Stochastic Processes and their Applications 8 (1978) 1–9.

ON AN OPTIMAL ASYMPTOTIC PROPERTY OF THE MAXIMUM LIKELIHOOD ESTIMATOR OF A PARAMETER FROM A STOCHASTIC PROCESS

C.C. HEYDE

CSIRO Division of Mathematics and Statistics, P.O. Box 1965, Canberra City, A.C.T. 2601, Australia

Received 13 October 1977
Revised 24 April 1978

This paper is concerned with the estimation of a parameter of a stochastic process on the basis of a single realization. It is shown, under suitable regularity conditions, that the maximum likelihood estimator is the best consistent asymptotically normal estimator in the sense of having minimum asymptotic variance. It also produces the best limiting probability of concentration in symmetric intervals. An application is given for the problem of estimating the mean of the offspring distribution in a Galton–Watson branching process.

Estimation for stochastic processes
maximum likelihood
best uniformly asymptotically normal estimator
limiting probability of concentration
martingale limit theorems

Introduction

In the case of estimating a parameter on the basis of observations which are independent and identically distributed or come from a stationary ergodic Markov chain, it is known that the maximum likelihood estimator (MLE) enjoys certain optimality properties. One of the most significant of these is a result of Schmetterer [13] (extending one of Rao [10]) that, subject to suitable regularity conditions, the MLE is the best consistent continuously asymptotically normal estimator in the sense of having minimum asymptotic variance. In this paper we shall obtain an analogous result for the much more complex case where the sample is from a general stochastic process. The result is an augmented and corrected version of that presented in Heyde [6]. We also show that the MLE produces the best asymptotic probability of concentration in symmetric intervals. The results follow from ideas of Weiss and Wolfowitz [15].

We consider a sample $X_1, X_2, \dots, X_n$ of consecutive observations from some stochastic process whose distribution depends on a single parameter, θ, $\theta \in \Theta$, Θ being an open interval. Let $L_n(\theta)$ be the likelihood function associated with

R. Maller et al. (eds.), *Selected Works of C.C. Heyde*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4419-5823-5_44, © Springer Science+Business Media, LLC 2010

$X_1, \dots, X_n$ and suppose that $L_n(\theta)$ is differentiable with respect to θ and $E_\theta(\mathrm{d}\log L_n(\theta)/\mathrm{d}\theta)^2 < \infty$ for each n. Suppose in addition that if $P_n(X_1, \dots, X_n)$ $(=L_n(\theta))$ is the joint probability (density) function of $X_1, \dots, X_n$, then $\sum_{x_n} P_n(x_1, \dots, x_n)(\int P_n(x_1, \dots, x_n)\,\mathrm{d}x_n)$ can be differentiated twice with respect to θ under the summation (integration) sign. We shall write F_k for the σ-field generated by $X_1, \dots, X_k$, $k \geq 1$, take F_0 as the trivial σ-field and set $L_0 = 1$. Then, writing

$$\frac{\mathrm{d}\log L_n(\theta)}{\mathrm{d}\theta} = \sum_{i=1}^{n} u_i(\theta),$$

we have $E_\theta(u_i(\theta)\,|\,F_{i-1}) = 0$ a.s., so that $\{\mathrm{d}\log L_n(\theta)/\mathrm{d}\theta, F_n, n \geq 1\}$ is a (square integrable) martingale. In addition, we set

$$I_n(\theta) = \sum_{i=1}^{n} E_\theta(u_i^2(\theta)\,|\,F_{i-1})$$

and note that, under the conditions imposed above, $v_i(\theta) = \mathrm{d}u_i(\theta)/\mathrm{d}\theta$ satisfies

$$E_\theta(u_i^2(\theta)\,|\,F_{i-1}) = -E_\theta(v_i(\theta)\,|\,F_{i-1}) \text{ a.s.}$$

Also, writing

$$J_n(\theta) = \sum_{i=1}^{n} v_i(\theta)$$

we note that $\{J_n(\theta) + I_n(\theta), F_n, n \geq 1\}$ is a martingale.

The quantity $I_n(\theta)$ is a form of conditional information which reduces to the standard Fisher information in the case where the X_i are independent random variables. The rôle of $I_n(\theta)$ in the stochastic process estimation context is a vital one and aspects of this are discussed in Heyde [5], [6], Heyde and Feigin [7]. The importance of $I_n(\theta)$ is easily seen from the following expansion. We suppose that θ is the true parameter value. Then, we can use Taylor's expansion to write for $\theta' \in \Theta$,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta'}\log L_n(\theta') &= \sum_{i=1}^{n} u_i(\theta') \\ &= \sum_{i=1}^{n} u_i(\theta) + (\theta' - \theta)\sum_{i=1}^{n} v_i(\theta_n^*) \\ &= \sum_{i=1}^{n} u_i(\theta) - (\theta' - \theta) I_n(\theta) + (\theta' - \theta)(J_n(\theta_n^*) + I_n(\theta)) \qquad (1)\end{aligned}$$

where $\theta_n^* = \theta + \gamma(\theta' - \theta)$ with $\gamma = \gamma(n, \theta)$ satisfying $|\gamma| < 1$. Further, since

$$[I_n(\theta)]^{-1} \sum_{i=1}^{n} u_i(\theta) \xrightarrow{\text{a.s.}} 0$$

provided $I_n(\theta) \xrightarrow{\text{a.s.}} \infty$ as $n \to \infty$ (a martingale strong law, e.g. see Heyde and Feigin [7]), we see that the likelihood equation has a root $\hat{\theta}_n$ which is strongly consistent for θ if $I_n(\theta) \xrightarrow{\text{a.s.}} \infty$ and

$$\limsup_{n\to\infty} [I_n(\theta)]^{-1}|I_n(\theta)+J_n(\theta_n^*)| < 1 \text{ a.s.}$$

for $|\theta_n^* - \theta| < \delta$ sufficiently small. Suppose that these conditions hold.

In order to utilize most effectively the MLE from large samples it is necessary for $\sum_{i=1}^n u_i(\theta)$ to converge in distribution, when appropriately normalized, to some proper limit law. Furthermore, it can be arranged in many cases that this limit law is normal, which is most convenient for confidence interval purposes. The most effective general norming seems to be provided by $I_n^{1/2}(\theta)$ and, indeed, $[I_n(\theta)]^{-1/2}\sum_{i=1}^n u_i(\theta)$ converges in distribution to normality under quite wide ranging circumstances (e.g. Hall [2], Scott [14]).

Now suppose that

$$[I_n(\theta)]^{-1/2} \sum_{i=1}^n u_i(\theta) \xrightarrow{d} N(0,1).$$

If, in addition to the above conditions, $J_n(\theta_n^{**}) = -I_n(\theta)(1+o(1))$ in probability as $n\to\infty$ for $\theta_n^{**} = \theta + s(\hat{\theta}_n - \theta)$ with any s satisfying $|s| \le 1$, we have from (1) that

$$I_n^{1/2}(\theta)(\hat{\theta}_n - \theta) \xrightarrow{d} N(0,1).$$

Then, if $T_n = T_n(X_1, \ldots, X_n)$ is any consistent estimator of θ for which $I_n^{1/2}(\theta)(T_n - \theta) \xrightarrow{d} N(0, \beta^2(\theta))$ where $\beta(\theta)$ is bounded and continuous in θ, we shall show, under suitable additional regularity conditions, that $\beta^2(\theta) \ge 1$. That is, the MLE is optimal within this class since $\beta(\theta) \equiv 1$ when T_n is the MLE.

Of course it is desirable if comparisons can be made between the MLE and those other consistent estimators for which $I_n^{1/2}(\theta)(T_n - \theta)$ converges in distribution to a proper law (not necessarily normal) or perhaps does not even converge. Such comparisons have been made in the case where the $\{X_i\}$ form a stationary ergodic Markov chain by using the concept of asymptotic efficiency in the Wolfowitz sense (see Roussas [12, Chapter 5]). Minor modifications of the standard theory as presented in [12] leads to comparisons of the kind

$$\begin{aligned}\lim_{n\to\infty} P_\theta(-c < (E_\theta I_n(\theta))^{1/2}(\hat{\theta}_n - \theta) < b) \ge \\ \ge \limsup_{n\to\infty} P_\theta(-c + w(\theta) < (E_\theta I_n(\theta))^{1/2}(T_n - \theta) < W(\theta) + b)\end{aligned}$$

for arbitrary positive b and c and certain $W(\theta) \ge w(\theta)$. This inequality holds under conditions which are quite similar to those of this paper but (at present) require that $I_n(\theta)(E_\theta I_n(\theta))^{-1} \xrightarrow{p} 1$ as $n\to\infty$. It is, however, the cases where this last condition is not satisfied that pose the real interest and challenge in the treatment of stochastic process estimation. Our arguments in this paper are principally concerned with the commonly occurring circumstances under which $I_n(\theta)(E_\theta I_n(\theta))^{-1} \xrightarrow{p} \eta(\theta)$ as $n \to \infty$, η being a random variable in general.

These difficulties can be avoided if we restrict consideration to symmetric concentration intervals. Then, using results of Weiss and Wolfowitz [15], we shall show that

$$\lim_{n\to\infty} P_\theta(-c < (E_\theta I_n(\theta))^{1/2}(\hat{\theta}_n - \theta) < c) \geq$$
$$\geq \limsup P_\theta(-c < (E_\theta I_n(\theta))^{1/2}(T_n - \theta) < c) \quad \text{for all } c > 0.$$

Principal result

We suppose that $(X_1, \ldots, X_n)$ possesses a density $P_n(X_1, \ldots, X_n)(=L_n(\theta))$, which is continuous in θ, with respect to a σ-finite measure μ_n and that the following two assumptions are satisfied.

Assumption 1. (i) $I_n(\theta) \xrightarrow{\text{a.s.}} \infty$, $I_n(\theta)/E_\theta I_n(\theta) \xrightarrow{p} \eta(\theta)(>0$ a.s.) for some r.v. η, and $J_n(\theta)/I_n(\theta) \xrightarrow{p} -1$ as $n \to \infty$, the convergences in probability being uniform in compacts of θ.

(ii) $[I_n(\theta)]^{-1/2} \,\mathrm{d} \log L_n(\theta)/\mathrm{d}\theta \xrightarrow{d} N(0, 1)$ (mixing), continuously in θ.

Assumption 2. For $\delta > 0$, supose $|\theta_n - \theta_0| \leq \delta/(E_{\theta_0} I_n(\theta_0))^{1/2}$. Then,

(i) $E_{\theta_n} I_n(\theta_n) = E_{\theta_0} I_n(\theta_0)(1 + \mathrm{o}(1))$ as $n \to \infty$, and

(ii) $I_n(\theta_n) = I_n(\theta_0)(1 + \mathrm{o}(1))$ a.s. as $n \to \infty$,

(iii) $J_n(\theta_n) = J_n(\theta_0) + \mathrm{o}(I_n(\theta_0))$ a.s. as $n \to \infty$.

These assumptions are not severe and conditions may be imposed directly on the stochastic process to ensure them. The mixing convergence in Assumption 1 calls for some comment. Here the mixing convergence means

$$\lim_{n\to\infty} E_\theta\left(\exp\left\{\mathrm{i}t(I_n(\theta))^{-1/2} \frac{\mathrm{d} \log L_n(\theta)}{\mathrm{d}\theta}\right\} \mid B\right) = \exp -\tfrac{1}{2}t^2$$

for all real t and all measurable B with $P(B) > 0$. The two parts of Assumption 1 ensure that for $g(x, y)$ any continuous function of two variables,

$$g\left\{(I_n(\theta))^{-1/2} \frac{\mathrm{d} \log L_n(\theta)}{\mathrm{d}\theta}, \frac{I_n(\theta)}{E_\theta I_n(\theta)}\right\} \xrightarrow{d} g(N(0, 1), \eta(\theta))$$

where N and η are independent (e.g. Eagleson [1]). Further, mixing versions of the standard martingale central limit results hold without additional conditions (McLeish [8, p. 628], Scott [14]).

We are now in a position to establish the following result.

Theorem. *Fix $C>0$, let $\theta_n=\theta+2C(E_\theta I_n(\theta))^{-1/2}$ and write $P_\theta^{(n)}$ for the probability measure corresponding to the likelihood $L_n(\theta)$. Suppose that T_n is any estimator of θ which satisfies the condition that for any $\theta\in\Theta$,*

$$\lim_{n\to\infty}\{P_\theta^{(n)}((E_\theta I_n(\theta))^{1/2}(T_n-\theta)>-C)-P_{\theta_n}^{(n)}((E_\theta I_\theta(\theta))^{1/2}(T_n-\theta_n)>-C)\}=0.$$

If there is a maximum likelihood estimator $\hat{\theta}_n$ which is consistent for any $\theta\in\Theta$, then under Assumptions 1 *and* 2,

$$\lim_{n\to\infty}P_\theta^{(n)}(-C<(E_\theta I_n(\theta))^{1/2}(\hat{\theta}_n-\theta)<C)$$
$$\geqslant\lim_{n\to\infty}\sup P_\theta^{(n)}(-C<(E_\theta I_n(\theta))^{1/2}(T_n-\theta)<C).$$

Furthermore, if $I_n^{1/2}(\theta)(T_n-\theta)\xrightarrow{d}N(0,\gamma^2(\theta))$ (mixing), continuously in θ, where $\gamma(\theta)$ is bounded, then $\gamma^2(\theta)\geqslant 1$.

The result of this theorem resolves the question of the best limiting probability of concentration in symmetric intervals but not in the general case. For the general case, I.V. Basawa and D.J. Scott in a preprint entitled "Efficient estimation for stochastic processes" have obtained the following inequality:

$$\lim_{n\to\infty}P_\theta^{(n)}(-\delta_1<(E_\theta I_n(\theta))^{1/2}(T_n-\theta)<\delta_2)\leqslant M_\theta(\delta_2)-M_\theta(-\delta_1),$$

when the limit exists, where

$$M_\theta(\delta_2)=P(\delta_2N(0,1)\eta^{1/2}(\theta)-\tfrac{1}{2}\delta_2^2\eta(\theta)\leqslant c_2)$$

and c_2 is the (unique) median of $\delta_2N(0,1)\eta^{1/2}(\theta)+\frac{1}{2}\delta_2^2\eta(\theta)$, η and N being independent, while

$$M_\theta(-\delta_1)=P(-\delta_1N(0,1)\eta^{1/2}(\theta)-\tfrac{1}{2}\delta_1^2\eta(\theta)\leqslant c_1)$$

where c_1 is the (unique) median of $-\delta_1N(0,1)\eta^{1/2}(\theta)+\frac{1}{2}\delta_1^2\eta(\theta)$. They also show that the upper bound in the inequality is not attained for the MLE $\hat{\theta}_n$ unless $\eta(\theta)=1$ a.s. The methods used are similar to those employed in Roussas [12], Chapter 5.

Proof of Theorem. The results follow from a straightforward application of Theorem 3.1 of Weiss and Wolfowitz [15]. We just need to verify Conditions A and B involved therein.

From (1) we have

$$0=\sum_{i=1}^{n}u_i(\theta)+(\hat{\theta}_n-\theta)J_n(\theta_n^*), \tag{2}$$

where $\theta_n^* = \theta + \gamma(n, \theta)(\hat{\theta}_n - \theta)$ with $|\gamma(n, \theta)| < 1$, so that Assumptions 1 and 2 give

$$\lim_{n\to\infty} P_\theta^{(n)}((E_\theta I_n(\theta))^{1/2}(\hat{\theta}_n - \theta) < y) = \lim_{n\to\infty} P_{\theta_n}^{(n)}((E_\theta I_n(\theta))^{1/2}(\hat{\theta}_n - \theta) < y) \quad (3)$$

$$= \mathbf{P}(\eta^{1/2}(\theta)N(0, 1) < y)$$

for any y, $-\infty < y < \infty$, where η and N are independent. This verifies Condition A.

To check Condition B we define

$$\Lambda_n = \log [L_n(\theta_n)/L_n(\theta)] \text{ on } \{L_n(\theta_n)/L_n(\theta) > 0\}$$

and suppose that Λ_n is arbitrarily (but measurably) defined on $\{L_n(\theta_n)L_n(\theta) = 0\}$. Then we may use Taylor's expansion to write

$$\Lambda_n = (\theta_n - \theta) \sum_{i=1}^n u_i(\theta) + \tfrac{1}{2}(\theta_n - \theta)^2 J_n(\theta_n^{**})$$

for $\theta_n^{**} \in [\theta, \theta_n]$, and using (2),

$$\Lambda_n = -2C(E_\theta I_n(\theta))^{-1/2}(\hat{\theta}_n - \theta)J_n(\theta_n^*) + 2C^2(E_\theta I_n(\theta))^{-1}J_n(\theta_n^{**}).$$

Thus, employing Assumptions 1 and 2 again,

$$\{\Lambda_n < 0\} = \{\hat{\theta}_n < \theta + C(E_\theta I_n(\theta))^{-1/2}(1 + o_p(1))\}$$

(the $o_p(1)$ denoting a term which tends in probability to zero as $n \to \infty$) while

$$\lim_{n\to\infty} P_\theta^{(n)}(\Lambda_n = 0) = \lim_{n\to\infty} P_{\theta_n}^{(n)}(\Lambda_n = 0) = 0,$$

so that Condition B is satisfied. The first part of the result of the theorem then follows from Theorem 3.1 of [15].

To establish the second part of the theorem we note that the mixing form of convergence results prescribed for T_n, together with Assumption 1, ensures that

$$\lim_{n\to\infty} P_\theta^{(n)}(-C < (E_\theta I_n(\theta))^{1/2}(T_n - \theta) < C) =$$

$$= \mathbf{P}(-C < \gamma(\theta)\eta^{1/2}(\theta)N(0, 1) < C),$$

where η and N are independent and the required result follows in view of (3).

Some remarks on applications

Various structural requirements need to be imposed on the process $\{X_i\}$ in order to allow the various assumptions to be checked. The imposition of stationarity, for example, leads to useful simplification in the requirements. However, we shall here restrict attention to the case where the stochastic process $\{X_i\}$ is a time-homogeneous Markov process whose distribution belongs to a *conditional exponential family*.

The concept of a conditional exponential family, which is discussed in some detail in Heyde and Feigin [7], is a generalization of that of the exponential family for the case of independent X_i. For $\{X_i\}$ belonging to a conditional exponential family,

$$\frac{\mathrm{d} \log L_n(\theta)}{\mathrm{d}\theta} = I_n(\theta)(\hat{\theta}_n - \theta) \tag{4}$$

and $I_n(\theta) = \phi(\theta) \sum_{i=1}^n H(X_{i-1})$ where ϕ does not involve the X_i and H does not involve θ. Furthermore, (Heyde and Feigin [7]), the result (4) holds if and only if the conditional probability (density) function of X_k given X_{k-1}, $f(X_k \mid X_{k-1}, \theta)$ satisfies

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \log f(x \mid y, \theta) = \phi(\theta) H(y)[m(x, y) - \theta]$$

where

$$E_\theta[m(X_i, X_{i-1}) \mid F_{i-1}] = \theta \text{ a.s.}$$

For the case of the conditional exponential family we easily see that Assumption 2 holds if $\phi(\theta)$ is continuous and differentiable.

As a particular example we shall consider the estimation of the mean θ of the offspring distribution of a supercritical Galton–Watson branching process. Here we have $1 < \theta = E_\theta(X_1 \mid X_0 = 1)$ and we shall suppose that $\sigma^2 = \mathrm{var}_\theta(X_1 \mid X_0 = 1) < \infty$. In this case it is known (Heyde and Feigin [7]) that the conditional exponential family consists of the family of power series distributions. These are the distributions for which

$$P_j = \mathbf{P}(X_1 = j \mid X_0 = 1) = a_j \lambda^j [f(\lambda)]^{-1}, \quad j = 0, 1, 2, \ldots, \lambda > 0,$$

where $a_j \geqslant 0$ and $f(\lambda) = \sum_{j=0}^\infty a_j \lambda^j$. Then,

$$\theta = \lambda f'(\lambda)[f(\lambda)]^{-1}, \qquad \sigma^2 = [(\mathrm{d}/\mathrm{d}\theta) \log \lambda]^{-1}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \log f(x \mid y, \theta) = \sigma^{-2}(x - y\theta),$$

so that

$$\frac{\mathrm{d} \log L_n(\theta)}{\mathrm{d}\theta} = \sigma^{-2} Y_{n-1}(\hat{\theta}_n - \theta)$$

where $Y_k = \sum_{j=0}^k X_j$ and $\hat{\theta}_n = (Y_n - Y_0) Y_{n-1}^{-1}$. For this family, $\theta = \theta(\lambda)$ is known to be a non-negative monotone increasing function of λ (Patil [9]) so that the parametrization can be equally well expressed in terms of θ.

Suppose $X_0 = 1$ and $\mathbf{P}(X_1 = 0) = 0$ for definiteness and convenience. The latter gives $X_n \xrightarrow{\text{a.s.}} \infty$ as $n \to \infty$. If $\mathbf{P}(X_1 = 0) > 0$, it is well known that $X_n \xrightarrow{\text{a.s.}} \infty$ on the non-extinction set and the results which we shall describe hold conditionally on non-extinction.

It follows from Theorem 3 of Heyde [4] that $Y_n(E_\theta Y_n)^{-1} \xrightarrow{\text{a.s.}} \eta(\theta)$ with η non-degenerate and a.s. positive. Further,

$$I_n(\theta) = \sigma^{-2}(\theta) Y_{n-1},$$
$$J_n(\theta) = (\mathrm{d}\sigma^2(\theta)/\mathrm{d}\theta)(Y_n - 1 - \theta Y_{n-1}) - \sigma^{-2}(\theta) Y_{n-1},$$

and Assumption 1(i) is satisfied since convergence in probability holds uniformly for $\theta \geqslant 1+\varepsilon$, any $\varepsilon > 0$, via calculation of $E_\theta(Y_n(E_\theta Y_n)^{-1} - \eta)^2$ (e.g. Harris [3], Lemma 7.1). That

$$(E_\theta I_n(\theta))^{-1/2} \frac{\mathrm{d} \log L_n(\theta)}{\mathrm{d}\theta} \xrightarrow{d} N(0, 1) \quad \text{(mixing)}$$

follows from Corollary 2 of Scott [14] and uniformity of convergence for $\beta \geqslant 1+\varepsilon$, any $\varepsilon > 0$, can be gleaned from a detailed examination of the proofs of Theorem 1 and Corollary 2 of the same paper. Assumption 1(ii) is therefore satisfied. Also, $\sigma^2(\theta)$ is continuous in θ and differentiable so that Assumption 2 holds, as noted previously.

The application of our theorem in this context substantially clarifies Heyde's [5] investigation of the class of estimators which are called asymptotically efficient for θ.

Acknowledgment

Helpful comments from P.D. Feigin, P. Jagers and a referee on an earlier version of this paper are gratefully acknowledged.

References

[1] G.K. Eagleson, Some simple conditions for limit theorems to be mixing, Teoriya Veroyat. i ee Primenen 21 (1976) 653–660.
[2] P.G. Hall, Martingale invariance principles, Ann. Probab. 5 (1977) 875–887.
[3] T.E. Harris, Branching processes, Ann. Math. Statist. 19 (1948) 474–494.
[4] C.C. Heyde, Extension of a result of Seneta for the supercritical Galton–Watson process, Ann. Math. Statist. 41 (1970) 739–742.
[5] C.C. Heyde, Remarks on efficiency in estimation for branching processes, Biometrika 62 (1975) 49–55.
[6] C.C. Heyde, An optimal property of maximum likelihood with application to branching process estimation, Bull. Int. Statist. Inst. 47 (1977) (to appear).
[7] C.C. Heyde and P.D. Feigin, On efficiency and exponential families in stochastic process estimation, in: G.P. Patil, S. Kotz and J.K. Ord, eds., Statistical Distributions in Scientific Work (Reidel, Dordrecht, Vol. 1, 1975) 227–240.

(8) D.L. McLeish, Dependent central limit theorems and invariance principles, Ann. Probab. 2 (1974) 620–628.
[9] G.P. Patil, Certain properties of the generalized power series distributions, Ann. Inst. Statist. Math. 14 (1962) 179–182.
[10] C.R. Rao, Criteria of estimation in large samples, Sankhyā, Ser. A, 25 (1963) 189–206.
[11] C.R. Rao, Linear statistical Inference and its Applications (Wiley, New York, 1965).
[12] G.G. Roussas, Contiguity of Probability Measures (Cambridge University Press, Cambridge, 1972).
[13] L. Schmetterer, On the asymptotic efficiency of estimates, in: F.N. David, ed., Research Papers in Statistics. Festschrift for J. Neyman (Wiley, New York, 1966).
[14] D.J. Scott, A central limit theorem for martingales and an application to branching processes, Stoch. Processes Appl. 6 (1977) 241–252.
[15] L. Weiss and J. Wolfowitz, Generalized maximum likelihood estimators, Theory Prob. Appl. 11 (1966) 58–81.