Ann. Inst. Statist. Math.
Vol. 40, No. 2, 297–309 (1988)

ASYMPTOTIC EXPANSIONS OF POSTERIOR EXPECTATIONS, DISTRIBUTIONS AND DENSITIES FOR STOCHASTIC PROCESSES

MARTIN CROWDER

Department of Mathematics, University of Surrey, Guildford Surrey GU2 5XH, U.K.

(Received March 30, 1987; revised September 28, 1987)

Abstract. Asymptotic expansions are derived for Bayesian posterior expectations, distribution functions and density functions. The observations constitute a general stochastic process in discrete or continuous time.

Key words and phrases: Asymptotic expansions, Bayesian approach, inference for stochastic processes, asymptotic posterior normality.

1. Preliminaries

Suppose that Y_n represents data (univariate or multivariate, discrete or continuous) collected up to stage n. In discrete time Y_n will represent observations $\{y_1,\ldots, y_n\}$ and in continuous time Y_n will represent a trace $\{y_t: t\leq n\}$. Let $p(Y_n|\theta)$, the joint probability function or density of Y_n, be a function of a single continuous parameter θ, and let $L_{n\theta}=\ln p(Y_n|\theta)$ denote the log-likelihood function. Asymptotic ($n\to\infty$) expansions, in descending powers of $V_n=-L''_{n\hat{\theta}_n}$ (where dashes denote differentiation with respect to θ and $\hat{\theta}_n$ is the maximum likelihood estimator), will be derived for:

(i) $E_nq(\theta)=E[q(\theta)|Y_n]$, the posterior expectation of a given function $q(\theta)$ of θ;

(ii) $P_n(\xi)=P[\theta\leq\hat{\theta}_n+\xi V_n^{-1/2}|Y_n]$, the posterior distribution function of $V_n^{1/2}(\theta-\hat{\theta}_n)$;

(iii) $p_n(\xi)=dP_n(\xi)/d\xi$, the posterior density of $V_n^{1/2}(\theta-\hat{\theta}_n)$.

The prior distribution is $p(\theta)$, proper or improper, and the parameter space Θ is defined as $\{\theta: p(\theta)>0\}$. Only the "regular" case is considered here in which $L'_{n\hat{\theta}_n}=0$, $\hat{\theta}_n$ maximizes $L_{n\theta}$ globally and $\hat{\theta}_n$ is an interior point of Θ.

The work here is an extension of that of Johnson (1967, 1970) to which it owes much though from which significant departures have to be made. Johnson derives expansions for $P_n(\xi)$ in the i.i.d. Exponential family case (Johnson (1967)), in the general i.i.d. case (Johnson (1970)), and for data comprising a stationary Markov chain (Johnson (1970)). The latter paper also gives expansions for $E_n\theta^r$ for integer r, posterior percentiles of θ, and normalizing transformations of θ.

Walker (1969) shows that $P_n \to \Phi$, the standard normal distribution function, under certain conditions. Dawid (1970) relaxes these conditions to derive the result for a much wider class of distributions, including those where the range of y depends on θ; it is noted below that one of his novel features, in essence, comes close to one of the vital ingredients here. Borwanker *et al.* (1971) show that $P_n \to \Phi$ for data comprising a stationary Markov chain, give the corresponding result for $E_n\theta^r$, and apply their results to regular Bayes estimation. Heyde and Johnstone (1979) prove that $P_n \to \Phi$ for general stochastic processes. A related expansion is given in Lemma 1 of Bickel *et al.* (1985) in connection with the optimality of maximum likelihood estimators and Bayes' risks.

The notation and main results are given in Section 2, with details and proofs relegated to Section 4. Some illustrations, applications and discussion form the substance of Section 3.

The motivation for the present work has both a theoretical and a practical side. For the former, in trying to justify certain asymptotic results for inference with stochastic processes it was found that an order of magnitude was required for the errors. For instance, the previous works show that $|P_n - \Phi| = o_p(1)$ but provide no more explicit order except in the i.i.d. case of Johnson (1970). Also, an error bound was required for expectations $E_n q_n(\theta)$, not just for P_n. On the more practical side, evaluation of $E_n q(\theta)$ for various $q(\theta)$ of interest, e.g., in random effects nonlinear regression models, can involve intractable integrals. Progress is being made by various workers on the numerical evaluation of such integrals, and asymptotic estimates such as those here can make a contribution to this.

2. Main results

The expansions are derived as in Johnson (1970) by applying Laplace's method to integrals of type $\int h(\theta) \exp(L_{n\theta} - L_{n\hat{\theta}_n}) d\theta$. The quantities $E_n q(\theta)$, $P_n(\xi)$ and $p_n(\xi)$ are expressible as ratios of such integrals with $h(\theta) = p(\theta)$ or $p(\theta)q(\theta)$ as appropriate. In the integrand $L_{n\theta} - L_{n\hat{\theta}_n}$ is expanded as $-(1/2)V_n s^2 + V_n s^3 K_{ns}$ with $s = \theta - \hat{\theta}_n$; the first term determines the normal form of the integral as $V_n \to \infty$. The detailed problems arise from obtaining a suitable Taylor series for $h(\theta)\exp(V_n s^3 K_{ns})$ in s, and ensuring negligibility of extraneous and remainder terms. The results in the theorem below, where $M+1$ is the number of terms in the expansion, are quoted in terms of

$$J_{nM}(h, \xi V_n^{-1/2}) = \sum_{l=0}^{M} A_l(h, \xi) V_n^{-(l+1)/2} / l! \ ,$$

$$A_l(h, \xi) = \sum_{j=0}^{l} \binom{l}{j} K_{l-1,j}(h, 0) \mu_{l+2j}(\xi) \ ,$$

$$K_{lk}(h, s) = d^l\{h(\hat{\theta}_n + s)K_{ns}^k\}/ds^l ,$$

$$\mu_l(\xi) = (2\pi)^{-1/2}\int_{-\infty}^{\xi} t^l \exp\left(-\frac{1}{2}t^2\right)dt .$$

The conditions under which the results hold are as follows. All limits are taken as $n\to\infty$. Weak form statements are used throughout, convergence in probability being easier to verify than convergence almost surely. For instance, $V_n\to\infty$ in P_0-probability means that for any $K>0$, $P_0(V_n\le K)\to 0$, P_0 being probability determined by the true parameter θ_0. For a strong form of the theorem the conditions must be converted to strong form in the obvious way. C_n will denote a (possibly shrinking) neighbourhood of $\hat{\theta}_n$: $C_n=\{\theta:|\theta-\hat{\theta}_n|<c_n^{-1}\}\cap\Theta$ where c_n is non-decreasing; C_n is a stochastic neighbourhood centred on $\hat{\theta}_n$, and $\bar{C}_n=\Theta-C_n$. $L_{n\theta}^{(j)}$ is the j-th derivative of $L_{n\theta}$.

C1: $V_n\to\infty$ and limsup $c_n^2\ V_n^{-1}$ ln $V_n<1/(M+2)$ in P_0-probability. The second inequality restricts the growth rate of c_n; it holds automatically if $c_n\not\to\infty$.

C2: $L'_{n\hat{\theta}_n}=0$ and $P_0[|\hat{\theta}_n|\le K]\to 1$ for some $K<\infty$.

Thus the m.l.e. $\hat{\theta}_n$ is a turning point of $L_{n\theta}$, and is eventually bounded.

C3: $P_0[p^{(M+1)}(\theta)$ and $q^{(M+1)}(\theta)$ are continuous on $C_n]\to 1$.

Usually, C2 and C3 will hold because $\hat{\theta}_n \xrightarrow[P_0]{} \theta_0$ (consistency), and $p^{(M+1)}(\theta)$ and $q^{(M+1)}(\theta)$ are continuous in some (fixed) neighbourhood of θ_0.

C4: (a) $P_0[L_{n\theta}^{(M+1)}$ is continuous on $C_n]\to 1$;

(b) $P_0[|L'''_{n\theta}/V_n|\le K_1|\theta-\hat{\theta}_n|^{\delta_1-1}$ on $C_n]\to 1$ for some $\delta_1>0$, $K_1^{1/\delta_1}\le c_n$;

(c) $P_0[|(L_{n\theta}^{(j)}-L_{n\hat{\theta}_n}^{(j)})/V_n|\le K_3$ on $C_n]\to 1$ for $j=3,\ldots, M+3$, $K_3<\infty$.

These are the "core" conditions requiring good behaviour of $L_{n\theta}$-derivatives on C_n.

C5: For some $\delta_2\in[0, 1)$

(a) $P_0[\sup_{\bar{C}_n}(L_{n\theta}-L_{n\hat{\theta}_n})/\ln V_n\le-(1/2)(M+2)/(1-\delta_2)]\to 1$;

(b) $P_0[\int_{\bar{C}_n}|h(\theta)|\exp\{\delta_2(L_{n\theta}-L_{n\hat{\theta}_n})\}d\theta\le K_2]\to 1$ for some $K_2<\infty$, for $h(\theta)=p(\theta)$ and $p(\theta)q(\theta)$.

Negligibility of the integral outside C_n is ensured by C5. Also (a) implies that $L_{n\theta}$ is globally maximized within C_n.

THEOREM 2.1. *Under conditions* C1-C5, $|E_nq(\theta)-J_{nM}(qp, \infty)/J_{nM}(p, \infty)|$, $|P_n(\xi)-J_{nM}(p,\xi V_n^{-1/2})/J_{nM}(p,\infty)|$ *and* $|p_n(\xi)-(d/d\xi)J_{nM}(p,\xi V_n^{-1/2})/J_{nM}(p, \infty)|$ *are each* $O(V_n^{-(M+2)/2})$ *with* P_0*-probability*$\to 1$.

The proof of the theorem is given in Section 4. For $M=0$ the result of Heyde and Johnstone (1979) is recovered, plus the error of relative magnitude $V_n^{1/2}$. However, the extra information is bought with conditions on $L'''_{n\theta}$ in C4 rather than just on $L''_{n\theta}$.

3. Examples

The purpose here is briefly to illustrate how the conditions C1–C5 may be approached in applications. First, as a benchmark, the i.i.d. case can be dealt with by noting that Johnson's (1970) conditions imply those here; he has a proper prior $p(\theta)$, $q(\theta)=1$, c_n=constant (b/δ_2 in his paper) and V_n replaced by n, essentially.

Of the examples following, the second and third are of mildly pathological nature. There is a non-ergodic case, where the limiting distribution of V_n is non-degenerate, and a case of the shrinking interval. Such were part of the detective work of Sweeting (1980).

3.1 *An i.n.i.d. case: Poisson loglinear model*

The observations Y_i ($i=1,\ldots, n$) are independent Poisson with means λ_i satisfying $\ln \lambda_i=x_i\theta$, x_i being an explanatory variable. The log-likelihood and its derivatives are $L_{n\theta}=\sum_{i=1}^{n}[-\exp(x_i\theta)+x_iy_i\theta-\ln y_i!]$, $L'_{n\theta}=\sum_{i=1}^{n}[x_iy_i-x_i\exp(x_i\theta)]$, and $L^{(j)}_{n\theta}=-\sum_{i=1}^{n}x_i^j\exp(x_i\theta)$ for $j\geq 2$. Assuming that $\Sigma x_i^2=O(n)$, it is standard that $\hat{\theta}_n$ is consistent. Then $V_n=\Sigma x_i^2\exp(x_i\hat{\theta}_n)=O(n)$ and C1 holds if $c_n^2\ln n/n\to 0$, say. C2 and C3 hold if $p^{(M+1)}(\theta)$ and $q^{(M+1)}(\theta)$ are continuous on C_n. For C4 (a) is true, (b) holds because $|L'''_{n\theta}/V_n|=O(n)/O(n)$ for $\theta\in C_n$, and for (c) $|(L^{(j)}_{n\theta}-L^{(j)}_{n\hat{\theta}_n})/V_n|=O(n)/O(n)$ also. For C5(a) we have $L_{n\theta}-L_{n\hat{\theta}_n}=-\{(\theta-\hat{\theta}_n)^2/2\}\Sigma x_i^2\cdot\exp(x_i\theta_i)$ where $|\theta_i-\hat{\theta}_n|<|\theta-\hat{\theta}_n|$. Thus for $c_n^{-1}\leq|\theta-\hat{\theta}_n|\leq K$, $L_{n\theta}-L_{n\hat{\theta}_n}\leq -c_n^{-2}O(n)/2$; this continues to hold also for $|\theta-\hat{\theta}_n|>K$ since $L''_{n\theta}<0\ \forall\ \theta$ means that $L_{n\theta}-L_{n\hat{\theta}_n}$ is convex. Hence $(L_{n\theta}-L_{n\hat{\theta}_n})/\ln V_n\leq -c_n^{-2}(n/\ln n)$ for $\theta\in\bar{C}_n$, and this $\to-\infty$ (assumed above) which suffices for C5(a). Finally, C5(b) is a mild restriction on $h(\theta)$, just requiring that

$$\int_{|s|\geq c_n^{-1}}|h(\hat{\theta}_n+s)|\exp\{-\delta_2\Sigma e^{x_i\hat{\theta}_n}[e^{x_is}-1-x_is]\}ds\leq K_2\ ,$$

with P_0-probability$\to 1$.

3.2 *A non-ergodic process in discrete time*

Anderson (1959) investigated the autoregressive model $y_i=\theta y_{i-1}+e_i$ ($i=1, 2,\ldots$) where y_0 is fixed, i.e., the results are conditional upon y_0. He derived the asymptotic distributions of the least-squares estimator of θ when $|\theta|>1,=1$, and <1, and where the errors e_i are uncorrelated with mean 0 and variance σ^2 We will take $|\theta|>1$ and the e_i's as independent $N(0, 1)$. Then $L_{n\theta}=-(1/2)\,n\cdot\ln(2\pi)-(1/2)\sum_{i=1}^{n}(y_i-\theta y_{i-1})^2$, $L'_{n\theta}=\Sigma y_{i-1}(y_i-\theta y_{i-1})$, and $L''_{n\theta}=-\Sigma y_{i-1}^2=-V_n$. Let $U_n=\Sigma e_iY_{i-1}$, then $\hat{\theta}_n=\theta_0+U_n/V_n$ and $L_{n\theta}-L_{n\hat{\theta}_n}=-(1/2)V_n(\theta-\hat{\theta}_n)^2$. Anderson's results show that $\theta_0^{-2n}V_n$ is asymptoticaly distributed as $Z_n^2/\theta_0^2\ (\theta_0^2-1)$ where $Z_n\xrightarrow[D]{}N(\theta_0y_0, 1)$; hence $V_n\xrightarrow[P_0]{}\infty$ (for any y_0) and C1 is satisfied if $nc_n^2\theta_0^{-2n}\to 0$. Also

$\hat{\theta}_n$ is consistent, so C2 and C3 hold provided $p^{(M+1)}(\theta)$ and $q^{(M+1)}(\theta)$ are continuous. Condition C4 is met trivially because $L_{n\theta}^{(j)}$ is zero for $j \geq 3$. For C5(a) we have, for $\theta \in \bar{C}_n$,

$$(L_{n\theta} - L_{n\hat{\theta}_n})/\ln V_n = -\frac{1}{2}(\theta - \hat{\theta}_n)^2 V_n/\ln V_n \leq -\frac{1}{2} c_n^{-2} V_n/\ln V_n ,$$

so the condition holds if $nc_n^2\theta_0^{-2n} \to 0$ as for C1. Part (b) of C5 requires only that

$$\int_{|s| \geq c_n^{-1}} |h(\hat{\theta}_n + s)| \exp\left\{ -\frac{1}{2}\delta_2 V_n s^2 \right\} ds \leq K_2 ,$$

with P_0-probability$\to 1$.

3.3 *A process in continuous time*

In a nonhomogeneous Poisson process let Y_n be the number of events occurring during $(0, n]$ with rate function $\lambda_{n\theta} = \theta(1+n)^{\theta-1}$, where $\theta > 0$. Thus Y_n is Poisson with mean $\mu_{n\theta} = (1+n)^\theta - 1$ and $L_{n\theta} = -\mu_{n\theta} + Y_n \ln \mu_{n\theta} - \ln Y_n!$, $\mu_{n\hat{\theta}_n} = Y_n$, $\hat{\theta}_n = \ln(1+Y_n)/\ln(1+n)$, and $V_n = [\ln(1+n)]^2 (Y_n+1)^2/Y_n$. Since $\theta_0 > 0$, $\mu_{n0} \to \infty$, $Y_n/\mu_{n0} \xrightarrow[P_0]{} 1$ and $\hat{\theta}_n \xrightarrow[P_0]{} \theta_0$. Also, assuming continuity of $p^{(M+1)}(\theta)$ and $q^{(M+1)}(\theta)$ on $\theta > 0$, conditions C2 and C3 are met.

Condition C4(a) holds for each M. For C4(b) we have

$$L_{n\theta}''' = [\ln(1+n)]^3 (\mu_{n\theta} + 1)\{-1 + Y_n(\mu_{n\theta} + 2)/\mu_{n\theta}^3\} , \tag{3.1}$$

using $\mu_{n\theta}^{(j)} = (\mu_{n\theta} + 1)[\ln(1+n)]^j$ for $j \geq 1$. It follows that

$$|(\theta - \hat{\theta}_n)^{1-\delta_1} L_{n\theta}'''/V_n| \underset{P_0}{\sim} |s|^{1-\delta_1}[\ln(1+n)](1+n)^s .$$

Taking $\delta_1 = 1/2$ and $c_n = [\ln(1+n)]^2$, this last expression is bounded by $\exp\{[\ln(1+n)]^{-1}\} \to 1$ as required by C4(b). Similarly for condition C4(c) with $j=3$ we have from (3.1) $|L_{n\theta}''' - L_{n\hat{\theta}_n}'''|/V_n \underset{P_0}{\sim} [\ln(1+n)]|1-(1+n)^s| < 1$ on C_n as required. Higher derivatives may be checked similarly.

Condition C1 is met with $c_n = [\ln(1+n)]^2$ since $c_n^2 V_n^{-1} \ln V_n \xrightarrow[P_0]{} 0$.

For condition C5 we have $L_{n\theta} - L_{n\hat{\theta}_n} = -(Y_n+1)a(u) + b_n(u)$ where $u = s \cdot \ln(1+n)$, $a(u) = e^u - 1 - u$, and $b_n(u) = -u + Y_n \ln[(Y_n + 1 - e^{-u})/Y_n]$. Since $u > -\ln(Y_n+1)$ necessarily, and $e^{-u} > 0$, $b_n(u) < \ln(Y_n+1) + Y_n \ln(1+Y_n^{-1})$. Also $\inf_{\bar{C}_n} a(u) \sim (1/2) c_n^{-2}$. Hence $\sup_{\bar{C}_n} (L_{n\theta} - L_{n\hat{\theta}_n})/\ln V_n \xrightarrow[P_0]{} -\infty$ and C5(a) is satisfied. For C5(b) note that $L_{n\theta} - L_{n\hat{\theta}_n} \underset{P_0}{\sim} -(1+n)^{\theta_0} a(u)$ on $\bar{C}_n$, with $a(u) = (1+n)^s - 1 - s \cdot \ln(1+n)$, so the integrability constraint on $h(\theta)$ is very mild in practice.

Note finally that Heyde and Johnstone's (1979) condition A5 fails here because $L''_{n\theta}/V_n+1 \underset{P_0}{\sim} -(1+n)^{\theta-\hat{\theta}_n} \underset{P_0}{\nrightarrow} 0$ unless a shrinking interval with $c_n^{-1} \cdot \ln(1+n) \to 0$ is used.

4. Details and proofs

A sequence of lemmas is given, leading up to the proof of Theorem 2.1. To cut clutter, the conditions in the lemmas are given in non-stochastic form. The "integral" referred to throughout this section is

$$I_n(h, \xi) = (2\pi)^{-1/2} \int_{-\infty}^{\hat{\theta}_n+\xi} h(\theta)\exp(L_{n\theta} - L_{n\hat{\theta}_n})d\theta$$
$$= (2\pi)^{-1/2} \int_{-\infty}^{\xi} h(\hat{\theta}_n + s)\exp\left(-\frac{1}{2}V_n s^2 + V_n s^3 K_{ns}\right)ds ,$$

where $s=\theta-\hat{\theta}_n$.

Lemma 4.1 gives the basic Taylor expansion of the integrand on C_n, and Lemma 4.2 shows that the integral over $\bar{C}_n$ is negligible. Lemma 4.3 assembles the form of the expansion using the results of Lemmas 4.1 and 4.2. Lemmas 4.4 and 4.5 convert the conditions on $L_{n\theta}^{(j)}$ into the bound on $K_{M+1,k}(h, s)$ crucial in the proof of Lemma 4.1; they represent a major component of the extra work required in generalizing from the i.i.d. case (Johnson (1970)) and from the case $M=0$ (Heyde and Johnstone (1979)). Lemma 4.6 fills in some details between Lemma 4.3 and the theorem to tidy up the proof of the latter.

LEMMA 4.1. *Taylor expansion of integrand on* $C_n=\{\theta:|\theta-\hat{\theta}_n|<c_n^{-1}\} \cap \Theta$. *Assume*

(i) $V_n\to\infty$, $|\hat{\theta}_n|\leq K$, *and* $h^{(M+1)}(\theta)$ *continuous on* C_n;

(ii) *For* $\theta \in C_n$, $n\geq N_1$, $K_1<\infty$ (a) $L_{n\theta}^{(M+1)}$ *is continuous*, (b) $|L_{n\theta}'''/V_n| \leq K_1|\theta-\hat{\theta}_n|^{\delta_1-1}$ *for some* $\delta_1>0$, (c) $|(L_{n\theta}^{(j)}-L_{n\hat{\theta}_n}^{(j)})/V_n|\leq K_3$ *for* $j=3,\ldots, M+3$. *Then, for each* N *and* $\theta \in C_n$,

$$h(\hat{\theta}_n + s)\exp(L_{n\theta} - L_{n\hat{\theta}_n})$$
$$= \exp\left(-\frac{1}{2}V_n s^2\right)\left\{\sum_{j=0}^{M} \sum_{k=0}^{N} K_{jk}(h, 0)V_n^k s^{j+3k}/j!k! + R_{ns}\right\} ,$$

where

$$s = \theta - \hat{\theta}_n, \quad K_{jk}(h, s) = d^j\{h(\hat{\theta}_n + s)K_{ns}^k\}/ds^j,$$

$$K_{ns} = \left(L_{n\theta} - L_{n\hat{\theta}_n} + \frac{1}{2}V_n s^2\right)/V_n s^3 ,$$

and

$$(4.1)\qquad |R_{ns}| < K_{MN}|s|^{M+1}\sum_{k=0}^{N}(V_ns^2)^k/k! + H_1|V_ns^{2+\delta_1}|^{N+1}\exp\left(\frac{1}{6}V_ns^2\right),$$

for some $H_1<\infty$ *and* $K_{MN}<\infty$, *and provided* $c_n\le K_1^{1/\delta_1}$.

PROOF. First, for each N and some $\alpha_1\in(0,1)$,

$$(4.2)\qquad \exp\left(L_{n\theta}-L_{n\hat\theta_n}+\frac{1}{2}V_ns^2\right)=\exp(V_ns^3K_{ns})$$
$$=\sum_{k=0}^{N}[V_ns^3K_{ns}]^k/k! + [V_ns^3K_{ns}]^{N+1}\exp(\alpha_1V_ns^3K_{ns})/(N+1)!\,.$$

Now, from (i) and (ii)(a) we have the Taylor expansion

$$(4.3)\qquad h(\hat\theta_n+s)K_{ns}^k=\sum_{j=0}^{M}K_{jk}(h,0)s^j/j! + K_{M+1,k}(h,\alpha_{2k}s)s^{M+1}/(M+1)!\,,$$

for some $\alpha_{2k}\in(0,1)$, $k=0,1,\ldots,N$. Then from (4.2) and (4.3)

$$(4.4)\qquad |R_{ns}|=\left|h(\hat\theta_n+s)\exp(V_ns^3K_{ns})-\sum_{j=0}^{M}\sum_{k=0}^{N}K_{jk}(h,0)s^j(V_ns^3)^k\Big/j!k!\right|$$
$$=\Bigg|\sum_{k=0}^{N}(V_ns^3)^kK_{M+1,k}(h,\alpha_{2k}s)s^{M+1}/k!(M+1)!$$
$$+h(\hat\theta_n+s)[V_ns^3K_{ns}]^{N+1}\exp(\alpha_1V_ns^3K_{ns})/(N+1)!\Bigg|\,.$$

But, by Lemmas 4.4 and 4.5, (i) and (ii)(c) imply $|s^kK_{M+1,k}(h,\alpha_{2k}s)|\le K'_{MN}<\infty$ for $k=0,\ldots,N$. Also, (i) implies $|h(\hat\theta_n+s)|\le H<\infty$ on C_n, and (ii)(b) implies

$$|K_{ns}|=\left|\frac{1}{6}s^3L'''_{n\theta_1}/V_ns^3\right|\le\frac{1}{6}K_1|s|^{\delta_1-1}\,,$$

for $|\theta_1-\hat\theta_n|<c_n^{-1}$ and $|s|<c_n^{-1}$. Hence (4.4) is bounded by

$$\sum_{k=0}^{N}(V_ns^2)^kK'_{MN}|s|^{M+1}/k!(M+1)!$$
$$+H\left|\frac{1}{6}K_1V_ns^{2+\delta_1}\right|^{N+1}\exp(\alpha_1V_n|s^3K_{ns}|)/(N+1)!\,.$$

Finally, (4.1) follows by noting that $\alpha_1V_n|s^3K_{ns}|<(1/6)K_1|s|^{\delta_1}V_ns^2\le(1/6)V_ns^2$ for $|s|<c_n^{-1}$. □

The shrinking neighbourhood, C_n here, is used, in slightly different forms, by Sweeting ((1980), condition C2), Dawid ((1970), condition C12), and Brown (1985). Heyde and Johnstone ((1979), condition A5) use a non-shrinking neighbourhood (with $c_n=\delta$).

$K_{jk}(h,\ 0)$ is expressible as $[\partial^{j+k}\{h(\hat{\theta}_n+s)\exp(wK_{ns})\}/\partial s^j\partial w^k]_{s=w=0}$. This corresponds to the form used by Johnson ((1970), equation 2.16). In the expansion there the theory of functions of two complex variables is applied but an explicit bound for the error, R_{ns} here, is not given.

LEMMA 4.2. *Behaviour of the integral on* $\bar{C}_n$.
Assume: *There exists* $\delta_2\in[0,\ 1)$ *s.t. for* $n\geq N_2$

(i) $\sup_{\bar{C}_n}(L_{n\theta}-L_{n\hat{\theta}_n})/\ln V_n\leq-(1/2)\,(M+2)/(1-\delta_2)$,

(ii) $\int_{\bar{C}_n}|h(\theta)|\exp[\delta_2(L_{n\theta}-L_{n\hat{\theta}_n})]d\theta\leq K_2$.

Then $\int_{\bar{C}_n}|h(\theta)|\exp(L_{n\theta}-L_{n\hat{\theta}_n})d\theta\leq K_2 V_n^{-(M+2)/2}$ *for* $n\geq N_2$.

PROOF. From (i) $(L_{n\theta}-L_{n\hat{\theta}_n})\leq\delta_2(L_{n\theta}-L_{n\hat{\theta}_n})^{-(M+2)/2}\ln V_n$.
Hence

$$\int_{\bar{C}_n}|h(\theta)|\exp(L_{n\theta}-L_{n\hat{\theta}_n})d\theta\leq\int_{\bar{C}_n}|h(\theta)|\exp[\delta_2(L_{n\theta}-L_{n\hat{\theta}_n})]\cdot\exp\left[-\frac{1}{2}(M+2)\ln V_n\right]d\theta$$

$$\leq K_2 V_n^{-(M+2)/2}.\qquad\square$$

Condition (i) is akin to Wolfowitz' (1949) property which he proves for the i.i.d. case in his comment accompanying Wald's (1949) classic note on consistency of the m.l.e. Other relatives of (i) appear in Johnson ((1970), Lemma 2.3), Dawid ((1970), condition C7), and Heyde and Johnstone ((1979), condition A4).

Condition (ii) may fail for $n<N_2$ because the factor $\exp(L_{n\theta}-L_{n\hat{\theta}_n})$ in the integrand does not dock the tails of $h(\theta)=p(\theta)q(\theta)$ heavily enough for convergence. This can occur when $p(\theta)$ is improper and $q(\theta)$ has heavy tails. In (i) and (ii) a fraction δ_2 of $(L_{n\theta}-L_{n\hat{\theta}_n})$ is used to discipline the integral, leaving the other $(1-\delta_2)$ to achieve the bound. On the other hand, with a proper prior (ii) may often hold with $\delta_2=0$; this will be true when $E_0q(\theta)$, the prior expectation of $q(\theta)$, is finite.

LEMMA 4.3. *Let the assumptions of Lemmas* 4.1 *and* 4.2 *be met with* $c_n\geq K_1^{1/\delta_1}$ *and* $\limsup(c_n^2V_n^{-1}\ln V_n)<1/(M+2)$. *Then*

$$I_n(h,\ \xi_nV_n^{-1/2})=\sum_{l=0}^{M}A_l(h,\ \xi_n)V_n^{-(l+1)/2}/l!\ +\ O(V_n^{-(M+2)/2})\ ,$$

uniformly in $|\xi_n|\leq c_n^{-1}\ V_n^{1/2}$ *where*

$$A_l(h, \xi) = \sum_{j=0}^{l} \binom{l}{j} K_{l-j,j}(h, 0)\mu_{l+2}(\xi) ,$$
$$\mu_l(\xi) = (2\pi)^{-1/2} \int_{-\infty}^{\xi} t^l \exp\left(-\frac{1}{2} t^2\right) dt .$$

PROOF. From Lemma 4.1, using the bound for R_{ns},

$$(4.5) \qquad I_n(h, \xi_n V_n^{-1/2}) = I_n(h, -c_n^{-1}) + (2\pi)^{-1/2} \int_{-c_n^{-1}}^{\xi_n V_n^{-1/2}} \exp\left(-\frac{1}{2} V_n s^2\right) \cdot \left\{ \sum_{j=0}^{M} \sum_{k=0}^{N} K_{jk}(h, 0)\, V_n^k s^{j+3k}/j!k! + R_{ns} \right\} ds ,$$

and

$$\left| \int_{-c_n^{-1}}^{\xi_n V_n^{-1/2}} \exp\left(-\frac{1}{2} V_n s^2\right) R_{ns} ds \right| = O(V_n^{-(M+2)/2} + V_n^{-1/2-\delta_1(N+1)/2}) ,$$

uniformly in ξ_n; this bound reduces to $O(V_n^{-(M+2)/2})$ on taking $N \geq (M+1)/\delta_1 - 1$. By Lemma 4.2, $I_n(h, -c_n^{-1}) = O(V_n^{-(M+2)/2})$. Also

$$(2\pi)^{-1/2} \int_{-c_n^{-1}}^{\xi_n V_n^{-1/2}} \exp\left(-\frac{1}{2} V_n s^2\right) s^{j+3k} ds = V_n^{-(j+3k+1)/2} \{\mu_{j+3k}(\xi_n) - \mu_{j+3k}(-c_n^{-1} V_n^{1/2})\} ,$$

and

$$|\mu_{j+3k}(-c_n^{-1} V_n^{1/2})| \leq (2\pi)^{-1/2} \int_{-\infty}^{-c_n^{-1} V_n^{1/2}} |t|^{j+3k} \exp\left(-\frac{1}{2} t^2\right) dt$$
$$\leq \exp\left[-\frac{1}{2}(1-\alpha) c_n^{-2} V_n\right] (2\pi)^{-1/2} \int_{-\infty}^{-c_n^{-1} V_n^{1/2}} |t|^{j+3k} \exp\left(-\frac{1}{2} \alpha t^2\right) dt$$
$$\leq V_n^{-(M+2)/2} O(1) ,$$

where $(M+2) c_n^2 V_n^{-1} \ln V_n < 1 - \alpha < 1$ for sufficiently large n. Hence (4.5) yields

$$I_n(h, \xi_n V_n^{-1/2}) = O(V_n^{-(M+2)/2}) + \sum_{j=0}^{M} \sum_{k=0}^{N} [K_{jk}(h, 0)/j!k!]\, V_n^k V_n^{-(j+3k+1)/2} \mu_{j+3k}(\xi_n) ,$$

and the result is obtained by collecting summation terms of like magnitude. □

The maximum allowable shrinkage rate for C_n, involving $\sqrt{\ln V_n / V_n}$, appears in Dawid ((1970), condition C12) with $M=0$ and $V_n = O(n)$ for the i.i.d. case. It is interesting that this rate emerges inevitably in the proof of Lemma 4.3 here, for general stochastic processes, whereas Dawid's purpose is quite different, namely to deal with the awkward nonregular case where the range of the observations depends on θ.

LEMMA 4.4. *With notation as in Lemma* 4.1, *the condition*

$$\sup_{C_n} |s^k K_{M+1,k}(h, s)| \le K_k < \infty \quad \text{for} \quad k = 1, 2, \dots$$

holds if (i) $h^{(M+1)}(\theta)$ *is continuous on* C_n, *and* (ii) $\sup_{C_n} |sK_{ns}^{(j)}| \le K_0 < \infty$ *for* $j=0,\dots, M+1$.

PROOF. By Leibnitz' formula

$$(4.6) \qquad \begin{aligned} |s^k K_{M+1,k}(h, s)| &= \left| s^k \sum_{j=0}^{M+1} \binom{M+1}{j} h^{(M+1-j)}(\hat{\theta}_n + s)(\partial^j K_{ns}^k / \partial s^j) \right| \\ &\le \sum_{j=0}^{M+1} \binom{M+1}{j} H |s^k \partial^j K_{ns} / \partial s^j| \ , \end{aligned}$$

since (i) implies $|h^{(M+1-j)}(\hat{\theta}_n + s)| \le H$ for $|s| \le c_n^{-1}$ and $j=0,\dots, M+1$. Assume, for induction, that (ii) implies $|s^l \partial^j K_{ns}^l / \partial s^j| \le 2^{(l-1)(M+1)} K_0^l$ for $l=1,\dots, k-1, j=0,\dots, M+1$, $|s| \le c_n^{-1}$; this is obviously true for $k=2$. Applying Leibnitz' again,

$$\begin{aligned} |s^k \partial^j K_{ns}^k / \partial s^j| &= \left| \sum_{i=0}^{j} \binom{j}{i} (sK_{ns}^{(j-i)})(s^{k-1} \partial^i K_{ns}^{k-1} / \partial s^i) \right| \\ &\le \sum_{i=0}^{j} \binom{j}{i} K_0 2^{(k-2)(M+1)} K_0^{k-1} < 2^{(k-1)(M+1)} K_0^k \ . \end{aligned}$$

Hence (4.6) becomes

$$|s^k K_{M+1,k}(h, s)| < \sum_{j=0}^{M+1} \binom{M+1}{j} 2^{(k-1)(M+1)} H K_0^k = 2^{k(M+1)} H K_0^k \ ,$$

as required. □

LEMMA 4.5. *With notation as in Lemma* 4.1

$$\sup_{C_n} |sK_{ns}^{(j)}| \le 2^{j-1} \sup_{C_n} |L_{n\theta}^{(j+2)} - L_{n\hat{\theta}_n}^{(j+2)}| / V_n \quad \text{for} \quad j \ge 1 \ .$$

PROOF. By Leibnitz' formula

$$(4.7) \qquad \begin{aligned} V_n K_{ns}^{(j)} &= d^j \{ V_n / 2s + (L_{n\theta} - L_{n\hat{\theta}_n}) / s^3 \} / ds^j \\ &= \frac{(-1)^j j! V_n}{2s^{j+1}} + \sum_{l=0}^{j} \binom{j}{l} L_{n\theta}^{(l)} \frac{(j-l+2)!(-1)^{j-l}}{2s^{j-l+3}} \end{aligned}$$

$$- L_{n\hat{\theta}_n} \frac{(j+2)!(-1)^j}{2s^{j+3}} .$$

The middle term of (4.7), after Taylor expansion to suitably chosen order, becomes

$$\text{(4.8)} \quad \sum_{l=0}^{j} \binom{j}{l} \frac{(j-l+2)!(-1)^{j-l}}{2s^{j-l+3}} \left\{ \sum_{k=0}^{j-l+1} \frac{s^k}{k!} L_{n\hat{\theta}_n}^{(l+k)} + \frac{s^{j-l+2}}{(j-l+2)!} L_{n\theta_l}^{(j+2)} \right\}$$

$$= \frac{1}{2} \sum_{v=0}^{j} \sum_{u=0}^{v} \binom{j}{v-u} \frac{(j-v+u+2)!(-1)^{j-v+u}}{u!s^{j-v+3}} L_{n\hat{\theta}_n}^{(v)}$$

$$+ \frac{1}{2s^2} \sum_{l=0}^{j} \binom{j}{l} (-1)^{j-l}(j-l+2) L_{n\hat{\theta}_n}^{(j+1)} + \frac{1}{2s} \sum_{l=0}^{j} \binom{j}{l} (-1)^{j-l} L_{n\theta_l}^{(j+2)} ,$$

where $|\theta_l - \hat{\theta}_n| < |\theta - \hat{\theta}_n|$ for each l, and the summation $\sum_{l=0}^{j} \sum_{k=0}^{j-l}$ has been rearranged as $\sum_{v=0}^{j} \sum_{u=0}^{v}$ with $v=l+k$, $u=k$. For $j>1$ the middle term in (4.8) is zero and, since $\sum_{u=0}^{v} \binom{v}{u} (-1)^u u^r = 0$ for $v>r$ (Feller (1971), Chapter 7, equation (1.7) with $i=0$, $a_j = j^s$, $0 \le s < r$), the first term is

$$\text{(4.9)} \qquad \frac{1}{2} \sum_{v=0}^{j} \frac{j!(-1)^{j-v}}{s^{j-v+3} v!} L_{n\hat{\theta}_n}^{(v)} \left\{ \sum_{u=0}^{v} \binom{v}{u} (-1)^u (j-v+u+1)(j-v+u+2) \right\}$$

$$= \frac{(-1)^j}{2s^{j+3}} \{(j+2)! L_{n\hat{\theta}_n} + j! s^2 L''_{n\hat{\theta}_n}\} .$$

For $j=1$ the first and second terms of (4.8) together yield (4.9). Hence, using (4.8) and (4.9), (4.7) becomes

$$V_n K_{ns}^{(j)} = \frac{(-1)^j j! V_n}{2s^{j+1}} + \frac{(-1)^j}{2s^{j+3}} \{(j+2)! L_{n\hat{\theta}_n} + j! s^2 L''_{n\hat{\theta}_n}\}$$

$$+ \frac{(-1)^j}{2s} \sum_{l=0}^{j} a_{jl} L_{n\theta_l}^{(j+2)} - L_{n\hat{\theta}_n} \frac{(j+2)!(-1)^j}{2s^{j+3}}$$

$$= \frac{(-1)^j}{2s} \sum_{l=0}^{j} a_{jl} L_{n\theta_l}^{(j+2)} ,$$

where $a_{jl} = \binom{j}{l} (-1)^l$, so $\sum_{l=0}^{j} a_{jl} = 0$ for $j \ge 1$. Thus, finally,

$$|sK_{ns}^{(j)}| = \left| (1/2) V_n^{-1} \sum_{l=0}^{j} a_{jl} (L_{n\theta_l}^{(j+2)} - L_{n\hat{\theta}_n}^{(j+2)}) \right|$$

$$\leq (1/2) V_n^{-1} \sum_{l=0}^{j} |a_{jl}| \sup_{C_n} | L_{n\theta}^{(j+2)} - L_{n\hat{\theta}_n}^{(i+2)} | \ . \qquad \square$$

LEMMA 4.6. *Under the assumptions and notation of Lemma* 4.3

(a) $I_n(h, \infty) = \sum_{l=0}^{M} A_l(h, \infty) V_n^{-(l+1)/2} / l! + O(V_n^{-(M+2)/2})$,

(b) $\partial I_n(h, \xi V_n^{-1/2}) / \partial \xi = \sum_{l=0}^{M} [\partial A_l(h, \xi) / \partial \xi] V_n^{-(l+1)/2} / l! + O(V_n^{-(M+2)/2})$.

PROOF. (a) Take $\xi_n = c_n^{-1} V_n^{1/2}$. Then, as in the proof of Lemma 4.3, $\mu_j(\infty) - \mu_j(\xi_n) = O(V_n^{-(M+2)/2})$. Hence $A_l(h, \infty) - A_l(h, \xi_n) = O(V_n^{-(M+2)/2})$. It follows from Lemma 4.3 that $I_n(h, \xi_n V_n^{-1/2}) = \sum_{l=0}^{M} A_l(h, \infty) V_n^{-(l+1)/2} / l! + O(V_n^{-(M+2)/2})$. Finally, by Lemma 4.2, $I_n(h, \infty) - I_n(h, \xi_n V_n^{-1/2}) = O(V_n^{-(M+2)/2})$.

(b) It has to be verified that differentiation with respect to ξ does not affect the order of magnitude of the error term in Lemma 4.3 with $\xi_n = \xi$, i.e., that

$$\frac{\partial}{\partial \xi} \int_{c_n^{-1}}^{\xi V_n^{-1/2}} \exp\left(-\frac{1}{2} V_n s^2\right) R_{ns} ds = O(V_n^{-(M+2)/2}) \ ,$$

for $|\xi| \leq c_n^{-1} V_n^{1/2}$. But the left hand side is bounded by

$$\begin{aligned} \left| e^{-\xi^2/2} R_{n,\xi V_n^{-1/2}} \right| &< \exp\left(-\frac{1}{2}\xi^2\right) \Big\{ K_{MN} |\xi V_n^{-1/2}|^{M+1} \sum_{k=0}^{N} \xi^{2k} / k! \\ &\qquad + H_1 | V_n (\xi V_n^{-1/2})^{2+\delta_1} |^{N+1} \exp(\xi^2 / 6) \Big\} \\ &= O(V_n^{-(M+2)/2} + V_n^{-\delta_1 (N+1)/2}) \ , \end{aligned}$$

since $\sup |\exp(-\xi^2/6) \xi^a| = O(1)$ for $a > 0$; also the choice $N \geq (M+2)/\delta_1 - 1$ ensures the required order of magnitude. $\square$

PROOF OF THEOREM 2.1. It is easily verified that conditions C1–C5 are just the stochastic versions of those of Lemma 4.3, incorporating those of Lemmas 4.1 and 4.2. The three results are proved as follows:

$$\begin{aligned} 1. \quad E_n q(\theta) &= \int_{-\infty}^{\infty} q(\theta) p(\theta | Y_n) d\theta = \int_{-\infty}^{\infty} q(\theta) p(\theta) p(Y_n | \theta) d\theta / p(Y_n) \\ &= \int_{-\infty}^{\infty} q(\theta) p(\theta) \exp(L_{n\theta}) d\theta / \int_{-\infty}^{\infty} p(\theta) \exp(L_{n\theta}) d\theta \\ &= I_n(qp, \infty) / I_n(p, \infty) \ . \end{aligned}$$

The result now follows from Lemma 4.6(a).

$$2. \quad P_n(\xi) = \int_{-\infty}^{\hat{\theta}_n + \xi V_n^{-1/2}} p(\theta | Y_n) d\theta = \int_{-\infty}^{\hat{\theta}_n + \xi V_n^{-1/2}} p(\theta) p(Y_n | \theta) d\theta / p(Y_n)$$

$$=\int_{-\infty}^{\hat{\theta}_n+\xi V_n^{-1/2}} p(\theta)\exp(L_{n\theta})d\theta \Big/ \int_{-\infty}^{\infty} p(\theta)\exp(L_{n\theta})d\theta$$
$$=I_n(p, \xi V_n^{-1/2})/I_n(p, \infty)\,.$$

The result follows from application of Lemma 4.3 with $\xi_n=\xi$ to the numerator, and Lemma 4.6(a) to the denominator.

3. The fact that $p_n(\xi)=dP_n(\xi)/d\xi$ has the indicated expansion with error $O(V_n^{-(M+2)/2})$ follows from Lemma 4.6(b). □

REFERENCES

Anderson, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations, *Ann. Math. Statist.*, **30**, 676–687.

Bickel, P. J., Gotze, F. and Van Zwet, W. R. (1985). A simple analysis of third-order efficiency of estimates, *Proc. Berk. Conf. in Honour of Jerzey Neyman and Jack Kiefer*, Vol. II, 749–768.

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

Brown, B. M. (1985). Multiparameter linearization theorems, *J. Roy. Statist. Soc. Ser. B*, **47**, 323–331.

Dawid, A. P. (1970). On the limiting normality of posterior distributions, *Proc. Camb. Phil. Soc.*, **67**, 625–633.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. II, 2nd ed., Wiley, New York.

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

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

Johnson, R. A. (1970). Asymptotic expansions associated with posterior distributions, *Ann. Math. Statist.*, **41**, 851–864.

Sweeting, T. J. (1980). Uniform asymptotic normality of the maximum likelihood estimator, *Ann. Statist.*, **8**, 1375–1381.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate, *Ann. Math. Statist.*, **20**, 595–601.

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

Wolfowitz, J. (1949). On Wald's proof of the consistency of the maximum likelihood estimate, *Ann. Math. Statist.*, **20**, 601–602.