The Annals of Statistics
1992, Vol. 20, No. 1, 580–589

ASYMPTOTIC ANCILLARITY AND CONDITIONAL INFERENCE FOR STOCHASTIC PROCESSES

BY TREVOR J. SWEETING

University of Surrey

Simple conditions on the observed information ensure asymptotic normality of the conditional distributions of the randomly normed score statistic and maximum likelihood estimator given a suitable asymptotically ancillary statistic. In particular, asymptotic normality holds conditional on any asymptotically ancillary statistic asymptotically equivalent to observed information. The results apply to inference from a general stochastic process and are of particular relevance in the case of nonergodic models.

1. Introduction. Under certain conditions it is shown that the conditional distributions of the randomly normed score statistic and maximum likelihood estimator, given any asymptotically ancillary statistic which contains data asymptotically equivalent to observed information, are asymptotically normal. These results are particularly relevant in the case of nonergodic models for dependent data. Efron and Hinkley (1978) discuss approximate ancillarity and conditionality in relation to observed information for i.i.d. observations. In Sweeting (1986) conditional limit results are obtained for a supercritical branching process. The mode of conditional convergence used in that paper is uniform convergence in compact subsets with respect to the conditioning variable, which leads to considerations of equicontinuity difficult to generalize to arbitrary models. In the present paper we use a weaker mode of *probability* conditional convergence; see Sweeting (1989) for a general discussion.

In Section 2 we formulate the problem and discuss the importance of general conditional results of this type. Regularity conditions are stated, the asymptotic ancillarity concept is introduced and the main results presented in Section 3. Proofs are given in Section 4.

2. Formulation of the conditional problem. Let Θ be an open subset of $\mathbb{R}^k$, $k \geq 1$, and $(P_\theta, \theta \in \Theta)$ be a family of probability measures on a measurable space $(\Omega, \mathscr{B})$. Let $\mathscr{B}_t \subset \mathscr{B}$ be a nondecreasing sequence of sub σ-fields, where $t > 0$ is a discrete or continuous parameter and let $P_{t\theta}$ be the restriction of P_θ to $\mathscr{B}_t$. Assume that, for each t and $\theta \in \Theta$, $P_{t\theta} \ll \lambda_t$, a σ-finite measure, with density $p_t(\theta)$ such that the set $\{p_t(\theta) > 0\}$ is independent of θ. Then, without loss of generality, λ_t can be assumed to be supported on this set and hence $l_t(\theta) = \log p_t(\theta)$ exists a.e. (λ_t). Further assume that the second-

Received May 1988; revised July 1991.

AMS 1980 *subject classifications*. Primary 62F12; secondary 62M99.

Key words and phrases. Asymptotic conditional inference, asymptotic ancillarity, nonergodic models, maximum likelihood estimator, score statistic.

order partial derivatives of $p_t(\theta)$ exist in a neighbourhood of the true value $\theta_0 \in \Theta$. It may be possible to extend the ideas in this paper to more general settings, not requiring the existence of dominating measures λ_t, or using weaker differentiability assumptions, for example, but this has not been investigated.

Let M_k be the space of all $k \times k$ matrices. The norm in M_k will be $|A| = (\operatorname{tr} A^T A)^{1/2}$. $A^{1/2}(A^{T/2})$ will denote the left (right) Cholesky square root (CSR) of the positive definite matrix A. Use of the CSR leads to tidier results than the symmetric square root; see, for example, Fahrmeir (1988). Properties of the CSR are used in the sequel without further comment.

Let $U_t(\theta) = l_t'(\theta)$ be the $k \times 1$ score statistic, $l_t''(\theta)$ the $k \times k$ matrix of second-order derivatives and $J_t(\theta) = -l_t''(\theta)$ the observed information matrix. Define $X_t(\theta) = B_t^{-1/2}U_t(\theta)$, $W_t(\theta) = B_t^{-1/2}J_t(\theta)B_t^{-T/2}$, where (B_t) is a sequence of positive definite matrices in M_k satisfying $B_t^{-1} \to 0$; and write $X_t = X_t(\theta_0)$, $W_t = W_t(\theta_0)$. It was shown in Sweeting (1980) that, under suitable conditions on $W_t(\theta)$, $(X_t, W_t) \Rightarrow (W^{1/2}Z, W)$ under θ_0, where Z is a standard normal random variable independent of W. (Here $\Rightarrow$ denotes weak convergence.) Using properties of the CSR and the continuous mapping theorem, we obtain $(Z_t, W_t) \Rightarrow (Z, W)$ under θ_0, where $Z_t = \{J_t(\theta_0)\}^{-1/2}U_t(\theta_0)$. The convergence in distribution of W_t suggests that we could regard W_t as asymptotically ancillary for θ. If Z_t retains its asymptotic normal distribution conditionally on W_t, then the conditionality principle would dictate basing inferences on this distribution. A similar argument applies to the randomly-normed ML estimator, when it exists. Under further nonlocal conditions, conditional inference would also formally agree with asymptotic Bayesian inference for every smooth prior on Θ; see, for example, Sweeting and Adekola (1987). Similar remarks have been made in Feigin and Reiser (1979), Sweeting (1980, 1986), Dawid (1991) and elsewhere. These ideas are made precise in the next section. Although the main motivation is to obtain limit theorems conditional on observed information, our results apply to any asymptotically ancillary statistic which includes data asymptotically equivalent to observed information. See McCullagh (1984), Basawa and Brockwell (1984) and Jensen (1987) for some related work.

To illustrate the importance of our results, we conclude this section by applying the conditional limit theorem obtained in Sweeting (1986) for the supercritical branching process. Assume a geometric offspring distribution with unknown mean θ and suppose that the unconditional sampling distribution of the asymptotic pivot $Y_t = B_t^{T/2}(\hat{\theta}_t - \theta)$ is used to construct a 95% confidence interval for θ, where B_t is Fisher's information. The asymptotic distribution of Y_t here is Student t on 2 degrees of freedom and the resulting interval is $\{\theta: |Y_t| \le 4.30\}$. From Sweeting (1986) there exists a statistic V_t asymptotically equivalent to observed information and asymptotically ancillary for θ and it is therefore important to examine the conditional coverage of the above confidence interval. Asymptotic results are given in Table 1 for a range of values of W_t, the ratio of observed to expected information, selected to represent the central 90% of the unit exponential distribution, which is the asymptotic distribution of W_t here. The poor conditional performance of the

TABLE 1
Conditional coverage of 95% confidence interval based on Y_t

Ratio of observed to expected information	0.05	0.20	1.00	2.00	3.00
Conditional coverage of 95% C.I.	66.4	94.6	99.998	$100-1.2 \times 10^{-7}$	$100-9.1 \times 10^{-12}$

confidence interval based on Y_t is well illustrated in Table 1. The results in Sweeting (1986) guarantee that the conditional coverage of the corresponding interval based on the randomly-normed pivotal quantity, however, is asymptotically 0.95.

3. Definitions and main results. Define the $B_t^{-T/2}$ neighbourhoods of θ_0 by

$$N_t(\theta_0, c) = \{\theta \in \Theta: |B_t^{T/2}(\theta - \theta_0)| < c\}$$

for all $c > 0$ and let $\phi_t^s = \theta_0 + B_t^{-T/2}s$, $s \in \mathbb{R}^k$. We impose the following conditions on the normalized information $W_t(\theta)$. In conditions D1 and D2 below, s_0 is a fixed positive number.

D1. There exists a random element W of M_k with $W > 0$ a.s. (θ_0) such that for all $|s| < s_0$,

$$W_t \Rightarrow W$$

under the (ϕ_t^s)-sequence of distributions.

D2. For all $c > 0$ and $|s| < s_0$,

$$\sup_{\theta \in N_t(\theta_0, c)} |W_t(\theta) - W_t| \to_p 0$$

under the (ϕ_t^s)-sequence of distributions.

P1. There exists a random element W of M_k with $W > 0$ a.s. (θ_0) such that

$$W_t \to_p W$$

under θ_0.

Finally, D1* will denote condition D1 strengthened to uniform weak convergence in $\{|s| < s_0\}$. These conditions hold for many standard problems. Conditions D1, D2 are similar to C1, C2, respectively, in Sweeting (1980), the main difference being the weakening to convergence in $B_t^{-T/2}$-neighbourhoods. No further Lindeberg-type conditions are assumed in the sequel. The first result is a version of Theorem 1 in Sweeting (1980).

THEOREM 3.1. *Under conditions* D1 *and* D2, *for all* $|s| < s_0$,

$$(X_t, W_t) \Rightarrow (W^{1/2}Z + Ws, W)$$

under the (ϕ_t^s)*-sequence of distributions, where* Z *is a standard normal random vector in* $\mathbb{R}^k$ *independent of* W.

We turn now to the question of asymptotic ancillarity. It is argued in Sweeting (1986) that the definition of asymptotic ancillarity of a sequence (A_t) of statistics should be related to the asymptotic constancy of the likelihood based on A_t in shrinking neighbourhoods $N_t(\theta_0, s)$ of θ_0. This is a local property of the distribution of A_t and it is not implied by, nor does it imply, convergence in distribution of A_t. Further discussion of the asymptotic ancillarity concept used here, and of related work, appears in Sweeting (1986).

We will permit extension of the basic space $(\Omega, \mathscr{B}, P_\theta)$ to $(\Omega^*, \mathscr{B}^*, P_\theta^*)$, the product space formed with an arbitrary probability space $(\Omega_0, \mathscr{B}_0, \lambda_0)$, and take $\mathscr{B}_t^* = \mathscr{B}_t \times \mathscr{B}_0$, $t > 0$. Let $(\mathscr{A}_t) \subset (\mathscr{B}_t^*)$ be a sequence of sub σ-fields (that is, $\mathscr{A}_t \subset \mathscr{B}_t^*$ for each t) and let $Q_{t\theta}$ be the restriction of $P_{t\theta}^*$ to $\mathscr{A}_t$. Without loss of generality we take the dominating measures λ_t to be *probability* measures [cf. Halmos and Savage (1949)]. Let μ_t be a dominating measure for the family $(Q_{t\theta}, \theta \in \Theta)$; the restriction of $\lambda_t^* = \lambda_t \times \lambda_0$ to $\mathscr{A}_t$, for example, is one such measure. Let $f_t(\theta)$ be a density of $Q_{t\theta}$ with respect to μ_t. From Theorem 3.1 we see that the dominant region of the likelihood function is a $B_t^{-T/2}$-neighbourhood of θ_0 and we make the following definition.

DEFINITION 3.2. The sequence of sub σ-fields $(\mathscr{A}_t) \subset (\mathscr{B}_t^*)$ is asymptotically ancillary for θ if, for some $s_0 > 0$ and all $|s| < s_0$,

$$f_t(\phi_t^s)/f_t(\theta_0) \to_p 1$$

under θ_0.

Note that this definition does not depend on the particular choice of $f_t(\theta)$. Also, the set $S_t = \{f_t(\theta) > 0\}$ is independent of θ and all dominating measures of $Q_{t\theta}$ are equivalent on S_t. Since $P_{\theta_0}(S_t) = 1$, it follows that the definition is independent of the choice of dominating measure μ_t. A sequence (A_t) of $\mathscr{B}_t^*$-measurable functions is asymptotically ancillary if the sequence of sub σ-fields induced by (A_t) is asymptotically ancillary. Convergence for all $s \in \mathbb{R}^k$ in Definition 3.2 might appear more natural, but the weaker definition given here suffices for the present purpose. Our concept of asymptotic ancillarity is similar to that used in Sweeting (1986), where the mode of convergence is *uniform* convergence in compact subsets of the common support of $Q_{t\theta}$.

DEFINITION 3.3. The sequence $(\tilde{J}_t)$ of $\mathscr{B}_t^*$-measurable functions in M_k is equivalent to observed information if, for some $s_0 > 0$ and all $|s| < s_0$,

$$B_t^{-1/2}\left[\tilde{J}_t - J_t(\phi_t^s)\right]B_t^{-T/2} \to_p 0$$

under (ϕ_t^s).

DEFINITION 3.4. The sequence $(\mathscr{A}_t) \subset (\mathscr{B}_t^*)$ of sub σ-fields is fully informative if there exists a sequence $(\tilde{J}_t)$ of $(\mathscr{A}_t)$-measurable functions in M_k equivalent to observed information.

Let $\mathscr{A}$, $\mathscr{A}_t$, $t > 0$, be sub σ-fields with $\mathscr{A}_t \subset \mathscr{B}_t^*$. The notation $X_t|\mathscr{A}_t \Rightarrow_p X|\mathscr{A}$ will mean $E(u(X_t)|\mathscr{A}_t) \to_p E(u(X)|\mathscr{A})$ for every bounded continuous function u. The next theorem is the main result of the paper.

THEOREM 3.5. *Let $(\mathscr{A}_t) \subset (\mathscr{B}_t^*)$ be a fully informative asymptotically ancillary sequence. Then, under conditions* P1 *and* D2,

$$(X_t, W_t)\big|\mathscr{A}_t \Rightarrow_p (W^{1/2}Z, W)\big|W$$

under θ_0, where Z is a standard normal random vector in $\mathbb{R}^k$ independent of W.

COROLLARY 3.6. *Let $(\mathscr{A}_t) \subset (\mathscr{B}_t^*)$ be a fully informative asymptotically ancillary sequence and $(\tilde{J}_t)$ a sequence of $(\mathscr{B}_t^*)$-measurable functions equivalent to observed information. Then, under conditions* P1 *and* D2,

$$\tilde{J}_t^{-1/2}U_t(\theta_0)|\mathscr{A}_t \Rightarrow_p Z$$

under θ_0 and there exist local maxima $(\hat{\theta}_t)$ of $l_t(\theta)$ for which

$$\tilde{J}_t^{T/2}\big(\hat{\theta}_t - \theta_0\big)\big|\mathscr{A}_t \Rightarrow_p Z$$

under θ_0, where Z is a standard normal random vector in $\mathbb{R}^k$.

If no fully informative asymptotically ancillary sequence exists, the above results are vacuous. We have the following result, however.

LEMMA 3.7. *Assume the likelihood possesses a unique local maximum with probability tending to one under θ_0. Then, under conditions* D1* *and* D2, *there exists an asymptotically ancillary sequence (J_t^*) of statistics equivalent to observed information.*

A partial converse is provided by the following lemma.

LEMMA 3.8. *Under conditions* P1 *and* D2, *condition* D1 *is necessary for the existence of an asymptotically ancillary sequence (J_t^*) equivalent to observed information.*

4. Proofs. Let $C(S)$ be the class of nonnegative bounded continuous functions on the metric space S and $C_0(S)$ the subclass of functions in $C(S)$ with compact support.

LEMMA 4.1. *Let $(\Omega, \mathscr{B}, \lambda)$ be a probability space and (X_n, Y_n), $n = 0, 1, \dots,$ be measurable mappings from $(\Omega, \mathscr{B})$ to $\mathbb{R}^k \times \mathbb{R}^l$. Suppose that, for some*

$s_0 > 0$ and all $|s| < s_0$, $v \in C_0(\mathbb{R}^l)$,

$$E\big(e^{s^T X_n} v(Y_n)\big) \to E\big(e^{s^T X_0} v(Y_0)\big) < \infty$$

as $n \to \infty$. Then $(X_n, Y_n) \Rightarrow (X_0, Y_0)$.

REMARK. Note that finiteness of $E(e^{s^T X_n})$ is not necessary here.

PROOF. If $E(v(Y_0)) > 0$, then, for n sufficiently large, we can define the probability measure $dP_n^v = v(Y_n)\,d\lambda / E(v(Y_n))$. Now, for $|s| < s_0$,

$$\begin{aligned} E_n^v(e^{s^T X_n}) &= E\big(e^{s^T X_n} v(Y_n)\big)/E(v(Y_n)) \\ &\to E\big(e^{s^T X_0} v(Y_0)\big)/E(v(Y_0)) = E_0^v(e^{s^T X_0}), \end{aligned}$$

where E_n^v denotes expectation under P_n^v and hence $P_n^v \Rightarrow P_0^v$. Therefore $E(u(X_n)v(Y_n)) \to E(u(X_0)v(Y_0))$ for every $u \in C_0(\mathbb{R}^k)$ and the desired weak convergence follows by approximation of indicators of finite rectangles. □

PROOF OF THEOREM 3.1. First note that $\phi_t^s \in \Theta$ eventually. Write $p_t = p_t(\theta_0)$, $q_t^s = p_t(\phi_t^s)$. A second-order Taylor expansion about $\theta = \theta_0$ gives

$$(1)\quad l_t(\phi_t^s) = l_t(\theta_0) + (\phi_t^s - \theta_0)^T l_t'(\theta_0) + \tfrac{1}{2}(\phi_t^s - \theta_0)^T l_t''(\psi_t^s)(\phi_t^s - \theta_0),$$

where $\psi_t^s = \alpha_t^s \theta_0 + (1 - \alpha_t^s)\phi_t^s$ for some α_t^s with $0 < \alpha_t^s < 1$. Taking exponentials in (1) and rearranging gives

$$(2)\qquad e^{s^T X_t} p_t = \exp\big[\tfrac{1}{2} s^T W_t(\psi_t^s) s\big] q_t^s.$$

Write $D_t = \sup_{\theta \in N_t(\theta_0, s_0)} |W_t(\theta) - W_t|$, $\Delta_t^s = W_t(\psi_t^s) - W_t$ and $v \in C_0(M_k \times \mathbb{R})$. Multiply (2) through by $v(W_t, D_t)$ and take expectations w.r.t. λ_t to give

$$(3)\qquad E_t^0\big(e^{s^T X_t} v(W_t, D_t)\big) = E_t^s\big(\exp\big[\tfrac{1}{2} s^T (W_t + \Delta_t^s) s\big] v(W_t, D_t)\big),$$

where E_t^s denotes expectation under $P_{t\phi_t^s}$. Since $|\Delta_t^s| \le D_t$ and $v \in C_0(M_k \times \mathbb{R})$, the integrand on the right in (3) is in $C_0(M_k^2 \times \mathbb{R})$ and D1, D2 imply that the right-hand side converges to

$$E\big(\exp\big[\tfrac{1}{2} s^T W s\big] v(W, 0)\big) = E\big(\exp\big[\tfrac{1}{2} s^T W^{1/2} Z\big] v(W, 0)\big),$$

where Z is a standard normal vector in $\mathbb{R}^k$ independent of W. Since this convergence holds for all $|s| < s_0$, it follows from Lemma 4.1 that $(X_t, W_t) \Rightarrow (W^{1/2}Z, W)$ under θ_0.

Finally, let $L_t^s = q_t^s / p_t$. Again from (2), $L_t^s = \exp[s^T X_t - \frac{1}{2} s^T (W_t + \Delta_t^s) s]$. Since $\Delta_t^s \Rightarrow 0$ under θ_0, preservation of weak convergence under continuous mapping gives

$$(X_t, W_t, L_t^s) \Rightarrow (W^{1/2}Z, W, L^s),$$

where $L^s = \exp[s^T W^{1/2} Z - \frac{1}{2} s^T W s]$. Since $E(L^s) = 1$, the result now follows by contiguity; see, for example, Theorem A.2.2 in Basawa and Scott (1983). □

We shall make use of the following relation. Let $(\Omega, \mathscr{B}, \lambda)$ be a probability space, $P \ll \lambda$ with density p, X a $\mathscr{B}$-measurable real function and $\mathscr{A} \subset \mathscr{B}$ a sub σ-field. Then

$$E_\lambda(Xp|\mathscr{A}) = E_p(X|\mathscr{A})f, \tag{4}$$

where f is the density of the restriction of P to $\mathscr{A}$ w.r.t. the restriction of λ to $\mathscr{A}$. Note in particular that $f = E_\lambda(p|\mathscr{A})$. The next lemma is a conditional version of Lemma 4.1.

LEMMA 4.2. *Let $(\Omega, \mathscr{B}, \lambda)$ be a probability space, $(\mathscr{A}_n, n = 0, 1, \dots,)$ a sequence of sub σ-fields of $(\Omega, \mathscr{B})$ and (X_n, Y_n), $n = 0, 1, \dots,$ measurable mappings from $(\Omega, \mathscr{B})$ to $\mathbb{R}^k \times \mathbb{R}^l$. Suppose that, for some $s_0 > 0$ and all $|s| \le s_0$, $v \in C_0(\mathbb{R}^l)$,*

$$E\Big(e^{s^T X_n} v(Y_n)\big|\mathscr{A}_n\Big) \to_p E\Big(e^{s^T X_0} v(Y_0)\big|\mathscr{A}_0\Big) < \infty \quad a.s.$$

as $n \to \infty$. Then $(X_n, Y_n)|\mathscr{A}_n \Rightarrow_p (X_0, Y_0)|\mathscr{A}_0$.

PROOF. Define P_n^v as in the proof of Lemma 4.1. Relation (4) gives

$$E\Big(e^{s^T X_n} v(Y_n)\big|\mathscr{A}_n\Big) = E_n^v\Big(e^{s^T X_n}\big|\mathscr{A}_n\Big) E\big(v(Y_n)\big|\mathscr{A}_n\big).$$

It follows that $E_n^v(e^{s^T X_n}|\mathscr{A}_n) \to_p E_0^v(e^{s^T X_0}|\mathscr{A}_0)$ and hence $E_n^v(u(X_n)|\mathscr{A}_n) \to_p E_0^v(u(X_0)|\mathscr{A}_0)$ for all $u \in C_0(\mathbb{R}^k)$, as in Corollary 3 of Sweeting (1989). Therefore $E(u(X_n)v(Y_n)|\mathscr{A}_n) \to_p E(u(X_0)v(Y_0)|\mathscr{A}_0)$ and this easily gives the corresponding convergence for finite rectangles. Since the class of (X_0, Y_0)-continuity rectangles with rational coordinates is a countable convergence-determining class, the result now follows from Theorem 9 in Sweeting (1989). □

LEMMA 4.3. *Let $(\Omega, \mathscr{B}, \lambda)$, $\mathscr{A}_n$, (X_n, Y_n), $n = 0, 1, \dots,$ be as in Lemma 4.2. Suppose that $X_n|\mathscr{A}_n \Rightarrow_p X_0|\mathscr{A}_0$ and $Y_n \to_p Y_0$, a constant. Then $(X_n, Y_n)|\mathscr{A}_n \Rightarrow_p (X_0, Y_0)|\mathscr{A}_0$.*

PROOF. Let $u \in C(\mathbb{R}^k)$, $v \in C(\mathbb{R}^l)$. Then

$$\big|E\big(u(X_n)v(Y_n)|\mathscr{A}_n\big) - v(Y_0)E\big(u(X_n)|\mathscr{A}_n\big)\big| \le \|u\| E\big(|v(Y_n) - v(Y_0)|\,|\mathscr{A}_n\big).$$

Now $v(Y_n) \to_p v(Y_0)$ and hence $E(|v(Y_n) - v(Y_0)|) \to 0$ since v is bounded. But $E(|v(Y_n) - v(Y_0)|) = E(E(|v(Y_n) - v(Y_0)|\,|\mathscr{A}_n))$, so $E(|v(Y_n) - v(Y_0)|\,|\mathscr{A}_n) \to_p 0$. The result follows since $E(u(X_n)|\mathscr{A}_n) \to_p E(u(X_0)|\mathscr{A}_0)$. □

PROOF OF THEOREM 3.5. Since $(\mathscr{A}_t)$ is fully informative, there exists a sequence $(\tilde{J}_t)$ of $\mathscr{A}_t$-measurable statistics equivalent to observed information. Write $\tilde{W}_t = B_t^{-1/2} \tilde{J}_t B_t^{-T/2}$. It follows from D2 that $\tilde{\Delta}_t^s = W_t(\psi_t^s) - \tilde{W}_t \to_p 0$ under (ϕ_t^s), where (ψ_t^s) is as in the proof of Theorem 3.1. Define $\tilde{D}_t = \sup_{\theta \in N_t(\theta_0, s_0)} |W_t(\theta) - \tilde{W}_t|$ and let $v \in C_0(M_k)$. Multiply (2) through by $v(\tilde{D}_t)$

and take conditional expectations w.r.t. λ_t^* relative to $\mathscr{A}_t$ to give

$$E_{\lambda_t^*}\big(e^{s^T X_t} v(\tilde{D}_t) p_t \big| \mathscr{A}_t\big) = E_{\lambda_t^*}\big(\exp\big[\tfrac{1}{2} s^T W_t(\psi_t^s) s\big] q_t^s \big| \mathscr{A}_t\big).$$

Relation (4) now gives

$$\text{(5)}\qquad \begin{aligned} &E_t^{*0}\big(e^{s^T X_t} v(\tilde{D}_t)\big| \mathscr{A}_t\big) \\ &\quad = \exp\big[\tfrac{1}{2} s^T \tilde{W}_t s\big] E_t^{*s}\big(\exp\big[\tfrac{1}{2} s^T \tilde{\Delta}_t^s s\big] v(\tilde{D}_t)\big| \mathscr{A}_t\big)\{f_t(\phi_t^s)/f_t(\theta_0)\}, \end{aligned}$$

where E_t^{*s} denotes expectation under $P_{t\phi_t^s}^*$. Since $\tilde{D}_t \Rightarrow 0$ under (ϕ_t^s), it follows from Lemma 4.3 that $\tilde{D}_t | \mathscr{A}_t \to_p 0$ and, since $|\tilde{\Delta}_t^s| \le \tilde{D}_t$,

$$\text{(6)}\qquad E_t^{*s}\big(\exp\big[\tfrac{1}{2} s^T \tilde{\Delta}_t^s s\big] v(\tilde{D}_t)\big| \mathscr{A}_t\big) \to_p v(0)$$

under (ϕ_t^s). Denote by V_t the difference between the left- and right-hand sides of (6). Then, using the $\mathscr{A}_t$-measurability of V_t, for all $\varepsilon > 0$,

$$\begin{aligned} P_{t\theta_0}^*(|V_t| > \varepsilon) &\le P_{t\theta_0}^*(|V_t| > \varepsilon,\, f_t(\theta_0) \le 2 f_t(\phi_t^s)) + P_{t\theta_0}^*(f_t(\theta_0) > 2 f_t(\phi_t^s)) \\ &\le 2 P_{t\phi_t^s}^*(|V_t| > \varepsilon) + P_{t\theta_0}^*(|\{f_t(\phi_t^s)/f_t(\theta_0)\} - 1| > 1) \to 0 \end{aligned}$$

from (6) and the asymptotic ancillarity of $(\mathscr{A}_t)$. The convergence in (6) therefore also holds under θ_0.

It now follows from P1, (5) and the asymptotic ancillarity of $(\mathscr{A}_t)$ that

$$E_t^{*0}\big(e^{s^T X_t} v(\tilde{D}_t)\big| \mathscr{A}_t\big) \to_p E\big(e^{s^T W^{1/2} Z}\big| W\big) v(0)$$

under θ_0, where Z is a standard normal vector in $\mathbb{R}^k$ independent of W. Lemma 4.2 now gives $X_t | \mathscr{A}_t \Rightarrow_p W^{1/2} Z | W$. Since $\tilde{W}_t$ is $\mathscr{A}_t$-measurable, it is easily seen that this implies $(X_t, \tilde{W}_t) | \mathscr{A}_t \Rightarrow_p (W^{1/2} Z, W) | W$. Finally, since $\tilde{W}_t - W_t \to_p 0$, Lemma 4.3 gives

$$\big(X_t, \tilde{W}_t, \tilde{W}_t - W_t\big)\big| \mathscr{A}_t \Rightarrow_p (W^{1/2} Z, W, 0) \big| W$$

and the result follows from Theorem 10 in Sweeting (1989). □

PROOF OF COROLLARY 3.6. All convergence statements are taken to be under θ_0. Write $\tilde{W}_t = B_t^{-1/2} \tilde{J}_t B_t^{-T/2}$ and note that $\tilde{J}_t^{-1/2} U_t = \tilde{W}_t^{-1/2} X_t$. Since $\tilde{W}_t - W_t \to_p 0$, Lemma 4.3 and Theorem 10 in Sweeting (1989) along with Theorem 3.5 give

$$\text{(7)}\qquad \big(X_t, \tilde{W}_t\big)\big| \mathscr{A}_t \Rightarrow_p (W^{1/2} Z, W) \big| W.$$

The first result now follows, again from Theorem 10 in Sweeting (1989).

For the second result, Theorem 3.5 implies the existence of a sequence $(\hat{\theta}_t)$ of local maxima of $l_t(\theta)$ for which $Y_t = B_t^{T/2}(\hat{\theta}_t - \theta_0)$ is stochastically bounded. This follows from a nonuniform version of Lemma 4 in Sweeting (1980) and the proof is omitted. A nonuniform version of Theorem 2 in Sweeting (1980) then gives $X_t - \tilde{W}_t Y_t \to_p 0$. Lemma 4.3 and (7) now give $\tilde{W}_t^{T/2} Y_t | \mathscr{A}_t \Rightarrow_p Z$ and the result follows from $\tilde{W}_t^{T/2} B_t^{T/2} = \tilde{J}_t^{T/2}$. □

PROOF OF LEMMA 3.7. From Theorem 3.1, (X_t) is stochastically bounded under (ϕ_t^s). A modification of Lemma 4 in Sweeting (1980) gives the existence for each s of a sequence $(\hat{\theta}_t^s)$ of local maxima of $l_t(\theta)$ such that $Y_t^s = B_t^{T/2} \cdot (\hat{\theta}_t^s - \phi_t^s)$ is stochastically bounded under (ϕ_t^s). It follows from D2 that $(J_t(\hat{\theta}_t))$ is a sequence of statistics equivalent to observed information, where $(\hat{\theta}_t)$ is the unique local maximum guaranteed with θ_0-probability tending to 1.

Now let $(\tilde{J}_t)$ be any sequence equivalent to observed information and write $\tilde{W}_t = B_t^{-1/2}\tilde{J}_t B_t^{-T/2}$. Let H be any distribution with bounded uniformly continuous density h with respect to Lebesgue measure on M_k. Then we can construct a $\mathscr{B}_0$-measurable function ξ in M_k with distribution H. Let $\varepsilon > 0$ and write $\tilde{W}_t^\varepsilon = \tilde{W}_t + \varepsilon\xi$. Then $\tilde{W}_t^\varepsilon$ has density $f_t^\varepsilon(\theta) = \int h(\{w - x\}/\varepsilon)\, dG_{t\theta}(x)$, where $G_{t\theta}$ is the distribution of $\tilde{W}_t$ under θ. From D1* it follows that $f_t^\varepsilon(\phi_t^s) \to f^\varepsilon(\theta_0)$ uniformly in $|s| < s_0$, where $f^\varepsilon(\theta_0) = \int h(\{w - x\}/\varepsilon)\, dG(x)$ and G is the distribution of W under θ_0. Moreover since the functions $h(\{w - x\}/\varepsilon)$ are equicontinuous in x, the convergence is uniform in w. Therefore there exists a sequence $\varepsilon_t \to 0$ such that, for all $|s| < s_0$, $f_t^{\varepsilon_t}(\phi_t^s)/f_t^{\varepsilon_t}(\theta_0) \to 1$ uniformly in w, which implies the probability convergence of this ratio to unity. Thus $W_t^* = \tilde{W}_t^{\varepsilon_t}$ is asymptotically ancillary and hence so is $J_t^* = B_t^{1/2} W_t^* B_t^{T/2}$. Furthermore since $W_t^* - \tilde{W}_t \to_p 0$ under (ϕ_t^s), J_t^* is equivalent to observed information. □

PROOF OF LEMMA 3.8. Take $\mathscr{A}_t$ to be the σ-field generated by J_t^*. Then Theorem 3.5 implies the unconditional limit $(X_t, W_t) \Rightarrow (W^{1/2}Z, W)$ under θ_0. The contiguity argument given at the end of the proof of Theorem 3.1 now yields $W_t \Rightarrow W$ under (ϕ_t^s), which is condition D1. □

Acknowledgments. The author wishes to thank the referees for their detailed comments on earlier drafts of this paper.

REFERENCES

BASAWA, I. V. and BROCKWELL, P. J. (1984). Asymptotic conditional inference for regular nonergodic models with an application to autoregressive processes. *Ann. Statist.* **12** 161–171.

BASAWA, I. V. and SCOTT, D. J. (1983). *Asymptotic Optimal Inference for Nonergodic Models. Lecture Notes in Statist.* **17**. Springer, Berlin.

DAWID, A. P. (1991). Fisherian inference in likelihood and prequential frames of reference. *J. Roy. Statist. Soc. Ser. B* **53** 79–109.

EFRON, B. and HINKLEY, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information (with discussion). *Biometrika* **65** 457–487.

FAHRMEIR, L. (1988). A note on asymptotic testing theory for nonhomogeneous observations. *Stochastic Process. Appl.* **28** 267–273.

FEIGIN, P. D. and REISER, B. (1979). On asymptotic ancillarity and inference for Yule and regular nonergodic processes. *Biometrika* **66** 279–283.

HALMOS, P. R. and SAVAGE, L. J. (1949). Application of the Radon–Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Statist.* **20** 225–241.

JENSEN, J. L. (1987). On asymptotic expansions in non-ergodic models. *Scand. J. Statist.* **14** 305–318.

McCullagh, P. (1984). Local sufficiency. *Biometrika* **71** 233–244.

Sweeting, T. J. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *Ann. Statist.* **8** 1375–1381. [Correction: *Ann. Statist.* (1982) **10** 320–321.]

Sweeting, T. J. (1986). Asymptotic conditional inference for the offspring mean of a supercritical Galton–Watson process. *Ann. Statist.* **14** 925–933.

Sweeting, T. J. (1989). On conditional weak convergence. *J. Theoret. Probab.* **2** 461–474.

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

Department of Mathematical and
Computing Sciences
University of Surrey
Guildford, Surrey GU2 5XH
England