analogue of the Edgeworth expansion.

All the works in this area restricted the attention to the regular cases only, except in [56] where an exponential limit to the posterior was obtained for a class of densities with discontinuities. This raises the question whether for general parametric families, the posterior converges (not necessarily to a normal distribution) after centering (not necessarily at the MLE) and scaling. The following complete answer in terms of the limiting process $Z(\cdot)$ was found in [35, 31].

THEOREM 4. *The posterior distribution converges to a limiting form if and only if* $Z(\cdot)$ *satisfies*

$$\frac{Z(u)}{\int Z(w)dw} = g(u+W) \tag{3}$$

for some random variable W *and a probability density* g*. In this case, Bayes estimates (MLE also for regular cases) can be taken as the centering and the limiting density is given by* $g(\cdot)$*.*

There are several implications of Theorem 4. First, it implies that in the regular cases, the posterior converges to a normal limit. Here observations need not be i.i.d., so it is a very general form of the Bayesian CLT. Moreover, in a sense it explains the Bernstein-von Mises phenomenon. The theorem implies that it is the limiting process $Z(\cdot)$ which is responsible for the limiting behaviour of the posterior. The local asymptotic normality (see [41]) for the regular models brings in the normal limit here. Secondly, the theorem implies that in most of the nonregular cases of practical importance (except for the case considered in [56]), posterior cannot converge to a limit. The nonregular cases where we have a limit include the families $U(0,\theta)$, location family of the exponential distribution and truncation models. The cases where we cannot have a limit include the families $U(\theta,\theta+1)$, $U(\theta,2\theta)$, location family of gamma or Weibull and change-point models. Sometimes there is an additional regular-type parameter (often a scale parameter) associated with a nonregular family (e.g., location-scale family of the exponential distribution). Ghosal and Samanta [33] considered this type of models and showed that question of existence of the limit solely depends on the type of nonregularity. Further, the marginal posterior for the regular component is always asymptotically normal.

In the cases where Theorem 4 concludes the existence of a limit of the posterior distribution, usually stronger results can be obtained by direct methods. An asymptotic expansion analogous to [43], with an exponential (instead of normal) leading term, is obtained in [34]. On the other hand, even if there is no limiting form of the posterior, useful approximation may still be found. In the change-point problem, such an approximation is obtained in [32].

In some situations, the dimension of the parameter increases to infinity with the sample size. It is important, from data-analytic point of view, whether normal approximation for the posterior is still valid. This problem is much harder than the case of fixed dimensional parameter. Under certain growth restrictions on the dimension, normal approximation holds [24, 25, 26].

REFERENCES

1. Antoniak, C. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2** 1152–1174.

2. Barron, A. R. (1986). Discussion of "On the consistency of Bayes estimates" by P. Diaconis and D. Freedman. *Ann. Statist.* **14** 26–30.

3. Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis.* Second ed. Springer-Verlag, New York.

4. Bernstein, S. (1917). *Theory of Probability* (in Russian).

5. Bickel, P. and Yahav, J. (1969). Some contributions to the asymptotic theory of Bayes solutions. *Z. Wahrsch. Verw. Gebiete* **11** 257–275.

6. Blackwell, D. (1973). Discreteness of Ferguson selections. *Ann. Statist.* **1** 356–358.

7. Blackwell, D. and Dubins, L. (1962). Merging of opinions with increasing opinions. *Ann. Math. Statist.* **33** 882–886.

8. Borwanker, J. D., Kallianpur, G. and Prakasa Rao, B. L. S. (1971). The Bernstein-von Mises theorem for stochastic processes. *Ann. Math. Statist.* **42** 1241–1253.

9. Breiman, L., Le Cam, L. and Schwartz. L. (1965). Consistent estimates and zero-one sets. *Ann. Math. Statist.* **35** 157–161.

10. Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley.

11. Chao, M. T. (1970). The asymptotic behaviour of Bayes estimators. *Ann. Math. Statist.* **41** 601–609.

12. Chen, C. F. (1985). On asymptotic normality of limiting density functions with Bayesian implications. *J. Roy. Statist. Soc. Ser. B* **47** 540–546.

13. Clarke, B. and Barron, A. R. (1994). Jeffreys' prior is asymptotically least favorable under entropy risk. *J. Statist. Plan. Inf.* **41** 37–60.

14. Clarke, B. and Ghosh, J. K. (1995). Posterior convergence given the mean. *Ann. Statist.* **23** 2116–2144.

15. Dawid, A. P. (1970). On the limiting normality of posterior distribution. *Proc. Canad. Phil. Soc. Ser. B* **67** 625–633.

16. Diaconis, P. and Freedman, D. (1986). On the consistency of Bayes estimates (with discussion). *Ann. Statist.* **14** 1–67.

17. Diaconis, P. and Freedman, D. (1986). On inconsistent Bayes estimates of location. *Ann. Statist.* **14** 68–87.

18. Doksum, K. (1974). Tail-free and neutral random probabilities and their posterior distributions. *Ann. Probab.* **2** 183–201.

19. Doob, J. L. (1948). Application of the theory of martingales. *Coll. Int. du CNRS, Paris,* 22–28.

20. Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.

21. Ferguson, T. S. (1974). Prior distribution on spaces of probability measures. *Ann. Statist.* **2** 615–629.

22. Freedman, D. (1963). On the asymptotic behavior of Bayes estimates in the discrete case I. *Ann. Math. Statist.* **34** 1386–1403.

23. Freedman, D. (1965). On the asymptotic behavior of Bayes estimates in the discrete case II. *Ann. Math. Statist.* **36** 454–456.

24. Ghosal, S. (1996). Asymptotic normality of posterior distributions for exponential families with many parameters. Technical Report, Indian Statistical Institute.

25. Ghosal, S. (1997). Normal approximation to the posterior distribution in generalized linear models with many covariates. *Mathematical Methods of Statistics* **6** 332–348.

26. Ghosal, S. (1998). Asymptotic normality of posterior distributions in high dimensional linear models. *Bernoulli* (To appear).

27. Ghosal, S., Ghosh, J. K. and Ramamoorthi, R. V. (1995). Consistent Bayesian inference about a location parameter with Polya tree priors. Technical Report, Michigan State University.

28. Ghosal, S., Ghosh, J. K. and Ramamoorthi, R. V. (1997). Non-informative priors via sieves and packing numbers. In *Advances in Statistical Decision Theory and Applications.* (S. Panchapakesan and N. Balakrishnan, Eds.), 119–132, Birkhauser, Boston, 1997.

29. Ghosal, S., Ghosh, J. K. and Ramamoorthi, R. V. (1997). Posterior consistency of Dirichlet mixtures in density estimation. Technical Report, Vrije Universiteit, Amsterdam.

30. Ghosal, S., Ghosh, J. K. and Ramamoorthi, R. V. (1998). Manuscript under preparation.

31. Ghosal, S., Ghosh, J. K. and Samanta, T. (1995). On convergence of posterior distributions *Ann. Statist.* **23** 2145–2152.

32. Ghosal, S., Ghosh, J. K. and Samanta, T. (1996). Approximation of the posterior distribution in a change-point problem. *Ann. Inst. Statist. Math.* (To appear).

33. Ghosal, S. and Samanta, T. (1995). Asymptotic behaviour of Bayes estimates and posterior distributions in multiparameter nonregular cases. *Math. Methods Statist.* **4** 361–388.

34. Ghosal, S. and Samanta, T. (1995). Asymptotic expansions of posterior distributions in nonregular cases. *Ann. Inst. Statist. Math.* **49** 181–197.

35. Ghosh, J. K., Ghosal, S. and Samanta, T. (1994). Stability and convergence of posterior in non-regular problems. In *Statistical Decision Theory and Related Topics V* (S. S. Gupta and J. O. Berger, Eds.), Springer-Verlag, 183–199.

36. Ghosh, J. K. and Ramamoorthi R. V. (1994). *Lecture notes on Bayesian asymptotics.* Under preparation.

37. Ghosh, J. K. and Ramamoorthi R. V. (1994). Consistency of Bayesian inference for survival analysis with or without censoring. Preprint.

38. Ghosh, J. K., Sinha, B. K. and Joshi, S. N. (1982). Expansion for posterior probability and integrated Bayes risk. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta and J. O. Berger, Eds.), Academic Press, 403–456.

39. Heyde, C. C. and Johnstone, I. M. (1979). On asymptotic posterior normality for stochastic processes. *J. Roy. Statist. Soc., Ser B*, **41** 184–189.

40. Ibragimov, I. A. and Has'minskii, R. Z. (1973). Asymptotic behaviour of some statistical estimators II. Limit theorems for the a posteriori density and Bayes' estimators. *Theory Probab. Appl.* **XVIII** 76–91.

41. Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer-Verlag, New York.

42. Johnson, R. A. (1967). On asymptotic expansion for posterior distribution. *Ann. Math. Statist.* **38** 1899–1906.

43. Johnson, R. A. (1970). Asymptotic expansions associated with posterior distribution. *Ann. Math. Statist.* **42** 1241–1253.

44. Laplace, P. S. (1774). Mémoire sur la probabilité de causes par les évenements. *Mem. Acad. R. Sci. Presentés par Divers Savanas.* **6**, 621–656. (Translated in *Stat. Sci.* **1** 359–378.)

45. Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modelling. *Ann. Statist.* **20** 1222–1235.

46. Lavine, M. (1994). More aspects of Polya tree distributions for statistical modelling. *Ann. Statist.* **22** 1161–1176.

47. Le Cam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *University of California Publications in Statistics* **1** 277–330.

48. Le Cam, L. (1958). Les properties asymptotiques des solutions de Bayes. *Publ. Inst. Statist. Univ. Paris* **7** 17–35.

49. Le Cam, L. (1970). Remarks on the Bernstein-von Mises theorem. Unpublished manuscript.

50. Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New York.

51. Lenk, P. J. (1988). The logistic normal distribution for Bayesian nonparametric, predictive densities. *J. Amer. Statist. Assoc.* **83** 509–516.

52. Lenk, P. J. (1991). Towards a predictable Bayesian nonparametric density estimator. *Biometrika* **78** 531–543.

53. Leonard, T. (1973). A Bayesian method for histograms. *Biometrika* **60** 297–308.

54. Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *Ann. Statist.* **12** 351–357.

55. Mauldin, R. D., Sudderth, W. D. and Williams, S. C. (1992). Polya trees and random distributions. *Ann. Statist.* **20** 1203–1221.

56. Samanta, T. (1988). *Some Contributions to the Asymptotic Theory of Estimation in Non-regular Cases.* Unpublished Ph. D. Thesis, Indian Statistical Institute.

57. Schwartz, L. (1965). On Bayes procedures. *Z. Wahrsch. Verw. Gebiete* **4**, 10–26.

58. Sweeting, T. J. and Adekola, A. O. (1987). Asymptotic posterior normality for stochastic processes. *J. Roy. Statist. Soc. Ser. B*, **49** 215–222.

59. Tierney, L. and Kadane, J. B. Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.* **81** 82–86.

60. von Mises, R. (1931). *Wahrscheinlichkeitsrechnung.* Springer, Berlin.

61. Walker, A. M. (1969). On the asymptotic behaviour of posterior distributions. *J. Roy. Statist. Soc. Ser. B* **31** 80–88.

62. West, M. (1992). Modelling with mixtures. In *Bayesian Statistics* **4** (J. M. Bernardo *et al.*, Eds.) Oxford University Press, 503–524.

63. Woodroofe, M. (1992). Integrable expansions for posterior distributions for one-parameter exponential families. *Statistica Sinica* **2** 91–111.

Division of Theoretical Statistics
and Mathematics
Indian Statistical Institute
203 Barrackpore Trunk Road
Calcutta 700 035
India