Improper priors, posterior asymptotic normality and conditional inference

D.A.S, Fraser, P. McDunnough, N.Taback
Department of Statistics
University of Toronto
Toronto, Ontario
Canada M5S 1A1

June 28, 2000

Abstract

The asymptotic normality of the posterior distribution of a vector parameter based on observing a stochastic process is established without the assumption of a proper prior. Applications to conditional-inference are discussed.

Keywords: Asymptotic normality, Conditional inference, Bayesian methods, Likelihood, Structural Distribution, Stochastic processes.

1. Introduction

There has been a modest amount of lierature on the posterior asymptotic normality of a scaler parameter based on a proper prior distribution. Heyde and Johnstone (1979) extended the proof of Walker (1969) to stochastic processes. A proof of the multiparameter case can be found in Johnstone (1978). Sweeting and Adekola (1987) extended the regularity conditions found in Heyde and Johnstone (1979) to cover a wider class of processes, again based on a proper prior.

Posterior asymptotic normality where the prior distribution is not assumed to be proper was first examined in Brenner, Fraser and McDunnough (1982) for a sample from a scalar location parameter model. Fraser and McDunnough (1984) extended the result to a general model with a scalar parameter and mention that the result should hold in the mutiparameter case. Indeed, the literature does not seem to contain a proof of the multiparameter case without the assumption of a proper prior. The purpose of this paper is to establish the asymptotic normality of the posterior distribution of a vector parameter based on observing a stochastic process without the assumption of a proper prior.

Posterior distributions or normalized likelihood functions have an intimate connection with the confidence distribution generated by a transformation or structural model using conditional inference. Theorem 3.1 in Brenner, Fraser and McDunnough (1982) gives conditions under which convergence almost surely of the suitably normalized likelihood function or the posterior to the standard multivariate normal distribution is sufficient for almost sure convergence of the standardized conditional-inference distribution to the standard multivariate normal. In a transformation model the likelihood function can be normalized with respect to an invariant measure on the parameter space, as a natural non-informative prior. If the parameter space is a compact group then the invariant measure will be proper, but for location, location-scale, regression models and others with noncompact groups, this prior is improper. Consider the classical regression model $\mathbf{y} = X\beta + \sigma\mathbf{z}$ where $\mathbf{z}$ is $N_k(\mathbf{0}, \mathbf{I})$ with the parameter $\theta = (\beta, \sigma)$; the strict asymptotic normality of the posterior of $\widehat{\Sigma}_n^{-1/2}(\theta - \widehat{\theta}_n)$ does not seem to be available in the literature, but is the concern of this paper.

2. Asymptotic Normality of the Posterior Distribution

2.1 Preamble Let the parameter space $\Theta = \Re^k$, be k-dimensional Euclidean space. We find it convenient to use the norm

$$\|x\| = \max_{1 \le j \le k} |x_j|$$

on $\Re^k$ rather than the Euclidean norm. If A is a linear transformation on $\Re^k$ defined by the matrix $[A_{ij}]$, then the norm of A will be taken as

$$\|A\| = \max_{1 \le i \le k} \sum_{j=1}^{k} |A_{ij}|$$

We then have that $\|Ax\| \le \|A\| \, \|x\|$ for all $x \in \Re^k$. It also follows from the Cauchy-Schwartz inequality that $\|x'Ax\| \le \|x'\| \, \|Ax\| \le \|A\| \, \|x\|^2$, for all $x \in \Re^k$, where x' denotes the transpose of the x.

Let $\widehat{\theta}_n = \widehat{\theta}$ be the maximum likelihood estimate of θ. The likelihood function from $\mathbf{x} = (x_1, ..., x_n)$ having density $f_n(\mathbf{x}|\theta)$ with respect to a $\sigma-$finite measure not dependent on θ is given by $L_n(\theta) = cf_n(\mathbf{x}|\theta)$ with log-likelihood $l_n(\theta) = a + \log f_n(\mathbf{x}|\theta)$ where a is an arbitrary constant.

2.2 Assumptions The following assumptions concerning the model and a prior $w(\theta)$ are closely related to those in Fraser and McDunnough (1984), and Johnstone (1978).

Assumption 1. $\int w(v)L_n(v)dv < \infty$ and $w(\theta)$ is continuous and positive at the true θ.

Assumption 2. $l_n(\theta)$ is twice continuously differentiable, and $\det(-l_n''(\theta)) \in (0, \infty)$, where

$$l_n''(\theta) = \left\{ \frac{\partial^2}{\partial\theta_i \partial\theta_j} l_n(\theta) \right\}$$

is the matrix of second-order derivatives of l_n at θ. Define $\Sigma_n(\theta) = \{-l_n''(\theta)\}^{-1}$ and assume that both $\Sigma_n \to 0$ and $\widehat{\Sigma}_n \to 0$, where $\widehat{\Sigma}_n = \Sigma_n(\widehat{\theta})$.

Assumption 3. For every $\delta > 0$

$$\overline{\lim_{n\to\infty}} \sup_{s:\|s-\theta\| \ge \delta} \left\| \Sigma_n^{-1}(\theta) \right\|^{-1} \{l_n(s) - l_n(\theta)\} \overset{a.s}{<} 0$$

Assumption 4 For every $\epsilon > 0$ there exists a $\delta > 0$ such that,

$$\overline{\lim_{n\to\infty}} \sup_{s:||s-\theta||\leq\delta} ||\Sigma_n'^{\frac{1}{2}}(\theta)\{l_n''(s) - l_n''(\theta)\}\Sigma_n^{\frac{1}{2}}(\theta)|| \overset{a.s}{<} \epsilon$$

Assumption 3 is a modified version of Assumption I' in Fraser and McDunnough (1984). Assumption 4. ensures that in a neighbourhood of its maximum the likelihood function has a multivariate normal form, that is,

$$\frac{L_n(\widehat{\theta} + \widehat{\Sigma}^{\frac{1}{2}} t)}{L_n(\widehat{\theta})} \overset{a.s}{\to} \exp\left(-\frac{1}{2}t't\right) \tag{1}$$

For this last statement expand l_n in a Taylor series about $\widehat{\theta}$ to obtain

$$l_n(\widehat{\theta} + \widehat{\Sigma}^{\frac{1}{2}} t) - l_n(\widehat{\theta}) = -\frac{1}{2}t'\widehat{\Sigma}_n'^{\frac{1}{2}}\{-l_n''(\widetilde{\theta})\}\widehat{\Sigma}_n^{\frac{1}{2}} t \ , \tag{2}$$

where $\widetilde{\theta}$ lies somewhere on the line joining $\widehat{\theta}$ to $\widehat{\theta} + \widehat{\Sigma}^{\frac{1}{2}} t$. Assumption 4. then gives (1).

2.3. The Main Result

Theorem 1. *Let $\{X_t, t \in T\}$ be a stochastic process and suppose that we observe a realization $\mathbf{x} = (x_1, ..., x_n)$ having density $f_n(\mathbf{x}|\theta)$ with respect to a $\sigma-$finite measure not dependent on θ. If assumptions 1,2,3,4 hold then*

$$\frac{\det\left(\widehat{\Sigma}_n^{\frac{1}{2}}\right) w(\widehat{\theta} + \widehat{\Sigma}_n^{\frac{1}{2}} t) L_n(\widehat{\theta} + \widehat{\Sigma}_n^{\frac{1}{2}} t)}{\int w(u) L_n(u) du} \overset{a.s}{\to} (2\pi)^{-k/2} e^{-t't/2}$$

Proof: From Assumption 1. it follows that $w(\widehat{\theta} + \widehat{\Sigma}^{\frac{1}{2}} t)/w(\widehat{\theta}) \overset{a.s}{\to} 1$. Combining this with (1) we see that it suffices to show that

$$\frac{\int w(u) L_n(u) du}{\det\left(\widehat{\Sigma}_n^{\frac{1}{2}}\right) w(\widehat{\theta}) L_n(\widehat{\theta})} \overset{a.s}{\to} (2\pi)^{k/2} \tag{3}$$

Towards this end we note we note that

$$\frac{\int w(u) L_n(u) du}{\det\left(\widehat{\Sigma}_n^{\frac{1}{2}}\right) w(\widehat{\theta}) L_n(\widehat{\theta})} = \int \frac{w(\widehat{\theta} + \widehat{\Sigma}_n^{\frac{1}{2}} t) L_n(\widehat{\theta} + \widehat{\Sigma}_n^{\frac{1}{2}} t)}{w(\widehat{\theta}) L_n(\widehat{\theta})} dt \ .$$

For any bounded set $A \subset \Re^k$ the dominated-convergence theorem yields

$$\int_A \frac{w(\widehat{\theta} + \widehat{\Sigma}_n^{\frac{1}{2}} t) L_n(\widehat{\theta} + \widehat{\Sigma}_n^{\frac{1}{2}} t)}{w(\widehat{\theta}) L_n(\widehat{\theta})} dt \overset{a.s}{\to} \int_A e^{-t't/2} dt$$

As in Fraser and McDunnough (1984) we define sets

$$\begin{aligned} B_{\delta,n} &= \left\{ t : \left\| \widehat{\theta} + \widehat{\Sigma}_n^{\frac{1}{2}} t - \theta \right\| < \delta \right\} \cap A^c \\ B_n &= \left\{ t : \left\| \widehat{\theta} + \widehat{\Sigma}_n^{\frac{1}{2}} t - \theta \right\| > \delta \right\} \cap A^c \end{aligned}$$

with $\delta > 0$. $B_{\delta,n}$ is the region outside the bounded subset $A \subset \Re^k$, but in a δ-neighbourhood of the true parameter; while B_n is the region outside a bounded subset and not in a neighbourhood of the true parameter. Towards (3) we note that it now suffices to show that

$$\lim_{A \to \Re^k} \lim_{n \to \infty} \int_{B_{\delta,n}} \frac{w(\widehat{\theta} + \widehat{\Sigma}_n^{\frac{1}{2}} t) L_n(\widehat{\theta} + \widehat{\Sigma}_n^{\frac{1}{2}} t)}{w(\widehat{\theta}) L_n(\widehat{\theta})} dt \overset{a.s}{=} 0 \tag{4}$$

and

$$\lim_{n \to \infty} \int_{B_n} \frac{w(\widehat{\theta} + \widehat{\Sigma}_n^{\frac{1}{2}} t) L_n(\widehat{\theta} + \widehat{\Sigma}_n^{\frac{1}{2}} t)}{w(\widehat{\theta}) L_n(\widehat{\theta})} dt \overset{a.s}{=} 0 \tag{5}$$

If we can show that

$$\lim_{A \to \Re^k} \lim_{n \to \infty} \int_{B_{\delta,n}} \frac{L_n(\widehat{\theta} + \widehat{\Sigma}_n^{\frac{1}{2}} t)}{L_n(\widehat{\theta})} dt \overset{a.s}{=} 0 \tag{6}$$

then (4) follows immediately. Towards (6) let $R_n = \Sigma_n(l_n''(\widetilde{\theta}) - l_n''(\widehat{\theta}))$ and $t \in B_{\delta,n}$; then the multivariate version of Taylor's theorem yields

$$\begin{aligned} l_n(\widehat{\theta} + \widehat{\Sigma}_n^{\frac{1}{2}} t) - l_n(\widehat{\theta}) + \frac{1}{2} \|t\|^2 &= \frac{1}{2} t' (\widehat{\Sigma}_n^{1/2})' \Sigma_n^{-1} R_n \widehat{\Sigma}_n^{1/2} t \\ &\leq \frac{1}{2} \left\| t' (\widehat{\Sigma}_n^{1/2})' (l_n''(\widetilde{\theta}) - l_n''(\widehat{\theta})) (\widehat{\Sigma}_n^{1/2}) t \right\| \\ &\leq \frac{1}{2} \|t'\| \left\| (\widehat{\Sigma}_n^{1/2})' \{ l_n''(\widetilde{\theta}) - l_n''(\widehat{\theta}) \} (\widehat{\Sigma}_n^{1/2}) \right\| \|t\| \overset{a.s}{<} \frac{\epsilon}{2} t' t \end{aligned}$$

where the last inequality follows from Assumption 4. Thus,

$$\frac{L_n(\widehat{\theta} + \widehat{\Sigma}_n^{\frac{1}{2}} t)}{L_n(\widehat{\theta})} \overset{a.s}{<} \exp(-c_\epsilon t' t)$$

where the constant $c_\epsilon > 0$ for any $\epsilon > 0$ and < 1. Finally for suitable $\delta > 0$ we have that,

$$\lim_{A \to \Re^k} \lim_{n \to \infty} \int_{B_{\delta,n}} \frac{L_n(\widehat{\theta} + \widehat{\Sigma}_n^{\frac{1}{2}} t)}{L_n(\widehat{\theta})} dt \overset{a.s}{<} \lim_{A \to \Re^k} \lim_{n \to \infty} \int_{B_{\delta,n}} \exp(-c_\epsilon t' t) = 0$$

In anticipation of (5), we write

$$\frac{L_n(\widehat{\theta}+\widehat{\Sigma}_n^{\frac{1}{2}}t)}{L_n(\widehat{\theta})}=\frac{L_1(\widehat{\theta}+\widehat{\Sigma}_n^{\frac{1}{2}}t)}{L_1(\widehat{\theta})}\frac{L_{n-1}(\widehat{\theta}+\widehat{\Sigma}_n^{\frac{1}{2}}t)}{L_{n-1}(\widehat{\theta})}$$

where $L_{n-1}\propto f(x_1,\ldots,x_n;\theta)/f(x_1;\theta)$ and observe that

$$\overline{\lim_{n\to\infty}}\sup_{s:\|s-\theta\|\geq\delta}\left\|\Sigma_n^{-1}(\theta)\right\|^{-1}[l_{n-1}(s)-l_{n-1}(\theta)]\leq\overline{\lim_{n\to\infty}}\sup_{s:\|s-\theta\|\geq\delta}\left\|\Sigma_n^{-1}(\theta)\right\|^{-1}[l_n(s)-l_n(\theta)]\overset{a.s}{<}0.$$

Then an application of Assumption 3 yields

$$\frac{L_{n-1}(\widehat{\theta}+\widehat{\Sigma}_n^{\frac{1}{2}}t)}{L_{n-1}(\widehat{\theta})}\overset{a.s}{\leq}\exp\left(-d\left\|\widehat{\Sigma}_n^{-1}\right\|\right)\ ,$$

forall $t\in B_n$ and for a constant $d>0$ (see Fraser and McDunnough, (1984). It then follows from Assumption 1. that

$$\begin{aligned}\int_{B_n}\frac{w(\widehat{\theta}+\widehat{\Sigma}_n^{\frac{1}{2}}t)L_n(\widehat{\theta}+\widehat{\Sigma}_n^{\frac{1}{2}}t)}{w(\widehat{\theta})L_n(\widehat{\theta})}dt &\leq \frac{e^{-d\|\widehat{\Sigma}_n^{-1}\|}}{w(\widehat{\theta})L_1(\widehat{\theta})}\int_{B_n}w(\widehat{\theta}+\widehat{\Sigma}_n^{\frac{1}{2}}t)L_1(\widehat{\theta}+\widehat{\Sigma}_n^{\frac{1}{2}}t)dt\\ &\leq \frac{e^{-d\|\widehat{\Sigma}_n^{-1}\|}}{w(\widehat{\theta})L_1(\widehat{\theta})\det\left(\widehat{\Sigma}_n^{1/2}\right)}\int w(u)L_n(u)du\overset{a.s}{\to}0\end{aligned}$$

from which (6) follows.

Corollary The confidence distribution of $\widehat{\Sigma}^{-1/2}(\theta-\widehat{\theta})$ is almost surely asymptotically normal; that is

$$P_{\theta_0}(\widehat{\Sigma}^{-1/2}(\theta-\widehat{\theta})\overset{d}{\to}N_k(\mathbf{0},\mathbf{I}))=1$$

where θ_0 is the true value of θ.

Proof: The multivariate analog of a theorem due to Scheffé (1947) yields

$$\int_{\widehat{\theta}+\widehat{\Sigma}_n^{\frac{1}{2}}t_2}^{\widehat{\theta}+\widehat{\Sigma}_n^{\frac{1}{2}}t_1}g_n(\theta)d\theta\overset{a.s}{\to}\int_{t_2}^{t_1}(2\pi)^{-k/2}\exp(-\|t\|^2/2)dt$$

where, $g_n(\theta)=w(\theta)L_n(\theta)/\int w(t)L_n(t)dt$.

REFERENCES

Brenner, D; Fraser, D.A.S., and McDunnough, P. (1982). On asymptotic normality of likelihood and conditional analysis. *Can. J. Statist.* **10**, 163–172.

Fraser, D.A.S., and McDunnough, P. (1984). Further remarks on asymptotic normality of likelihood and conditional analysis. *Can. J. Statist.* **12**, 183–190.

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

Johnstone, I.M. (1978). M.Sc. Thesis, Australian National University.

Scheffé, H. (1947). A useful convergence theorem for probability distributions. *Ann. Math. Statist.* **18**, 434–438.

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

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