Annales Academiæ Scientiarum Fennicæ
Series A. I. Mathematica
Volumen 17, 1992, 29–38

ON QUASI-SCORE PROCESSES

A.A. Gushchin

Steklov Mathematical Institute, Vavilov Street 42, 117966 Moscow GSP-1, Russia

Abstract. In this paper we consider some aspects of quasi-likelihood methods used in estimation of parameters of stochastic processes.

1. Let P_θ be the distribution of an observed stochastic process or, more generally, let $(P_\theta, \theta \in \Theta \subseteq \mathbf{R}^k)$ be a family of probability measures on a measurable space $(\Omega, \mathscr{F})$ with a filtration $(\mathscr{F}_t, t \geq 0)$, and let Θ be an open subset of $\mathbf{R}^k$. In what follows we do not suppose that our model is fully specified. This means that, if we know the true value θ of the parameter, we do not know exactly the measure P_θ and we can only say that P_θ belongs to some family $\mathscr{P}_\theta$ of probability measures on $(\Omega, \mathscr{F})$.

Let us present some typical examples. Examples 2 and 4 are especially important. Examples 1 and 3 are particular cases of Examples 2 and 4, respectively. They are presented in order to illustrate the main ideas in a simple framework.

Example 1. We observe a stochastic sequence $(X_n, n \geq 1)$ of the form

$$X_n = \theta X_{n-1} + \varepsilon_n, \qquad X_0 = 0, \tag{1}$$

where $\theta \in \Theta \subseteq \mathbf{R}^1$, ε_n are independent random variables with zero mean and finite variance σ^2 (the distribution of ε_n may depend on n and θ). Here $\Omega = \{(x_1, \dots, x_n, \dots), x_n \in \mathbf{R}^1\}$, $\mathscr{F}_t = \sigma\{X_n, n \leq t\}$, $\mathscr{F} = \bigvee_t \mathscr{F}_t$, $\mathscr{P}_\theta$ consists of all possible distributions of the sequence (X_n) satisfying (1) under the above assumptions on the distribution of the sequence (ε_n).

Example 2. We observe a d-dimensional stochastic process $X = (X_t, t \geq 0)$ of the form

$$X_t = \int_0^t f_s(\theta)\, d\lambda_s + m_t(\theta),$$

where (λ_t) is a real, increasing, right-continuous, predictable process with $\lambda_0 = 0$, $\big(f_t(\theta)\big)$ is a d-dimensional predictable process, and $\big(m_t(\theta)\big)$ is a d-dimensional locally square integrable martingale with $m_0(\theta) = 0$ and the quadratic characteristic $\langle m(\theta)\rangle_t = \int_0^t a_s(\theta)\, d\lambda_s$, where $\big(a_s(\theta)\big)$ is a predictable process with values in the space of symmetric, non-negative definite $d \times d$ matrices. Predictability and the martingale property are considered with respect to the filtration generated by X. In the present case Ω, $\mathscr{F}$, $(\mathscr{F}_t)$ and $\mathscr{P}_\theta$ are constructed as in Example 1.

doi:10.5186/aasfm.1992.1705

Example 3. We observe a diffusion-type process $X = (X_t, t \geq 0)$ of the form

$$dX_t = a_t(\theta; X, Y)\, dt + dW_t, \qquad X_0 = 0,$$

where W is a Wiener process, the process Y is also observable, but its distribution is unknown.

Analogously, we can consider a counting process X with the intensity depending on θ and another observable process Y, see Greenwood and Wefelmeyer [8].

Example 4. We observe a d-dimensional semimartingale X with a triplet $\big(B(\theta), C(\theta), \nu(\theta)\big)$ of predictable characteristics. The family $\mathscr{P}_\theta$ consists of all solutions to the corresponding martingale problem. More formally: a filtered space $\big(\Omega, \mathscr{F}, (\mathscr{F}_t, t \geq 0)\big)$, a d-dimensional cadlag adapted process X on $\big(\Omega, \mathscr{F}, (\mathscr{F}_t)\big)$, a triplet $\big(B(\theta), C(\theta), \nu(\theta)\big)$, $\theta \in \Theta \subseteq \mathbf{R}^k$ (a candidate for the characteristics of X, see Jacod and Shiryaev [15, III-2-3]) and an initial condition $P_{\theta,0}$, $\theta \in \Theta$, on $(\Omega, \mathscr{F}_0)$ are given. Then $\mathscr{P}_\theta = \mathscr{S}\big(\mathscr{F}_0, X \mid P_{\theta,0}; B(\theta), C(\theta), \nu(\theta)\big)$, see Definition III-2-4 in Jacod and Shiryaev [15].

2. Let us describe a general method of parameter estimation in the present setting. We follow, in many respects, the paper of Godambe and Heyde [7]. This method includes the standard methods of estimation: maximum likelihood, least-squares, etc., under mild regularity conditions.

Let $G = G_t(\theta; \omega)$ be an $\mathbf{R}^k$-valued function of three arguments: observation $\omega \in \Omega$ (we will omit this argument in what follows), time $t \geq 0$, and parameter $\theta \in \Theta$. It is supposed that, for every fixed θ, the process $G(\theta) = \big(G_t(\theta), t \geq 0\big)$ is a k-dimensional locally square integrable martingale with respect to each measure P_θ from $\mathscr{P}_\theta$, and $G_0(\theta) = 0$ for each θ. The function G is usually called an estimating function or an estimating process. The estimator $\hat{\theta}_t$ corresponding to G is the solution of the estimating equation $G_t(\theta) = 0$ (the terms 'M-estimator' or 'martingale estimator' are often used for such estimators). The problems of consistency and asymptotic normality of M-estimators as well as some other estimators were considered, in particular, in the series of papers of Chitashvili, Lazrieva and Toronjadze in a rather general scheme, see Lazrieva, Toronjadze and Chitasvili [2–4] and references given there.

Here we want to compare estimators corresponding to different estimating processes. Let $k = 1$ for simplicity. Suppose, for each θ, the process $\dot{G}(\theta) = \big(\dot{G}_t(\theta), t \geq 0\big)$ (a dot stands for differentiation with respect to θ) is a special semimartingale with respect to every $P_\theta \in \mathscr{P}_\theta$ with a canonical decomposition

$$\dot{G}_t(\theta) = M_{G,t}(\theta) + \overline{G}_t(\theta),$$

where $M_G(\theta)$ is a P_θ-local martingale and $\overline{G}(\theta)$ is predictable with finite variation. (It should be noted that the processes $M_G(\theta)$, $\overline{G}(\theta)$ (and $\big\langle G(\theta)\big\rangle$ later) may depend, generally speaking, on the choice of P_θ in the class $\mathscr{P}_\theta$, so our notation is slightly ambiguous.) Then, under rather broad assumptions,

$$\langle G(\theta)\rangle_t^{-1/2}\overline{G}_t(\theta)(\hat{\theta}_t-\theta)\longrightarrow N(0,1) \tag{2}$$

in law under P_θ as $t\to\infty$.

The asymptotic criterion of optimality of estimating processes proposed by Godambe and Heyde [7] is based on (2). Let $\mathscr{G}$ be a class of estimating processes satisfying (2) for each θ and $P_\theta\in\mathscr{P}_\theta$. Then a process G is optimal in $\mathscr{G}$ if it maximizes the limit of the expression

$$\frac{\overline{G}_t^2(\theta)}{\langle G(\theta)\rangle_t}. \tag{3}$$

But typically, if the class $\mathscr{G}$ is large enough, there exists a process $Q\in\mathscr{G}$ such that

$$\overline{G}_t(\theta)=-\langle G(\theta),Q(\theta)\rangle_t \tag{4}$$

for all $t\geq 0$, $G\in\mathscr{G}$, $\theta\in\Theta$ and $P_\theta\in\mathscr{P}_\theta$. (Nothing changes if we multiply Q, or the right-hand side of (4), by any non-zero constant, but it is more convenient to take this constant to be -1.) Then the process Q maximizes the ratio (3) for every fixed t because of the Kunita–Watanabe inequality

$$\frac{\overline{G}_t(\theta)^2}{\langle G(\theta)\rangle_t}=\frac{\langle G(\theta),Q(\theta)\rangle_t^2}{\langle G(\theta)\rangle_t}\leq\langle Q(\theta)\rangle_t.$$

Thus, the estimator corresponding to the process Q is optimal in the sense of (2).

Another a so-called fixed sample optimality criterion, going back to the paper of Godambe [6], is also considered by Godambe and Heyde [7]. They suggest maximizing

$$\frac{\left(\mathbf{E}_\theta\dot{G}_t(\theta)\right)^2}{\mathbf{E}_\theta G_t(\theta)^2} \tag{5}$$

at fixed t ($\mathbf{E}_\theta$ stands for the expectation under P_θ). But if $M_G(\theta)$ is a martingale, $G(\theta)$ is a square integrable martingale under P_θ and (4) is fulfilled for Q. Moreover,

$$\mathbf{E}_\theta\dot{G}_t(\theta)=\mathbf{E}_\theta\overline{G}_t(\theta)=-\mathbf{E}_\theta\langle G(\theta),Q(\theta)\rangle_t,$$

$$\mathbf{E}_\theta G_t(\theta)^2=\mathbf{E}_\theta\langle G(\theta)\rangle_t,$$

and
$$\frac{\left(\mathbf{E}_\theta \dot{G}_t(\theta)\right)^2}{\mathbf{E}_\theta G_t(\theta)^2} = \frac{\left(\mathbf{E}_\theta \langle G(\theta), Q(\theta)\rangle_t\right)^2}{\mathbf{E}_\theta \langle G(\theta)\rangle_t} \le \mathbf{E}_\theta \langle Q(\theta)\rangle_t,$$
by the Kunita–Watanabe inequality. Thus, a process Q satisfying (4) is optimal in the sense of the fixed sample criterion as well.

Instead of (2) we can also consider another type of asymptotic behaviour of an estimator:

$$(6) \qquad \left(\mathbf{E}_\theta G_t(\theta)^2\right)^{-1/2} \left(\mathbf{E}_\theta \dot{G}_t(\theta)\right)(\hat{\theta}_t - \theta) \longrightarrow N(0,1)$$

in law under P_θ as $t \to \infty$. It is usually valid for ergodic models, under more restrictive assumptions than (2). Here we are interested in maximizing the expression (5) as well.

It should be noted that another asymptotic behaviour of M-estimators (even maximum likelihood estimators) than the one stated in (2) and (6) is possible, see the review of Barndorff-Nielsen and Sørensen [1]. The argument showing that the process Q satisfying (4) is optimal does not lead to the desired result in this case.

In the multi-parameter case ($k > 1$) we use the following result instead of (2):

$$(2^*) \qquad \langle G(\theta)\rangle_t^{-1/2} \overline{G}_t(\theta)(\hat{\theta}_t - \theta) \longrightarrow N(0, I_k),$$

and
$$(\hat{\theta}_t - \theta)^T \overline{G}_t(\theta)^T \langle G(\theta)\rangle_t^{-1} \overline{G}_t(\theta)(\hat{\theta}_t - \theta) \longrightarrow \chi^2_k$$
in law under P_θ as $t \to \infty$. Here
$$\langle G(\theta)\rangle_t = \left(\langle G_i(\theta), G_j(\theta)\rangle_t, i,j \le k\right),$$
$$\dot{G}_t(\theta) = \left(\partial G_{i,t}(\theta)/\partial\theta_j, i,j \le k\right),$$
$\overline{G}_t(\theta)$ is defined as in the one-dimensional case, and T stands for transposition.

Analogously to the case $k = 1$, if there exists $Q \in \mathscr{G}$ such that

$$(4^*) \qquad \overline{G}_t(\theta) = -\langle G(\theta), Q(\theta)^T\rangle_t$$

for all $t \ge 0$, $G \in \mathscr{G}$, $\theta \in \Theta$ and $P_\theta \in \mathscr{P}_\theta$, then
$$\overline{G}_t(\theta)^T \langle G(\theta)\rangle_t^{-1} \overline{G}_t(\theta) \le \overline{Q}_t(\theta)^T \langle Q(\theta)\rangle_t^{-1} \overline{Q}_t(\theta) = \langle Q(\theta)\rangle_t$$
(here $\langle G(\theta), Q(\theta)^T\rangle_t = \left(\langle G_i(\theta), Q_j(\theta)\rangle_t, i,j \le k\right)$ and $A \le B$ means that the matrix $B - A$ is non-negative definite). Thus, the process Q satisfying (4*) maximizes (in the sense of the partial ordering indicated above) the matrix expression

$$(3^*) \qquad \overline{G}_t(\theta)^T \langle G(\theta)\rangle_t^{-1} \overline{G}_t(\theta)$$

for all $t \geq 0$, and Q is optimal in the class $\mathscr{G}$ in the sense of the limiting behaviour (2*) of the estimator corresponding to it. In the same way it can also be proved that Q maximizes the expression

$$(5^*)\qquad \big(\mathbf{E}_\theta \dot{G}_t(\theta)\big)^T \big(\mathbf{E}_\theta G_t(\theta) G_t(\theta)^T\big)\big(\mathbf{E}_\theta \dot{G}_t(\theta)\big).$$

3. Here we shall confine our attention on the relations (4) and (4*). Estimators corresponding to estimating processes will not be considered.

Let $\mathscr{G}$ be a class of estimating processes G for which the process $\overline{G}(\theta)$ is well-defined for all $\theta \in \Theta$. A process $Q \in \mathscr{G}$ will be called an optimal estimating process or a quasi-score process if it satisfies (4) ($k = 1$) or (4*) ($k > 1$) (see a justification later on for the second term). How can the optimal estimating process be found?

It is clear that, to prove (4*), it suffices to check (4) for every component G_i of the function $G = (G_1, \dots, G_k)^T$ and for every direction θ_j, $1 \leq j \leq k$. So we shall consider the case $\Theta \subseteq \mathbf{R}^1$ only.

Let us first assume that the class $\mathscr{P}_\theta$ contains only one measure P_θ for each θ. Then under certain differentiability (with respect to θ) conditions on the family (P_θ) the score process $V(\theta) = \big(V_t(\theta), t \geq 0\big)$ can be defined as the derivative (with respect to θ) of the logarithm of the likelihood ratio process or by using some other differentiability concepts (for example, see the definition of local (in time) differentiability of the family (P_θ) in Jacod [13]). It was noticed by Feigin [5] that the score process $V(\theta)$ is a P_θ-local martingale. We suppose here that $V(\theta)$ is a P_θ-locally square integrable martingale. A remarkable fact is that under certain regularity conditions on the elements of $\mathscr{G}$

$$\overline{G}_t(\theta) = -\big\langle G(\theta), V(\theta)\big\rangle_t$$

for all $t \geq 0$, $G \in \mathscr{G}$.

The arguments to verify this are easy. Let P be a probability measure such that $P_{\theta|\mathscr{F}_t} \ll P|_{\mathscr{F}_t}$ for all $t \geq 0$, $\theta \in \Theta$, and let $Z(\theta) = \big(Z_t(\theta), t \geq 0\big)$ be the density process of P_θ with respect to P. Then $V_t(\theta) = \dot{Z}_t(\theta)/Z_t(\theta)$. Since $G(\theta)$ is a P_θ-local martingale ($G(\theta) \in \mathscr{M}_{\mathrm{loc}}(P_\theta)$) we have $G(\theta)Z(\theta) \in \mathscr{M}_{\mathrm{loc}}(P)$. This relation implies $\dot{G}(\theta)Z(\theta) + G(\theta)\dot{Z}(\theta) \in \mathscr{M}_{\mathrm{loc}}(P)$. Dividing the latter process by $Z(\theta)$ we get $\dot{G}(\theta) + G(\theta)V(\theta) \in \mathscr{M}_{\mathrm{loc}}(P_\theta)$ (in particular, if $G(\theta) \equiv 1$ we see that $V(\theta) \in \mathscr{M}_{\mathrm{loc}}(P_\theta)$). So the predictable process with finite variation $\overline{G}(\theta) + \big\langle G(\theta), V(\theta)\big\rangle$ is, by the definitions of the processes $\overline{G}(\theta)$ and $\big\langle G(\theta), V(\theta)\big\rangle$, also a P_θ-local martingale. Hence, it is equal to zero identically.

Now let us proceed to the general case. Fix a mapping $\pi\colon \theta \to P_\theta \in \mathscr{P}_\theta$ (for instance, in Example 1 fix a distribution of "the error" ε_n). If this mapping is smooth enough the score process $V^\pi(\theta)$ can be defined and

$$(7)\qquad \overline{G}_t(\theta) = -\big\langle G(\theta), V^\pi(\theta)\big\rangle_t$$

for all $t \geq 0$, $G \in \mathscr{G}$, $\theta \in \Theta$. This fact is very important not only for finding the optimal estimating process but also for other reasons. For example, it is well-known that

$$V_t^\pi(\theta) = \int_0^t \frac{d\langle G(\theta), V^\pi(\theta)\rangle_s}{d\langle G(\theta)\rangle_s}\, dG_s(\theta) + \tilde{V}_t^\pi(\theta),$$

where the P_θ-locally square integrable martingale $\tilde{V}^\pi(\theta)$ is "orthogonal" to the process $G(\theta)$, i.e., $\langle G(\theta), \tilde{V}^\pi(\theta)\rangle \equiv 0$. According to (7),

$$V_t^\pi(\theta) = -\int_0^t \frac{d\overline{G}_s(\theta)}{d\langle G(\theta)\rangle_s}\, dG_s(\theta) + \tilde{V}_t^\pi(\theta),$$

and the first term on the right-hand side is "the projection" of the process $V^\pi(\theta)$ on "the stable subspace" generated by $G(\theta)$, see Jacod [12]. This "projection" does not depend on the mapping π (but it may depend on P_θ together with $\overline{G}(\theta)$ and $\langle G(\theta)\rangle$). Moreover, by the Kunita–Watanabe inequality

$$\langle V^\pi(\theta)\rangle_t \geq \frac{\langle G(\theta), V^\pi(\theta)\rangle_t^2}{\langle G(\theta)\rangle_t} = \frac{\overline{G}_t(\theta)^2}{\langle G(\theta)\rangle_t}. \tag{8}$$

This inequality gives a lower bound for the Fisher information which is, in fact, the expectation of the left-hand side of (8). As mentioned above, the right-hand side of (8) is maximal in the class $\mathscr{G}$ if G is the optimal estimating process. This inequality is connected with the Cramér–Rao inequality, see Gushchin [9].

Now, suppose the optimal estimating process Q exists. By (4) and (7) we then get

$$\langle G(\theta), V^\pi(\theta) - Q(\theta)\rangle \equiv 0$$

for all $G \in \mathscr{G}$. Thus, the optimal estimating process $Q(\theta)$, provided it exists, is "the projection" (for any fixed θ) of the process $V^\pi(\theta)$ on the set $\mathscr{G}(\theta) = \{G(\theta), G \in \mathscr{G}\}$. More formally: let $\mathscr{G}^*(P_\theta)$ be the smallest subspace in the space $\mathscr{M}^2_{\mathrm{loc}}(P_\theta)$ of P_θ-locally square integrable martingales which is stable under stochastic integration and contains $\mathscr{G}(\theta)$, then the projection of $V^\pi(\theta)$ on $\mathscr{G}^*(P_\theta)$ is well-defined, it belongs to $\mathscr{G}(\theta)$ and coincides with $Q(\theta)$.

Thus, the optimal estimating process, provided it exists, can be found as follows: choose a mapping $\pi\colon \theta \to P_\theta$ such that (7) is valid, then $Q(\theta)$ can be obtained as the projection of $V^\pi(\theta)$ on $\mathscr{G}^*(P_\theta)$. Of course, one needs to prove independently that the found projection satisfies (4).

It should be noted that, typically, the mapping π can be chosen in such a way that the process $V^\pi(\theta)$ already belongs to $\mathscr{G}(\theta)$.

The above arguments explain the use of the term 'quasi-score process' with respect to Q. The term 'quasi-score function' has been usually used for concrete estimating functions in particular cases, while the property (4) or (4*) has been used (explicitly or implicitly) to prove the optimality of the quasi-likelihood function in some or other sense, see Thavaneswaran and Thompson [17], Hutton and Nelson [11], Godambe and Heyde [7], Sørensen [16], Jacod [14].

Example 1 (continued). Let

$$\mathscr{G} = \left\{ G : G_n(\theta) = \sum_{k=1}^{n} c_k(\theta)(X_k - \theta X_{k-1}) \Big(= \sum_{k=1}^{n} c_k(\theta)\varepsilon_k \Big) \right\}$$

where $c_k(\theta)$ is a function of $(X_1, \ldots, X_{k-1})$. Here

$$\dot{G}_n(\theta) = \sum_{k=1}^{n} \dot{c}_k(\theta)\varepsilon_k - \sum_{k=1}^{n} c_k(\theta)X_{k-1}, \qquad \overline{G}_n(\theta) = -\sum_{k=1}^{n} c_k(\theta)X_{k-1},$$

and

$$Q_n(\theta) = \frac{1}{\sigma^2} \sum_{k=1}^{n} X_{k-1}(X_k - \theta X_{k-1})$$

is the quasi-score process. The corresponding quasi-likelihood estimator coincides with the least squares estimator. The quasi-score process coincides with the score process if the errors have a Gaussian distribution.

Example 2 (continued). It is natural to take

$$\mathscr{G} = \left\{ G : G_t(\theta) = \int_0^t \alpha_s(\theta)\, dX_s - \int_0^t \alpha_s(\theta) f_s(\theta)\, d\lambda_s \Big(= \int_0^t \alpha_s(\theta)\, dm_s(\theta) \Big) \right\}$$

where $\alpha(\theta)$ is a predictable process with values in the space of $k \times d$ matrices such that $G(\theta)$ and $\overline{G}(\theta)$ are well-defined. Here

$$\overline{G}_t(\theta) = -\int_0^t \alpha_s(\theta) \dot{f}_s(\theta)\, d\lambda_s,$$

and it is easy to see that the quasi-score process is defined by

$$Q_t(\theta) = \int_0^t \dot{f}_s(\theta)^T a_s^{-1}(\theta)\, dX_s - \int_0^t \dot{f}_s(\theta)^T a_s^{-1}(\theta) f_s(\theta)\, d\lambda_s, \tag{9}$$

see Hutton and Nelson [11]. The quasi-score process coincides with the score process if, for example, $\lambda_t = t$ and $\big(m_t(\theta)\big)$ is a standard Wiener process.

Example 3 (continued). Let

$$\mathscr{G} = \Big\{ G : G_t(\theta) = \int_0^t \alpha_s(\theta; X, Y)\, dX_s - \int_0^t \alpha_s(\theta; X, Y) a_s(\theta; X, Y)\, ds \\ \Big(= \int_0^t \alpha_s(\theta; X, Y)\, dW_s \Big) \Big\}$$

where $\alpha(\theta; X, Y)$ is a predictable process such that $G(\theta)$ and $\overline{G}(\theta)$ are well-defined. Here the quasi-score proces

$$Q_t(\theta) = \int_0^t \dot{a}_s(\theta; X, Y)\, dX_s - \int_0^t \dot{a}_s(\theta; X, Y) a_s(\theta; X, Y)\, ds$$

coincides with the score process if Y is a deterministic process.

Example 4 (continued). Let

$$\mathscr{G} = \Big\{ G : G_t(\theta) = \int_0^t \alpha_s(\theta)\, dX_s^c(\theta) + \int_0^t \int \gamma_s(\theta; x)(\mu - \nu^\theta)(ds, dx) \Big\},$$

where X^c is the continuous martingale part of X, μ is the random measure associated with the jumps of X, $\alpha(\theta)$ and $\gamma(\theta)$ are predictable functions such that $G(\theta)$ and $\overline{G}(\theta)$ are well-defined. The form of the quasi-score function was indicated by Sørensen [16] if X is quasi-left continuous and by Jacod [14] for the general case. The quasi-score function is of the same form as the score function in the situation when the corresponding martingale problem has a unique solution, and the density process and the score process can be computed.

Sørensen [16] gave some examples showing that quasi-likelihood estimators based on the Hutton–Nelson quasi-score process (9) can be unsatisfactory in comparison with maximum likelihood estimators. This can be easily explained. In Example 2, if the distribution of the error term $(m_t(\theta))$ is known, the corresponding score process cannot be in the class $\mathscr{G}$ as defined in Example 2. In this case the maximum likelihood estimator has better properties than the quasi-likelihood one. On the other hand, if the distribution of the error term is unspecified (as in Example 1 where we only know the first two moments of the distribution of ε_n) we cannot use the class $\mathscr{G}$ as defined in Example 4 since the estimating function from this class can depend not only on θ but also on the distribution of errors. In this situation the function Q defined by (9) is the best possible. It should be noted that additional information about the distribution of errors allows us to enlarge the class of estimating functions and to improve the quality of estimation, see, e.g., Heyde [10].

Finally, we note that Chitashvili, Lazrieva and Toronjadze [4] recently considered another approach to the problem discussed here in a scheme generalizing the partial likelihood scheme.

4. In conclusion let us note that the modern martingale theory makes it possible to formulate rather weak conditions for the validity of the property (7) and the properties (4) and (4*) in particular cases. Jacod [14; Proposition 3.17 and the proof of Theorem 3.19] proved (4*) in the situation of Example 4. The equality (7) is proved in the same paper [14; Theorem 2.31 and the proof of Theorem 3.5], provided the existence of a process Y with bounded jumps such that for each θ Y is a P_θ-semimartingale with canonical decomposition

$$Y = G(\theta) + H(\theta), \tag{10}$$

$G(\theta) \in \mathscr{M}^2_{\mathrm{loc}}(P_\theta)$, $H(\theta)$ is a predictable process with finite variation. The assumption that Y has bounded jumps was weakened by Gushchin [9, Theorem 2]. Analogously, the conditions in Proposition 3.17 in Jacod [14] can be weakened. It should be noted that Theorem 2 in Gushchin [9] can be generalized to estimating processes not necessarily restricted by the relation (10). These results will be published elsewhere.

References

[1] Barndorff-Nielsen, O.E., and M. Sørensen: Asymptotic likelihood theory for stochastic processes. A review. - Research Report 162, Department of Theoretical Statistics, Aarhus University, 1989. (To appear in Internat. Statist. Rev.).

[2] Chitashvili, R.J., N.L. Lazrieva, and T.A. Toronjadze: On asymptotic estimation theory under model disturbance in a general scheme of statistical models. - Probability Theory and Mathematical Statistics, Trudy Tbilisi Math. Inst. 92, 1989, 106–145 (Russian).

[3] Chitashvili, R.J., N.L. Lazrieva, and T.A. Toronjadze: Asymptotic theory of M-estimators in general statistical models. Part 1: On asymptotic behaviour of estimators under model disturbance. - CWI Report BS-R9019, Centre for Mathematics and Computer Science, Amsterdam, 1990.

[4] Chitashvili, R.J., N.L. Lazrieva, and T.A. Toronjadze: Asymptotic theory of M-estimators in general statistical models. Part 2: On asymptotic behaviour of estimators in the presence of nuisance parameters. - CWI Report BS-R9020, Centre for Mathematics and Computer Science, Amsterdam, 1990.

[5] Feigin, P.D.: Maximum likelihood estimation for continuous-time stochastic processes. - Adv. in Appl. Probab. 8, 1976, 712–736.

[6] Godambe, V.P.: An optimum property of regular maximum likelihood estimation. - Ann. Math. Statist. 31, 1960, 1208–1212.

[7] Godambe, P.E., and C.C. Heyde: Quasi-likelihood and optimal estimation. - Internat. Statist. Rev. 55, 1987, 231–244.

[8] Greenwood, P.E., and W. Wefelmeyer: On optimal estimating functions for partially specified counting process models. - Preprints in Statistics 123, Mathematical Institute, University of Cologne, 1989.

[9] Gushchin, A.A.: Improvement and generalization of the Cramér–Rao inequality for a filtered space. - Probability Theory and Mathematical Statistics. Proceedings of the Fifth Vilnius Conference, Vol. 1, Utrecht/Vilnius, VSP BV/Mokslas, 1990, 480–489.

[10] HEYDE, C.C.: On combining quasi-likelihood estimating functions. - Stochastic Process. Appl. 25, 1987, 281–287.

[11] HUTTON, J.E., and P.I. NELSON: Quasi-likelihood estimation for semimartingales. - Stochastic Process. Appl. 22, 1986, 245–257.

[12] JACOD, J.: Calcul stochastique et problèmes de martingales. - Lecture Notes in Mathematics 714. Springer-Verlag, Berlin–Heidelberg–New York, 1979.

[13] JACOD, J.: Une application de la topologie d'Emery: le processus information d'un modèle statistique filtre. - Séminaire de Probabilités XXIII, Lecture Notes in Mathematics 1372. Springer-Verlag, Berlin–Heidelberg–New York, 1989.

[14] JACOD, J.: Regularity, partial regularity, partial information process, for a filtered statistical model. - Probab. Theory Related Fields (to appear).

[15] JACOD, J., and A.N. SHIRYAEV: Limit theorems for stochastic processes. - Springer-Verlag, Berlin–Heidelberg–New York, 1987.

[16] SØRENSEN, M.: On quasi-likelihood for semimartingales. - Research Report 171, Department of Theoretical Statistics, Aarhus University, 1988.

[17] THAVANESWARAN, A., and M.E. THOMPSON: Optimal estimation for semimartingales. - J. Appl. Probab. 23, 1986, 409–417.