ASYMPTOTIC BEHAVIOR OF POSTERIOR DISTRIBUTIONS FOR RANDOM PROCESSES UNDER INCORRECT MODELS

Keigo Yamada
Japan UNIVAC Research Institute
Tokyo DIA Building, 1-28-38
Shinkawa, Chuo-ku, Tokyo 104 Japan

Abstract

In this paper, the asymptotic behavior of posterior distributions on parameters contained in random processes is examined when the specified model for the densities is not necessarily correct. Uniform convergence of likelihood functions in some way is shown to be a sufficient condition for the posterior distributions to be asymptotically confined to a set (Theorem 1). For ergodic stationary Markov processes uniform convergence of likelihood functions is established by the ergodic theorem for Banach-valued stationary processes (Proposition 1). A sufficient condition for the uniform convergence is also shown for general random processes (Proposition 2). These results are used to analyze the asymptotic behavior of posterior distributions on parameters contained in linear systems under incorrect models (Example 1 and 2).

1. INTRODUCTION. Let $\{X_n\}$, $n=1, 2, \cdots$ be a family of random variables defined on a probability space $(\Omega, \mathcal{B}, P)$. A model is given which specifies that the joint density of random variables $X_1, \cdots, X_n$ is one of the densities $f_n(x_1, x_2, \cdots, x_n \mid \theta)$, where the indexing parameter θ takes its values in the parameter space Θ, assumed to be a compact metric space. π denotes a prior distribution on $(\Theta, \mathcal{B}(\Theta))$, where $\mathcal{B}(\Theta)$ is the Borel σ-field of Θ, and π_n denotes the corresponding posterior distribution of the parameter given $X_1, X_2, \cdots, X_n$. Thus, for any $A \in \mathcal{B}(\Theta)$,

$$(1)\quad \pi_n A = \int_A f_n(X_1, \cdots, X_n \mid \theta)\, d\pi(\theta) \Big/ \int_\Theta f_n(X_1, \cdots, X_n \mid \theta)\, d\pi(\theta).$$

In this paper we study the asymptotic behavior of the sequence $\{\pi_n\}$ under the situation that the joint density of $\{X_n\}$ need not correspond to any of the densities in the specified model. Such an analysis was done by Berk [1] when $\{X_n\}$ are identically and independently distributed $(i.i.d.)$.

It is, however, desirable to do the same kind of analysis for more general cases since most of the stochastic processes we encounter in practical problems are not $i.i.d.$.

As was shown in Berk [1], when the process $\{X_n\}$ is $i.i.d.$, uniform convergence of the likelihood functions $f_n(X_1, \cdots, X_n \mid \theta)$ in some way ensures that the posterior distribution for the parameter θ is asymptotically confined to a set (which is called the asymptotic carrier by Berk). In Theorem 1, it is shown that the same thing is true when $\{X_n\}$ are not necessarily $i.i.d.$

In general, it is impossible to determine the asymptotic carrier since the true density for the observed process is not known. We can, however, analyze to some extent the asymptotic behavior of posterior distributions under a misspecified (incorrect) model by investigating the property of the asymptotic carrier.

In section 3, the uniform convergence of likelihood functions is established for ergodic stationary Markov processes using the ergodic theorem in Banach space, and an example of the analysis is given. A sufficient condition for the uniform convergence is given for general processes in Section 4. These results are then applied to the analysis of the asymptotic behavior of posterior distributions on parameters involved in multi-input, multi-output linear systems when the model is incorrect.

2. CONVERGENCE OF POSTERIOR DISTRIBUTIONS. We assume the following:

(A1) For any n and $\theta \in \Theta$, $f_n(x_1, \cdots, x_n \mid \theta)$ is jointly Borel-measurable.

(A2) $f_n(X_1, \cdots, X_n \mid \theta) > 0$ with probability one.

(A3) For any nonempty open set $A \in \mathcal{B}(\Theta)$, $\pi(A) > 0$.

As was indicated in Introduction, the following theorem states that, if the likelihood functions $f_n(X_1, \cdots, X_n \mid \theta)$ converges uniformly in θ in some way, then the posterior distribution $\{\pi_n\}$ defined in (1) is asymptotically confiend to a set.

Theorem 1. Assume (A1 - 3). Suppose that, for a continuous function $\eta(\theta)$ defined on Θ,

(2) $$\sup_{\theta \in \Theta} \left| \frac{1}{n} \log f_n(X_1, \cdots, X_n \mid \theta) - \eta(\theta) \right| \to 0 \qquad \text{in probability,}$$

then, for any open set $A \in \mathcal{B}(\Theta)$ which contains the asymptotic carrier A_0,

(3) $$\pi_n A \to 1 \qquad \text{in probability}$$

where A_0 is defined as $A_0 = \{\theta \,;\, \eta^* = \eta(\theta),\ \eta^* = \sup_{\theta \in \Theta} \eta(\theta)\}$

Remark 1. Since $\eta(\cdot)$ is continuous on the compact set Θ, η^* is finite and A_0 is not empty.

Proof. The proof follows the method given by Berk [1]. It is sufficient to prove

$$L_n A = \frac{\pi_n A^c}{\pi_n A} \to 0 \qquad \text{in probability}$$

where A^c is complement of the set A.

We should note that $\pi_n A > 0$ for all n with probability one because of (A2 - 3).

Now

$$L_n A = \int_{A^c} f_n(X_1, \cdots, X_n \mid \theta)\, d\pi(\theta) \Big/ \int_A f_n(X_1, \cdots, X_n \mid \theta)\, d\pi(\theta)$$

$$= \int_{A^c} (\exp \frac{1}{n} \log f_n(X_1, \cdots, X_n \mid \theta))^n d\pi(\theta) \Big/ \int_A (\exp \frac{1}{n} \log f_n(X_1, \cdots, X_n \mid \theta))^n d\pi(\theta).$$

We shall show that, for any $A \in \mathscr{B}(\Theta)$,

$$(4) \qquad (\int_A (\exp \frac{1}{n} \log f_n(X_1, \cdots, X_n \mid \theta))^n d\pi(\theta))^{\frac{1}{n}} \to \sup_{\theta \in A} \exp \eta(\theta)$$

in probability.

By the condition (2) in the theorem,

$$(5) \qquad \sup_{\theta \in A} \left| \frac{1}{n} \log f_n(X_1, \cdots, X_n \mid \theta) - \eta(\theta) \right| \to 0$$

in probability.

Hence we have

$$(6) \qquad \sup_{\theta \in A} \left| \exp \frac{1}{n} \log f_n(X_1, \cdots, X_n \mid \theta) - \exp(\theta) \right| \to 0$$

in probability.

In fact, defining $F_n(\omega;\theta)$ by

$$F_n(\omega;\theta) = \frac{1}{n} \log f_n(X_1, \cdots, X_n \mid \theta),$$

$$(7) \qquad \begin{aligned} &\sup_{\theta \in A} | \exp F_n(\omega;\theta) - \exp \eta(\theta) | \\ &\leq \sup_{\theta \in A} |F_n(\omega;\theta) - \eta(\theta)| \exp\{|\eta(\theta)| + |F_n(\omega;\theta) - \eta(\theta)|\} \\ &\leq \sup_{\theta \in A} |F_n(\omega;\theta) - \eta(\theta)| \cdot \exp\{\sup_{\theta \in A} |\eta(\theta)| + \sup_{\theta \in A} |F_n(\omega;\theta) - \eta(\theta)|\} \end{aligned}$$

Since $\exp(\cdot)$ is continuous, using (5) and Theorem 6 in 3, II in Gihman - Skorohod [2],

$$\exp\{\sup_{\theta \in A} |F_n(\omega;\theta) - \eta(\theta)|\} \to 1$$

in probability.

Hence again, by noting $\exp\{\sup_{\theta \in A} |\eta(\theta)|\} < \infty$ the last term in (7) converges to 0 in probability and (6) follows.

Now by Minkowski's inequality

$$\left| (\int_A (\exp F_n(\omega;\theta))^n d\pi(\theta))^{\frac{1}{n}} - (\int_A (\exp \eta(\theta))^n d\pi(\theta))^{\frac{1}{n}} \right|$$

$$\leq (\int_A | \exp F_n(\omega;\theta) - \exp \eta(\theta) |^n d\pi(\theta))^{\frac{1}{n}}$$

$$\leq \sup_{\theta \in \Theta} | \exp F_n(\omega;\theta) - \exp\eta(\theta) | \to 0$$

in probability.

On the other hand

$$\left(\int_A (\exp\eta(\theta))^n d\pi(\theta) \right)^{\frac{1}{n}} \to \sup_{\theta \in A} \exp\eta(\theta)$$

(see Yoshida ([8], Theorem 1, 3, I, P. 34)). Combining these results, (4) follows.

Now

$$\lim_{n\to\infty} L_n A = \lim_{n\to\infty} \left[\frac{\sup_{\theta \in A^c} \exp\eta(\theta)}{\sup_{\theta \in A} \exp\eta(\theta)} \right]^n$$

in probability.

By noting that A^c is compact, $\eta(\cdot)$ is continuous and that $A \supset A_0$,

$$0 \leq \sup_{\theta \in A^c} \exp\eta(\theta) / \sup_{\theta \in A} \exp\eta(\theta) < 1$$

and hence

$$L_n A \to 0 \qquad \text{in probability.}$$

Q.E.D.

Remark 1. The above proof shows that, if the convergence in (2) holds with probability one, then the convergence in (3) holds with probability one.

Remark 2. Suppose

$$\frac{1}{n} \mathrm{E} \log f_n(X_1, \cdots, X_n | \theta) \to \eta(\theta)$$

for each θ. Let $f_n(X_1, \cdots, X_n | \theta_0)$ be the true density of $\{X_n\}$. Then we have $\eta(\theta_0) \geq \eta(\theta)$. In fact,

$$\begin{aligned} \eta(\theta_0) - \eta(\theta) &= \lim_{n\to\infty} \frac{1}{n} (\mathrm{E} \log f_n(X_1, \cdots, X_n | \theta_0) \\ &\quad - \mathrm{E} \log f_n(X_1, \cdots, X_n | \theta)) \\ &= \lim_{n\to\infty} \frac{1}{n} \mathrm{E} \log \frac{f_n(X_1, \cdots, X_n | \theta_0)}{f_n(X_1, \cdots, X_n | \theta)} \geq 0. \end{aligned}$$

It is well known (Kullback [3]) that

$$\mathrm{E} \log \frac{f_n(X_1, \cdots, X_n | \theta_0)}{f_n(X_1, \cdots, X_n | \theta)} \geq 0$$

Theorem 1 shows that the asymptotic behavior of the posterior distributions under a specified model can be analyzed by using $\eta(\theta)$ once likelihood functions converges uniformly to $\eta(\theta)$ in a manner defined in (2). For $i.i.d.$ random variables $\{X_n\}$, Berk [1] established the condition (2) by using the strong law of large numbers for Banach-valued $i.i.d.$ random variables. By the similar idea, we can show that the condition (2) holds for ergodic stationary Markov processes by the ergodic theorem for Banach-valued stationary processes. This will be done in the next section.

3. THE CASE OF ERGODIC STATIONARY MARKOV PROCESSES. In this section we treat the case where the process $\{X_n\}$, $n=1, 2, \cdots$ is an ergodic stationary Markov process. We shall show that, under a specified model described soon, the condition (2) in Theorem 1 is satisfied for this class of stochastic processes.

Let $f(y|x,\theta_0)$ be the transition probability density of the process $\{X_n\}$ characterized by a parameter θ_0. $f(x|\theta_0)$ denotes the density of the random variable X_1. Then, given a parameter set Θ which is a compact metric space,

$$f_n(X_1,\cdots,X_n|\theta)=f(X_1|\theta)\prod_{i=2}^{n} f(X_i|X_{i-1},\theta),\, n=1,2,\cdots$$

is the likelihood functions of $\{X_n\}$ defined on the parameter set Θ. We shall adopt the functions $f_n(x_1,\cdots,x_n|\theta), n=1,2,\cdots$ as a model for the densities of the process $\{X_n\}$. The following assumptions are made:

(B1) $f(\cdot,\cdot|\cdot)$ is jointly measurable and, for each fixed (x, y), $f(y|x,\cdot)$ is continuous. $f(\cdot|\theta)$ is measurable for each $\theta\in\Theta$.

(B2) There exists a measurable function $K(y|x)$ such that

$$E\,K(X_2|X_1)<\infty \qquad \text{and} \qquad |\log f(y|x,\theta)|\leq K(y|x).$$

Let us define

$$\begin{aligned}\eta(\theta)&=E\log f(X_2|X_1,\theta)\\ &=\int \log f(y|x,\theta)\cdot f(y|x,\theta_0)\cdot f(x|\theta_0)\,dy\,dx.\end{aligned} \tag{8}$$

Note that under (B1 - 2) $E\log f(X_2|X_1,\cdot)$ exists in the sense of Bochner's integral. Then we have

Proposition 1. Under the assumption (B1) and (B2),

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\log f_n(X_1,\cdots,X_n|\theta)-\eta(\theta)\right|\to 0$$

with probability one.

Proof. We use the ergodic theorem for Banach-valued stationary processes. Let $C(\Theta)$

be the space of all continuous functions defined on Θ with the sup-norm. Since

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}f(X_1\mid\theta)\right|\to 0 \qquad \text{with probability one,}$$

it suffices to prove

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=2}^{n+1}\log f(X_i\mid X_{i-1},\theta)\right|\to 0 \tag{9}$$

with probability one.

By the assumption (B1) and separability of the space $C(\Theta)$ with the aid of Lemma 2.2.1 of Padgett-Taylor [4], $Y_i=\log f(X_{i+1}\mid X_i,\cdot)$, $i=1,\cdots\cdots$ are random variables in $C(\Theta)$. Hence if we can show that the process $\{Y_n\}$, $n=1,2,\cdots$ is an ergodic stationary process in Banach space $C(\Theta)$, then the ergodic theorem for Banach-valued stationary processes (see Parthasarathy [5]) asserts (9) and the conclusion follows.

To show the stationarity of the process $\{Y_n\}$, it is sufficient to prove, for example,

$$P\{\omega;(Y_1(\omega),\cdots,Y_k(\omega))\in A\}=P\{\omega;(Y_2(\omega),\cdots,Y_{k+1}(\omega))\in A\} \tag{10}$$

for any k and $A\in\mathscr{B}(C^k(\Theta))$ where $C^k(\Theta)$ is the product space of k copies of $C(\Theta)$ This can be done by the same method as in Lemma 2.3.4 of Padget-Taylor [4].

The set

$$U=\{\{x\in C^k(\Theta),f(x)<b\}:f\in C^k(\Theta)^*\text{and}\,b\in R^1\},$$

where $C^k(\Theta)^*$ is the dual space of $C^k(\Theta)$, is a family of unicity for the Borel-field $\mathscr{B}(C^k(\Theta))$ (P. 25 of Padget-Taylor [4]), and it suffices to show that (10) holds for any $A\in U$. Now,

for $B=\{x\in C^k(\Theta),f(x)<b\}\in U$,

$$P\{\omega;(Y_1,\cdots,Y_k)\in B\}=P\{\omega;f(Y_1,\cdots,Y_k)<b\}.$$

Since $Y_i=\log f(X_{i+1}\mid X_i,.)$, $f(Y_1,\cdots,Y_k)$ is a function of $X_1,\cdots,X_{k+1}$, *i.e.*,

$$f(Y_1,\cdots,Y_k)=g(X_1,\cdots,X_{k+1}).$$

g is a composite function of g_1 and g_2 where

$$g_1:(x_1,\cdots,x_{k+1})\in R^{k+1}\to(f(x_2\mid x_1,\cdot),\cdots,f(x_{k+1}\mid x_k,\cdot))\in C^k(\Theta)$$

$$g_2:\omega\in\Omega\to(X_1(\omega),\cdots,X_{k+1}(\omega))\in R^{k+1}$$

The assumption (B1) and separability of $C^k(\Theta)$ ensure measurability of g_1, and hence g is measurable. With this fact and stationarity of the process $\{X_n\}$, $g(X_1,\cdots,X_{k+1})$ and $g(X_2,\cdots,X_{k+2})$ have the same distribution. Hence

$$P\{\omega\,;\,f(Y_1,\cdots,Y_k)<b\}$$

$$=P\{\omega\,;\,g(X_1,\cdots,X_{k+1})<b\}$$

$$=P\{\omega\,;\,g(X_2,\cdots,X_{k+2})<b\}$$

$$=P\{\omega\,;\,f(Y_2,\cdots,Y_{k+1})<b\}$$

$$=P\{\omega\,;\,(Y_2,\cdots,Y_{k+1})\in B\}.$$

This establishes (10) and, hence, stationarity of $\{Y_n\}$.

Finally ergodicity of $\{Y_n\}$ is also proved by showing that for every $A\in\mathscr{B}(C^{k+1}(\Theta))$, $k=1,2,\cdots$,

$$\text{(11)}\qquad \frac{1}{N}\sum_{n=1}^{N}\chi_A(Y_n,\cdots,Y_{n+k})\to P(\omega\,;\,(Y_1,\cdots,Y_{k+1})\in A)$$

with probability one where χ_A is the characteristic function of the set A. Since $\chi_A(Y_n,\cdots,Y_{n+k})$ is a function of $\{X_n\}$, *i.e.*,

$$\chi_A(Y_n,\cdots,Y_{n+k})=g(X_n,\cdots,X_{n+k},X_{n+k+1}),$$

just as before $g(\cdot)$ is a measurable function. Hence the process $\{Z_n\}$, where $Z_n=g(X_n,\cdots,X_{n+k+1})$, is ergodic and stationary, and by the ergodic theorem

$$\frac{1}{N}\sum_{n=1}^{N}\chi_A(Y_n,\cdots,Y_{n+k})=\frac{1}{N}\sum_{n=1}^{N}g(X_n,\cdots,X_{n+k+1})$$

$$\to E\ g(X_1,\cdots,X_{k+2})$$

$$=E\chi_A(Y_1,\cdots,Y_{k+1})$$

$$=P\{\omega\,;Y_1,\cdots,Y_{k+1})\in A\}.$$

This shows the relation (11). Q.E.D.

Remark 3. From Remark 2 we have $\eta(\theta_0)\geqq\eta(\theta)$. If the following condition

$$(*):\qquad f(y\,|\,x,\theta_1)\neq f(y\,|\,x,\theta_2)\ a.e.\quad \theta_1\neq\theta_2$$

and $f(y\,|\,x,\theta_0)$ and $f(x\,|\,\theta_0)$ are both positive, holds then $\eta(\theta_0)=\eta(\theta)$ implies $\theta_0=\theta$.

Example 1. Let us consider a first order ergodic stationary Markov process $\{X(n)\}$ given by

$$X(n)=\Phi_0 X(n-1)+G_0 w(n-1),\ n=\cdots,-1,0,1,\cdots$$

where $X(n)\in R^d$ and $\{w(n)\}$ is a k-dimensional vector valued $i.i.d.$ random sequence with normal distribution $N(O,I_k)$, I_k identity matrix. Φ_0 and G_0 are unknown $d\times d$ and $d\times k$ matrices respectively, and we assume that absolute values of all eigenvalues of Φ_0 lie in a unit circle. We investigate the asymptotic behavior of posterior distributions of the parameter $\theta=(\Phi,G)$ on a compact set Θ which need not contain the true parameter $\theta_0=(\Phi_0,G_0)$ The transition density of the process $\{X(n)\}$ is given by

$$f(y|x,\theta_0)=((2\pi)^d|G_0G_0'|)^{-\frac{1}{2}}\exp\{-\frac{1}{2}(y-\Phi_0x)'(G_0G_0')^{-1}(y-\Phi_0x)\}$$

where we assumed nonsingularity of the matrix G_0G_0' . The density $f(x|\theta_0)$ of $X(n)$ is normal $N(O,\Gamma)$ where Γ is given by

$$\Gamma=G_0G_0'+\Phi_0G_0G_0'\Phi_0'+\Phi_0^2\ G_0G_0'\ (\Phi_0')^2+\cdots.$$

Now the function $\eta(\theta)$ defined by (8) is

$$\begin{aligned}\eta(\theta)&=-\frac{1}{2}\log(2\pi)^d|GG'|\\&\quad-\frac{1}{2}\int(y-\Phi x)'(GG')^{-1}(y-\Phi x)f(y|x;\theta_0)f(x|\theta_0)\,dydx\\&=-\frac{1}{2}\log(2\pi)^d|GG'|-\frac{1}{2}\mathrm{trace}[(GG')^{-1}G_0G_0']\\&\quad-\frac{1}{2}\int x'(\Phi_0-\Phi)'(GG')^{-1}(\Phi_0-\Phi)xf(x|\theta_0)\,dx\\&=-\frac{1}{2}\log(2\pi)^d|GG'|-\frac{1}{2}\mathrm{trace}[(GG')^{-1}G_0G_0']\\&\quad-\frac{1}{2}\mathrm{trace}(\Phi_0-\Phi)'(GG')^{-1}(\Phi_0-\Phi)\Gamma\end{aligned}\tag{12}$$

where we are assuming nonsingularity of GG' for all $\theta=(\Phi,G)\in\Theta$.
As we have shown in Remark 2

$$\begin{aligned}\eta(\theta_0)&=-\frac{1}{2}\log(2\pi)^d|G_0G_0'|-\frac{1}{2}d\\&\geqq\eta(\theta),\ \forall\theta\in\Theta.\end{aligned}$$

The asymptotic carrier A_0 for the parameter set Θ can be calculated by using (12) and this enables us various kinds of analysis for the asymptotic behavior of posterior distributions. For example, let the parameter set Θ be such that $\Theta=\{(\Phi,G_1)$,

$\Phi\in\Theta_1,\ G_1\neq G_0\}$ where Θ_1 is a compact set regarding the parameter Φ and contains Φ_0. For this case the asymptotic carrier A_0 contains only one point (Φ_0, G_1) since $\mathrm{trace}(\Phi_0-\Phi)'(GG')^{-1}(\Phi_0-\Phi)\Gamma=0$ if and only if $\Phi=\Phi_0$. Hence even if the specified model does not include the true density, *i.e.*, $G_1\neq G_0$, as far as the parameter Φ is concerned the posterior distributions on Φ converge to the true point Φ_0.

4. A SUFFICIENT CONDITION FOR GENERAL CASES. Returning to the general case, let $\{X_n\}, n=1,\cdots,$ be a random sequence. Given a joint density model $f_n(x_1,\cdots,x_n|\theta)$, $\theta\in\Theta$, $n=1,2,\cdots$ for the process $\{X_n\}$, we have the following proposition regarding the uniform convergence of likelihood functions.

Proposition 2. Assume that

(i) $\quad \mathrm{Var}(\frac{1}{n}\log f_n(X_1,\cdots,X_n|\theta))\rightarrow 0$

uniformly in θ

(ii) $\quad \mathrm{E}(\frac{1}{n}\log f_n(X_1,\cdots,X_n|\theta))\rightarrow\eta(\theta)$

uniformly in θ, then the condition (2) in Theorem 1 is satisfied

Proof. Application of Chebyshev's inequality easily shows the result. In fact, for an arbitrary number $c>0$,

$$P\{\omega;\sup_{\theta\in\Theta}|\frac{1}{n}\log f_n(X_1,\cdots,X_n|\theta)-\eta(\theta)|>c\}$$
$$\leq P\{\omega;\sup_{\theta\in\Theta}|\frac{1}{n}\log f_n(X_1,\cdots,X_n|\theta)-\frac{1}{n}\mathrm{E}\log f_n(X_1,\cdots,X_n|\theta)|>\frac{c}{2}\}$$
$$+P\{\omega;\sup_{\theta\in\Theta}|\frac{1}{n}\mathrm{E}\log f_n(X_1,\cdots,X_n|\theta)-\eta(\theta)|>\frac{c}{2}\}$$

By Chebyshev's inequality

$$\leq\frac{4}{c^2}\mathrm{E}\{\sup_{\theta\in\Theta}|\frac{1}{n}\log f_n(X_1,\cdots,X_n|\theta)-\frac{1}{n}\mathrm{E}\log f_n(X_1,\cdots,X_n|\theta)|\}^2$$
$$+\frac{4}{c^2}\{\sup_{\theta\in\Theta}|\frac{1}{n}\mathrm{E}\log f_n(X_1,\cdots,X_n|\theta)-\eta(\theta)|\}^2$$

By Fatou's lemma, the first term in the above equation is less than

$$\frac{4}{c^2}\{\sup_{\theta\in\Theta}\mathrm{E}|\frac{1}{n}\log f_n(X_1,\cdots,X_n|\theta)-\frac{1}{n}\mathrm{E}\log f_n(X_1,\cdots,X_n|\theta)|\}^2$$

Hence the conclusion follows. Q.E.D.

Remark 4. (a) In Proposition 2 if $\eta(\theta)$ is continuous, then we can apply Theorem 1. (b) When the process $\{X_n\}$ is an independent (but not necessarily identically distributed) sequence, the density model is given by

$$f_n(x_1, \cdots, x_n \mid \theta) = f_1(x_1 \mid \theta) \cdots f_n(x_n \mid \theta), \quad \theta \in \Theta$$

where $f_i(x_i \mid \theta)$ is the density model for random variable X_i.
In this case the assumption in Proposition 2 takes the form:

(i) $\frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var} \log f_i(X_i \mid \theta) \to 0$ uniformly in θ.

(ii) $\frac{1}{n} \sum_{i=1}^{n} E \log f_i(X_i \mid \theta) \to \eta(\theta)$ uniformly in θ.

Example 2. Let us consider the following linear system:

$$X_{n+1} = \Phi X_n + G U_n, \quad X_1 \text{ ; given}$$

$$Y_n = H X_n + \xi_n$$

where $X_n \in R^d, Y_n, \xi_n \in R^l, U_n \in R^m$ and matrices Φ, G, H have appropriate dimensions. $\{U_n\}$ is a given control sequence and we assume that $\{\xi_n\}$ is an independent and normally distributed sequence with mean zero and covariance matrix Γ. The unknown parameter θ consists of Φ, G, H and Γ. The process $\{Y_n\}$ is clearly independent but not identically distributed. Φ_0, G_0, H_0 and Γ_0 denote the true parameter. Then Y_n has a normal distribution $N(H_0 X_n(\theta_0), \Gamma_0)$ where $X_n(\theta_0)$ is the state vector corresponding to the true parameter θ_0, and the density model is given by

$$f_n(Y_1, \cdots, Y_n \mid \theta) = \prod_{i=1}^{n} f_i(Y_i \mid \theta)$$

where

$$f_i(Y_i \mid \theta) = ((2\pi)^l |\Gamma|)^{-\frac{1}{2}} \exp\{-\frac{1}{2}(Y_i - H X_i(\theta))' \Gamma^{-1} (Y_i - H X_i(\theta))\}$$

and $X_i(\theta)$ is the state vector corresponding to the parameter θ.
We assume the following:

(C1) Φ_0 and $\Phi \in \Theta$ are stable matrices, *i.e.*, absolute values of all eigenvalues of these matrices lie in a unit circle.

(C2) The control sequence is uniformly bounded, *i.e.*, $|U_u| \leq K$. and, it has the property such that

$$\frac{1}{N} \Sigma (H_0 X_n(\theta_0) - H X_n(\theta))(H_0 X_n(\theta) - H X_n(\theta))'$$

converges to a function of θ uniformly in θ.

Then conditions (i) and (ii) in Proposition 2 are satisfied. To show this, first we note that the state vector $X_n(\theta)$ is uniformly bounded, *i.e.*, $|X_n(\theta)| \leq M$ for all n and $\theta \in \Theta \cup \{\theta_0\}$. In fact, since

$$X_n(\theta) = GU_{n-1} + \Phi GU_{n-2} + \Phi^2 GU_{n-3} + \cdots + \Phi^{n-1} GU_0 + \Phi^n X_0 \ ,$$

$$|X_n(\theta)| \leq ||G|| \, |U_{n-1}| + ||\Phi|| \, ||G|| \, |U_{n-2}| + \cdots + ||\Phi||^{n-1} ||G|| \, |U_0| + ||\Phi||^n |X_0|$$

where the matrix norm $||A||$ is defined by $||A|| = \sup_{|x| \leq 1} |Ax|$.
The right hand side of the above inequality is less than

$$\begin{aligned} & K \, ||G|| \, (1 + ||\Phi|| + ||\Phi||^2 + \cdots \quad) + |X_0| \\ & = K \, ||G|| \frac{1}{1 - ||\Phi||} + |X_0| \leq M \end{aligned} \tag{13}$$

Note that by (C1) $||\Phi|| < 1$. The last inequality in (13) is due to the compactness of the set Θ, and the uniform boundedness of $X_n(\theta)$ follows. Since $\{Y_n\}$ are independent, according to Remark 4 (b), let us calculate $\mathrm{Var}\ \log f_n(Y_n | \theta)$.
Since

$$\begin{aligned} \log f_n(Y_n|\theta) = & \log((2\pi)^l |\Gamma|)^{-\frac{1}{2}} - \frac{1}{2} (H_0 X_n(\theta_0) - HX_n(\theta))' \Gamma^{-1} (H_0 X_n(\theta_0) - HX_n(\theta)) \\ & - \xi_n' \Gamma^{-1} (H_0 X_n(\theta_0) - HX_n(\theta)) \\ & - \frac{1}{2} \xi_n' \Gamma^{-1} \xi_n \quad , \end{aligned}$$

we have

$$\begin{aligned} \mathrm{Var}\ \log f_n(Y_n|\theta) = & \mathrm{E}\,(\xi_n' \Gamma^{-1} (H_0 X_n(\theta_0) - HX_n(\theta)) + \frac{1}{2} \xi_n' \Gamma^{-1} \xi_n \\ & - \frac{1}{2} \mathrm{E}\, \xi_n' \Gamma^{-1} \xi_n)^2 \\ & \leq 2\,\mathrm{E}\,(\xi_n' \Gamma^{-1} (H_0 X_n(\theta_0) - HX_n(\theta)))^2 + 2\mathrm{Var}(\frac{1}{2} \xi_n' \Gamma^{-1} \xi_n) \\ & \leq M \end{aligned}$$

where M does not depend on θ.
The last inequality comes from finiteness of the moment of ξ_n and uniform boundedness of $H_0 X_n(\theta_0) - HX_n(\theta)$. Now the condition (i) in Remark (b) is easily checked by noting that

$$\begin{aligned} \mathrm{Var}(\frac{1}{2} \log f_n(Y_1, \cdots, Y_n | \theta) = & \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}\ \log f_i(Y_i | \theta) \\ & \leq \frac{M}{n} \to 0 \qquad \text{uniformly in } \theta . \end{aligned}$$

Similarly, by the condition (C2),

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E} \log f_n(Y_n \mid \theta) \to \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} \mathrm{E} \log f_i(Y_i \mid \theta) = \eta(\theta)$$

uniformly in θ.

Thus by Proposition 2 we have

$$\sup_{\theta\in\Theta} \left| \frac{1}{n} \log f_n(Y_1,\cdots,Y_n \mid \theta) - \eta(\theta) \right| \to 0$$

in probability and $\eta(\theta)$ is given by

$$\eta(\theta) = \log((2\pi)^{l}|\Gamma|)^{-\frac{1}{2}} - \frac{1}{2}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{n}(H_0 X_n(\theta_0) - HX_n(\theta))'\Gamma^{-1}$$

$$(H_0 X_n(\theta_0) - HX_n(\theta)) - \frac{1}{2}\mathrm{trace}\ \Gamma^{-1}\Gamma_\theta$$

As was done in Example 1 in Section 3, we shall investigate the asymptotic behavior of posterior distributions on (Φ, G, H) under the condition that Γ is arbitrarily fixed, $i.e.$ $\Gamma=\Gamma_1$. Let the parameter set Θ be such that $\Theta=\Gamma_1\times\Theta_1$ where Θ_1 is a parameter set regarding (Θ, G, H) and we shall assume that Θ_1 contains the true parameter(Φ_0, G_0, H_0). Clearly $\eta(\theta_1)=\sup_{\theta\in\Theta}\eta(\theta)$ where $\theta_1=(\Gamma_1, \Phi_0, G_0, H_0)$ and hence the asymptotic carrier A_0 contains θ_1. But in this case $\eta(\theta_1)=\eta(\theta)$ does not imply $\theta_1=\theta$ since, for all non-singular matrix T, $\theta=(\Gamma_1, T\Phi_0T^{-1}, TG_0, H_0T^{-1})$ gives the same values to $\eta(\theta)$. When the input U_n and the output Y_n are both one dimensional, a necessary and sufficient condition on the input sequence $\{U_n\}$ for the asymptotic carrier A_0 to contain only one point, $i.e.$ $\theta_1=(\Gamma_1, \Phi_0, G_0, H_0)$ is known (Aoki and Yue [6]) under the condition that (Φ, H) has a cannonical observable form and (Φ, G) is a controllable pair.

We shall consider more general cases. To simplify the analysis, we assume (C3): The control sequence $\{U_n\}$ is a uniformly bounded $i.i.d$ random process and

$$\mathrm{E}\,U_iU_j' = \begin{cases} W & i=j\ ; W:\ \text{positive definite} \\ 0 & i=j \end{cases}$$

Furthermore, since we are only concerned with the asymptotic behavior of posterior distributions, we assume that the time index n of $\{Y_n\}$ tends to infinite past, $i.e.$, $n=\cdots, -1, 0, 1\cdots$.

Then since

$$X_n = GU_n + \Phi GU_{u-1} + \Phi^2 GU_{n-2} + \cdots$$

and $\{U_n\}$ is a uniformly bounded $i.i.d$ sequence, by (C1) $\{X_n\}$ is an ergodic stationary process and so is $\{Y_n\}$.

Now $\{Y_n\}$ are no more independent, but since $\xi_n = Y_n - H_0X_n(\theta_0)$ is an independent and normally distributed sequence, the likelihood function $f_n(Y_1, \cdots, Y_n \mid \theta)$ of $Y_1, \cdots, Y_n$ is given by

$$f_n(Y_1, \cdots, Y_n \mid \theta) = \prod_{i=1}^{n} ((2\pi)^{l} |\Gamma|)^{-\frac{1}{2}} \exp\{-\frac{1}{2}(Y_i - HX_i(\theta))' \Gamma^{-1}(Y_i - HX_i(\theta))\}, \tag{14}$$

and we shall assume that this function $f_n(Y_1, \cdots, Y_n \mid \theta)$ is to be the density model for $\{Y_n\}$. Let us investigate the asymptotic behavior of $(1/n)\log f_n(Y_1, \cdots, Y_n \mid \theta)$ directly without using Proposition 2. Since

$$\frac{1}{n}\log f_n(Y_1, \cdots, Y_n \mid \theta) = \log((2\pi)^{l}|\Gamma|)^{-\frac{1}{2}} - \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}(H_0X_i(\theta_0) - HX_i(\theta))'\Gamma^{-1}(H_0X_i(\theta_0) - HX_i(\theta)) - \frac{1}{n}\sum_{i=1}^{n}\xi_i'\Gamma^{-1}(H_0X_i(\theta_0) - HX_i(\theta)) - \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\xi_i'\Gamma^{-1}\xi_i,$$

we shall apply the ergodic theorem to each term of the above equation. Noting that by the same discussion as in Section 2 $\{X_n(\theta)\}$ is an ergodic stationary process in Banach space $C(\Theta)$, by the ergodic theorem in Banach space we have

$$\sup_{\theta \in \Theta} \left| \frac{1}{n}\sum_{i=1}^{n} \xi_n' \Gamma^{-1}(H_0X_n(\theta_0) - HX_n(\theta)) \right| \to 0$$

with probability one.
Here we used the fact that ξ_n and $H_0X_n(\theta_0) - HX_n(\theta)$ are independent.
Similarly

$$\sup_{\theta \in \Theta} \left| \frac{1}{n}\sum_{i=1}^{n} \xi_i' \Gamma^{-1}\xi_i - \operatorname{trace}\Gamma^{-1}\Gamma_0 \right| \to 0$$

with probability one,
and

$$\sup_{\theta \in \Theta} \left| \frac{1}{n}\sum_{i=1}^{n}(H_0X_i(\theta_0) - HX_i(\theta))'\Gamma^{-1}(H_0X_i(\theta_0) - HX_i(\theta)) - E(H_0X_1(\theta_0) - HX_1(\theta))'\Gamma^{-1}(H_0X_1(\theta_0) - HX_1(\theta)) \right| \to 0$$

with probability one.
Hence we have

$$\sup_{\theta \in \Theta} \left| \frac{1}{n}\log f_n(Y_1, \cdots, Y_n \mid \theta) - \eta(\theta) \right| \to 0 \quad \text{with probability one}$$

where

$$\eta(\theta) = \log((2\pi)^{l}|\Gamma|)^{-\frac{1}{2}} - \frac{1}{n}\ \mathrm{trace}\,\Gamma^{-1}\Gamma_0$$
$$-\frac{1}{2}\mathrm{E}(H_0X_1(\theta_0) - HX_1(\theta))'\Gamma^{-1}(H_0X_1(\theta_0) - HX_1(\theta))$$

Let us calculate

$$\mathrm{E}(H_0X_1(\theta_0) - HX_1(\theta))'\Gamma^{-1}(H_0X_1(\theta_0) - HX_1(\theta))$$
$$= \mathrm{trace}\,\Gamma^{-1}\mathrm{E}(H_0X_1(\theta_0) - HX_1(\theta))(H_0X_1(\theta_0) - HX_1(\theta))'.$$

Since

$$HX_1(\theta) = HGU_0 + H\Phi GU_{-1} + H\Phi^2GU_{-2} + \cdots$$

and by (C3), we have

$$\mathrm{E}(H_0X_1(\theta_0) - HX_1(\theta))(H_0X_1(\theta_0) - HX_1(\theta))'$$
$$= \sum_{i=0}^{\infty}(H_0\Phi_0^iG_0 - H\Phi^iG)W(H_0\Phi_0^iG_0 - H\Phi^iG)'$$

Hence

(15)
$$\eta(\theta) = \log((2\pi)^{l}|\Gamma|)^{-\frac{1}{2}} - \frac{1}{2}\ \mathrm{trace}\,\Gamma^{-1}\Gamma^0$$
$$-\frac{1}{2}\mathrm{trace}\,\Gamma^{-1}\sum_{i=0}^{\infty}(H_0\Phi_0^iG_0 - H\Phi^iG)W(H_0\Phi_0^iG_0 - H\Phi^iG)'.$$

Let $\eta(\theta_1) = \eta(\theta)$ where $\theta_1 = (\Gamma_1, \Theta_0, G_0, H_0)$. Then

$$H_0\Phi_0^iG_0 = H\Phi^iG \qquad \text{for all} \quad i \geqq 0.$$

Thus under the assumption

(C4): For all $\theta \in \Theta \cup \{\theta_0\}$ the system

$$\begin{cases} X_n = \Phi X_{n-1} + GU_{n-1} \\ Y_n = HX_n \end{cases}$$

is a minimal realization,

there exists a non-singular matrix T such that

(16) $\Phi = T\Phi_0T^{-1},\ G = TG_0,\ H = H_0T^{-1}$

(See Brockett [7]).

As we have already shown, if the parameter (Φ, G, H) is completely unknown, $\eta(\theta_0)=\eta(\theta)$ does not imply $\theta_0=\theta$. Motivated by this fact, we shall consider the case where some of elements of (Φ_0, G_0, H_0) are known a priori so that the following condition (C5) holds:

(C5): $\theta\in\Theta$, which satisfies (16), is equal to θ_1.

Then under this assumption $\eta(\theta_1)=\eta(\theta)$ implies $\theta_1=\theta$.

We summarize here the obtained result.

Proposition 3. Consider the linear system:

$$X_n=\Phi X_{n-1}+GU_n$$

$$Y_n=HX_n+\xi_n \qquad n=\cdots,-1,0,1,\cdots$$

Under the condition (C1) and (C3),

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\log f_n(Y_1,\cdots,Y_n|\theta)-\eta(\theta)\right|\to 0$$

with probability one

where $f_n(Y_1\cdots,Y_n|\theta)$ and $\eta(\theta)$ are given in (14) and (15).

For a case where $\Theta=\Gamma_1\times\Theta_1$ (defined earlier), under further assumptions (C4) and (C5), $\eta(\theta_1)=\eta(\theta)$ implies $\theta_1=\theta$ and hence the asymptotic carrier A_0 contains only one point $\theta_*=(\Gamma_1, \Phi_0, G_0, H_0)$

References

[1] R.H. Berk, Limiting behavior of posterior distributions when the model is incorrect, Ann. Math. Statist., 37 (1966), 51-58

[2] I.I. Gikhman and A.V. Skorphod, "Introduction to the Theory of Random Processes," W.B. Sounders Co., Philadelphia (1969)

[3] S. Kullback, "Information and Statistics," Wiley, New York (1959)

[4] W.J. Padgett and R.L. Taylar, "Laws of Large Numbers for Normed Linear Spaces and Certain Fréchet Spaces," Lecture Notes in Mathematics, Springer-Verlag, Berlin (1973)

[5] K.R. Parthasarathy, "Probability Measures on Metric Spaces," Academic Press, New York (1967)

[6] M. Aoki and P.C. Yue, On certain convergence questions in system identification, SIAM. J. Control, 8 (1970)

[7] R.W. Brockett, "Finite Dimensional Linear Systems," Wiley, New York (1970)

[8] K. Yoshida, "Functional Analysis," Academic Press, New York (1968)