The Annals of Statistics
1995, Vol. 23, No. 6, 2145–2152

ON CONVERGENCE OF POSTERIOR DISTRIBUTIONS

BY SUBHASHIS GHOSAL,[1] JAYANTA K. GHOSH[2] AND TAPAS SAMANTA

Indian Statistical Institute

A general (asymptotic) theory of estimation was developed by Ibragimov and Has'minskii under certain conditions on the normalized likelihood ratios. In an earlier work, the present authors studied the limiting behaviour of the posterior distributions under the general setup of Ibragimov and Has'minskii. In particular, they obtained a necessary condition for the convergence of a suitably centered (and normalized) posterior to a constant limit in terms of the limiting likelihood ratio process. In this paper, it is shown that this condition is also sufficient to imply the posterior convergence. Some related results are also presented.

1. Introduction. We consider the general setup of Ibragimov and Has'minskii (1981) (henceforth abbreviated as IH) that includes the regular cases and also a wide variety of nonregular cases. A general (asymptotic) theory of estimation was developed in IH, where the problem was reduced to the study of the properties of a suitably normalized likelihood ratio. IH obtained the asymptotic properties of estimates under certain conditions on the normalized likelihood ratios (see Section 2). In the recent paper of Ghosh, Ghosal and Samanta (1994) (GGS hereafter), it was shown that under the general setup of IH, one can obtain useful information on the asymptotic behaviour of posterior distributions as well. In particular, GGS obtained a necessary condition for the convergence of a suitably centered (and normalized) posterior distribution to a constant limit in terms of the limiting likelihood ratio process. The main purpose of this paper is to show that this condition is also sufficient to imply the posterior convergence. This result is potentially applicable to many situations involving stochastic processes and some nonregular cases. In particular, it implies an in-probability version of the Bernstein–von Mises theorem in an extremely general form. Apart from this, we show that the first two conditions of IH always imply posterior consistency. A very general result on the asymptotic equivalence of the Bayes estimates and the maximum likelihood estimate (MLE) in the regular cases is also proved under the present setup. While this result is quite well known in many particular cases or under conditions much stronger than ours, it seems to be new in this general form.

Received June 1994; January 1995.

[1]Research supported by the National Board of Higher Mathematics, Department of Atomic Energy, Bombay, India.
[2]Research partly supported by NSF Grant DMS-93-07727.
AMS 1991 *subject classifications.* Primary 62F15, 62F25.
Key words and phrases. Asymptotics, Bernstein–von Mises theorem, Bayes estimates, convergence of posterior, likelihood ratio process.

2. Convergence of posterior distributions. Let $\{(\mathscr{X}^n, \mathscr{A}^n), P_\theta^n;\ \theta \in \Theta\}$ be a sequence of statistical experiments generated by observations $X^n \in \mathscr{X}^n$, where $\Theta \subset \mathbb{R}^d$, $d \ge 1$, is a Borel set with nonempty interior and P_θ^n admits a density $p^n(x^n; \theta)$. For a fixed θ_0 in the interior of Θ, the *likelihood ratio process* (LRP) is defined by

$$Z_n(u) = Z_{n,\theta_0}(u) = \frac{p^n(x^n; \theta_0 + \varphi_n u)}{p^n(x^n; \theta_0)}, \qquad u \in U_n,$$

where $U_n := \varphi_n^{-1}(\Theta - \theta_0)$ and φ_n is an appropriate normalizing factor.

The general theory of IH has been developed under the conditions stated below.

CONDITIONS (IH).

(IH1) For some $M > 0$, $m_1 \ge 0$ and $\alpha > 0$,

$$E_{\theta_0}|Z_n^{1/2}(u_1) - Z_n^{1/2}(u_2)|^2 \le M(1 + R^{m_1})\|u_1 - u_2\|^\alpha,$$

for all $u_1, u_2 \in U_n$ satisfying $\|u_1\| \le R$ and $\|u_2\| \le R$.

(IH2) For all $u \in U_n$,

$$E_{\theta_0} Z_n^{1/2}(u) \le \exp[-g_n(\|u\|)],$$

where $\{g_n\}$ is a sequence of real-valued functions on $[0, \infty)$ satisfying the following: (a) for a fixed $n \ge 1$, $g_n(y) \uparrow \infty$ as $y \uparrow \infty$; (b) for any $N > 0$,

$$\lim_{\substack{y \to \infty \\ n \to \infty}} y^N \exp[-g_n(y)] = 0.$$

(IH3) The finite-dimensional distributions of $\{Z_n(u)\colon u \in \mathbb{R}^d\}$ converge to those of a stochastic process $\{Z(u)\colon u \in \mathbb{R}^d\}$.

Below, all the probability statements will refer to the "true parameter" θ_0.

Let Π be the class of (possibly improper) prior densities on Θ which are continuous and positive at θ_0 and have polynomial majorants. For example, Jeffreys' prior in the regular case is an element of Π. Let $\mathscr{L}$ be the class of continuous loss functions $l\colon \mathbb{R}^d \to [0, \infty)$ satisfying the conditions of Theorem I.10.2 of IH. This class of loss functions is sufficiently general to include all loss of the form $l(x) = \|x\|^p$, $p \ge 1$. Below, we shall consider only priors $\pi \in \Pi$ and loss functions $l \in \mathscr{L}$. Set

$$\xi_n(u) = \frac{Z_n(u)\pi(\theta_0 + \varphi_n u)}{\int_{U_n} Z_n(v)\pi(\theta_0 + \varphi_n v)\, dv},$$

which is the posterior density of the normalized parameter $u = \varphi_n^{-1}(\theta - \theta_0)$ with respect to a prior $\pi \in \Pi$, and also set $\xi(u) = Z(u)/(\int_{\mathbb{R}} Z(v)\, dv)$. Let $\pi_n(\theta \in A | X^n)$ denote the posterior probability of a set $A \subset \Theta$ given X^n.

DEFINITION 1. We say that the posterior is *strongly consistent* if, for any neighbourhood V of θ_0, $\lim_{n\to\infty} \pi_n(\theta \notin V|X^n) = 0$ a.s.

The posterior is called *weakly consistent* if $\pi_n(\theta \notin V|X^n) \to_p 0$.

In GGS, it was shown that the posterior is asymptotically free of prior if one has posterior consistency (GGS, Theorem 2.1). We here observe that posterior consistency always holds under conditions (IH1) and (IH2).

PROPOSITION 1. *Assume Conditions* (IH1) *and* (IH2) *and consider a prior* $\pi \in \Pi$. *Then the posterior is weakly consistent. If further* $\sum_{n=1}^{\infty}\|\varphi_n\|^s < \infty$ *for some* $s > 0$, *then the posterior is strongly consistent* (*provided the almost sure convergence is meaningful*).

PROOF. Let V be a neighbourhood of θ_0 and let $r > 0$ be such that the open ball of radius r around θ_0 is contained in V. Then by Lemma I.5.2 of IH, for any $N > 0$, there is a constant C_N such that

$$E[\pi_n(\theta \notin V|X^n)] \le E\left[\int_{\|u\|>r/\|\varphi_n\|} \xi_n(u)\,du\right] \le C_N r^{-N}\|\varphi_n\|^N.$$

The result is now immediate. □

DEFINITION 2. An $\mathbb{R}^d$-valued statistic $\hat{\theta}_n$ is called a *proper centering* if, for all sets A in the Borel sigma-field $\mathscr{B}^d$ on $\mathbb{R}^d$, there exist numbers $Q(A)$ such that

$$\sup\left\{\left|\pi_n\left(\varphi_n^{-1}\left(\theta - \hat{\theta}_n\right) \in A|X^n\right) - Q(A)\right| : A \in \mathscr{B}^d\right\} \to_p 0. \tag{1}$$

A statistic $\hat{\theta}_n$ is called a *semiproper centering* if, for each $A \in \mathscr{B}^d$,

$$\pi_n\left(\varphi_n^{-1}\left(\theta - \hat{\theta}_n\right) \in A|X^n\right) \to_p Q(A). \tag{2}$$

A statistic $\hat{\theta}_n$ is called *compatible* (with the posterior) if $(\varphi_n^{-1}(\hat{\theta}_n - \theta_0), \xi_n(\cdot))$, as a random element in $\mathbb{R}^d \times L^1(\mathbb{R}^d)$, converges in law.

REMARK 1. In view of Theorem 2.1 of GGS, if any of the above statements in Definition 2 holds for some prior $\pi \in \Pi$, then it holds for any other prior in Π.

The following result characterizes the existence of a posterior limit.

THEOREM 1. *Assume Conditions* (IH). *If either a proper centering or a compatible semiproper centering* $\hat{\theta}_n$ *exists, then there exists a random variable* W *such that* (a) $\varphi_n^{-1}(\hat{\theta}_n - \theta_0) \to_d W$ *and* (b) *for almost all* $u \in \mathbb{R}^d$, $\xi(u - W)$ *is nonrandom.*

Conversely, if (b) *holds for some random variable* W, *then any Bayes estimate* (*for a loss* $l \in \mathscr{L}$ *and a prior* $\pi \in \Pi$) *works as a compatible proper centering.*

The necessity part was established in GGS (Theorems 2.4 and 2.5). We here prove the sufficiency part. Let $\tilde{\theta}_n$ be a Bayes estimate with respect to a loss $l \in \mathscr{L}$ and a prior $\pi \in \Pi$. By Remark 1, it is enough to consider the posterior with respect to the same prior π. Also, set $\psi_n(s) = \int_{\mathbb{R}^d} l(s - u)\xi_n(u)\,du$ and $\psi(s) = \int_{\mathbb{R}^d} l(s-u)\xi(u)\,du$. By the assumptions made on the loss function l, the random function $\psi(s)$ attains its absolute minimum at a unique point τ.

We first establish the following result, which is also of independent interest.

PROPOSITION 2. *Under Conditions* (IH), *any Bayes estimate is compatible.*

PROOF. We shall show that

$$\left(\varphi_n^{-1}\left(\tilde{\theta}_n - \theta_0\right), \xi_n(\cdot)\right) \to_d (\tau, \xi(\cdot)). \tag{3}$$

By the arguments used in the proof of Theorem I.10.2 of IH, it suffices to show that, for all $M > 0$,

$$(\psi_n(\cdot|M), \xi_n(\cdot)) \to_d (\psi(\cdot|M), \xi(\cdot)) \tag{4}$$

as random elements in $C[-M, M]^d \times L^1(\mathbb{R}^d)$; here $\psi_n(\cdot|M)$ and $\psi(\cdot|M)$ stand for the restrictions of $\psi_n(\cdot)$ and $\psi(\cdot)$, respectively, on $[-M, M]^d$, and $C[-M, M]^d$ denotes the space of continuous functions on $[-M, M]^d$ with the uniform metric. From Theorem I.10.2 of IH and Theorem 2.3 of GGS, respectively, we know that $\{\psi_n(\cdot|M)\}$ and $\{\xi_n(\cdot)\}$ are tight; hence it suffices to verify the convergence of finite-dimensional distributions. Let $s_1, \dots, s_m \in \mathbb{R}^d$ and $A_1, \dots, A_k \in \mathscr{B}^d$. We have to show that

$$\begin{aligned}(\psi_n(s_1), \dots, \psi_n(s_m), \xi_n(A_1), \dots, \xi_n(A_k))\\ \to_d (\psi(s_1), \dots, \psi(s_m), \xi(A_1), \dots, \xi(A_k));\end{aligned} \tag{5}$$

here $\xi_n(A)$ and $\xi(A)$ stand for $\int_A \xi_n(u)\,du$ and $\int_A \xi(u)\,du$, respectively. By the arguments used in the proof of Theorem I.10.2 of IH, (5) follows from Theorem A.1 of the Appendix. □

PROOF OF THEOREM 1. By the given condition, $\xi(u) = g(u + W)$, where g is a fixed probability density. Let c be the unique minimizer of the function $\int_{\mathbb{R}^d} l(s-u)g(u)\,du$. Then $\tau = W + c$ and hence, without loss of generality, we can assume that $W = \tau$. The posterior density of $v := \varphi_n^{-1}(\theta - \tilde{\theta}_n)$ is given by $\pi_n^*(v|X^n) = \xi_n(v + \tau_n)$, where $\tau_n = \varphi_n^{-1}(\tilde{\theta}_n - \theta_0)$. By Proposition 2 and Lemma A.1 of the Appendix, in the space $L^1(\mathbb{R}^d)$, we have

$$\left(\pi_n^*(v|X^n) \colon v \in \mathbb{R}^d\right) \to_d \left(g(v) \colon v \in \mathbb{R}^d\right). \tag{6}$$

The result is now immediate since $g(\cdot)$ is a nonrandom element. □

Theorem 1 has a wide range of applicability since no particular structure (like i.i.d. or regularity) is assumed. The theorem is valid irrespective of the form of the limit; it may be nonnormal as in Examples 2 and 4.

EXAMPLE 1. If the families of distributions satisfy the LAN condition, then the posterior converges and the limit is a normal probability. Examples of such regular cases include the independent homogeneous case, independent nonhomogeneous case, nonlinear regression model, Gaussian white noise, the case with almost smooth densities (see IH) and planar Gibbsian point processes [Mase (1992)].

EXAMPLE 2. Consider i.i.d. observations with a common density $f(\cdot;\theta)$ possessing r jumps at points $a_1(\theta),\dots,a_r(\theta)$ and satisfying the conditions of IH (Chapter V, page 242). Let $p_i(\theta)$ and $q_i(\theta)$ denote, respectively, the right-hand and left-hand limits of the density at $x = a_i(\theta)$, $i = 1,\dots,r$. Assume that either

$$(q_i(\theta) = 0 \text{ and } a_i'(\theta) > 0) \quad \text{or} \quad (p_i(\theta) = 0 \text{ and } a_i'(\theta) < 0) \quad \forall i = 1,\dots,r \tag{7}$$

or

$$(q_i(\theta) = 0 \text{ and } a_i'(\theta) < 0) \quad \text{or} \quad (p_i(\theta) = 0 \text{ and } a_i'(\theta) > 0) \quad \forall i = 1,\dots,r. \tag{8}$$

Then the family is locally asymptotically exponential in the sense of IH (Chapter V, page 276) and hence condition (b) of Theorem 1 holds. If case (7) holds, the limiting posterior distribution has density $c(\theta)\exp[c(\theta)(x + b(\theta))]$ supported on $x < -b(\theta)$, where $c(\theta) = \sum_{i=1}^{r}(p_i(\theta) - q_i(\theta))a_i'(\theta) > 0$ and $b(\theta)$ is the unique minimizer of $h(s) := \int_{-\infty}^{0} l(s-u)c(\theta)\exp[c(\theta)u]\,du$. If case (8) holds, the limit can similarly be identified. Important examples of this kind include location shift of an exponential density, $U(0,\theta)$ and so forth.

EXAMPLE 3. Consider a multiparameter family of densities which is nonregular with respect to a real parameter θ and "smooth" with respect to the other parameters, say, φ (e.g., φ may be a scale parameter). Ghosal and Samanta (1994) verified Conditions (IH) for such a family and showed that the limiting LRP factorizes into two independent processes. If φ is the parameter of interest, proceeding as in Theorem 1 and using the results of Ghosal and Samanta (1994), one can see that the marginal posterior of φ is asymptotically normal.

EXAMPLE 4. For the multiparameter family considered in Example 3, the joint posterior of θ and φ converges if the nonregularity is of the type mentioned in Example 2. The limit in this case is the product of the limit obtained in Example 2 and a normal probability.

It is interesting to note that in all the above examples, there exist finite-dimensional asymptotically sufficient statistics of dimension the same as that of the parameter.

We now examine whether the Bayes estimate in Theorem 1 can be replaced by the MLE. Let condition (IH1) be replaced by the following:

(IH1)′ There exist numbers $m \geq \alpha > d$ and $M_1, m_1 \geq 0$ such that

$$E|Z_n^{1/m}(u_1) - Z_n^{1/m}(u_2)|^m \leq M_1(1 + R^{m_1})\|u_1 - u_2\|^{\alpha},$$

for all $u_1, u_2 \in U_n$ satisfying $\|u_1\|, \|u_2\| \leq R$.

THEOREM 2. *Assume conditions* (IH1)′, (IH2) *and* (IH3). *Let* $Z_n(\cdot)$ *and* $Z(\cdot)$ *have continuous sample paths and* $Z(\cdot)$ *attain its maximum at a unique point* $\hat{u}$. *Then, as* $n \to \infty$,

$$(9) \qquad \left(\varphi_n^{-1}\left(\hat{\theta}_n - \theta_0\right), \varphi_n^{-1}\left(\tilde{\theta}_n - \theta_0\right)\right) \to_d (\hat{u}, \tau),$$

where $\hat{\theta}_n$ *stands for the MLE,* $\tilde{\theta}_n$ *for the Bayes estimate and* τ *is as defined earlier.*

PROOF. For any function f on $\mathbb{R}^d$, let $f(\cdot; M)$ denote the restriction of $f(\cdot)$ on $[-M, M]^d$. By arguments used in the proofs of Theorems I.10.1 and I.10.2 of IH, it is enough to show that, as random elements in $C[-M, M]^d \times C[-M, M]^d$,

$$(10) \qquad (Z_n(\cdot; M), \psi_n(\cdot; M)) \to_d (Z(\cdot; M), \psi(\cdot; M)).$$

Since tightness has already been verified, it remains to prove the convergence of finite-dimensional distributions. However, this follows from an obvious modification of the arguments given in IH (page 108) and Theorem A.1 of the Appendix. □

Theorem 2 has a useful consequence in the LAN, LAMN and LAQ situations [see Jeganathan (1982), Le Cam (1986) and Le Cam and Yang (1990) for definitions].

COROLLARY 1. *Assume the conditions of Theorem* 2 *and suppose further that the limiting LRP is of the form*

$$(11) \qquad Z(u) = \exp[u'\Delta - (1/2)u'\Sigma u],$$

where Δ *is a random vector and* Σ *is an almost surely positive definite random matrix. Then*

$$(12) \qquad \varphi_n^{-1}\left(\tilde{\theta}_n - \hat{\theta}_n\right) \to_p 0,$$

that is, the MLE and the Bayes estimates are asymptotically equivalent.

PROOF. By Theorem 2, $\varphi_n^{-1}(\tilde{\theta}_n - \hat{\theta}_n) \to_d (\tau - \hat{u})$. If (11) is satisfied, Anderson's lemma implies that $\tau = \hat{u} = \Sigma^{-1}\Delta$, and hence (12) holds. □

We now consider the special situation where both the hypotheses of Theorem 2 and the LAN condition are satisfied. In this case, Σ is a nonrandom matrix and Δ is distributed as $N_d(0, \Sigma)$. Then condition (b) of Theorem 1 holds, and hence we can have a limit of the posterior with the Bayes estimate as a proper centering. By Corollary 1, one can now replace the Bayes estimate by the MLE. Moreover, the limit of the posterior is $N_d(0, \Sigma^{-1})$. The same conclusion can be reached by a more direct route. By arguments similar to those in Theorem 2, one can show that *the MLE is compatible* and then one can obtain the result following the proof of Theorem 1. Thus we obtain an *in-probability* version of the well-known Bernstein–von Mises theorem in a much more general setting.

APPENDIX

LEMMA A.1. *For $f \in L^1(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$, define f_x by $f_x(y) = f(y - x)$. Then the mapping $(x, f) \mapsto f_x$, from $\mathbb{R}^d \times L^1(\mathbb{R}^d)$ into $L^1(\mathbb{R}^d)$, is continuous in x and is an isometry in f, and so is jointly continuous.*

For a proof, see Rudin [(1974), Theorem 9.5, page 183].

THEOREM A.1. *Let $\xi_n(t)$, $n \geq 1$, and $\xi(t)$ be measurable random functions defined on a compact set $F \subset \mathbb{R}^d$, and let $w(t)$ be a measurable function on F. Assume that the following conditions are satisfied*:

(a) $\sup_{n \geq 1} E(\int_F |w(t)|\,|\xi_n(t)|\,dt) < \infty$;
(b) *there exist $H, \alpha > 0$ such that* $\sup_{n \geq 1} E|\xi_n(t) - \xi_n(s)| \leq H\|t - s\|^{\alpha}$;
(c) *finite-dimensional distributions of $\xi_n(t)$ converge to those of $\xi(t)$.*

Then, for any $t_1, \dots, t_k \in \mathbb{R}^d$,

$$\left(\xi_n(t_1), \dots, \xi_n(t_k), \int_F w(t)\,\xi_n(t)\,dt\right) \to_d \left(\xi(t_1), \dots, \xi(t_k), \int_F w(t)\,\xi(t)\,dt\right).$$

The proof is a minor modification of that of Theorem I.A.22 of IH.

REFERENCES

GHOSAL, S. and SAMANTA, T. (1994). Asymptotic behaviour of Bayes estimates and posterior distributions in multiparameter nonregular cases. *Mathematical Methods of Statistics* To appear.

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 5* (S. S. Gupta and J. O. Berger, eds.) 183–199. Springer, New York.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer, New York.

JEGANATHAN, P. (1982). On the asymptotic theory of estimation when the limit of the log likelihood ratios is mixed normal. *Sankhyā Ser. A* **44** 173–212.

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.

LE CAM, L. and YANG, G. L. (1990). *Asymptotics in Statistics: Some Basic Concepts*. Springer, New York.

MASE, S. (1992). Uniform LAN condition of planar Gibbsian point processes and optimality of maximum likelihood estimates of soft-core potential functions. *Probab. Theory Related Fields* **92** 51–67.

RUDIN, W. (1974). *Real and Complex Analysis*, 2nd ed. McGraw-Hill, New York.

SUBHASHIS GHOSAL
JAYANTA K. GHOSH
DIVISION OF THEORETICAL STATISTICS
AND MATHEMATICS
INDIAN STATISTICAL INSTITUTE
203, B. T. ROAD
CALCUTTA 700 035
INDIA

TAPAS SAMANTA
COMPUTER SCIENCE UNIT
INDIAN STATISTICAL INSTITUTE
203, B. T. ROAD
CALCUTTA 700 035
INDIA