PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 18, Fasc. 1 (1998), pp. 19–31

THE ASYMPTOTIC CONSISTENCY AND EFFICIENCY OF FIXED-SIZE SEQUENTIAL CONFIDENCE SETS

BY

ROMAN RÓŻAŃSKI (WROCŁAW)

Abstract. In the paper, a sequential confidence set based on an estimation process of a multivariate parameter is constructed. Under the assumption that the estimation process scaled by an increasing positive process has an asymptotic distribution it is proved that the sequential confidence set is asymptotically consistent and asymptotically efficient. The results are applied to the sequential confidence sets based on maximum likelihood estimators of a multivariate parameter in the iid case and in the exponential class of processes.

1. Introduction. Let $Y_1, Y_2, \ldots, Y_n$ be a sequence of independent, identically distributed random variables with $EY = \mu$ and $\operatorname{Var} Y = \sigma^2 < \infty$. If σ^2 is known, then the fixed sample size confidence interval with endpoints $\bar{X}_n \pm d$ has an approximate coverage probability $1-\alpha$ and prescribed width $2d$, where $d \approx \sigma u_\alpha / \sqrt{n}$ and $\Phi(u_\alpha) = 1-\alpha/2$. When σ is unknown, one can, of course, use a consistent estimator $\hat{\sigma}_n$ of σ obtaining the $1-\alpha$ approximate coverage probability but the width of the confidence interval, although tending to 0 almost surely, is not fixed (approximately). One can conclude that there does not exist any fixed sample size procedure for which the confidence interval, defined above, achieves approximately the $1-\alpha$ coverage probability and the prescribed width $2d$. In Chow and Robbins [3], a sequential confidence interval has been defined so that it is asymptotically consistent (the coverage probability converges to $1-\alpha$) and asymptotically efficient (the ratio of the expected random sample size to the best fixed sample size converges to 1 as the width of the confidence interval tends to 0). The results and ideas of [3] have been extended to other models both parametric and non-parametric. In Grambsch [7] a sequential fixed-width confidence interval for an unknown parameter based on the maximum likelihood estimator was considered. In the paper the stopping time has been defined under the assumption that the Fisher information as a function of the unknown parameter can be derived. Under the assumption, the asymptotical consistency of the sequential confidence interval has been proved.

In [8] Grambsch extends her results concerned with the asymptotical consistency to the multidimensional case. The authoress proposes a stopping time based on estimate of the smallest eigenvalue of the Fisher information matrix. Thus we must still know the functional dependence of the Fisher information matrix on the unknown parameter. The results have been applied to the logistic regression problem. In [15] Yu has proved that the sequential fixed-width confidence interval based on the maximum likelihood estimator (in the iid case) of a one-dimensional parameter and the stopping time defined through the empirical Fisher information is asymptotically consistent and asymptotically efficient. In Glynn and Whitt [6] the most general model, connected with some applications to stochastic simulation, has been considered. Under the assumption that the estimation process satisfies some version of functional limit law the authors prove that the sequential fixed-area confidence set defined by them is asymptotically consistent.

The paper is organized as follows. In Section 2 the results of Glynn and Whitt [6] is generalized. Namely, the assumption on the functional limit law is replaced by the one that the estimation process scaled by an increasing positive process has some asymptotical distribution. A sequential fixed-area confidence set has been defined. The confidence set is proved to be asymptotically consistent and under some additional assumption asymptotically efficient as well (with respect to the scaling process).

In Section 3 the results contained in Yu [15] are extended to the multidimensional case. A sequential fixed-area confidence set which is based on the maximum likelihood estimator (in the iid case) of an unknown multivariate parameter is defined. The appropriate stopping time is defined through the empirical Fisher matrix. For such a confidence set, asymptotical consistency and asymptotical efficiency are proved. It is worth noting that the shape of the confidence set may be chosen in an arbitrary way, it may be either rectangle or ellipsoid or another one.

In Section 4 the problem of sequential confidence set estimation for the exponential class of multivariate process is considered. It is a wide class of processes containing diffusion processes, Markov processes with a finite set of states and counting processes. We apply the results of Section 2 to get asymptotical consistency and asymptotical efficiency of the sequential confidence set for the class of processes.

In Section 5 some remarks concerned with other possible applications of the results are given. Finally, Section 6 contains proofs of propositions and theorems formulated in previous sections.

2. Fixed-width sequential confidence sets. General statement. Let $X = \{X(t), t \geqslant 0\}$ be an R^d-valued process called the *estimation process* for estimating the unknown parameter $\theta \in R^d$. The time parameter t may be either discrete or continuous. For instance, $\{X(t)\}$ may be a sequence of maximum likelihood estimators of the parameter θ, where t denotes the size of a sample. Assume that the estimation process X satisfies the following conditions:

CONDITION 2.1. *There exists a non-singular $d \times d$ matrix Γ, a constant γ and a stochastic process $S(t)$ whose almost all realizations are non-negative continuous functions converging to ∞ as $t \to \infty$. In addition, assume that there exists an R^d-valued random variable Y with distribution F_Y such that*

$$S(t)^{\gamma}\big(X(t)-\theta\big) \Rightarrow \Gamma Y \quad \text{as } t \to \infty.$$

Typically, Condition 2.1 is satisfied for $\gamma = \frac{1}{2}$ and a normally distributed random variable Y.

Additionally, we assume that the process $X(t)$ is uniformly continuous in probability, which means that

CONDITION 2.2 (see Anscombe [1] and Gut [9]). *For every ε there exists a δ such that*

$$P\big(\max_{0<h<t\delta} |X(t+h)-X(t)| > \varepsilon\big) < \varepsilon$$

for all t greater than or equal to some t_0.

To construct a confidence set for θ on the approximate level $1-\alpha$ we assume that there exists a bounded set A for which

CONDITION 2.3. *$P(Y \in A) = 1-\alpha$ and $P(Y \in \partial A) = 0$, where ∂A is the boundary of A.*

We assume throughout the paper that $m(A) > 0$, where m is the Lebesgue measure in R^d. Let

$$C_\Gamma(t) = X(t) - S(t)^{-\gamma}\Gamma A.$$

The following proposition shows that $C_\Gamma(t)$ is an appropriate confidence set achieving the level $1-\alpha$ as $t \to \infty$.

PROPOSITION 2.1. *If Condition 2.1 is satisfied, then*

$$P\big(\theta \in C_\Gamma(t)\big) \to 1-\alpha \quad \text{as } t \to \infty.$$

Unfortunately, the matrix Γ is also unknown. Therefore, it must be estimated. Assume that there exists an estimator $\Gamma(t)$ of Γ which is weakly consistent. This means that $\Gamma(t) \Rightarrow \Gamma$ as $t \to \infty$. Let

$$C(t) = X(t) - S(t)^{-\gamma}\Gamma(t)A.$$

Then we have the following

PROPOSITION 2.2. *If $\Gamma(t) \Rightarrow \Gamma$ and Conditions 2.1 and 2.3 hold, then*

$$P\big(\theta \in C(t)\big) \to 1-\alpha \quad \text{as } t \to \infty.$$

Let us turn now to the definition of a sequential stopping rule for construction of an approximate sequential fixed-size confidence set. Let

$$T(\varepsilon) = \inf\{t \geqslant 0:\ m\big(C(t)\big)^{1/d} + a(t) \leqslant \varepsilon\},$$

where $a(t)$ is a strictly positive stochastic process decreasing monotonically to 0 as $t \to \infty$ and satisfying $a(t) = o\big(S(t)^{-\gamma}\big)$ almost surely. The process $a(t)$ guarantees that the stopping rule $T(\varepsilon)$ does not terminate too early. We have $T(\varepsilon) \geqslant t(\varepsilon)$, where

$$t(\varepsilon) = \inf\{t \geqslant 0:\ a(t) \leqslant \varepsilon\},$$

and thus $T(\varepsilon) \geqslant t(\varepsilon) \to \infty$ as $\varepsilon \to 0$. To prove theorems about asymptotic validity of the stopping rule $T(\varepsilon)$ let us assume that the following strong consistency condition holds:

CONDITION 2.4. *There exists a strongly consistent estimator $\Gamma(t)$ of the matrix Γ, i.e. $\Gamma(t) \to \Gamma$ almost surely as $t \to \infty$.*

Under the conditions given above we can show the following theorem:

THEOREM 2.1. *If Conditions 2.1–2.4 hold and $t \to \infty$ or $\varepsilon \to 0$, then*

(i) $S(t)^{\gamma}\big(m\big(C(t)^{1/d}\big) + a(t)\big) \to m(\Gamma A)^{1/d}$ *almost surely,*

(ii) $\varepsilon^{1/\gamma} S\big(T(\varepsilon)\big) \to m(\Gamma A)^{1/\gamma d}$ *almost surely,*

(iii) $\varepsilon^{-1}\Big(m\big(C\big(T(\varepsilon)\big)\big)\Big)^{1/d} \to 1$ *almost surely,*

(iv) $\varepsilon^{-1}\big[X\big(T(\varepsilon)\big) - \theta\big] \Rightarrow m(\Gamma A)^{-1/d}\,\Gamma Y$,

(v) $P\big(\theta \in C\big(T(\varepsilon)\big)\big) \to 1 - \alpha$.

The results of Theorem 2.1 resemble those of Theorem 1 in Glynn and Whitt [6]. The difference is that the assumption (2.1) in [6], concerned with the functional central limit theorem for the estimation process, is replaced by Conditions 2.1 and 2.2.

The result (ii) of Theorem 2.1 can be strengthened. Namely, under some additional assumption on the estimator $\Gamma(t)$ and the sequence $a(t)$, the following theorem asserts that the expected value of $S\big(T(\varepsilon)\big)$ is asymptotically equivalent to the optimal non-random size of the $1-\alpha$ confidence set.

THEOREM 2.2. *If $E\sup_t \|\Gamma(t)^{1/\gamma}\| < \infty$ and $a(t) = o\big(S(t)^{-1}\big)$, then*

$$\varepsilon^{1/\gamma} ES\big(T(\varepsilon)\big) \to m(\Gamma A)^{1/\gamma d} \quad \text{as } \varepsilon \to 0.$$

Remark 2.1. Here, we define $\|\Gamma\|$ as $\sum_i \sum_j |\gamma_{i,j}|$.

3. Sequential fixed-width confidence sets associated with the maximum likelihood estimation of a multivariate parameter. Let $U_1, U_2, \ldots, U_n, \ldots$ be a sequence of independent, identically distributed random variables with density function $f(u, \theta)$ depending on an unknown parameter $\theta \in R^d$. Assume that the following Cramer regularity conditions hold:

CONDITION 3.1. *For almost all u, all $i, k, l = 1, 2, \ldots, d$ and all $\theta \in R^d$ the derivatives*

$$\frac{\partial}{\partial \theta_i} \log f(u, \theta), \quad \frac{\partial^2}{\partial \theta_i \partial \theta_k} \log f(u, \theta), \quad \frac{\partial^3}{\partial \theta_i \partial \theta_k \partial \theta_l} \log f(u, \theta)$$

exist. Moreover, assume that there exist functions G_i, $H_{k,l}$, $N_{k,l}$ such that

$$\left| \frac{\partial}{\partial \theta_i} f(u, \theta) \right| < G_i(u), \quad \left| \frac{\partial^2}{\partial \theta_k \partial \theta_l} f(u, \theta) \right| < H_{k,l}(u)$$

and

$$\left| \frac{\partial^2}{\partial \theta_i \partial \theta_k} \log f(u, \theta) \right| < N_{i,k}(u),$$

where

$$E_\theta G_i(U)/f(U, \theta) < \infty, \quad E_\theta H_{k,l}(U)/f(U, \theta) < \infty$$

and

$$E_\theta N_{i,k}(U) \log N_{i,k}(U) < \infty.$$

CONDITION 3.2. *For each $i, k, l = 1, 2, \ldots, d$ there exist functions $J_{i,k,l}$ such that for every θ and for almost all u*

$$\left| \frac{\partial^3}{\partial \theta_i \partial \theta_k \partial \theta_l} \log f(u, \theta) \right| < J_{i,k,l}(u) \quad \text{and} \quad \sup_\theta E_\theta J_{i,k,l}(U) < \infty.$$

CONDITION 3.3. *For every θ the matrix*

$$\Sigma(\theta) = \left[E_\theta \frac{\partial}{\partial \theta_j} \log f(U, \theta) \frac{\partial}{\partial \theta_k} \log f(U, \theta) \right]$$

is positive definite.

Under these conditions it is well known that the maximum likelihood estimator $\hat{\theta}_n$ of the parameter θ is strongly consistent and asymptotically normal, i.e. $\hat{\theta}_n \to \theta$ almost surely and

$$\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow \mathcal{N}(0, \Sigma_\theta^{-1}) \quad \text{as } n \to \infty.$$

Putting $S(t) = n$ and $X(t) = \hat{\theta}_n$ we infer that Condition 2.1 holds with $\gamma = \frac{1}{2}$, $\Gamma = (\Sigma(\theta)^{-1})^{1/2}$, and Y being a random variable $\mathcal{N}(0, I)$ distributed. Moreover, it is easy to see that the sequence $\hat{\theta}_n$ satisfies Condition 2.2, which means that the sequence of maximum likelihood estimators of θ is uniformly continuous in probability.

Let A be a set for which Condition 2.3 is satisfied and $C_\Gamma(n) = \hat{\theta}_n - n^{-1/2} \Gamma A$, where $\Gamma = (\Sigma^{-1}(\theta))^{1/2}$. Since the matrix $\Sigma(\theta)$ depends on the unknown parameter θ, the matrix Γ is also unknown. Therefore, as an estimator of Γ we take $\Gamma(n) = (\hat{\Sigma}^{-1}(\hat{\theta}_n))^{1/2}$, where

$$\hat{\Sigma}(\hat{\theta}_n) = \left[-\frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial^2}{\partial \theta_j \partial \theta_k} \log f(U_i, \theta) \right)_{\theta = \hat{\theta}_n} \right],$$

where $j, k = 1, 2, \ldots, d$. Consequently, we can define the sequence

$$C(n) = \hat{\theta}_n - n^{-1/2} \Gamma(n) A$$

of confidence sets and the following stopping time:

$$T(\varepsilon) = \inf\{n > 0: \ m(C(n))^{1/d} + a(n) \leqslant \varepsilon\},$$

where, as in Section 2, m denotes the Lebesgue measure in R^d and $a(n) = o(n^{-1/2})$. Now, applying Theorems 2.1 and 2.2 we obtain the following results:

THEOREM 3.1. *If Conditions* 3.1–3.3 *hold and* $n \to \infty$ *or* $\varepsilon \to 0$, *then*

(i) $n^{\gamma}\big(m(C(n)^{1/d}) + a(n)\big) \to m(\Gamma A)^{1/d}$ *almost surely,*

(ii) $\varepsilon^{1/\gamma} T(\varepsilon) \to m(\Gamma A)^{1/\gamma d}$ *almost surely,*

(iii) $\varepsilon^{-1}\big(m(C(T(\varepsilon)))\big)^{1/d} \to 1$ *almost surely,*

(iv) $\varepsilon^{-1}[\hat{\theta}_{T(\varepsilon)} - \theta] \Rightarrow m(\Gamma A)^{-1/d} \Gamma Y$,

(v) $P\big(\theta \in C(T(\varepsilon))\big) \to 1 - \alpha$.

THEOREM 3.2. *If Conditions* 3.1–3.3 *hold and* $a(n) = o(n^{-1})$, *then*

$$\varepsilon^{1/\gamma} ET(\varepsilon) \to m(\Gamma A)^{1/\gamma d} \quad \text{as } \varepsilon \to 0.$$

4. Fixed-area sequential confidence sets based on the exponential family of stochastic processes. Let $(\Omega, \mathscr{F})$ be a measurable space with a class of probability measures $(P_\theta, \theta \in \Theta)$, $\Theta \subset R^d$, and with a stochastic process Y. By $\mathscr{F}_t$ we denote the right-continuous filtration generated by the process Y observed on the interval $[0, t]$. Let P_θ^t be the restriction of P_θ to $\mathscr{F}_t$. The process Y is supposed to belong to the exponential class of processes, which means we assume that there exists a measure P on $(\Omega, \mathscr{F})$ such that, for all $t > 0$ and $\theta \in \Theta$, $P_\theta^t \ll P^t$ and

$$L_t(\theta) = \frac{dP_\theta^t}{dP_t} = \exp\big(\theta^T A_t - \kappa(\theta) S_t\big), \tag{1}$$

where T denotes the operation of transposition. The density $L_t(\theta)$ is the likelihood function for the process Y observed on the interval $[0, t]$. The process A is a d-dimensional vector process which is supposed to be right-continuous with limits from the left. It is also assumed that the process S_t is a non-decreasing predictable process for which $S_0 = 0$, $S_t \to \infty$ as $t \to \infty$, and almost all realizations of the process S_t are continuous functions. The exponential class of processes defined by formula (1) contains many important classes of processes including exponential families of diffusions, exponential families of counting processes and Markov processes with finite state-space (see Küchler and Sørensen [10], Stefanov [14], Sørensen [13]). Now, assume that the following conditions hold:

CONDITION 4.1. *The function $\kappa(\cdot)$ in (1) is a steep strictly convex function* (see Küchler and Sørensen [10] for the definition of a steep function), *$P_\theta(A_t/S_t \in \partial C) = 0$, where C denotes the closed convex support of the random variable B_u/u ($B_u = A_{\tau(u)}$ and $\tau(u) = \inf\{t\colon S_t \geqslant u\}$). Moreover, suppose that $\theta \in \Theta$.*

CONDITION 4.2. *There exists an increasing non-random function $\phi_\theta(t)$ such that under P_θ we have $S_t/\phi_\theta(t) \to \eta^2(\theta)$ in probability as $t \to \infty$, where $\eta^2(\theta)$ is a strictly positive finite random variable.*

CONDITION 4.3. *The function $\nabla_\theta \kappa$ is a homeomorphism from R^d onto R^d.*

Using a random change of time argument, Küchler and Sørensen [10] have proved (see also Stefanov [14]) that under Conditions 4.1–4.3 the maximum likelihood estimator $\hat{\theta}_t$ of the parameter θ exists. The estimator is uniquely determined and

$$\hat{\theta}_t = \nabla_\theta^{-1} \kappa(A_t/S_t).$$

Moreover, under P_θ, $\hat{\theta}_t \to \theta$ almost surely and

$$\sqrt{S_t}(\hat{\theta}_t - \theta) \Rightarrow \mathcal{N}(0, \Gamma^2) = \Gamma Y,$$

where Y is a random variable with the $\mathcal{N}(0, I)$ distribution and

$$\Gamma^2 = \Sigma^{-1}, \quad \Sigma(\theta) = \left[\frac{\partial^2}{\partial\theta_j \partial\theta_k} \kappa(\theta)\right]_{j,k=1}^{d}$$

To estimate the unknown matrix Γ we apply

$$\Gamma(t) = \left(\Sigma^{-1}(\hat{\theta}_t)\right)^{1/2}$$

which is obviously a strongly consistent estimator of Γ. Thus we can turn to the construction of a sequential confidence set. Let A be a set for which Condition 2.3 is satisfied and let

$$C(t) = \hat{\theta}_t - S(t)^{-1/2} \Gamma(t) A$$

be a family of confidence sets. Defining the stopping time

$$T(\varepsilon) = \inf\left\{t \geqslant 0\colon\ m\big(C(t)\big)^{1/d} + a(t) \leqslant \varepsilon\right\},$$

where $a(t)$ is a strictly positive stochastic process decreasing monotonically almost surely to 0 as $t \to \infty$ and satisfying $a(t) = o\big(S(t)^{-1/2}\big)$, we get the following

THEOREM 4.1. *Under Conditions* 4.1–4.3 *the stopping time $T(\varepsilon)$ and the confidence sets $C(t)$ and $C\big(T(\varepsilon)\big)$ have all the properties formulated in Theorem* 2.1.

Moreover, under some additional assumptions the first order efficiency of the stopping time $T(\varepsilon)$ can be proved. Namely, we have

THEOREM 4.2. *If* $a(t) = o\big(S(t)\big)$ *and there exists* $C > 0$ *such that*

$$\|\Sigma(\theta)\| \leqslant C \|\nabla_\theta \kappa(\theta)\|^p \tag{2}$$

for some $0 \leqslant p \leqslant 2$, *then* $\varepsilon^2 ES\big(T(\varepsilon)\big) \to m(\Gamma A)^{2/d}$ *as* $\varepsilon \to 0$.

Remark 4.1. Actually, the condition given by formula (2) includes many important examples and it is not too restrictive. One could see that if $\theta \in R$, then all functions $\kappa(\theta)$ which are majorized by polynomials of θ or an exponential function of θ satisfy (2). The same is true for multivariate generalizations of one-dimensional models.

EXAMPLE 1 (see Sørensen [12] and Erlandsen and Sørensen [5] for details and practical applications of the model considered). Let Y_t be a solution to the following stochastic differential equation:

$$dY_t = \big(\theta\mu_t(Y) + \nu_t(Y)\big)dt + \sigma_t(Y)\,dW_t, \tag{3}$$

where $\theta \in \Theta \subset R$, W_t is a Wiener process, and μ_t, ν_t, σ_t are known non-anticipating functionals which satisfy conditions of the Lipschitz type guaranteeing the existence and uniqueness of a solution to the stochastic equation (3) for all $\theta \in \Theta$. Let us define the following stochastic processes:

$$A_t = \int_0^t \tilde{\sigma}_s(Y)^2 \mu_s(Y)\,dY_s - \int_0^t \tilde{\sigma}_s(Y)^2 \mu_s(Y)\,\nu_s(Y)\,ds \tag{4}$$

and

$$S_t = \int_0^t \tilde{\sigma}_s(Y)^2 \mu_s(Y)^2\,ds, \tag{5}$$

where

$$\tilde{\sigma}_t(y) = \begin{cases} \sigma_t(y)^{-1} & \text{if } \sigma_t(y) \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Assume now that $P_\theta\big(\sigma_t(Y) = 0\big) = 0$, $P_\theta(S_t < \infty) = 1$ for all $t \geqslant 0$ and all $\theta \in \Theta$. Then the measures P_θ^t are mutually equivalent. If we choose $P^t = P_{\tilde{\theta}}^t$ for some $\tilde{\theta}$, then the Radon–Nikodym derivatives are of the form (1) with A_t and S_t given by (4) and (5) and with $\kappa(\theta) = \frac{1}{2}\theta^2$. In addition, assume that $P(\lim_{t\to\infty} S_t = \infty) = 1$ and that Condition 4.2 holds. Note that, obviously, Conditions 4.1 and 4.3 are satisfied. The maximum likelihood estimator $\hat{\theta}_t$ takes the form $\hat{\theta}_t = A_t S_t^{-1}$.

Let $A = [-u_\alpha, u_\alpha]$, where $\Phi(u_\alpha) = 1 - \alpha/2$. Since in the example $\ddot{\kappa}(\theta) = 1$ the confidence sets $C_\Gamma(t) = C(t)$ take the following form:

$$C(t) = [\hat{\theta}_t - S(t)^{-1/2} u_\alpha,\ \hat{\theta}_t + S(t)^{-1/2} u_\alpha].$$

By Theorems 4.1 and 4.2, the stopping time $T(\varepsilon)$ and the confidence set $C(T(\varepsilon))$ may be a useful tool for estimation and hypotheses testing as well.

EXAMPLE 2. Let N_t be a counting process with intensity $\lambda_t = \varrho U_t$, where U_t is a positive predictable stochastic process with finite expectation for all t. Then the likelihood function for the process N is given by (1) with $\theta = \log\varrho$, $\kappa(\theta) = e^{\theta}-1$, $A_t = N_t$ and $S_t = \int_0^t U_s\,ds$. Assuming that $\int_0^\infty U_s\,ds = \infty$ we see that all Conditions 4.1–4.3 are satisfied. The maximum likelihood estimator for θ is $\hat{\theta}_t = \log A_t S_t^{-1}$ and the estimator $\Gamma(t) = \exp(-\hat{\theta}_t)$. Then, if $[-u_\alpha, u_\alpha]$ is as in Example 1, the confidence set $C(t)$ takes the form

$$C(t) = \left[\log\left(\frac{N_t}{S_t}\right) - \frac{u_\alpha}{\sqrt{N_t}},\ \log\left(\frac{N_t}{S_t}\right) + \frac{u_\alpha}{\sqrt{N_t}}\right].$$

Further, by Theorems 4.1 and 4.2 we have the desired asymptotic validity and efficiency of the first order for the stopping time $T(\varepsilon)$ and the confidence set $C(T(\varepsilon))$.

5. Concluding remarks. It is worth noting that the results of Section 2 could be applied to the cases when the partial likelihood method of estimation is used. In the recent paper by Martinsek [11] a sequential confidence ellipsoid for the multivariate parameter of logistic regression has been constructed. The confidence set has been proved to be asymptotically consistent and asymptotically efficient. Similar results but in a simpler way could be proved by using the results of Section 2. It is also possible to obtain asymptotically consistent and efficient confidence sets based on M-estimators.

It is also worth of interest to construct such confidence sets for estimation of the intensity function in a multiplicative model of a counting process. The results will be presented in the forthcoming paper.

6. Proofs.

Proof of Proposition 2.1. The matrix Γ is non-singular, and thus

$$P(\theta \in C_\Gamma(t)) = P(\Gamma^{-1} S(t)^\gamma (X(t)-\theta) \in A).$$

By Condition 2.1, we have

$$\Gamma^{-1} S(t)^\gamma (X(t)-\theta) \Rightarrow \Gamma^{-1}\Gamma Y = Y \quad \text{as } t \to \infty,$$

and since $P(Y \in \partial A) = 0$, we obtain

$$P(\Gamma^{-1} S(t)^\gamma (X(t)-\theta) \in A) \to P(Y \in A) = 1-\alpha \quad \text{as } t \to \infty,$$

which completes the proof of the proposition.

Proof of Proposition 2.2. By Condition 2.1 and Theorem 4.4 of Billingsley [2], the random vector

$$[\Gamma(t), S(t)^\gamma (X(t)-\theta)] \Rightarrow [\Gamma, Y] \quad \text{as } t \to \infty.$$

Applying the continuous mapping theorem (Theorem 5.1 of Billingsley [2]) we obtain

$$\Gamma(t)^{-1} S(t)^{\gamma}\big(X(t)-\theta\big) \Rightarrow \Gamma^{-1}\Gamma Y = Y \quad \text{as } t \to \infty.$$

Further the proof goes along the lines of the proof of Proposition 2.1.

Proof of Theorem 2.1. The proof of the theorem goes along the lines of that of Theorem 1 in Glynn and Whitt [6], and therefore it is omitted.

Proof of Theorem 2.2. Using Theorem 2.1 (ii) it is enough to prove that $\varepsilon^{1/\gamma} S\big(T(\varepsilon)\big)$ is uniformly integrable with respect to ε. Let $V(t) = m\big(C(t)\big)^{1/d} + a(t)$. Note that, by the spatial invariance and the scaling properties of the Lebesgue measure m, we have

$$m\big(X(t)-S(t)^{-\gamma}\Gamma(t)A\big)^{1/d} = S(t)^{-\gamma} m\big(\Gamma(t)A\big)^{1/d} = S(t)^{-\gamma}\big(\det\big(\Gamma(t)m(A)\big)\big)^{1/d}.$$

By the definition of $T(\varepsilon)$ we obtain $V\big(T(\varepsilon)-1\big) > \varepsilon$. Since $E\sup_t \|\Gamma(t)^{1/\gamma}\| < \infty$ and $a(t) = o\big(S(t)^{-1}\big)$, we can write

$$\begin{aligned}\sup_{\varepsilon}\int_B \varepsilon^{1/\gamma} S\big(T(\varepsilon)\big)\,dP &\leqslant \sup_{\varepsilon}\int_B S\big(T(\varepsilon)\big)\,V\big(T(\varepsilon)-1\big)^{1/\gamma}\,dP \\ &\leqslant M\sup_{\varepsilon}\int_B \big(\det\big(\Gamma\big(T(\varepsilon)-1\big)m(A)\big)\big)^{1/d\gamma}\,dP \\ &\leqslant M\sup_{\varepsilon}\int_B \big(\sup_t\big(\|\Gamma(t)^{1/\gamma}\|\big)m(A)\big)^{1/d\gamma}\,dP \to 0 \quad \text{as } P(B)\to 0.\end{aligned}$$

Consequently, $\varepsilon^{1/\gamma} S\big(T(\varepsilon)\big)$ is uniformly integrable with respect to ε, which by Theorem 2.1 (ii) completes the proof of Theorem 2.2.

Proof of Theorem 3.1. Obviously, under Conditions 3.1–3.3, Conditions 2.1–2.3 are in force. To check the validity of Condition 2.4 let us note that, by the strong consistency of the maximum likelihood estimator $\hat{\theta}_n$ and the strong law of large numbers,

$$\hat{\Gamma}_n = \hat{\Sigma}(\hat{\theta}_n) = \left[-\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial^2}{\partial\theta_j\,\partial\theta_k}\log f(U_i,\theta)\right)_{\theta=\hat{\theta}_n}\right]$$

converges almost surely to the matrix $\Sigma(\theta)$ as $n \to \infty$. The continuity of square root and inverse operation for matrices imply that $\hat{\Gamma}_n$ converges almost surely to $\Gamma = (\Sigma^{-1})^{1/2}$ as $n \to \infty$. Finally, the theorem is a straightforward consequence of Theorem 2.1.

Proof of Theorem 3.2. To prove the theorem it is enough to show that $E_\theta \sup_n \|\Gamma(n)^2\| < \infty$. In the case considered we have $\Gamma(n)^2 = \hat{\Sigma}^{-1}(\hat{\theta}_n)$, where

$$\hat{\Sigma}(\hat{\theta}_n) = \left[-\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial^2}{\partial\theta_j\,\partial\theta_k}\log f(U_i,\theta)\right)_{\theta=\hat{\theta}_n}\right].$$

Denoting by $\lambda_{\min}(\Sigma)$ the smallest eigenvalue of a matrix Σ, let us note that

$$\sup_n \hat{\lambda}_{\min}(\hat{\Sigma}(\hat{\theta}_n)) \leqslant \sup_n \|\hat{\Sigma}(\hat{\theta}_n)\| \leqslant \sum_{j=1}^{d}\sum_{k=1}^{d} \sup_n \frac{1}{n}\sum_{i=1}^{n} \left|\left(\frac{\partial^2}{\partial\theta_j \partial\theta_k}\log f(U_i, \theta)\right)_{\theta=\hat{\theta}_n}\right|$$
$$\leqslant \sum_{j=1}^{d}\sum_{k=1}^{d} \sup_n \frac{1}{n}\sum_{i=1}^{n} N_{j,k}(U_i).$$

The random variables $N_{j,k}(U_i)$ are independent, identically distributed (with respect to (i)). Thus Condition 3.1 and Theorem 4.14 (Chow et al. [4]) imply that

$$E_\theta \sup_n \frac{1}{n}\sum_{i=1}^{n} N_{j,k}(U_i) < \infty.$$

Consequently, we can write

$$E_\theta \sup_n \hat{\lambda}_{\min}(\hat{\Sigma}(\hat{\theta}_n)) \leqslant \sum_{j=1}^{d}\sum_{k=1}^{d} E_\theta \sup_n \frac{1}{n}\sum_{i=1}^{n} N_{j,k}(U_i) < \infty,$$

which implies that the sequence $\hat{\lambda}_{\min}(\hat{\Sigma}(\hat{\theta}_n))$ is uniformly integrable. Since the estimator $\hat{\theta}_n$ is strongly consistent, it follows that $\hat{\lambda}_{\min}(\hat{\Sigma}(\hat{\theta}_n)) \to \lambda_{\min}(\Sigma(\theta))$ almost surely as $n \to \infty$, which together with uniform integrability gives

$$E_\theta \hat{\lambda}_{\min}(\hat{\Sigma}(\hat{\theta}_n)) \to \lambda_{\min}(\Sigma(\theta)).$$

Consequently,

$$E_\theta \sup_n \|\Gamma(n)^2\| \leqslant C E_\theta \sup_n \hat{\lambda}_{\max}(\Gamma(n)^2) = E_\theta \sup_n \frac{1}{\hat{\lambda}_{\min}(\hat{\Sigma}(\hat{\theta}_n))}$$
$$= \int_0^\infty P\left(\sup_n \frac{1}{\hat{\lambda}_{\min}(\hat{\Sigma}(\hat{\theta}_n))} > u\right) du = \int_0^\infty P\left(\inf_n \hat{\lambda}_{\min}(\hat{\Sigma}(\hat{\theta}_n)) < \frac{1}{u}\right) du$$
$$= \int_0^1 P\left(\inf_n \hat{\lambda}_{\min}(\hat{\Sigma}(\hat{\theta}_n)) < \frac{1}{u}\right) du + \int_1^\infty P\left(\inf_n \hat{\lambda}_{\min}(\hat{\Sigma}(\hat{\theta}_n)) < \frac{1}{u}\right) du < \infty,$$

which completes the proof.

Proof of Theorem 4.1. Let us define the stopping time

$$\tau_u = \inf\{t:\ S_t \geqslant u\}.$$

Since S is non-decreasing and continuous, the mapping $u \to \tau_u$ is strictly increasing and has an inverse function. Obviously, we have $S(\tau_u) = u$. Moreover, one can prove (see Stefanov [14] and Küchler and Sørensen [10]) that the process

$$B_u = A(S(\tau_u))$$

is a process with homogeneous independent increments for which $E_\theta B_u = \nabla_\theta \kappa(\theta) u$ and the covariance matrix

$$E_\theta\big(B_u - \nabla_\theta \kappa(\theta)\big)\big(B_u - \nabla_\theta \kappa(\theta)\big)^T = \Sigma(\theta)u = \left[\frac{\partial^2}{\partial\theta_j \partial\theta_k}\kappa(\theta)\right]_{j,k=1}^{d} u.$$

In fact, we can write that $\tau(S_t) = t$, and thus $A_t = B_{S(t)}$. Hence, all properties of the process A_t can be deduced by investigating the process B with changed random time. Consequently, using the random change time argument we obtain Conditions 2.1–2.4 from Conditions 4.1–4.3. Finally, Theorem 4.1 follows immediately from Theorem 2.1.

Proof of Theorem 4.2. Let $\lambda_{\min}(\Sigma)$ be the smallest eigenvalue of the matrix Σ. Then we can write

$$\sup_t \lambda_{\min}\big(\Sigma(\hat\theta_t)\big) \leqslant \sup_t \|\Sigma(\hat\theta_t)\| \leqslant C \sup_t \|\nabla_\theta \kappa(\hat\theta_t)\|^p.$$

Taking the expectation value on both sides and the formula for $\hat\theta_t$ we obtain

$$E \sup_t \lambda_{\min}\big(\Sigma(\hat\theta_t)\big) \leqslant C \sup_t \|\nabla_\theta \kappa(\hat\theta)_t\|^p$$

$$= CE \sup_t \|\nabla_\theta \nabla_\theta^{-1} \kappa(A_t/S_t)\|^p = CE \sup_t \|A_t/S_t\|^p.$$

In the same way as in the proof of Theorem 4.1 we can define the stopping time τ_u such that the process $B_u = A\big(S(\tau_u)\big)$ is a process with homogeneous independent increments. Moreover, the process

$$B_u - \nabla_\theta \kappa(\theta) u$$

is a square integrable martingale, which implies that

$$E \sup_t \|B_u/u\|^2 < \infty.$$

From the last statement we infer that also

$$E \sup_t \|A_t/S_t\|^p < \infty.$$

Finally, we have $E \sup_t \lambda_{\min}\big(\Sigma(\hat\theta_t)\big) < \infty$. Thus $\lambda_{\min}\big(\Sigma(\hat\theta_t)\big)$ is uniformly integrable with respect to t. The rest of the proof runs in the same way as the proof of Theorem 3.2.

REFERENCES

[1] F. J. Anscombe, *Sequential estimation*, J. Roy. Statist. Soc. Ser. B 15 (1953), pp. 1–21.

[2] P. Billingsley, *Convergence of Probability Measures*, Willey, New York 1968.

[3] Y. S. Chow and H. Robbins, *On the asymptotic theory of fixed-width sequential confidence intervals for the mean*, Ann. Math. Statist. 36 (1965), pp. 457–462.

[4] — and D. Siegmund, *Optimal Stopping. Great Expectations*, Willey, New York 1977.
[5] M. Erlandsen and M. Sørensen, *Statistical analysis of the variation of the oxygen concentration in a river by means of diffusion processes*, in: *Applied Statistics Symposium, Aarhus, January 1984*, L. S. Mortensen (Ed.), RECAU, Aarhus 1984.
[6] P. W. Glynn and W. Whitt, *The asymptotic validity of sequential stopping rules for stochastic simulations*, Ann. Appl. Probab. 2, No. 1 (1992), pp. 180–198.
[7] P. Grambsch, *Sequential sampling based on the observed Fisher information to guarantee the accuracy of the maximum likelihood estimator*, Ann. Statist. 11 (1983), pp. 68–77.
[8] — *Sequential maximum likelihood estimation with applications to logistic regression in case-control studies*, J. Statist. Plann. Inference 22 (1989), pp. 355–369.
[9] A. Gut, *Stopped Random Walks*, Springer, New York 1988.
[10] U. Küchler and M. Sørensen, *Exponential families of stochastic processes and Levy processes*, J. Statist. Plann. Inference 39 (1994), pp. 211–237.
[11] A. Martinsek, *Fixed-sized confidence regions for parameters of a logistic regression model*, Ann. Statist. 20, No. 4 (1992), pp. 1953–1969.
[12] M. Sørensen, *On maximum likelihood estimation in randomly stopped diffusion-type processes*, Internat. Statist. Rev. 51 (1983), pp. 93–110.
[13] — *On sequential maximum likelihood estimation for exponential families of stochastic processes*, ibidem 54 (1996), pp. 191–210.
[14] V. Stefanov, *Explicit limit results for minimal sufficient statistics and maximum likelihood estimators in some Markov processes: exponential families approach*, Ann. Statist. 23, No 4 (1995), pp. 1073–1101.
[15] K. F. Yu, *On fixed-width confidence intervals associated with maximum likelihood estimation*, J. Theoret. Probab. 2, No. 2 (1989), pp. 193–200.

Wrocław University of Technology
Wybrzeże Wyspiańskiego 27
PL-50-370 Wrocław, Poland
e-mail: rozanski@im.pwr.wroc.pl

Received on 9.5.1997