Journal of Mathematical Sciences, Vol. 121, No. 6, 2004

STOCHASTIC EQUATIONS IN THE PROBLEMS OF SEMIMARTINGALE PARAMETER ESTIMATION

T. Toronjadze UDC 519.218.1, 519.246.2, 517.988.63

Introduction

Stochastic equations, both differential and nondifferential, play an important role in many problems of statistics of random processes, in particular, in problems of estimating the unknown parameters of semimartingales. Many problems of the estimation theory are reduced to the investigation of questions connected with the solvability (both strong and weak) of stochastic differential equations (SDEs) and with the asymptotic behavior of solutions of stochastic estimational equations.

The formalization of a statistical problem and the desire to study it from the standpoint of generality which is implied by the very essence of a statistical problem make it necessary to consider new specific problems of stochastic analysis and use new methods of investigating previously known problems. Such problems include the investigation of special types of stochastic differential equations and properties of their solutions, and also the asymptotic behavior of the roots of estimational equations in the case of a model disturbance in various formulations. Using the methods of martingale theory, one can make a general statistical model with filtration and an important particular case of it, in which models are associated with semimartingales, the objects of the research.

This monograph is concerned with studies of this kind.

Chapter 1 deals with the structure of all solutions of the Carathéodory-type stochastic differential equation whose drift coefficient satisfies the well-known Carathéodory condition from the theory of ordinary differential equations.

The statistical problem that leads to such SDEs is the innovation problem for the nonlinear filtering (estimation). Our assumption is that the investigated processes are Itô processes. Our aim is to present such a process without losing information as a simpler and more convenient process for studying a diffusion-type process. In doing so, it should only be assumed that the drift coefficient of the desired Itô process is square integrable (with respect to the Lebesgue measure) with probability 1. It is desirable not to strengthen the principal assumption which is natural in many senses (in investigating the absolute continuity of measures of the corresponding processes, the structure of Itô functionals and diffusion-type processes and so on).

However, the problem in such a general formulation has turned out to be rather difficult and still remains unsolved.

Under various additional restrictions both on the structure of the processes considered and on the kind of dependence of the processes participating in the scheme, this problem has been solved by many authors. We will speak on this in more detail below when discussing the results of Chapter 2. Here, in Chapter 1, the drift process is assumed to be a random variable independent of a Wiener process participating in the scheme. Then the principal assumption holds trivially. Our aim is to confirm the validity of the hypothesis about the existence of innovations (under the above-mentioned assumption only) in this particular case.

However, it has turned out that the corresponding SDE (this problem is reduced to the proof of its strong solvability) has a singularity at the point $t = 0$ and, therefore, the well-known results from the SDE theory are not applicable in this case. Therefore, we have developed a theory of such SDEs and

Translated from Itogi Nauki i Tekhniki, Seriya Sovremennaya Matematika i Ee Prilozheniya. Tematicheskie Obzory. Vol. 111, Functional Analysis, 2002.

studied the global and local properties of solutions, as well as the structure of the integral funnel of all solutions and so on. As a result, we have obtained a solution of the considered statistical problem.

Chapter 2 deals with the construction of an innovation process for the observed component of a partially observable diffusion-type process in the one- and multidimensional cases. Such problems arise in the estimation theory of partial likelihood schemes, the nonlinear filtering theory, and the theory of stochastic control by incomplete data.

As mentioned above, the considered problem belongs to the range of previously existing problems which demand new methods for their investigation. In particular, such methods include the generalized Bayes formula, linear integral inequalities derived for the functional of "filter," the stochastic version of the Gronwall–Bellman lemma and so on. It is interesting to note that to solve even a one-dimensional problem, one should use the multidimensional version of the lemma mentioned above. The use of these methods made it possible to get rid of many assumptions like the assumption of smoothness on the coefficients of the scheme.

In Chapter 3 the robust estimators for statistical models associated with semimartingales are constructed. We consider models with shrinking contamination neighborhoods, where a sequence of alternative measures is contiguous to a sequence of basic measures. As the basic class of estimators, we consider the class of generalized CLAN (consistent, linear, asymptotically normal) estimators. Note that here stochastic equations also play an important role, in particular, in the construction of these estimators. One of the construction methods consists in studying the solvability of stochastic estimational equations and the asymptotic behavior of their solutions for the model disturbance. Thus, we studied the problem of the local limiting behavior of the roots of such equations by the appropriately generalized Dugue–Kramer–Le Breton method. We also studied the global limiting behavior of the roots of these equations and obtained the desired CLAN estimators. Using the results obtained, we construct B-robust estimators with respect to the risk functional determined by the asymptotic mean-square error.

In considering the general model of statistical experiments we give the definition of the notion of "shrinking contamination neighborhoods" and formulate the minimax optimization problem. We also develop methods for finding optimal score functions, which, as it turns out, are the Huber functions.

We investigate the problem of robustness in two stages. In stage 1 we study separately and in great detail an important particular case of the discrete time. The well-known special models of time series are discussed. In stage 2 we consider the general case associated with semimartingales.

Finally, in Chapter 4 we introduce and investigate the Robbins–Monro-type stochastic differential equation and, in particular, study the question whether the solution of this equation is convergent with probability 1. Many generalized schemes of stochastic approximation and recursive estimation can be reduced to equations of the considered type. In this context, we have obtained a theorem which includes as particular cases many familiar results. The question how the results obtained here are related to the previously known results is studied when treating the special cases.

Theorems and facts from the general theory of random processes used in the present work can be found in [8, 21, 39, 48, 52, 59, 63, 82, 88, 99, 130].

Chapter 1 deals with the following one-dimensional SDE:

$$d\xi_t = A(t, \xi_t)dt + dW_t, \quad 0 \le t \le T,\ \xi_0, \tag{0.1}$$

where the function $A(t, x)$: $[0, T] \times \mathbb{R}_1 \to \mathbb{R}_1$ is a Borel-measurable function with respect to a pair (t, x), $W = (W_t)$ is a standard Wiener process, and ξ_0 is a random variable, independent of W. We study the structure of all solutions of this SDE.

In Sec. 1.1 we prove (Theorem 1.1) that Eq. (0.1) has a pathwise unique strong solution if and only if it has a weak solution unique in law.

This result and a method of proof based on the Tanaka–Meyer formula and the Yamada–Watanabe theorem [37, 126] were at first published in 1981 [101]. An analogous approach was developed in [62, 80] in 1983 (see also [51, 84]).

The consideration of the example given at the end of Sec. 1.1 naturally leads to the question: what is a sufficiently wide class of conditions for the existence and uniqueness of a weak or strong solution of Eq. (0.1), without requiring, in general, that the measure corresponding to the constructed solution be necessarily absolutely continuous with respect to a Wiener measure? Moreover, it will be interesting to find conditions which would, perhaps, guarantee only the existence of weak and strong solutions of Eq. (0.1) and, in this connection, to consider the relation between the sets of these solutions, in other words, to study the structure of the class of all solutions of Eq. (0.1).

Note that Eq. (0.1) is equivalent to the ordinary differential equation (with a coefficient containing "stochastics")

$$\frac{d\eta_t}{dt} = \widetilde{A}(t, \eta_t), \quad \eta_0 = \xi_0, \tag{0.2}$$

where $\widetilde{A}(t, x) = A(t, x + W_t)$, $0 \le t \le T$, $x \in \mathbb{R}_1$.

As in the theory of ordinary (deterministic) differential equations, it is natural to study Eq. (0.2) by the following scheme: first to study the local structure of the solutions of (0.2) and then to find the conditions under which the solutions can be continued on the whole "time" interval.

The technique used for obtaining such theorems is the "truncation" of the coefficient of the equation and finding sufficiently simple conditions which guarantee the existence of global solutions for the equation with a "truncated" coefficient, which, obviously, leads to the existence theorems for local solutions of the initial equation (see, e.g., [9, 15]).

The Carathéodory conditions are the well-known conditions of this type.

C-conditions:

(1) the function $A(t, x)$ is measurable in t for any fixed $x \in \mathbb{R}_1$ and continuous in x for any fixed t, $0 \le t \le T$;

(2) there exists a Borel-measurable function $m(t)$, $m(t) \ge 0$, $0 \le t \le T$, such that for any $t \in [0, T]$ and $x \in \mathbb{R}_1$,

$$|A(t, x)| \le m(t), \quad \int_0^T m(t)\, dt < \infty. \tag{0.3}$$

Now we consider Eq. (0.2). Note that under the Carathéodory conditions in the theory of ordinary differential equations the following statements are basic:

(a) there exists a solution of (0.2), i.e., for every fixed $\omega \in \Omega$ there exists a continuous function $(\eta_t(\omega))$, $0 \le t \le T$, satisfying Eq. (0.2) with $\widetilde{A}(t, x) = A(t, x + W_t(\omega))$;

(b) there exist the so-called *maximal* and *minimal solutions* of (0.2) with a fixed initial condition;

(c) the cross section of the *integral funnel* of the solutions (i.e., the set of all solutions of Eq. (0.2) "starting" from one point) for every t represents a closed interval.

However, these statements do not provide the existence of a solution $(\eta_t(\omega))$, $0 \le t \le T$, $\omega \in \Omega$, of Eq. (0.2) at least measurable with respect to ω and, in particular, possessing the desirable property of $\mathcal{F}^{\eta_0} \vee \mathcal{F}_t^W$-measurability for every t, where $\mathcal{F}_t^W = \sigma(W_s, s \le t)$, or the existence of a strong solution of Eq. (0.2).

The natural analogue of statement (c), saying that for any $t \in [0, T]$, at least one strong solution passes through a point ξ (random variable) of a random interval $I_{t_0} := \{\xi\colon \underline{\xi}_{t_0} \le \xi \le \overline{\xi}_{t_0}\ (P\text{-a.s.})\}$, where $\underline{\xi}_{t_0}$ and $\overline{\xi}_{t_0}$ are the lower and upper bounds of the integral funnel section at the point $t = t_0$, is not true, or, roughly speaking, strong solutions do not fill up the interval I_{t_0}. Moreover, if we consider the class of all anticipating solutions (this notion is defined below) instead of the class of strong solutions of Eq. (0.2), then the above-given statements (a), (b), and (c) hold.

It is well known ([3,51,84]) that the continuous process $\xi = (\xi_t)$, $0 \le t \le T$, is a *strong solution* of SDE (0.1) on the given probability space $(\Omega, \mathcal{F}, P)$ and with respect to the fixed Wiener process $W = (W_t)$,

$0 \le t \le T$, and initial condition ξ_0, if the process ξ is adapted to the P-augmented filtration $(\mathcal{F}^{\xi_0} \vee \mathcal{F}_t^W)$, $0 \le t \le T$, and satisfies SDE (0.1) with the initial condition ξ_0.

If the solution process ξ is such that the random variable ξ_t is $(\mathcal{F}^{\xi_0} \vee \mathcal{F}_T^W)$-measurable for each t, $0 \le t \le T$, then we call the process ξ an *anticipating solution* of SDE (0.1).

Note that we introduce the notion of the anticipating solution only for equations of the type (0.1), i.e., for equations with unit diffusion coefficient.

The anticipating solution can be constructed, for example, simply as follows.

Let (ξ_t^1) and (ξ_t^2), $0 \le t \le T$, be two "distinct" strong solutions of Eq. (0.1) and let A be an event from $\mathcal{F}_T^W = \sigma(W_t,\, 0 \le t \le T)$. Then it is obvious that the process

$$\xi_t = \xi_t^1 I_{\{A\}} + \xi_t^2 I_{\{A^c\}}, \quad 0 \le t \le T, \tag{0.4}$$

where $I_{\{\cdot\}}$ is an indicator of the event $\{\cdot\}$, is the anticipating solution of Eq. (0.1) and each of its trajectories represents a trajectory of a strong solution (either (ξ_t^1) or (ξ_t^2)). This construction suggests only that every anticipating solution can be represented as a combination of strong solutions.

In Theorem 1.5, the above-mentioned hypothesis concerning the representation of anticipating solutions in the form of a combination of strong solutions acquires a strong sense. Namely, Theorem 1.5 proves that for every anticipating solution $\xi = (\xi_t)$, $0 \le t \le T$, of Eq. (0.1), a measurable functional $\Phi(t,\omega,\omega_1)$, $0 \le t \le T$, $\omega, \omega_1 \in C_{[0,T]}$ can be found, such that for any $\omega_1 \in C_{[0,T]}$ the stochastic process $\Phi(\cdot,\cdot,\omega_1)$ is a strong solution of (0.1) and

$$\xi_t(\omega) = \Phi(t,\omega,\omega), \quad 0 \le t \le T \quad (P\text{-a.s.}),$$

where P is a standard Wiener measure defined on the measure space $(C_{[0,T]}, \mathcal{B}_{[0,T]})$ of continuous functions.

This theorem, together with Theorems 1.6, 1.2, 1.3, and 1.4, gives a complete description of the integral funnel of all solutions of (0.1) in terms of strong solutions.

The structure of the solutions of the stochastic equations of general type has been studied in [128,129]. The application of the methods developed in these papers to SDE (0.1) leads to the fact that under conditions on $A(t,x)$ (e.g., $|A(t,x)| \le m(t)$, $\int_0^T m^{1+\varepsilon}(t)\,dt < \infty$, $\varepsilon > 0$) it becomes possible to present every weak solution (in a certain sense) in the form of a combination of anticipating solutions of Eq. (0.1). To be more precise, if

$$(\Omega, \mathcal{F}, F = (\mathcal{F}_t), P, W = (W_t), \xi = (\xi_t), 0 \le t \le T)$$

is a weak solution of Eq. (0.1), then one can find a combination of the objects

$$(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{F} = (\widetilde{\mathcal{F}}_t), \widetilde{P}, \widetilde{W} = (\widetilde{W}_t), \widetilde{\xi} = (\widetilde{\xi}_t), 0 \le t \le T, \alpha)$$

(α is a random variable) and a measurable functional $\Phi(t,x,\omega)$, $0 \le t \le T$, $x \in \mathbb{R}_1$, $\omega \in C_{[0,T]}$, continuous in t, such that the random variable α is independent of the Wiener process $\widetilde{W}$ and uniformly distributed on $[0,1]$, $\widetilde{\xi} = (\widetilde{\xi}_t)$, $0 \le t \le T$, and $\xi = (\xi_t)$, $0 \le t \le T$, have the same probability law, and for any $x \in [0,1]$, the process $\Phi(t,x,\widetilde{W})$ represents an anticipating solution of Eq. (0.1) (with $\widetilde{W}$ instead of W), and

$$\widetilde{\xi}_t = \Phi(t,\alpha,\widetilde{W}), \quad 0 \le t \le T \quad (P\text{-a.s.}).$$

Obviously, the result obtained in Theorem 1.5 (the representation of an anticipating solution of Eq. (0.1) in the form of a combination of strong solutions) together with that just described (the representation of a weak solution in the form of a combination of anticipating solutions) provides us with the representation of a weak solution in the form of a combination of strong solutions.

In Theorem 1.7, a sufficient condition for the uniqueness of the solution of (0.1) is given, such that it does not suggest an absolute continuity of the measure corresponding to the solution of (0.1) with respect to a Wiener measure.

In Sec. 1.3, applying the "truncation technique" to Eq. (0.1), we deduce existence and extension theorems for the local solutions of Eq. (0.1). Here (Proposition 1.2, Theorem 1.8 and its Corollary 1.1, and Theorems 1.9, 1.10, and 1.11) conditions which guarantee the possibility of the extension of the solutions to the whole "time" interval $[0,\infty)$ under consideration, as well as conditions of the uniqueness are given.

Note that in choosing the "truncation" function the local (i.e., with small t) properties of a Wiener process are substantially used, in particular, the Lévy law of the iterated logarithm [38].

A survey of the results as well as many aspects of the theory of SDEs close to those considered by us, can be found in [3, 47, 56, 121, 131, 132].

Section 1.4 deals with special cases and examples illustrating different aspects of the suggested approach.

The last section, Sec. 1.5, is devoted to the innovation problem for the process

$$d\xi_t = \theta\, dt + dW_t, \quad \xi_0,$$

where θ is a random variable independent of the process W.

The investigation of this statistical problem leads to the Carathéodory-type SDE, since the function $A(t,x)$ from (1.46) is an unbounded function near the point $t=0$.

In Sec. 1.5 we give a positive solution of this problem.

A number of papers were devoted to the problem of innovations in different schemes. We briefly describe here some of the main aspects and results.

Introduce a stochastic basis $(\Omega, \mathcal{F}, F = (\mathcal{F}_t),\ 0 \le t \le T,\ P)$ with a Wiener process $W = (W, F)$, and let a process $\beta = (\beta, F)$ be defined on it.

Let

$$\int_0^T |\beta_t|\, dt < \infty \quad (P\text{-a.s.})$$

and consider the Itô process $\xi = (\xi, F)$ with the differential

$$d\xi_t = \beta_t dt + dW_t, \quad \xi_0 = 0. \tag{0.5}$$

In the general case, the existence of an innovation process for ξ means that the following assertions hold.

I. (Representation). There exists a nonanticipating functional $m_t(X)$, $0 \le t \le T$, $x \in C_{[0,T]}$ such that the process

$$\overline{W} = (\overline{W}, F^{\xi}),$$

where

$$\overline{W}_t = \xi_t - \int_0^t m_s(\xi)\, ds, \tag{0.6}$$

is a Wiener process.

Here

$$F^{\xi} = (\mathcal{F}^{\xi}_t), \quad 0 \le t \le T, \tag{0.7}$$

where $\mathcal{F}^{\xi}_t = \sigma(\xi_s, s \le t)$ is an augmented [51] σ-algebra generated by the process ξ up to the moment t.

Analogously, denote

$$F^{\overline{W}} = \big(\mathcal{F}^{\overline{W}}_t\big), \quad 0 \le t \le T, \tag{0.8}$$

where $\mathcal{F}^{\overline{W}}_t = \sigma(\overline{W}_s, s \le t)$ is an augmented σ-algebra generated by the process $\overline{W}$ up to the moment t.

II. (Coincidence of augmented filtrations).

$$F^{\overline{W}} = F^{\xi} \pmod P,$$

i.e., the process $\overline{W}$ contains the same information as the process ξ.

Statement I was proved for the first time by Shiryaev [87] and Kailath [22, 49] under the assumption

$$\int_0^T E|\beta_t(\omega)|\, dt < \infty,$$

and later the condition was considerably weakened in the papers of Yershov [127], Toronjadze [96] (for the case where $\beta_t(\omega) = \theta(\omega)$ is a random variable), and Meyer [77], where under the assumption that

$$m_t = E(\beta_t \mid \mathcal{F}_t^\xi)$$

exists (for almost all t w.r.t. the Lebesgue measure), it was shown that

$$\int_0^T |m_t| \, dt < \infty \quad (P\text{-a.s.})$$

and statement I holds.

It is also known that if β and W are independent, then it is sufficient for statement I that

$$\int_0^T \beta_t^2 \, dt < \infty \quad (P\text{-a.s.}) \tag{0.9}$$

(see Lipster and Shiryaev [66]).

Cirelson [13] has constructed an example of bounded $\beta = (\beta, F^\xi)$ for which statement I holds automatically but statement II cannot take place.

Thus, for statement II to hold, additional assumptions on the type of the dependence of β on W should be introduced.

For example, it is shown in [66] that if (β, W) is a Gaussian system and condition (0.9) is satisfied, then statements I and II hold.

If, on the other hand, β is assumed to be independent of W, then the sufficient conditions (of boundedness) on β in statement II were successively weakened by different authors:

(a) Clark [14]: $|\beta_t(\omega)| \le c < \infty$ (P-a.s.), where c is a nonrandom constant;

(b) Allinger and Mitter [2]:

$$E \int_0^T \beta_t^2 \, dt < \infty;$$

(c) Chitashvili [11]:

$$\int_0^T E(\beta_t^2 \mid \mathcal{F}_t^\xi) \, dt < \infty \quad (P\text{-a.s.}).$$

There exists a conjecture that condition (0.9) is also sufficient for statement II to hold, but this conjecture has not yet been proved.

In the special case, where

$$\beta_t = \theta$$

is a random variable, and, therefore, condition (0.9) holds automatically, statement II is proved in Chapter 1, Sec. 1.5.

The problem of filtration and control of a partially observable random process naturally leads to the case of a special dependence of β on W arising under the assumption that the process ξ is the so-called observable component of a diffusion-type process (η, ξ).

Consider the process (η, ξ) which is the solution of the following SDE:

$$\begin{aligned} d\eta_t &= a_t(\xi, \eta)dt + b_t(\xi, \eta)dW_t, \quad \eta_0 = 0, \\ d\xi_t &= A_t(\xi, \eta)dt + c_t(\xi)dW_t, \qquad \xi_0 = 0, \end{aligned} \tag{0.10}$$

where W is a multidimensional Wiener process, the coefficients a and A and b and c are vector- and matrix-valued, respectively, nonanticipating functionals.

In such a scheme, the problem of innovations is solved for the following cases:

(d) the so-called conditionally Gaussian scheme (see [66]): there exist coefficients a, A, b, and c such that

$$\begin{aligned} a_t(x,y) &= a_t(x)y_t, \\ A_t(x,y) &= A_t(x)y_t, \\ b_t(x,y) &= b_t(x), \\ c_t(x) &= c_t(x), \end{aligned}$$

which satisfy the Lipschitz and boundedness conditions with respect to the variable x;

(e) Krylov ([54, 55]) has considered the case of a multidimensional diffusion-type process in the Markov case (i.e., for example,

$$A_t(x,y) = A_t(x_t, y_t)),$$

where the coefficients a, A, b, and c are sufficiently smooth and bounded.

Note that the existence of a "noise" correlation is not excluded here;

(f) Chitashvili [12] has studied the special case of scheme (0.10), the so-called triangular system (one-dimensional), where the coefficients a, A, b, and c satisfy the Lipschitz condition with respect to the two variables x and y and the condition of a linear growth.

The "noise" correlation is not excluded; however, specific additional conditions are imposed on it.

In Sec. 2.1, we prove the multidimensional stochastic version of the Gronwall–Bellman lemma (see, e.g., [66]).

In the next two sections we consider a one-dimensional (Sec. 2.2) and multidimensional (Sec. 2.3) partially observable diffusion type process of the form (0.10).

In Sec. 2.2 we consider the scheme given by SDE (2.1), and under the linear growth (in both space variables x and y) and Lipschitz conditions (with respect to the variable x) on the coefficients of the scheme we prove the existence of the innovation process $\overline{W}$ for the observable component ξ. In contrast to case (f), the Lipschitz condition on $a_t(x,y)$ and $A_t(x,y)$ with respect to the variable y is not required here.

In Sec. 2.3, we consider a multidimensional case and under the boundedness and the Lipschitz conditions on the coefficients a, A, b, and c, we construct an innovation process.

This result generalizes case (e) in two directions: first we consider a non-Markov case, and second, reject the smoothness conditions on the coefficients.

The results of Chapters 1 and 2 have been published in [95–103, 107, 108, 111].

The problems considered in Chapter 3 belong to the asymptotic theory of robust estimation for dependent observations.

The theory of robust estimation for the case of independent, identically distributed (i.i.d.) observations was investigated for the first time by Huber [34, 35] and developed by Hampel et al. [32]. The key role in this theory is played by the M-estimators introduced by Huber as generalizations of the maximal likelihood estimates (MLE). The M-estimators can be constructed as solutions of the stochastic estimational equation

$$\sum_{i=1}^{n} \psi(X_i, \theta) = 0, \tag{0.11}$$

where X_i, $i \geq 1$, are the i.i.d. observations with common density $f(x,\theta)$, $\theta \in \Theta$ is an unknown parameter to be estimated, the so-called score function $\psi(x,\theta)$ is such that $E_\theta \psi(X_1,\theta) = 0$ and, hence, the left-hand side of Eq. (0.11) is a martingale with respect to the measure generated by the density $f(x,\theta)$.

Proceeding from Eq. (0.11), one can construct the CLAN estimators [6, 112, 113], which are asymptotically equivalent to the M-estimators. This class of estimators is one of the basic classes in the theory of robust estimation.

The necessity of extension of the class MLE to the class of M- or CLAN-estimators arises because of the main assumption of the theory: the parametric family of distributions of observations is not exactly specified, and observations are assumed to be distributed with density from some neighborhood of the basic (core) density $f(x,\theta)$.

More frequently the so-called Huber's gross error (contamination) model is considered. In this case the "neighborhood" is given by the formula

$$\Phi_\varepsilon^H = \{\widetilde{f}(x,\theta) : \widetilde{f}(x,\theta) = (1-\varepsilon)f(x,\theta) + \varepsilon h(x,\theta)\}, \quad \varepsilon > 0, \tag{0.12}$$

where $h(x,\theta)$ is a density from some class H. The class H is specified depending on the statistical problem under consideration. Measures generated by densities $\widetilde{f}(x,\theta)$ are called alternative measures or, simply, alternatives, and H is called a class determining alternatives.

This model has a clear statistical meaning. Let X_i, $i \geq 1$, be i.i.d. observations with density $f(x,\theta)$, and W_i, $i \geq 1$, be i.i.d. observations with density $h(x,\theta)$. Consider the i.i.d. sequence of 0–1 random variables Z_i, $i \geq 1$, with

$$P(Z_i = 1) = \varepsilon, \quad i \geq 1.$$

If the sequences (X_i), (W_i), and (Z_i) are mutually independent, then the random variables

$$Y_i = (1 - Z_i)X_i + Z_i W_i \tag{0.13}$$

form i.i.d. observations with density $\widetilde{f}(x,\theta)$. Thus the observations (X_i) are "contaminated" by the observations (W_t).

Introduce a criterion of comparison of estimates based on a risk functional. Frequently, as this functional there occurs an asymptotic mean-square error and the estimator is called optimal if it is a minimax estimator with respect to the risk functional, where the maximum is taken over the class H, whereas the minimum is taken over the class $\Psi = \{\psi(x,\theta)\}$ of functions, which determine the estimators (in particular, over the class that determines the CLAN estimators).

In such a statement, MLEs are now not optimal. Optimal estimators are prescribed by the Huber functions, included in the class of functions $\psi(x,\theta)$ determining M-estimators.

The optimal score functions have the form

$$\psi^* = [l - \beta^*]_{-m^*}^{m^*}, \tag{0.14}$$

i.e., are centered, truncated, maximum likelihood scores, where β^* is the centering parameter and m^* is the truncation parameter.

In the case where the "radius" ε of the neighborhood Φ_ε^H depends on the time variable, i.e., $\varepsilon = \varepsilon_n$ and $\varepsilon_n \to 0$ as $n \to \infty$, we have the set of shrinking contamination neighborhoods $\{\Phi_{\varepsilon_n}^H\}$. If now the sequence of alternative measures $\{\widetilde{P}_\theta^n\}$ is contiguous to that of basic measures $\{P_\theta^n\}$, then we obtain the shrinking contamination neighborhoods with contiguous alternatives. The above-described scheme can be generalized at least in two principal directions: (1) passage to an infinite-dimensional parameter set, i.e., to semiparametric models and (2) passage to dependent observations ([33, 57, 70, 92]) (composition of these two directions is also possible).

Many aspects of the estimation theory dealing with semiparametric models for i.i.d. observations are presented in [7, 81, 124]. A number of papers [26–30] are devoted to questions of the asymptotic theory of estimation for semimartingale models.

In our work, the set of parameters Θ is one-dimensional and the observations are dependent.

As it turned out, when passing from the i.i.d. observations to the dependent observations, the contamination model given by formula (0.12) (call it contamination of measures) and the model described by formula (0.13) (call it contamination of trajectories or replacement model) do not coincide but, on the contrary, radically differ.

Contamination of measures for dependent discrete time observations, more precisely, for a stationary ergodic $AR(p)$ model, was first investigated by Künsch [57]. Under certain assumptions on the model, he constructed a so-called optimal B-robust ("B" is the abbreviation of "bias") estimate [32] and showed that this estimate is given by the Huber function. Contamination of trajectories was first considered by Martin and Yohai [70] for stationary ergodic time series. They have shown that in considering the process of moving average, the Huber function provides no robust estimates.

Our approach generalizes contamination of measures (0.12) and, therefore, continues the investigations started by Künsch. Moreover, we consider shrinking contamination neighborhoods with contiguous alternatives for statistical models with filtration associated with semimartingales.

Let us describe briefly the points concerning this model. First of all, we must clarify the notion of M-estimators, investigate their asymptotic properties, and then consider the problem of constructing these estimators (in particular, the CLAN estimators). To this end, in Sec. 3.1 we consider the general stochastic estimational equations. Investigating the limiting behavior of roots of these equations, we obtain the CLAN estimators. In Subsection 3.1.1 (Theorem 3.1) we investigate the question of asymptotic solvability and asymptotic behavior of roots of these equations in the neighborhood of a fixed point θ of the set of parameters Θ. All these classical questions for i.i.d. observations have been investigated by Dugue and Cramer [19]. Le Breton has considered the case of the diffusion processes.

Subsection 3.1.2 (Theorem 3.2) concerns the investigation of global asymptotic properties of the solutions. This scheme is less popular. We refer to the paper of Perlman [79]. Note only that the "local theorem" provides us with the solution of the identification problem of unknown parameter, whereas the "global theorem" allows us to construct consistent estimators.

Section 3.2 deals with the robust estimation problems in discrete time models and Sec. 3.3 with the general case.

The basic results obtained in these sections can be summarized as follows.

(a) The main notions of robust estimation theory are generalized and the main objects of this theory are introduced. On this basis, an optimization minimax problem is stated (with asymptotic mean-square error as a risk functional) for shrinking contamination neighborhoods with contiguous alternatives for statistical models associated with semimartingales under the integral representation property for martingales [67].

(b) Conditions are given under which optimal-in-the-minimax-sense score martingales (which determine optimal B-robust estimates) are defined by the Huber functions (Theorems 3.4 and 3.6).

Let us give a more detailed account of our approach. We consider an array scheme which is formalized by the consideration of a sequence of statistical models $\mathcal{E} = \{\mathcal{E}_n\}_{n\geq 1}$ (see Subsection 3.3.1). See also, e.g., [36]. Note that for every $n \geq 1$ the model $\mathcal{E}_n$ is assumed to be regular (see Subsection 3.3.2), while the very sequence $\mathcal{E}$ is ergodic (see Subsection 3.3.1). We have introduced the concept of CLAN estimators (see Subsection 3.3.2) and then, on the basis of the notion of an exponential martingale, we have introduced the set of shrinking contamination neighborhoods of a core sequence of measures (see Subsection 3.3.3) generalizing (0.12). We have also investigated the asymptotic properties of CLAN estimators under the sequence of alternative measures belonging to this set and obtained the "biased" estimates (Proposition 3.1).

Further, we investigate the question under what kind of conditions the sequence of alternative measures $\{\widetilde{P}_\theta^n\}$ is contiguous to the basic sequence of measures $\{P_\theta^n\}$ (see Subsection 3.3.3), establish an exact form for asymptotic distributions of CLAN estimators under such a sequence of measures (Proposition 3.2), and obtain an analogue of influence functionals (see Subsection 3.3.3 and (3.37)), which play an important role in all these problems.

We need this preparatory work in order to introduce and calculate the risk functional (see Subsection 3.3.4) and also to formulate our optimization problem (see Subsection 3.3.4). However, the risk functional $D(L, N, \theta)$, being the functional of a sequence of martingales determining the CLAN estimators and of martingales determining alternatives, is, in the general case, an implicit function of these sequences. Therefore, in the general case, it is impossible to obtain any constructive solution of the optimization problem and to construct an optimal sequence of score martingales.

Furthermore, we assumed that the martingales under consideration possess the integral representation property (see Subsection 3.3.3) and on the basis of this property we worked out the so-called approximation technique as follows: starting from relation (3.104), we first fix index $n \geq 1$ and then consider the optimization problem associated with the risk functional $D_n(L^n, N^n, \theta)$ (see (3.103)). Stochastic analysis (in the presence of the integral representation property of martingales) allows us to deduce an explicit formula for calculating $D_n(L^n, N^n, \theta)$ (see (3.125)) and for describing explicitly the classes Ψ_n^0 and $\mathcal{H}_n^R$ of functions for which the optimization minimax problem is stated and solved (see Subsection 3.3.8, 1, 2, (i), (ii)). As a result, we obtain a score martingale whose integrands in the integral representation are Huber functions (see Theorem 3.5).

Then we construct classes Ψ and $\mathcal{H}_\Psi$ of sequences of functions, which determine the score martingales, and the martingales determine the alternatives, respectively (see Subsection 3.3.8), in such a way that the score martingale, which is optimal for each fixed step n with respect to $D_n(L^n, N^n, \theta)$, would form a sequence of optimal martingales with respect to the risk functional $D(L, N, \theta)$ (see Theorem 3.6).

We end with the survey of Sec. 3.2.

This section deals with the discrete-time statistical models. We have singled this case out of the general ones, considered in Sec. 3.3, for the following reasons:

1. This case is one of the most important particular cases involving many known time-series models. A large number of papers are devoted to the investigation of various aspects of robust estimation for the i.i.d. observations and for various classes of time series. This allows us to compare earlier known results with those obtained by us (see Subsection 3.2.4). In the general case we lose this possibility.

2. The objects introduced in this section are simple and give way to the statistical interpretation.

3. Making use of the compact and simple objects under consideration, we present all proofs in detail.

4. The methods of proof of the basic theorems (see Theorems 3.3 and 3.4) are rather common and after slight changes and remarks can be applied to the general case. All these arguments used in Sec. 3.3, help us to avoid not substantial but cumbersome calculations.

In Subsection 3.2.2, we state and solve the optimization problem for the fixed nth step. The solution is the Huber function $\psi^n = [l^n - \beta^n]_{-m_n}^{m_n}$ (see Theorem 3.3). The equation for the optimal truncation level m_n^* is derived, studied, and used for the approximation. Moreover, this equation can be applied to the investigation of the differentiability of the optimal score function $\psi^{*,n}$ with respect to θ (see Subsection 3.2.5). Such equations have never been studied, even for i.i.d. case. This is a point of this subsection.

In Subsection 3.2.3, we introduce classes of sequences of score functions Ψ and alternatives $\mathcal{H}_\Psi$. The ergodicity conditions formulated in the definitions of these classes ensure optimality of the sequence $\psi^* = \{\psi^{*,n}\}_{n\geq 1}$, where $\psi^{*,n}$ is the optimal score function constructed at the fixed nth step (see Theorem 3.4). The ergodicity conditions in the definition of the class Ψ are rather involved because the centering parameter β^n is a nonlinear functional of the conditional distribution of maximum likelihood scores, $Q^{n,l}$. If this distribution is symmetric with respect to zero and hence $\beta^n = \beta(Q^{n,l}) = 0$, then the ergodicity conditions are simplified, i.e., the usual condition of weak convergence for averaged distributions is sufficient.

Subsection 3.2.4 is devoted to examples of various special models illustrating various aspects of the problem. Below we briefly discuss this point and indicate the relations to the known results of different authors.

The i.i.d. case in a more strict (uniform) setting was considered for the first time by Bickel [5,6].

Note that in Model I.2 we consider independent, nonidentically distributed observations, i.e., the simplest but nonstationary process. None of the earlier known approaches can be applied in this case. The stationary ergodic $AR(k)$ model for the contamination neighborhoods of a fixed "radius" $\varepsilon > 0$ was studied by Künsch [57]. He considered a multidimensional parameter, formulated the optimization problem in the Hampel setting and solved it, and, as a result, obtained the Huber function. Our approach to this case is described in Model II. Staab [92] considered the stationary ergodic $ARMA(p, q)$ models.

In Model III we show that Staab's approach is covered by ours. Case (b) of Model I.2, Model II, and Model III in Staab's setting, and cases 1(b) and 1(c) of Model IV illustrate the situation where the risk functional $D(\psi, H, \theta)$ can be expressed explicitly, and contamination is full. Case (a) of Model I.2 and case 1(a) of Model IV correspond to the situation where the minimax problem is reduced to the form (3.23), but contamination is not full. Model III in the general setting and case 1(d) of Model IV illustrate the situation where the above theory can be applied in its full capacity.

In the last subsection, Subsection 3.2.5, a method of constructing optimal CLAN estimators is given.

The results of this chapter have been published in [104, 109, 112–117, 119, 120].

In Chapter 4, the SDE of the form

$$dZ_t = H_t(Z_{t-})dK_t + M(dt, Z_{t-}), \quad Z_0 \tag{0.15}$$

is considered. Here $H_t(u)$ and $M(t,u)$, $t \in [0,\infty)$, $u \in \mathbb{R}_1$, are random fields with the following properties: for each $u \in \mathbb{R}_1$, the process $(H_t(u))_{t\geq 0}$ is predictable, the process $M(u) = (M(t,u))_{t\geq 0}$ is a locally square-integrable martingale, and $k = (K_t)_{t\geq 0}$ is an increasing predictable process. The family $M(u)$, $u \in \mathbb{R}_1$, is assumed to possess the integral representation properties of various types.

If $u = (u_t)_{t\geq 0}$ is a predictable process, we use the symbol $\int_0^t M(ds, u_s)$ for the notation of the corresponding stochastic integral and $M(dt, u_t)$ is its "differential." For instance, if $M(u) = f(u)\cdot m$, then $\int_0^t M(ds, u_s) := \int_0^t f(u_s)\, dm_s$ and $M(dt, u_t) = f(u_t)dm_t$. For details, see Sec. 4.1.

We call SDE (0.15) the Robbins–Monro-type (RM-type) SDE if the drift coefficient $H_t(u)$ satisfies the following conditions: for all $t \in [0,\infty)$ P-a.s.

$$\begin{aligned} &H_t(0) = 0, \\ &H_t(u)u < 0 \text{ for all } u \neq 0. \end{aligned}$$

SDE (0.15) naturally includes the RM stochastic approximation algorithms with martingale noises (see, e.g., [68, 72–76]). For example, if $H_t(u) = \gamma_t R(u)$ and $M(t,u) = \int_0^t \gamma_s\, dm_s$, where $\gamma = (\gamma_t)_{t\geq 0}$ is a nonnegative predictable process, $R(u)$ is a deterministic function (regression function) with $R(0) = 0$, $R(u)u < 0$, and $m = (m_t)_{t\geq 0}$ is some locally square-integrable martingale, SDE (0.15) gives the generalized RM procedure introduced in [72].

In the paper of Lazrieva and Toronjadze [110], the algorithm of constructing the recursive maximum likelihood estimation procedures for general statistical models with filtration was proposed. In the case of discrete time, this procedure is given in Example 1(a) below and is embedded in (0.15), while it is not covered by the generalized RM algorithm, although it should be mentioned that in the i.i.d. case the classical RM algorithm contains recursive estimation procedures ([1, 64, 78, 85, 86]).

Thus, the consideration of the RM-type SDE (0.15) allows us to study both stochastic approximation and recursive estimation procedures by a common approach.

The question of strong solvability of SDE (0.15) is well investigated (see, e.g., [23–25, 44, 71, 83, 84]). Assume that there exists a unique strong solution $Z = (Z_t)_{t\geq 0}$ of (0.15) on the whole interval $[0,\infty)$.

In Chapter 4 we study only the problem of P-a.s. convergence $Z_t \to 0$ as $t \to \infty$.

Our approach to this problem is based on two representations, standard and nonstandard, of the predictable bounded variation process $a = (A_t)_{t\geq 0}$ in the decomposition of the semimartingale $(Z_t^2)_{t\geq 0}$ in the form of the difference of two predictable increasing processes $A^1 = (A_t^1)_{t\geq 0}$ and $A^2 = (A_t^2)_{t\geq 0}$ and uses Theorem 4.1 on convergence sets of nonnegative semimartingales [65, 67].

Two groups of conditions, (I) and (II), connected with the standard and nonstandard representations are introduced in Sec. 4.2. On this basis, the main result, concerning the convergence $Z_t \to 0$ P-a.s. as $t \to \infty$ is formulated (see Theorem 4.2).

In the same section, the relationship between groups of conditions (I) and (II) are also investigated.

In the next section, Sec. 4.3, some simple sufficient conditions for (I) and (II) are given.

In the last section, Sec. 4.4, the series of examples illustrating the efficiency of all aspects of our approach is given.

The results of Chapter 4 have been published in [105, 106, 110, 118].

In conclusion, we note that in [4, 10, 16–18, 20, 41–43, 46, 53, 61, 68, 69, 89, 90, 94, 122, 123, 125] one can find many questions concerned with the statistics of random processes which are close in spirit to the problems presented in Chapters 3 and 4.

Chapter 1

STRUCTURE OF SOLUTIONS OF A ONE-DIMENSIONAL SDE WITH UNIT DIFFUSION COEFFICIENT

1.1. Regular Equations

Consider the following one-dimensional stochastic differential equation (SDE)

$$d\xi_t = A(t,\xi_t)dt + dW_t, \quad 0 \le t \le T, \quad \xi_0, \tag{1.1}$$

where $W = (W_t)$, $0 \le t \le T$, is a standard Wiener process, a function $A(t,x)$: $[0,T] \times \mathbb{R}_1 \to \mathbb{R}_1$ is Borel-measurable with respect to a pair of variables (t,x), and ξ_0 is an arbitrary real random variable (r.v.) independent of W.

An SDE is said to be *regular in law* (*weakly regular*) ([126]) if there exists at least one weak solution and if all solutions (which can be defined in different probability spaces) with the same initial distributions have the same probability law.

An SDE is said to be *strongly regular* ([126]) if there exists a strong solution which is pathwise unique.

The strong regularity implies the regularity in law.

Now we show that for SDE (1.1) the two notions are equivalent.

Recall the following facts.

Proposition 1.1. *Let $\xi^i = (\xi^i_t)$, $0 \le t \le T$, $i = 1,2$, be two Itô processes with differentials*

$$d\xi^i_t = a_i(t,\omega)dt + b_i(t,\omega)dW_t, \quad 0 \le t \le T, \quad \xi^i_0, \quad i = 1,2,$$

and let

$$\eta(t) = \max(\xi^1_t, \xi^2_t), \quad 0 \le t \le T.$$

Then

$$\eta(t) = \eta(0) + \int_0^t I_{\{\xi^1_s > \xi^2_s\}} d\xi^1_s + \int_0^t I_{\{\xi^2_s \ge \xi^1_s\}} d\xi^2_s + \frac{1}{2}\Lambda_t^{\xi^1-\xi^2}(0),$$

where $\Lambda^\xi_t(a)$ is a semimartingale (the Itô process) local time at the point $a \in \mathbb{R}_1$ ([51, 84]).

Proof. It immediately follows from the simple relation

$$\eta(t) = (\xi^1_t - \xi^2_t)^+ + \xi^2_t,$$

where $x^+ = \max(0,x)$, and the Tanaka–Meyer formula that if $X = (X_t)$, $0 \le t \le T$, is a continuous semimartingale, then

$$X_t^+ = X_0^+ + \int_0^t I_{(0,\infty)}(X_s)dX_s + \frac{1}{2}\Lambda^X_t(0).$$

□

Remark 1.1. It is well known [84] that

$$\Lambda^X_t(0) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^t I_{(0,\varepsilon)}(X_s)d\langle X\rangle_s. \tag{1.2}$$

The following simple lemma is given for the completeness.

Lemma 1.1. *Let ξ_1 and ξ_2 be two r.v. such that*

$$\xi_1 \geq \xi_2 \quad (P\text{-a.s.}),$$

and the distribution functions of r.v. ξ_1 and ξ_2 coincide:

$$F_{\xi_1}(x) = F_{\xi_2}(x) \quad \forall x \in \mathbb{R}_1.$$

Then $\xi_1 = \xi_2$ (P-a.s.).

Proof. Denote $\eta_i = f(\xi_i)$, $i = 1, 2$, where $f(x)$, $x \in \mathbb{R}_1$, is a strongly increasing bounded function, e.g., $f(x) = \arctan x$. Let $\zeta = \eta_1 - \eta_2$. Then, from the conditions of the lemma, we obtain

$$\zeta \geq 0, \quad E\zeta = 0.$$

Hence $\zeta = 0$ (P-a.s.). The assertion follows. □

Theorem 1.1 (of equivalence). *Equation* (1.1) *is strongly regular if and only if it is regular in law.*

Proof. The necessity is obvious. Now we prove the sufficiency.

Let SDE (1.1) be regular in law and let ξ^1 and ξ^2 be two weak solutions of (1.1) defined on the same probability space, and $\xi_0^1 = \xi_0^2$ (P-a.s.).

According to Proposition 1.1, it is easy to see that the continuous stochastic processes $\overline{\eta}(t) = \max(\xi_t^1, \xi_t^2)$ and $\underline{\eta}(t) = \min(\xi_t^1, \xi_t^2)$, $0 \leq t \leq T$, are solutions of (1.1) since $\Lambda_t^{\xi^1 - \xi^2}(0) = 0$ (see (1.2)). Due to the conditions of the theorem, these processes have the same probability law. But for each t, $0 \leq t \leq T$,

$$\underline{\eta}(t) \leq \xi_t^1, \xi_t^2 \leq \overline{\eta}(t) \quad (P\text{-a.s.})$$

and, therefore, according to Lemma 1.1,

$$\underline{\eta}(t) = \xi_t^1 = \xi_t^2 = \overline{\eta}(t) \quad (P\text{-a.s.}).$$

The desirable statement follows immediately if we use the Yamada–Watanabe theorem [126] claiming that the existence of a pathwise unique weak solution implies the existence of a unique strong solution. □

The approach used in Proposition 1.1 for the equation

$$d\xi_t = A(t, \xi_t)dt + B(t, \xi_t)dW_t, \quad 0 \leq t \leq T, \quad \xi_0 \tag{1.3}$$

results, for the maximum $\eta(t)$ of two solutions (ξ_t^1) and (ξ_t^2), $0 \leq t \leq T$, in the following formula:

$$\begin{aligned} \eta(t) = \eta(0) &+ \int_0^t (A(s, \eta(s))ds + B(s, \eta(s))dW_s) \\ &+ \lim_{\varepsilon \downarrow 0} \frac{1}{4\varepsilon} \int_0^t (B(s, \xi_s^1) - B(s, \xi_s^2))^2 I_{\{|\xi_s^1 - \xi_s^2| \leq \varepsilon\}} ds, \end{aligned}$$

from which it becomes obvious that the statement of Theorem 1.1 is valid for (1.3) if we assume that the function $B(t, x)$ satisfies the Hölder condition

$$|B(t, x) - B(t, y)| \leq b(t)|x - y|^\alpha, \quad \alpha \geq \frac{1}{2},$$

and

$$\int_0^T b^2(t)dt < \infty.$$

Note, however, that it is impossible to get rid of the assumptions on the function $B(t, x)$ in (1.3).

To see this, we refer to the well-known example due to Tanaka [37]

$$d\xi_t = B(\xi_t)dW_t, \quad 0 \leq t \leq T, \quad \xi_0 = 0,$$

where $B(x) = 1$ and $B(x) = -1$ if $x > 0$ and $x < 0$, respectively.

To illustrate Theorem 1.1, we consider a simple sufficient condition for the existence and uniqueness of the weak solution (and, hence, by Theorem 1.1 for the strong solution) of (1.1), having a measure absolutely continuous with respect to the measure of the process $\xi_t = W_t + \xi_0$, $0 \le t \le T$.

Example. Let for any t, $0 \le t \le T$, and $x \in \mathbb{R}_1$

$$|A(t,x)| \le K(1+|x|), \quad K = \text{const} < \infty.$$

More general conditions of this type are given in [66].

1.2. The Carathéodory-Type SDE. Existence Theorem. Description of the Integral Funnel

Assume for simplicity that $\xi_0 = 0$.

As the stochastic basis $(\Omega, \mathcal{F}, F = (\mathcal{F}_t), 0 \le t \le T, P)$, we consider $(C_{[0,T]}, \mathcal{B}_{[0,T]}, B = (\mathcal{B}_{[0,t]})$, $0 \le t \le T$, $P^T)$, with the measure space $(C_{[0,T]}, \mathcal{B}_{[0,T]})$ of continuous functions $\omega = \omega_t$, $0 \le t \le T$, $\omega_0 = 0$, a standard Wiener measure P^T and P^T-augmented filtration $B = (\mathcal{B}_{[0,t]})$, $0 \le t \le T$.

Denote by $W = (W_t)$, $0 \le t \le T$, the coordinate process $W_t(\omega) = \omega_t$, $0 \le t \le T$. Then, with the measure P^T the process W is a standard Wiener process. Further, let $F^W = (\mathcal{F}_t^W)$, $0 \le t \le T$, be the P^T-augmented filtration generated by the process W, $\mathcal{F}_t^W = \sigma(W_s, s \le t)$.

Let Ξ denote the class of all anticipating solutions of Eq. (1.1) and Ξ_s be the class of all strong solutions of this equation.

Assume that $A(t,x)$ satisfies the Carathéodory conditions (C-conditions above, see Introduction).

Theorem 1.2. *If C-conditions hold, then there exists an anticipating solution of Eq.* (1.1).

Proof. Denote $\eta(t) = \xi_t - W_t$, $0 \le t \le T$. We write Eq. (1.1) as

$$d\eta(t) = A(t, \eta(t) + W_t)dt, \quad 0 \le t \le T, \quad \eta(0) = 0, \tag{1.4}$$

and consider the sequence $\{\eta_j(t)\}$, $j \ge 1$, $0 \le t \le T$, given by the following relation: for any $j \ge 1$,

$$\eta_j(t) = \begin{cases} 0 & \text{if } 0 \le t \le \dfrac{T}{j}, \\ \int_0^{t-\frac{T}{j}} A(s, \eta_j(s) + W_s)ds & \text{if } \dfrac{T}{j} < t \le T. \end{cases} \tag{1.5}$$

We will need for our discussion the following two lemmas. Lemma 1.2 is of interest in itself.

Lemma 1.2. *Sequence* (1.5) *is* (*relatively*) *compact in* $C_{[0,T]}$.

Proof. Indeed, $\eta_1(t) = 0$ for any t, $0 \le t \le T$. If $j \ge 2$, then for any t, by (0.3) we have

$$|\eta_j(t)| \le \int_0^{\left(t-\frac{T}{j}\right)^+} |A(s, \eta_j(s) + W_s)|ds \le M\left(\left(t - \frac{T}{j}\right)^+\right) \le M(T) < \infty,$$

where $x^+ = \max(0,x)$, $M(t) = \int_0^t m(s)\,ds$, $0 \le t \le T$.

Further, for any $t_1, t_2 \in [0,T]$ and any $j \geq 1$,

$$
\begin{aligned}
|\eta_j(t_2) - \eta_j(t_1)| &\leq \left| \int_{\left(t_1 - \frac{T}{j}\right)^+}^{\left(t_2 - \frac{T}{j}\right)^+} |A(s, \eta_j(s) + W_s)| ds \right| \\
&\leq \left| \int_{\left(t_1 - \frac{T}{j}\right)^+}^{\left(t_2 - \frac{T}{j}\right)^+} m(s)\, ds \right| \\
&= \left| \int_0^{\left(t_2 - \frac{T}{j}\right)^+} m(s)\, ds - \int_0^{\left(t_1 - \frac{T}{j}\right)^+} m(s)\, ds \right| \\
&= \left| M\left(\left(t_2 - \frac{T}{j}\right)^+ \right) - M\left(\left(t_1 - \frac{T}{j}\right)^+ \right) \right|.
\end{aligned}
$$

Hence, sequence (1.5) is uniformly bounded and uniformly continuous; therefore, it is relatively compact in $C_{[0,T]}$. □

Let $\{x_n(\omega), \omega \in \Omega\}$, $n \geq 1$, be a sequence of random elements defined on some complete probability space $(\Omega, \mathcal{F}, P)$ with values in some complete separable metric space $(X, \sigma_p(X), \rho)$, where ρ is a metric in X and $\sigma_p(X)$ is a Borel σ-algebra generated by the metric ρ.

Lemma 1.3. *If the sequence $\{x_n(\omega), \omega \in \Omega\}_{n \geq 1}$ is a (relatively) compact (P-a.s.) in X, then there exists a sequence of random variables $\{\eta_j(\omega)\}_{j \geq 1}$, $n_j : \Omega \to \mathbb{N} = (1, 2, 3, ...)$ such that*

(1)

$$
\begin{aligned}
&n_1(\omega) < \infty,\ n_2(\omega) < \infty, ...,\ \ (P\text{-}a.s.), \\
&n_j(\omega) < n_{j+1}(\omega) \quad \forall j \geq 1\ \ (P\text{-}a.s.);
\end{aligned}
$$

(2) *for any $j \geq 1$, the random element $(x_{n_j(\omega)}(\omega),\ \omega \in \Omega)$ is $\mathcal{F}/\sigma_p(X)$-measurable;*

(3) *the subsequence $\{x_{n_j(\omega)}(\omega),\ \omega \in \Omega\}_{j \geq 1}$ of the sequence $\{x_n(\omega),\ \omega \in \Omega\}_{n \geq 1}$ converges with probability* 1.

Proof. By virtue of the (P-a.s.) compactness of the sequence $\{x_n(\omega),\ \omega \in \Omega\}_{n \geq 1}$, there exists a P-null set B, $P(B) = 0$, such that for any $\omega \in \Omega \setminus B$ the sequence $\{x_n(\omega)\}_{n \geq 1}$ is a (relatively) compact set. Further, let $\{y_m\}_{m \geq 1}$ be a countable dense set of elements of the space X.

Denote

$$
O_{m,\varepsilon} = \{x \in X : \rho(x, y_m) < \varepsilon\}, \quad \varepsilon > 0.
$$

By virtue of the compactness of the sequence $\{x_n(\omega)\}_{n \geq 1}$, one can choose a finite subcovering from the covering $\{O_{m,\frac{1}{2}}\}_{m \geq 1}$ of the sequence $\{x_n(\omega)\}_{n \geq 1}$. Therefore, there exists at least one ball, which contains an infinite number of elements of the sequence $\{x_n(\omega)\}_{n \geq 1}$ (this fact can be analytically written as follows: there exists a number $m = m(\omega) < \infty$ such that $\underline{\lim}_{n \to \infty} \rho(x_n, y_m) < 1/2$).

Now we define the number $m_1(\omega)$ by the relation

$$
m_1(\omega) = \min\{m : \underline{\lim}_{n \to \infty} \rho(x_n(\omega), y_m) < 1/2\}
$$

and set

$$
n_1(\omega) = \min\{n : \rho(x_n(\omega), y_{m_1(\omega)}) < 1/2\}.
$$

It is obvious that $m_1(\omega) < \infty$ and $n_1(\omega) < \infty$. Similarly, for any $j \geq 2$, we set

$$m_j(\omega) = \min\left\{m : \varliminf_{n\to\infty} \rho(x_n(\omega), y_m) < \frac{1}{2^j},\ \rho(y_m, y_{m_{j-1}(\omega)}) < \frac{1}{2^{j-1}}\right\},$$

$$n_j(\omega) = \min\left\{n > n_{j-1}(\omega) : \rho(x_n(\omega), y_{m_j(\omega)}) < \frac{1}{2^j}\right\}.$$

Note that for each $j \geq 1$, $m_j(\omega) < \infty$ (each subsequence of a compact sequence is compact itself), $n_j(\omega) < \infty$. Obviously, the sequence $\{n_j(\omega)\}_{j\geq 1}$ satisfies requirement (1) and random element $\{x_{n_j(\omega)}(\omega),\ \omega \in \Omega\}$ satisfies requirement (2) for each $j \geq 1$.

It remains to show that the sequence $\{x_{n_j(\omega)}(\omega),\ \omega \in \Omega\}_{j\geq 1}$ converges with probability 1. To this end it is sufficient to verify its (P-a.s.) fundamentality.

For each $m \geq 1$ we have

$$\rho(x_{n_j(\omega)}(\omega), x_{n_{j+m}(\omega)}(\omega)) \leq \sum_{i=j-1}^{\infty} \frac{1}{2^j} \to 0 \ \text{ as } \ j \to \infty \quad (P\text{-a.s.}).$$

$\square$

Before proving the theorem, we note that by virtue of Lemmas 1.2 and 1.3 just proved, there exists a subsequence $\{\eta_{j_k(\omega)}(\omega)\}_{k\geq 1}$ of sequence (1.5) such that

$$\lim_{n\to\infty} \eta_{j_k(\omega)}(t,\omega) = \eta(t,\omega), \quad 0 \leq t \leq T,$$

uniformly with respect to t. Thus, $\eta(t,\omega)$ is a continuous process.

Now we recall that $A(t,x)$ is a function continuous in the variable x, satisfying (0.3). Hence we have obtained the required statement by letting $k \to \infty$ (and using the Lebesgue-dominated convergence theorem) in the relation

$$\eta_{j_k}(t) = \int_0^t A(s, \eta_{j_k}(s) + W_s)\, ds - \int_{\left(t-\frac{T}{j_k}\right)^+}^t A(s, \eta_{j_k}(s) + W_s)\, ds.$$

$\square$

Remark 1.2. The solution $\xi = (\xi_t)$, $0 \leq t \leq T$, just constructed is, generally speaking, an anticipating solution of Eq. (1.1), since the sequence $\{j_k(\omega)\}_{k\geq 1}$ may depend on the whole trajectory of the Wiener process $W = (W_t)$, $0 \leq t \leq T$.

Let Ξ_1 and Ξ_2 be subsets of the set Ξ of anticipating solutions of Eq. (1.1).

Denote

$$\begin{cases} \overline{\xi}_t = \operatorname*{ess\,sup}_{\Xi_1}(\xi_t), & 0 \leq t \leq T, \\ \underline{\xi}_t = \operatorname*{ess\,inf}_{\Xi_2}(\xi_t), & 0 \leq t \leq T. \end{cases} \tag{1.6}$$

Theorem 1.3. *Stochastic processes* $\overline{\xi} = (\overline{\xi}_t)$, $0 \leq t \leq T$, *and* $\underline{\xi} = (\underline{\xi}_t)$, $0 \leq t \leq T$, *are anticipating solutions of Eq.* (1.1).

Proof. We consider again Eq. (1.4) and give the proof in terms of the process $\eta(t) = \xi_t - W_t$, $0 \leq t \leq T$.

We show, for example, that the process $\overline{\eta}(t) = \overline{\xi}_t - W_t$, $0 \leq t \leq T$, is an anticipating solution of Eq. (1.4).

In order to do this, we denote

$$\eta_m^{\max} = \max\{\eta_1(t), ..., \eta_m(t)\}, \quad 0 \leq t \leq T,$$

for any natural number m, where $\eta_i(t)$ for any $i = 1, ..., m$ is an anticipating solution of Eq. (1.4). Just similarly to what has been done in Proposition 1.1, we can show that the following lemma holds.

Lemma 1.4. *The stochastic process $\eta_m^{\max}(t)$, $0 \le t \le T$, is an anticipating solution of Eq.* (1.4).

Before proving the theorem, we note that for any $t_1, t_2 \in [0, T]$,

$$|\overline{\eta}(t_2) - \overline{\eta}|(t_1) \le \operatorname*{ess\,sup}_{\Xi_1} \left| \int_{t_1}^{t_2} |A(s, \eta(s) + W_s)| \, ds \right| \le |M(t_2) - M(t_1)|,$$

where $M(t) = \int_0^t m(s)\, ds$, $0 \le t \le T$, i.e., $\overline{\eta}(t)$, $0 \le t \le T$, is uniformly continuous with probability 1.

It is now obvious that for any $\varepsilon > 0$ there exist a finite partition $\{t_j\}^\varepsilon$, $j = 1, 2, ..., m$, of the interval $[0, T]$ and the anticipating solutions $\eta_j(t)$, $0 \le t \le T$, $j = 1, 2, ..., m$, of Eq. (1.4) such that

$$0 \le \overline{\eta}(t_j) - \eta_j(t_j) \le \varepsilon,$$

and for any $t_1, t_2 \in (t_j, t_{j+1})$, $j = 1, ..., m-1$,

$$|\overline{\eta}(t_2) - \overline{\eta}(t_1)| \le \varepsilon,$$
$$|\eta_m^{\max}(t_2) - \eta_m^{\max}(t_1)| \le \varepsilon.$$

Thus, for any $\varepsilon > 0$ and any t, $0 \le t \le T$,

$$|\overline{\eta}(t) - \eta_m^{\max}(t)| \le 3\varepsilon.$$

To complete the proof it is sufficient to pass to the limit in the relation

$$\eta_m^{\max}(t) = \int_0^t A(s, \eta_m^{\max}(s) + W_s)\, ds.$$

□

Assume that $\Xi_i = \Xi$, $i = 1, 2$, i.e., ess sup and ess inf in (1.6) are taken on the whole set Ξ of anticipating solutions of Eq. (1.1). In these cases, stochastic processes $\overline{\xi} = (\overline{\xi}_t)$ and $\underline{\xi} = (\underline{\xi}_t)$, $0 \le t \le T$, are called, respectively, the maximal and minimal solutions of Eq. (1.1).

Theorem 1.4. *The maximal solution $\overline{\xi} = (\overline{\xi}_t)$, $0 \le t \le T$, and minimal solution $\underline{\xi} = (\underline{\xi}_t)$, $0 \le t \le T$, are strong solutions of Eq.* (1.1).

Proof. Show, for example, that $\overline{\xi} = (\overline{\xi}_t)$, $0 \le t \le T$, is a strong solution of Eq. (1.1).

For any fixed s, $0 \le s \le T$, denote by $\Xi_{[0,s]}$ a class of all anticipating solutions (ξ_t^s), $0 \le t \le s$, of Eq. (1.1) defined on the interval $[0, s]$ (by Theorem 1.2, the class $\Xi_{[0,s]}$ is not empty), i.e., it is a set of the $\mathcal{F}_s^W$-measurable for every t, $0 \le t \le s$, stochastic processes (ξ_t^s), $0 \le t \le s$, such that

$$P^s\left(\xi_t^s = \int_0^t A(u, \xi_u^s)\, du + W_t,\ 0 \le t \le s\right) = 1,$$

where $P^s = P^T / \mathcal{B}_{[0,s]}$ is a restriction of the Wiener measure P^T to the σ-algebra $\mathcal{B}_{[0,s]}$.

Put

$$\overline{\xi}_t^s = \operatorname*{ess\,sup}_{\Xi_{[0,s]}} (\xi_t^s), \quad 0 \le t \le s, \quad 0 \le s \le T.$$

In such notation, the statement of the theorem takes the following form: for any l, $0 \le l \le T$, the random variable $\overline{\xi}_l^T$ is $\mathcal{F}_l^W$-measurable.

It is obvious that if we solve problem (1.1) consecutively in the interval $[0, l]$, with the initial condition $\xi_0 = 0$ (i.e., if we construct an anticipating solution (ξ_t^l), $0 \le t \le l$), then in the interval $[l, T]$ with the initial condition ξ_l^l, "sticking together" the constructed solutions, we obtain an anticipating solution on the whole interval $[0, T]$, coinciding with (ξ_t^l), $0 \le t \le l$, on the interval $[0, l]$.

In other words, any anticipating solution (ξ_t^s), $0 \le t \le s$, $0 \le s \le T$, of Eq. (1.1) can be extended to the whole interval $[0, T]$.

Hence, considering an anticipating solution $(\overline{\xi}_t^l)$, $0 \le t \le l$, on the interval $[0,l]$ and extending it to the whole interval $[0,T]$, we obtain an anticipating solution (ξ_t^T), $0 \le t \le T$, which by virtue of the definition of $(\overline{\xi}_t^T)$, $0 \le t \le T$, is such that

$$P^T(\xi_t^T \le \overline{\xi}_t^T,\ 0 \le t \le T) = 1. \tag{1.7}$$

Now, since

$$P^T(\xi_l^T = \overline{\xi}_l^l) = 1,$$

we obtain from (1.7)

$$P^T(\overline{\xi}_l^l \le \overline{\xi}_l^T) = 1.$$

We now have to prove that

$$P^T(\overline{\xi}_l^T \le \overline{\xi}_l^l) = 1.$$

To this end, we define on the space $\Omega = C_{[0,T]}$ the operator

$$\Lambda(\omega) = \begin{pmatrix} \Lambda_1 \\ \Lambda_2 \end{pmatrix}(\omega) = \begin{pmatrix} u \\ v \end{pmatrix}, \tag{1.8}$$

where $\omega \in C_{[0,T]}$, $u \in C_{[0,l]}$ $v \in C_{[l,T]}$, $\Lambda_1(\omega) = u$, $\Lambda_2(\omega) = v$,

$$u_t = \omega_t, \quad 0 \le t \le l,$$

and

$$v_t = \omega_t - \omega_l, \quad l \le t \le T.$$

Note that

$$\omega_t = \begin{cases} u_t, & 0 \le t \le l; \\ v_t + u_l, & l \le t \le T. \end{cases}$$

Denote by $\mathcal{B}_{[l,T]} = \sigma(\omega_t - \omega_l, l \le t \le T)$ the augmented σ-algebra generated by increments. Then, obviously, $\mathcal{B}_{[0,T]} = \mathcal{B}_{[0,l]} \vee \mathcal{B}_{[l,T]}$, $\mathcal{B}_{[0,l]}$ and $\mathcal{B}_{[l,T]}$ being independent under a probability measure P^T. Denote, further, by P^l a Wiener measure defined on the σ-algebra $\mathcal{B}_{[0,l]}$ and by $P^{l,T}$ a Wiener measure defined on $\mathcal{B}_{[l,T]}$. Obviously, the operator Λ provides a one-to-one measure-preserving transformation of the space $C_{[0,T]}$ onto $C_{[0,l]} \times C_{[l,T]}$, i.e.,

$$\Lambda : (C_{[0,T]}, \mathcal{B}_{[0,T]}, P^T) \rightleftarrows (C_{[0,l]}, \mathcal{B}_{[0,l]}, P^l) \times (C_{[l,T]}, \mathcal{B}_{[l,T]}, P^{l,T}),$$

and for any $B \in \mathcal{B}_{[0,T]}$,

$$P^T(B) = (P^l \times P^{l,T})(\Lambda B),$$

where $\Lambda B = \{\Lambda(\omega) : \omega \in B\}$.

Introduce the functional

$$\phi(t,u,v) = \overline{\xi}_t^T\left(\Lambda^{-1}\begin{pmatrix} u \\ v \end{pmatrix}\right), \quad 0 \le t \le T,\ u \in C_{[0,l]},\ v \in C_{[l,T]}, \tag{1.9}$$

where the operator Λ^{-1} is inverse to Λ. Obviously,

$$P^l \times P^{l,T}\Bigg((u,v) : \phi(t,u,v) = \int_0^t A(s,\phi(s,u,v))\,ds + W_t\left(\Lambda^{-1}\begin{pmatrix} u \\ v \end{pmatrix}\right), \quad 0 \le t \le l\Bigg) = 1.$$

From the definition of the operator Λ and the coordinate Wiener process W it follows that ($P^{l,T}$-a.s.) for every v

$$P^l\left(u : \phi(t,u,v) = \int_0^t A(s,\phi(s,u,v))\,ds + W_t(\Lambda_1^{-1}(u)),\ 0 \le t \le l\right) = 1.$$

Thus, ($P^{l,T}$-a.s.) for every v, the process $\phi(t,u,v)$, $0 \le t \le l$, belongs to the class of anticipating solutions $\Xi_{[0,l]}$ (obviously, if $t \in [0,l]$,

$$W_t\left(\Lambda^{-1}\begin{pmatrix}u\\v\end{pmatrix}\right) = W_t(\Lambda_1^{-1}(u)) = W_t(\omega)).$$

Hence ($P^{l,T}$-a.s.) for every v

$$P^l\left(u : \phi(l,u,v) \le \overline{\xi}_l^l\left(\Lambda^{-1}\begin{pmatrix}u\\v\end{pmatrix}\right)\right) = 1,$$

which implies that (P^T-a.s.) for every ω

$$P^T(\overline{\xi}_l^T \le \overline{\xi}_l^l \mid \Lambda_2)(\omega) = 1$$

(recall that the σ-algebras $\mathcal{B}_{[0,l]}$ and $\mathcal{B}_{[l,T]}$ are independent under the probability measure P^T). Finally, by averaging, we obtain

$$P^T(\overline{\xi}_l^T \le \overline{\xi}_l^l) = 1.$$

□

Theorem 1.5. *For every anticipating solution $\xi = (\xi_t)$, $0 \le t \le T$, of Eq.* (1.1), *there exists a measurable functional $\Phi(t,\omega,\omega_1)$ defined on the measure space*

$$([0,T] \times C_{[0,T]} \times C_{[0,T]}, \sigma([0,T]) \times \mathcal{B}_{[0,T]} \times \mathcal{B}_{[0,T]})$$

such that for any $\omega_1 \in C_{[0,T]}$, $\Phi(\cdot,\cdot,\omega_1) \in \Xi_s$, i.e., $\Phi(\cdot,\cdot,\omega_1)$ is a strong solution of Eq. (1.1) *and*

$$\xi_t(\omega) = \Phi(t,\omega,\omega), \quad 0 \le t \le T \quad (P\text{-a.s.}).$$

Proof. Given an anticipating solution $\xi = (\xi_t)$, $0 \le t \le T$, we denote by $(\xi_t^{\text{str}}(s,x))$, $s \le t \le T$, $0 \le s \le T$, $x \in \mathbb{R}_1$, a strong solution of Eq. (1.1), considered in the interval $[s,T]$, with the initial condition $\xi_s^{\text{str}}(s,x) = x$. By Theorem 1.4, such a solution exists. We fix t_1, $0 < t_1 < T$. Using the notation introduced in the proof of the previous theorem, we note that there exists an event $B \in \mathcal{B}_{[t_1,T]}$, $P^{t_1,T}(B) = 1$, such that for any $v \in B$,

$$P^{t_1}\left(u : \phi(t,u,v) = \int_0^t A(s,\phi(s,u,v))\,ds + W_t(\Lambda_1^{-1}(u)),\ 0 \le t \le t_1\right) = 1,$$

where the functional $\phi(t,u,v)$ is defined as in (1.9), i.e., $\phi(t,u,v) = \xi_t\left(\Lambda^{-1}\begin{pmatrix}u\\v\end{pmatrix}\right)$, $l := t_1$.

Now we consider the functional

$$\widetilde{\phi}(t,u,v) = \begin{cases} \phi(t,u,v) & \text{if } v \in B, \\ \xi_t^{\text{str}}(0,0) & \text{if } v \notin B, \end{cases} \tag{1.10}$$

where $u \in C_{[0,t_1]}$, $v \in C_{[t_1,T]}$ are given by (1.8) with $l := t_1$.

Define the following functional:

$$\widetilde{\Phi}(t,\omega,\omega_1) = \widetilde{\phi}(t,\Lambda_1(\omega),\Lambda_2(\omega_1)), \tag{1.11}$$

where $(\omega,\omega_1) \in C_{[0,T]} \times C_{[0,T]}$, $0 \le t \le T$, the operators Λ_1 and Λ_2 are defined by relation (1.8), $l := t_1$.

Further, we define the functional $\Phi^1(t,\omega,\omega_1)$ by the relations

$$\Phi^0(t,\omega,\omega_1) = \xi_t(\omega),$$

$$\Phi^1(t,\omega,\omega_1) = U(\Phi^0,t_1,t,\omega,\omega_1)$$

$$:= \begin{cases} \widetilde{\Phi}(t,\omega,\omega_1) & \text{if } 0 \le t \le t_1; \\ \xi_t(\omega) & \text{if } t > t_1 \text{ and } \widetilde{\Phi}(t_1,\omega,\omega_1) = \widetilde{\Phi}(t_1,\omega,\omega); \\ \xi_t^{\text{str}}(t_1,\widetilde{\Phi}(t_1,\omega,\omega_1)) & \text{if } t > t_1 \text{ and } \widetilde{\Phi}(t_1,\omega,\omega_1) \ne \widetilde{\Phi}(t_1,\omega,\omega). \end{cases} \tag{1.12}$$

It can be easily seen that:

(1) for any $\omega_1 \in \Omega$, the process $\Phi^1(\cdot,\cdot,\omega_1)$ is an anticipating solution of Eq. (1.1);
(2) for any fixed $\omega_1 \in \Omega$ and any t, $0 \le t \le T$,

$$\Phi^1(t,\cdot,\omega_1) \text{ is } \mathcal{F}_t^{\Phi^0}\text{-measurable}$$

and

$$\Phi^1(t,\cdot,\omega_1) \text{ is } \mathcal{F}_{t_1}^{W}\text{-measurable}$$

for $t \le t_1$;
(3) $\Phi^1(t,\omega,\omega) = \xi_t(\omega)$, $0 \le t \le T$ (P-a.s.).

For any $t_2 \ne t_1$, $0 \le t_2 \le T$, on the basis of the functional $\Phi^1(t,\omega,\omega_1)$, by the same technique as above, we can construct a functional $\Phi^2(t,\omega,\omega_1)$ such that properties (1) and (3) hold without any changes and property (2) is transformed into the following property:

(2′) for any fixed $\omega_1 \in \Omega$ and any t, $0 \le t \le T$, a functional $\Phi^2(t,\cdot,\omega_1)$ is $\mathcal{F}_t^{\Phi^1}$-measurable and $\Phi^2(t,\cdot,\omega_1)$ is $\mathcal{F}_{t_k}^{W}$-measurable for $t < t_k$, $k = 1, 2$.

Namely,

$$\Phi^2(t,\omega,\omega_1) = U(\Phi^1, t_1, t, \omega, \omega_1),$$

i.e., one is to proceed from the anticipating solution $(\Phi^1(t,\cdot,\omega_1))$, $0 \le t \le T$, of Eq. (1.1) instead of the given anticipating solution (ξ_t), $0 \le t \le T$ (see (1.10), (1.11), and (1.12)).

Consider a dense sequence of points $\{t_n\}_{n\ge1}$ of the interval $[0,T]$ and define a sequence of functionals $\{\Phi^n(t,\omega,\omega_1)\}_{n\ge1}$ recursively by the relation

$$\Phi^n(t,\omega,\omega_1) = U(\Phi^{n-1}, t_n, t, \omega, \omega_1), \quad n \ge 1,$$
$$\Phi^0(t,\omega,\omega_1) = \xi_t(\omega).$$

We show that

$$\Phi(t,\omega,\omega_1) = \inf_n \sup_{m\ge n} \Phi^m(t,\omega,\omega_1)$$

possesses the property required in the theorem.

The fact that $\Phi(t,\cdot,\omega_1)$ is measurable with respect to $\mathcal{F}_t^W$ follows from the following statement: for any $n \ge 1$, $\Phi^{n+1}(t,\cdot,\omega_1)$ is $\mathcal{F}_t^{\Phi^n}$-measurable and $\Phi^n(t,\cdot,\omega_1)$ is $\mathcal{F}_{t_k}^W$-measurable for every $t_k \ge t$, $k = 1, 2, ..., n$.

This implies that

$$\sup_{m\ge n} \Phi^m(t,\cdot,\omega_1)$$

possesses a similar property, and hence $\Phi(t,\cdot,\omega_1)$ is $\mathcal{F}_{t_k}^W$-measurable for every $t_k \ge t$, $k = 1, 2, ...$, i.e., $\Phi(t,\cdot,\omega_1)$ is $\mathcal{F}_{t+}^W$-measurable. But $\mathcal{F}_{t+}^W = \mathcal{F}_t^W$. Hence, $\Phi(t,\cdot,\omega_1)$ is $\mathcal{F}_t^W$-measurable for every t, $0 \le t \le T$.

Further, we note that for any $m = 1, 2, ...$ and for fixed ω_1, the process $\Phi^m(t,\cdot,\omega_1)$, $0 \le t \le T$, satisfies Eq. (1.1). Thus, by Theorem 1.3, the process $\Phi_n(t,\cdot,\omega_1) = \sup_{m\ge n} \Phi^m(t,\cdot,\omega_1)$, $0 \le t \le T$, and the process $\Phi(t,\cdot,\omega_1) = \inf_n \Phi_n(t,\cdot,\omega_1)$, $0 \le t \le T$, are solutions of Eq. (1.1) for any fixed ω_1.

Finally, by our construction,

$$\Phi(t,\omega,\omega) = \xi_t(\omega), \quad 0 \le t \le T \quad (P\text{-a.s.}).$$

□

Remark 1.3. If, in representation (0.4), the event $A \in \mathcal{B}_{[0,T]}$, then $\xi_t(\omega) = \Phi(t,\omega,\omega)$ (P-a.s.) with

$$\Phi(\cdot,\omega,\omega_1) = I_{\{A\}}(\omega_1)\xi^1_\cdot(\omega) + I_{\{A^c\}}(\omega_1)\xi^2_\cdot(\omega).$$

Now we pass to the description of the integral funnel of the solutions of Eq. (1.1).

Denote

$$V_t = \{\xi_t : ((\xi_t),\ 0 \le t \le T) \in \Xi\},$$
$$V_t^s = \{\xi_t : ((\xi_t),\ 0 \le t \le T) \in \Xi_s\},$$

for any t, $0 \le t \le T$, i.e., V_t (respectively, V_t^s) is a set of random variables ξ_t which represents a section of the integral funnel of the anticipating (respectively, strong) solution set of Eq. (1.1) at the point t. Note that in all cases, Eq. (1.1) is "solved" under the same initial conditions.

Let, further,

$$V_t(\omega) = \{\xi_t(\omega) : ((\xi_t),\ 0 \le t \le T) \in \Xi\},$$
$$V_t^s(\omega) = \{\xi_t(\omega) : ((\xi_t),\ 0 \le t \le T) \in \Xi_s\},$$

i.e., let $V_t(\omega)$ (respectively, $V_t^s(\omega)$) be a set of points from $\mathbb{R}_1$ such that through each of these points there passes a trajectory (with fixed ω) of at least one anticipating (respectively, strong) solution of (1.1).

In other words, $V_t(\omega)$ (respectively, $V_t^s(\omega)$) represents a section of V_t (respectively, of V_t^s) at the point ω.

It is obvious that $V_t^s \subset V_t$ and $V_t^s(\omega) \subset V_t(\omega)$ (P-a.s.).

Theorem 1.6. *For any t, $0 \le t \le T$:*

(1) a section V_t of the integral funnel of all anticipating solutions of Eq. (1.1) coincides with the subset H of all $\mathcal{F}_T^W$-measurable random variables such that with probability 1,

$$\underline{\xi}_t \le \eta \le \overline{\xi}_t,$$

where $\underline{\xi}_t$ and $\overline{\xi}_t$ are defined by (1.6), i.e.,

$$V_t = \{\eta, \eta \in H : \underline{\xi}_t \le \eta \le \overline{\xi}_t \ (P\text{-a.s.})\};$$

(2) $V_t(\omega) = V_t^s(\omega)$ (P-a.s.).

Proof. We fix t_0, $0 \le t_0 \le T$. Let a random variable $\eta \in H$, where H is a set of all $\mathcal{F}_T^W$-measurable random variables with

$$\underline{\xi}_{t_0} \le \eta \le \overline{\xi}_{t_0} \quad (P\text{a.s.}).$$

We show that: (1) there exists an anticipating solution $\xi = (\xi_t)$, $0 \le t \le T$, of Eq. (1.1) such that $\xi_{t_0} = \eta$; (2) for any fixed $\omega_0 \in \Omega \equiv C_{[0,T]}$ there exists a strong solution $\xi_t(\omega, \omega_0)$ of Eq. (1.1) such that

$$\xi_{t_0}(\omega_0, \omega_0) = \eta(\omega_0).$$

Indeed, it can be easily seen that if we consider the Carathéodory scheme with the "initial" condition $\zeta_{t_0} = \eta$, i.e.,

$$\zeta_t = -\int_t^{t_0} A(s, \zeta_s)\, ds + W_t - W_{t_0} + \eta, \quad \zeta_{t_0} = \eta, \tag{1.13}$$

and take the sequence

$$\eta_j(t) = \begin{cases} \eta \quad \text{if } \ t_0 - \dfrac{t_0}{j} \le t \le t_0, \\ -\displaystyle\int_{t+\frac{t_0}{j}}^{t_0} A(s, \eta_j(s) + W_s - W_{t_0})\, ds + \eta \quad \text{if } \ 0 \le t \le t_0 - \dfrac{t_0}{j}, \ j \ge 1, \end{cases}$$

as an approximated solution (i.e., denote $\eta(t) = \zeta_t - (W_t - W_{t_0})$ and rewrite Eq. (1.13) as

$$d\eta(t) = A(t, \eta(t) + W_t - W_{t_0})\, dt, \quad \eta(t_0) = \eta, \quad 0 \le t \le t_0,$$

see (1.4) and (1.5)), we can, similarly to Theorem 1.2, construct a solution (ζ_t), $0 \le t \le t_0$, of Eq. (1.13) such that $\zeta_{t_0} = \eta$. By the extension (see the proof of Theorem 1.4), we obtain a solution (ζ_t) of Eq. (1.13) defined on the whole interval $[0,T]$.

Further, we denote

$$\tau = \begin{cases} \sup\limits_{0 \le t \le t_0} \{t : \zeta_t = \overline{\xi}_t \text{ or } \zeta_t = \underline{\xi}_t\}, \\ 0 \quad \text{if} \quad \zeta_t \ne \overline{\xi}_t, \zeta_t \ne \underline{\xi}_t, \ 0 \le t \le t_0, \end{cases}$$

and set

$$\xi_t = \begin{cases} \underline{\xi}_t \text{ if } \ 0 \le t \le \tau, \zeta_\tau = \underline{\xi}_\tau, \tau > 0, \\ \overline{\xi}_t \text{ if } \ 0 \le t \le \tau, \zeta_\tau = \overline{\xi}_\tau, \tau > 0, \\ \zeta_t \text{ if } \ \tau \le t. \end{cases}$$

It is obvious that the process $\xi = (\xi_t)$, $0 \le t \le T$, $\xi_{t_0} = \eta$, constructed above is the solution of Eq. (1.1) desired in (1).

Further, it follows from Theorem 1.5 that the desired in (2) strong solution is

$$\xi_t(\omega, \omega_0) = \Phi(t, \omega, \omega_0),$$

where Φ is a functional from Theorem 1.5 constructed from the solution ξ of Eq. (1.1) which was obtained in (1).

Indeed,

$$\eta(\omega_0) = \xi_{t_0}(\omega_0) = \Phi(t_0, \omega_0, \omega_0) = \xi_{t_0}(\omega_0, \omega_0).$$

$\square$

Finally, we give a sufficient condition for the uniqueness of the solution of Eq. (1.1).

Theorem 1.7. *Let the function $A(t,x)$ together with the C-conditions (see (0.3)) satisfy the relation*

$$\int_0^T \inf_{y \in \mathbb{R}_1} \sup_{x \in \mathbb{R}_1} |A(t,x) - A(t,y)|^2 dt < \infty. \tag{1.14}$$

Then there exists a pathwise unique strong solution of SDE (1.1).

Proof. It is well known that for any $\varepsilon > 0$ there exists a measurable (with respect to t) function y_t, $0 \le t \le T$, such that

$$\int_0^T \sup_{x \in \mathbb{R}_1} |A(t,x) - A(t,y_t)|^2 dt \le \int_0^T \inf_{y \in \mathbb{R}_1} \sup_{x \in \mathbb{R}_1} |A(t,x) - A(t,y)|^2 dt + \varepsilon T < \infty. \tag{1.15}$$

Denote

$$\Gamma_t = \int_0^t \gamma_s ds,$$

where $\gamma_t = A(t, y_t)$, and define the stochastic process $\zeta = (\zeta_t)$ to be the equality

$$\zeta_t = \Gamma_t + W_t, \quad 0 \le t \le T.$$

Lemma 1.5. *If $P_{\overline{\xi}}$, $P_{\underline{\xi}}$, and P_ζ are measures on $(C_{[0,T]}, \mathcal{B}_{[0,T]})$ corresponding to the processes $\overline{\xi} = (\overline{\xi}_t)$, $\underline{\xi} = (\underline{\xi}_t)$, and $\zeta = (\zeta_t)$, $0 \le t \le T$, respectively, where $\overline{\xi}$ and $\underline{\xi}$ are strong solutions of Eq.* (1.1) *defined in Theorem* 1.4, *then*

$$P_{\overline{\xi}} \ll P_\zeta, \quad P_{\underline{\xi}} \ll P_\zeta.$$

Proof. Let us show, for example, that $P_{\overline{\xi}} \ll P_\zeta$. By (1.15),

$$P\left(\int_0^T (A(t, \overline{\xi}_t) - \gamma_t)^2 dt < \infty\right) = 1.$$

Hence (see [66]), $P_{\overline{\xi} - \Gamma} \ll P_W$. But Γ_t is a deterministic function and, therefore, $P_{\overline{\xi}} \ll P_{W+\Gamma} = P_\zeta$. $\square$

Lemma 1.6. *For any* $A \in \mathcal{B}_{[0,T]}$,

$$P_{\overline{\xi}}(A) = P_{\underline{\xi}}(A).$$

Proof. By the above lemma and the form of densities [66], we obtain

$$P_{\overline{\xi}}(A) = \int_A \frac{dP_{\overline{\xi}}}{dP_\zeta}(z)\, dP_\zeta(z) = \int_A \frac{dP_{\underline{\xi}}}{dP_\zeta}(z)\, dP_\zeta(z) = P_{\underline{\xi}}(A).$$

□

Getting back to the proof of the theorem, we note that if $\xi^1 = (\xi_t^1)$ and $\xi^2 = (\xi_t^2)$ are two arbitrary solutions of Eq. (1.1), then

$$P\big(\underline{\xi}_t \le \xi_t^1, \xi_t^2 \le \overline{\xi}_t,\ 0 \le t \le T\big) = 1,$$

which, together with Lemmas 1.6 and 1.1, leads to the relation

$$P\big(\underline{\xi}_t = \xi_t^1 = \xi_t^2 = \overline{\xi}_t,\ 0 \le t \le T\big) = 1.$$

□

Remark 1.4. A simpler than (1.14) condition

$$\int_0^T m^2(t)\, dt < \infty$$

will, certainly, ensure the uniqueness of the strong solution of Eq. (1.1), but already in the simple case $A(t,x) \equiv A(t)$, $\int_0^T A^2(t)\, dt = +\infty$, where Eq. (1.1) has a unique strong solution, this condition is not satisfied.

Remark 1.5. If we replace the symbol $\int_0^T$ in condition (1.14) by $\int_\varepsilon^T$, $\varepsilon > 0$, then there exists a unique strong solution $\xi_t^{\mathrm{str}}(\varepsilon, x)$ of Eq. (1.1) in the interval $[\varepsilon, T]$ with the initial condition $\xi_\varepsilon^{\mathrm{str}}(\varepsilon, x) = x$, $x \in \mathbb{R}_1$.

In this case, the functional $\Phi(t, \omega, \omega_1)$ from Theorem 1.5 can be constructed rather simply.

Namely,

$$\Phi(t, \omega, \omega_1) = \lim_{n\to\infty} \Phi^n(t, \omega, \omega_1),$$

where for any $n \ge 1$

$$\Phi^n(t, \omega, \omega_1) = \begin{cases} U\left(\Phi^{n-1}, \dfrac{1}{n}, t, \omega, \omega_1\right) & \text{if } t \le \dfrac{1}{n}; \\ \xi_t^{\mathrm{str}}\left(\dfrac{1}{n}, \Phi^n\left(\dfrac{1}{n}, \omega, \omega_1\right)\right) & \text{if } t > \dfrac{1}{n}, \end{cases} \qquad \Phi^0(t, \omega, \omega_1) = \xi_t(\omega).$$

1.3. The Carathéodory-Type SDE. Local Solutions: Existence and Extension Theorems

In this section, we consider Eq. (1.1) on the whole time interval $[0, \infty)$. For convenience we assume that $\xi_0 = 0$.

Fix $T > 0$ and let $g(t)$, $0 \le t \le T$, $g(0) = 0$ be a continuous decreasing function. Let, further,

$$\tau_u^g = \inf\{t > 0 : |u_t| \ge g(t)\} \wedge T \tag{1.16}$$

(with the usual convention $\inf \varnothing = \infty$), where $a \wedge b = \min(a, b)$, $u \in C_{[0,T]}$.

Let $\Xi^{\mathrm{loc},g}$ and $\Xi_s^{\mathrm{loc},g}$ denote the classes of *anticipating* and *strong local solutions* of Eq. (1.1), i.e., we say, e.g., that the continuous stochastic process $\xi = (\xi_t)$, defined, perhaps, only on a stochastic interval $[0, \tau_\xi^g]$, belongs to the class $\Xi_s^{\mathrm{loc},g}$ if $P(\tau_\xi^g > 0) = 1$ and the process $(\xi_{t\wedge\tau_\xi^g})$ is adapted to the filtration $(\mathcal{F}^W_{t\wedge\tau_\xi^g})$ and is such that for every t,

$$\xi_{t\wedge\tau_\xi^g} = \int_0^{t\wedge\tau_\xi^g} A(s, \xi_s)\, ds + W_{t\wedge\tau_\xi^g} \quad (P\text{-a.s.}),$$

or, equivalently,

$$\xi_t = \int_0^t A(s, \xi_s)\, ds + W_t \quad (t \le \tau_\xi^g, \quad P\text{-a.s.}).$$

The coefficient $A(t, x)$ of Eq. (1.1) is assumed to satisfy the following two conditions:

(1) the function $A(t, x)$ is measurable in t for any fixed $x \in \mathbb{R}_1$ and continuous in x for any fixed $t \ge 0$;
(2) there exists a continuous function $M(t)$, $M(t) \ge 0$, $0 \le t \le T$, $M(0) = 0$, such that for any t, $0 \le t \le T$,

$$\int_0^t \overline{A}(s, \alpha_s + M(s))\, ds \le M(t), \tag{1.17}$$

where

$$\alpha_t = (1 + \varepsilon) \left(2t \ln \ln \frac{1}{t} \right)^{1/2}, \quad \varepsilon > 0, \tag{1.18}$$

is the Kolmogorov–Khinchin–Lévy upper function [38], and

$$\overline{A}(t, x) = \sup_{|y| \le x} |A(t, y)|, \quad x \in \mathbb{R}_1. \tag{1.19}$$

Remark 1.6. Condition (0.3) implies (1.17). It is sufficient to set

$$M(t) = \int_0^t m(s)\, ds, \quad 0 \le t \le T.$$

Remark 1.7. The function $M(t)$ from (1.17) can be constructed as follows. For each $n \ge 1$, $t \in [0, T]$, we set

$$M^n(t) = \int_0^t \overline{A}(s, M^{n-1}(s) + \alpha_s)\, ds, \quad M(0) \equiv 0. \tag{1.20}$$

It is easy to see that $\forall t \in [0, T]$ and $n \ge 1$,

$$M^{n+1}(t) \ge M^n(t)$$

(since $\overline{A}(t, x) \ge 0$ and $\overline{A}(t, x) \uparrow x$). Hence, there exists

$$\lim_{n \to \infty} M^n(t) := M(t), \quad 0 \le t \le T \tag{1.21}$$

(finite or infinite). The function $\overline{A}(t, x)$ is left-continuous. Therefore, if we pass to the limit as $n \to \infty$ in (1.20), we obtain

$$\lim_{n \to \infty} M^n(t) = \int_0^t \overline{A}(s, \lim_{n \to \infty} M^n(s) + \alpha_s)\, ds,$$

i.e.,

$$M(t) = \int_0^t \overline{A}(s, M(s) + \alpha_s)\, ds,$$

where $M(t)$ is defined by (1.21).

If we now assume that the sequence $\{M^n\}_{n \ge 1}$ is bounded, $M^n(t) \le h(t)$ $\forall n \ge 1$, $t \in [0, T]$, and $h(t)$ is a finite function, then the function $M(t)$ from (1.21) satisfies all the desirable conditions.

Obviously, the function $M = M(t)$, $0 \le t \le T$, just constructed, is the minimal solution of inequality (1.17), and $M(t) \uparrow t$.

We set

$$\widetilde{A}(t, x) = \begin{cases} A(t, x) & \text{if } |x| \le \beta_t, \\ A(t, \beta_t \operatorname{sign} x) & \text{if } |x| > \beta_t, \end{cases} \tag{1.22}$$

where $\beta_t = \alpha_t + M(t)$, $0 \le t \le T$.

Along with Eq. (1.1), we consider an auxiliary equation

$$d\xi_t = \widetilde{A}(t, \xi_t)dt + dW_t, \quad 0 \le t \le T, \quad \xi_0 = 0 \tag{1.23}$$

(with "truncated" coefficient). Denote by $\widetilde{\Xi}$ and $\widetilde{\Xi}_s$ the classes of anticipating and strong solutions of Eq. (1.23). The following proposition will help to establish a relation between global and local solutions of Eq. (1.1).

Proposition 1.2. *If the process $\xi = (\xi_t)$, $0 \le t \le T$, belongs to the class $\widetilde{\Xi}$ ($\widetilde{\Xi}_s$, respectively), then $\xi = (\xi_t)$, $0 \le t \le T$, belongs to the class $\Xi^{\mathrm{loc},\beta}$ ($\Xi_s^{\mathrm{loc},\beta}$, respectively).*

Proof. Indeed,

$$\xi_t = \int_0^t \widetilde{A}(s, \xi_s)\, ds + W_t = \int_0^t A(s, \xi_s)\, ds + W_t$$

($t \le \tau_\xi^\beta$, P-a.s.). It is enough to show that

$$P(\tau_\xi^\beta > 0) = 1.$$

Denoting $\eta_t = \xi_t - W_t$, we obtain

$$\begin{aligned} \tau_\xi^\beta &= \inf\{t > 0 : |\xi_t| \ge \beta_t\} \wedge T \\ &= \inf\{t > 0 : |\eta_t + W_t| \ge M(t) + \alpha_t\} \wedge T \\ &\ge \inf\{t > 0 : |\eta_t| + |W_t| \ge M(t) + \alpha_t\} \wedge T \\ &\ge \inf\{t > 0 : |W_t| \ge \alpha_t\} \wedge T = \tau_W^\alpha \wedge T, \end{aligned}$$

since by (1.17), $|\eta(t)| \le M(t)$ $\forall t \in [0, T]$ (recall that the process ξ satisfies Eq. (1.23)). But by the law of the iterated logarithm,

$$P(\tau_W^\alpha > 0) = 1.$$

□

Now we note that for any t, $0 \le t \le T$, and $x \in \mathbb{R}_1$,

$$|\widetilde{A}(t, x)| \le \overline{A}(t, \beta_t) \equiv m(t),$$

i.e., $\widetilde{A}(t, x)$ satisfies the Carathéodory condition (0.3).

Thus, by virtue of Proposition 1.2, it suffices to substitute in the statements of Theorems 1.2–1.6 the phrase "local solution" instead of the word "solution" and the stochastic interval $[0, \tau_\xi^\beta]$ instead of the interval $[0, T]$, in order for these statements to be also valid for the case under consideration (i.e., under condition (1.17)).

The following theorem allows us to extend the solutions of Eq. (1.1) to the whole interval $[0, \infty)$.

Theorem 1.8. *Let the following two conditions be satisfied:*

(a) *for any s, $s \ge 0$, and x, $x \in \mathbb{R}_1$, there exist a point $T^{s,x}$, $T^{s,x} \ge T_0 > 0$ (where T_0 is a fixed point) and a continuous function $M^{s,x}(T) \ge 0$, $M^{s,x}(0) = 0$, defined on $[0, T^{s,x}]$ such that for any t, $0 \le t \le T^{s,x}$,*

$$\int_0^t \sup_{|y| \le M^{s,x}(u) + \alpha_u} |A(s + u, x + y)|\, du \le M^{s,x}(t); \tag{1.24}$$

(b) *for $s = 0$, there exist a point $T^{0,0}$, $T^{0,0} > 0$, and a continuous function $M^{0,0}(t) > 0$, $M^{0,0}(0) = 0$, defined on $[0, T^{0,0}]$, such that for any t, $0 \le t \le T^{0,0}$, inequality (1.24) holds only at the point $x = 0$.*

Proof. Denote $\beta_t^{s,x} = \alpha_t + M^{s,x}(t)$, $0 \le t \le T^{s,x}$, and define the sequence of hitting times $(\tau_n)_{n\ge 1}$ by the following recursive equalities:

$$\begin{aligned}
&\tau_1(\omega) = \inf\{t > 0 : |\omega_t| \ge \beta_t^{0,0}\} \wedge T^{0,0},\\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\\
&\tau_n(\omega) = \inf\{t > \tau_{n-1} : |\omega_t - \omega_{\tau_{n-1}}| \ge \beta_{t-\tau_{n-1}}^{\tau_{n-1},\omega_{\tau_{n-1}}}\} \wedge \{\tau_{n-1} + T^{\tau_{n-1},\omega_{\tau_{n-1}}}\},\\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots
\end{aligned} \tag{1.25}$$

For any $(s,x) \in [0,\infty) \times \mathbb{R}_1$, consider Eq. (1.1) on the interval $[s,\infty)$ with the initial condition $\xi_s = x$.

Then it follows from the above-stated results and from the assumptions of the theorem that there exists an anticipating solution $(\xi_t(s,x))$ of the equation such that $\xi_s(s,x) = x$ and

$$P\left\{\xi_t(s,x) = x + \int_s^t A(u, \xi_u(s,x))\, du + W_t - W_s,\ s \le t \le \tau_{s,x} + s\right\} = 0,$$

where

$$\tau_{s,x} = [\inf\{t > s : |\xi_t(s,x) - x| \ge \beta_{t-s}^{s,x}\} - s] \wedge T^{s,x}. \tag{1.26}$$

Let the stochastic process (ξ_t), $t \ge 0$, be given by the relation

$$\xi_t = \begin{cases} \xi_t(0,0) & \text{if } 0 \le t \le \tau_1,\\ \xi_t(\tau_n, \xi_{\tau_n}) & \text{if } \tau_n \le t \le \tau_{n+1}, \quad n \ge 1, \end{cases}$$

where in definition (1.25) of hitting times one has to substitute the values of the process ξ, i.e., $\tau_n \equiv \tau_n(\xi)$.

Now we note that

$$|\xi_t(s,x) - x - (W_t - W_s)| \le M^{s,x}(t-s) \quad (s \le t \le \tau_{s,x} + s,\ P\text{-a.s.}).$$

Hence

$$\tau_{s,x} \ge \tau_s^0 \quad (P\text{-a.s.}),$$

where

$$\tau_s^0 := [\inf\{t > s : |W_t - W_s| \ge \alpha_{t-s}\} - s] \wedge T^{0,0}.$$

By the definition, for all $n \ge 1$,

$$\tau_n = \tau_{n-1} + \tau_{n-1,\xi_{\tau_{n-1}}}, \quad \tau_0 = 0.$$

Let the sequence τ_n^* be given recursively by the following formula: for any $n \ge 1$,

$$\tau_n^* = \tau_{n-1}^* + \tau_{\tau_{n-1}^*}^0, \quad \tau_0^* = 0.$$

Now we recall that by the law of the iterated logarithm

$$P(\tau_s^0 > 0) = 1$$

and from the strong Markov property of a Wiener process $W = (W_t)$, the sequence $\{\tau_{\tau_{n-1}^*}^0\}_{n\ge 1}$ is an i.i.d. sequence. Hence

$$\tau_n^* = \sum_{i=1}^n \tau_{\tau_{i-1}^*}^0 \to \infty \quad \text{as} \quad n \to \infty \quad (P\text{-a.s.}).$$

But

$$P(\tau_n \ge \tau_n^*) = 1.$$

Thus,

$$\tau_n \to \infty \quad \text{as} \quad n \to \infty \quad (P\text{-a.s.}).$$

□

Remark 1.8. If we take a strong solution of Eq. (1.1) as the initial solution $\xi_t(s,x)$ which is involved in the proof of the theorem, then, obviously, the extension will also be strong.

Corollary 1.1. *Change condition* (a) *in the formulation of Theorem* 1.8 *as follows*:

(a′) *for some real number $H > 0$ and for any s, $0 \le s \le T$, and x, $|x| \le H$, there exist a real number $T^{s,x}$, $T^{s,x} \ge T^H > 0$, and a continuous function $M^{s,x}(t) \ge 0$, $M^{s,x}(0) = 0$, defined on $[0, T^{s,x}]$, such that for any t, $0 \le t \le T^{s,x}$, inequality* (1.24) *holds.*

Then the (*anticipating or strong*) *solution $\xi = (\xi_t)$ of Eq.* (1.1) *can be extended up to the moment τ_ξ^H, where τ_ξ^H is defined by* (1.16).

The proof is obvious.

Theorem 1.9. *Under the condition of Theorem* 1.8, *the equality*

$$\xi_t(\omega) = \Phi(t, \omega, \omega), \quad t \ge 0 \quad (P\text{-a.s.})$$

holds, where the functional $\Phi(t, \omega, \omega_1)$ has the properties described in Theorem 1.5.

Proof. We will use the notation introduced in the proof of the previous theorem and Theorem 1.5.

Let $\xi = (\xi_t)$, $t \ge 0$, be some anticipating solution of Eq. (1.1). For any $(s,x) \in [0,\infty) \times \mathbb{R}_1$, we will consider Eq. (1.1) in the interval $[s, \infty)$ with the initial condition $\xi_s = x$. Then from the conditions of the theorem and due to the above-stated results it follows that there exist:

(1) an anticipating solution of Eq. (1.1), $\xi_t(s,x)$, $s \le t \le \tau_{s,x} + s$, $\xi_s(s,x) = x$;

(2) a strong solution of Eq. (1.1), $\xi_t^{\mathrm{str}}(s,x)$, $s \le t \le \tau_{s,x} + s$, $\xi_s^{\mathrm{str}}(s,x) = x$;

(3) the functional $\Phi_{s,x}^\xi(t, \omega, \omega_1)$ with the properties stated in Theorem 1.5. The Markov moment $\tau_{s,x}$ is defined by (1.26).

Let $\Phi_1(t, \omega, \omega_1)$ denote a functional corresponding to the process $\xi = (\xi_t)$, $t \ge 0$, and coinciding with $\Phi_{0,0}^\xi(t, \omega, \omega_1)$ in the interval $0 \le t \le \tau_{0,\xi_0}$.

For any $n > 1$, we construct the functional $\Phi_n(t, \omega, \omega_1)$ in the following way. In the interval $0 \le t \le \tau_{n-1}(\Phi_{n-1})$, we set

$$\Phi_n(t, \omega, \omega_1) = \Phi_{n-1}(t, \omega, \omega_1).$$

If, otherwise, $t \ge \tau_{n-1}(\Phi_{n-1})$, then:

(1) for the ω_1 for which

$$\xi_{\tau_{n-1}(\Phi_{n-1})}(\omega) = \Phi_{n-1}(\tau_{n-1}(\Phi_{n-1}), \omega, \omega_1),$$

we set

$$\Phi_n(t, \omega, \omega_1) = \Phi^\xi_{\tau_{n-1}(\Phi_{n-1}), \xi_{\tau_{n-1}(\Phi_{n-1})}}(t, \omega, \omega_1);$$

(2) for the ω_1 for which

$$\xi_{\tau_{n-1}(\Phi_{n-1})}(\omega) \ne \Phi_{n-1}(\tau_{n-1}(\Phi_{n-1}), \omega, \omega_1),$$

we set

$$\Phi_n(t, \omega, \omega_1) = \xi_t^{\mathrm{str}}(\tau_{n-1}(\Phi_{n-1}), \Phi_{n-1}(\tau_{n-1}(\Phi_{n-1}), \omega, \omega_1)).$$

Similarly to the proof of the previous theorem, we can prove that

$$P(\tau_n(\Phi_n) \ge \tau_n^*) = 1$$

and $P(\tau_n^* \to \infty \text{ as } n \to \infty) = 1$. □

Theorem 1.10. *Let the conditions of Theorem* 1.8 *hold and let for any $s > 0$,*

$$\int_0^{T^{s,x}} \inf_{|z| \le \beta_u^{s,x}} \sup_{|y| \le \beta_u^{s,x}} |A(s+u, x+y) - A(s+u, x+z)|^2 du < \infty, \tag{1.27}$$

where $\beta_t^{s,x} = \alpha_t + M^{s,x}(t)$.

If $s = 0$, then inequality (1.27) *can take place only at the point $x = 0$.*

Then there exists a pathwise unique strong solution of Eq. (1.1) *on the whole interval* $[0,\infty)$.

Proof. It is similar to the proof of Theorem 1.8 and is omitted here. □

Corollary 1.2. *Change condition* (a) *in the formulation of Theorem* 1.10 *as follows*:

(a$''$) *for any real number* $H>0$ *and for any* x, $|x|\le H$, *and* s, $s>0$, *there exist a real number* $T^{s,x}\ge T^H>0$, *and a continuous function* $M^{s,x}(t)\ge 0$, $M^{s,x}(0)=0$, *defined on* $[0,T^{s,x}]$, *such that for any* t, $0\le t\le T^{s,x}$, *inequality* (1.27) *holds.*

Then there exists a local strong pathwise unique solution of Eq. (1.1) *satisfying the equality*

$$\xi_{t\wedge\tau_\xi^H}=\int_0^{t\wedge\tau_\xi^H}A(s,\xi_s)\,ds+W_{t\wedge\tau_\xi^H}\quad (P\text{-}a.s.),$$

for each $H>0$ (τ_ξ^H *is defined in* (1.16)), *and if*

$$\tau=\lim_{H\to\infty}\tau_\xi^H,$$

then

$$\lim_{t\uparrow\tau}|\xi_t|=+\infty\quad(\tau<\infty,\quad P\text{-}a.s.),$$

i.e., there exists a pathwise unique strong solution of Eq. (1.1) *defined up to the explosion time* τ.

Proof. It is obvious and is omitted here. □

Remark 1.9. Let the function $A(t,x):[0,T]\times\mathbb{R}_1\to\mathbb{R}_1$ be

(1$'$) Borel-measurable in the pair (t,x);

(2$'$)

$$\overline{\lim_{t\to 0}}\sup_{|x|\le\alpha_t}|A(t,x)|<\infty, \tag{1.28}$$

where α_t is defined by (1.18).

For each $t\in[0,T]$, $\varepsilon>0$, denote

$$\widetilde{\alpha}_t=\Big(1+\frac{\varepsilon}{2}\Big)\left(2t\ln\ln\frac{1}{t}\right)^{1/2},\qquad M(t)=\frac{\varepsilon}{2}\left(2t\ln\ln\frac{1}{t}\right)^{1/2}.$$

Obviously, $\alpha_t=\widetilde{\alpha}_t+M(t)$, and by (1.28),

$$\overline{\lim_{t\to 0}}\,\overline{A}(t,\alpha_t)<\infty, \tag{1.29}$$

where $\overline{A}(t,x)$ is given by (1.19).

Since the function $\overline{A}(t,\alpha_t)$ is bounded (see (1.29)) in the neighborhood of the point $t=0$, we obtain

$$\int_0^t\overline{A}(s,\alpha_s)\,ds=O(t)\quad\text{as}\quad t\to 0.$$

Hence, there exists a point $t_0>0$ such that $\forall t\in[0,t_0]$,

$$\int_0^t\overline{A}(s,\widetilde{\alpha}_s+M(s))\,ds=\int_0^t\overline{A}(s,\alpha_s)\,ds\le M(t). \tag{1.30}$$

Obviously,

$$\inf_{|y|\le\alpha_t}\sup_{|x|\le\alpha_t}|A(t,x)-A(t,y)|^2\le 4(\sup_{|x|\le\alpha_t}|A(t,x)|)^2=4\overline{A}^2(t,\alpha_t).$$

Thus, we obtain

$$\int_0^{t_0}\inf_{|y|\le\alpha_t}\sup_{|x|\le\alpha_t}|A(t,x)-A(t,y)|^2dt<\infty. \tag{1.31}$$

For each $t \in [0, t_0]$ and $x \in \mathbb{R}_1$, we denote

$$\widetilde{A}(t,x) = \begin{cases} A(t,x) & \text{if } |x| \le \alpha_t, \\ A(t, \alpha_t \operatorname{sign} x) & \text{if } |x| > \alpha_t, \end{cases}$$

and consider Eq. (1.23).

Since for any $x \in \mathbb{R}_1$, $|\widetilde{A}(t,x)| \le |\overline{A}(t,\alpha_t)|$ and

$$\varlimsup_{t\to 0} |\widetilde{A}(t,x)| \le \varlimsup_{t\to 0} |\overline{A}(t,\alpha_t)| < \infty,$$

we obtain

$$|\widetilde{A}(t,x)| \le c < \infty$$

for any $t \in [0, t_0]$, $x \in \mathbb{R}_1$.

Therefore, by [131], Eq. (1.23) (with the truncated coefficient) is strongly regular. On the other hand, condition (1.17) and condition (b) of Theorem 1.8 are satisfied in the interval $[0, t_0]$ (see (1.30) and (1.31)). Now, similarly to the proofs of Proposition 1.2 and Theorem 1.8 (where we do not use the continuity of the function $A(t,x)$ in the variable x), we can easily verify that the following theorem is valid.

Theorem 1.11. *Under conditions* $(1')$ *and* $(2')$, *Eq.* (1.1) *is locally strongly regular.*

1.4. Special Cases and Examples

Let $A(t,x)$ be a Borel-measurable function continuous with respect to x.

1. (a) There exists a constant $c < \infty$ such that $|A(t,x)| \le c$ for any $x \in \mathbb{R}_1$ and $t \in [0,T]$.

Then there exists a pathwise unique strong solution of Eq. (1.1) (see [75]).

(b) There exist constants H and $C_H < \infty$ such that $|A(t,x)| \le C_H$ for any t, $0 \le t \le T$, and x, $|x| \le H$.

Then there exists a pathwise unique local strong solution of Eq. (1.1) defined up to the moment τ_ξ^H (see (1.16)).

(c) If condition (b) is satisfied for all H, $H > 0$, then there exists a pathwise unique strong solution of (1.1) defined up to the explosion time $\tau = \lim\limits_{H\to\infty} \tau_\xi^H$, which is a generalization of the analogous statement for the case where $A(t,x)$ satisfies a local Lipschitz condition.

2. Let there exist a constant $r_0 > 0$ and a function $b(t,r)$ with $b(t,r) \ge 0$,

$$\int_0^T b(t,r)\,dt < \infty \quad \forall r,$$

and for $|x| \le r_0$,

$$|A(t,x)| \le b(t,r_0), \quad 0 \le t \le T.$$

Then the conditions of Theorems 1.8 and 1.10 are satisfied if for any $(s,x) \in [0,\infty) \times \mathbb{R}_1$, we take $T^{s,x} = T_0 > 0$ so small that $M(t) + \alpha_t \le r_0$, where

$$M(t) = \int_0^t b(s,r_0)\,ds.$$

If, in addition,

$$\int_0^T b^2(s,r)\,ds < \infty \quad \forall r,$$

then Eq. (1.1) is strongly regular on the whole interval $[0,\infty)$.

3. (a) Let there exist a constant $c > 0$ such that for each $t \in [0,\infty)$

$$|A(t,x)| \le c(1+|x|) \quad \forall x \in \mathbb{R}_1.$$

We now find the function $M^{s,x}(t)$ and the point T_0 such that conditions (1.24) and (1.27) be satisfied.

Indeed, it is sufficient to show that for each (s,x) there exists a function $M^{s,x}(t)$ such that

$$\int_0^t c(1+|x|+\alpha_u+M^{s,x}(u))\,du \le M^{s,x}(t) \tag{1.32}$$

for each $t\in[0,T_0]$, $T_0>0$, and

$$\int_0^{T_0} \overline{A}^2(s+u,x+\alpha_u+M^{s,x}(u))\,du<\infty.$$

We set

$$M^{s,x}(t):=At+B|x|t,$$

where $A>0$ and $B>0$ are some constants. From inequality (1.32) we obtain

$$ct+c|x|t+cK(t)+cA\frac{t^2}{2}+cB|x|\frac{t^2}{2}\le At+B|x|t, \tag{1.33}$$

where $K(t)=\int_0^t \alpha_s ds$.

Since inequality (1.33) is true at the point $t=0$, it is sufficient to require that the inequality

$$c+c|x|+c\alpha_t+cAt+cB|x|t\le A+B|x|$$

for the derivatives (in the variable t) of the left- and right-hand sides of inequality (1.33) be satisfied.

For this, it is sufficient to require that

$$c(1+\alpha_t+At)\le A,$$
$$c(1+Bt)\le B.$$

But the left-hand sides of the previous inequalities are increasing functions of t. Therefore, if we find the constants A and B from the equalities

$$c(1+\alpha_{T_0}+AT_0)=A,$$
$$c(1+BT_0)=B,$$

where $T_0>0$ is a constant, and take $T_0<\dfrac{1}{c}$, then the function

$$M^{s,x}(t)=\frac{c(1+\alpha_{T_0})}{1-cT_0}t+\frac{c}{1-cT_0}|x|t,\quad 0\le t\le T_0,$$

and the point T_0 satisfies all the requirements.

(b) More generally, let

$$|A(t,x)|\le a_t+b_t|x|, \tag{1.34}$$

where (a_t) and (b_t) are deterministic functions,

$$a_t\ge 0,\quad b_t\ge 0,\quad 0\le t\le T,$$
$$\int_0^T a_t dt<\infty,\quad \int_0^T b_t dt<\infty,\quad x\in\mathbb{R}_1.$$

It is easy to see that if we consider the equation

$$M^{s,x}(t)=\int_0^t a_{s+u}du+(|x|+c+V^{s,x})\int_0^t b_{s+u}du, \tag{1.35}$$

for a function $M^{s,x}$, where

$$c=\alpha_T,\quad V^{s,x}=\frac{\displaystyle\int_0^{T_0} a_{s+u}du+(|x|+c)\int_0^{T_0} b_{s+u}du}{1-\displaystyle\int_0^{T_0} b_{s+u}du}$$

(note that $V^{s,x} = M^{s,x}(T_0)$) and take $T_0 > 0$ so small that for any s, $0 \le s \le T$,

$$\int_0^{T_0} b_{s+u} du < 1,$$

then the function $M^{s,x}(t)$ and the point $T_0 > 0$ constructed in (1.35) satisfy all the requirements of inequality (1.24).

If, in addition, we require that

$$\int_0^T a_t^2 dt < \infty \quad \text{and} \quad \int_0^T b_t^2 dt < \infty, \tag{1.36}$$

then condition (1.27) is also satisfied.

4. Now we show that conditions (1.24) and (1.27) are essential.

(a) This example is very simple. Let $A(t,x) = \phi(t)$, i.e., it does not depend on the variable x. Then inequality (1.24) is satisfied if and only if

$$\int_0^{T_0} |\phi(t)| dt < \infty.$$

Indeed, if the last inequality is satisfied and we denote $M(t) = \int_0^t |\phi(s)\, ds$, then inequality (1.24) is satisfied. The inverse is also trivial. Hence, if inequality (1.24) is not satisfied, then Eq. (1.1) has no sense.

(b) Let

$$A(t,x) = 2a(t)|x|^{1/2} \operatorname{sign} x, \quad 0 \le t \le T, \quad x \in \mathbb{R}_1,$$

where $a(t) = \alpha_t^{-1/2} \alpha_t'$, where α_t is defined in (1.18) and α_t' is its derivative.

It is easy to see that $\overline{A}(t,x) = 2a(t)|x|^{1/2}$ (see (1.19)).

Rewrite inequality (1.24) in the following equivalent form:

$$\int_0^t \overline{A}(s, \beta_s)\, ds + \alpha_t \le \beta_t, \quad 0 \le t \le T,$$

where $\beta_t = \alpha_t + M(t)$ and show that $\beta_t = (3 + 2\sqrt{2})\alpha_t$, $0 \le t \le T$, satisfies the last inequality.

Indeed,

$$\begin{aligned}\int_0^t \overline{A}(s, \beta_s)\, ds + \alpha_t &= 2(3+2\sqrt{2})^{1/2} \int_0^t a(t) \alpha_t^{1/2} dt + \alpha_t \\ &= 2(3+2\sqrt{2})^{1/2} \int_0^t \alpha_s^{-1/2} \alpha_s' \alpha_s^{1/2} ds + \alpha_t \\ &= \left(2(3+2\sqrt{2})^{1/2} + 1\right) \alpha_t = (3+2\sqrt{2})\alpha_t = \beta_t.\end{aligned}$$

On the other hand,

$$\int_0^T 4a^2(t) \inf_{|z| \le \beta_t} \sup_{|y| \le \beta_t} \left| |y|^{1/2} \operatorname{sign} y - |z|^{1/2} \operatorname{sign} z \right|^2 dt = 4(3+2\sqrt{2}) \int_0^T (\alpha_s')^2 ds = \infty. \tag{1.37}$$

Hence, (1.27) is false.

Now we construct two different local strong solutions of Eq. (1.1).

Note that for each $t \in [0,T]$,

$$\int_0^t A(s, \alpha_s)\, ds - \alpha_t = \alpha_t, \tag{1.38}$$

$$\int_0^t A(s, -\alpha_s)\, ds + \alpha_t = \alpha_t. \tag{1.39}$$

Construct a sequence of processes $(\xi^{1,n})_{n\geq 0} = \{(\xi_t^{1,n}),\ 0 \leq t \leq T\}_{n\geq 0}$, putting

$$\begin{aligned} \xi_t^{1,0} &= \alpha_t, \\ \xi_t^{1,n} &= \int_0^t A(s, \xi_s^{1,n-1})\, ds + W_t, \quad n = 1, 2, ..., \end{aligned} \tag{1.40}$$

for each t.

By the law of the iterated logarithm ($t \leq \tau_W^\alpha$, P-a.s.),

$$\xi_t^{1,1} = \int_0^t A(s, \alpha_s)\, ds + W_s = \alpha_t + \alpha_t + W_t \geq \alpha_t = \xi_t^{1,0}.$$

Hence,

$$\xi_t^{1,n} = \int_0^t A(s, \xi_s^{1,n-1})\, ds + W_t \geq \int_0^t A(s, \xi_s^{1,n-2})\, ds + W_t = \xi_t^{1,n-1} \quad (t \leq \tau_W^\alpha,\ P\text{-a.s.})$$

for each $n \geq 2$. Thus, sequence (1.40) is $(t \leq \tau_W^\alpha,\ P\text{-a.s.})$ nondecreasing.

By induction it is easy to verify that

$$|\xi_t^{1,n}| \leq \beta_t, \quad 0 \leq t \leq T \quad (t \leq \tau_W^\alpha,\ P\text{-a.s.})$$

for each $n \geq 0$.

Indeed,

$$|\xi_t^{1,1}| \leq \int_0^t \overline{A}(s, \alpha_s)\, ds + \alpha_t \leq \int_0^t \overline{A}(s, \beta_s)\, ds + \alpha_t = \beta_t$$

$(t \leq \tau_W^\alpha$, P-a.s.), and if $|\xi_t^{1,n-1}| \leq \beta_t$ ($t \leq \tau_W^\alpha$, P-a.s.), then

$$|\xi_t^{1,n}| \leq \int_0^t \overline{A}(s, \xi_s^{1,n-1})\, ds + \alpha_t \leq \int_0^t \overline{A}(s, \beta_s)\, ds + \alpha_t = \beta_t,$$

as was required.

Hence, there exists a finite limit

$$\xi_t^1 = \lim_{n\to\infty} \xi_t^{1,n} \quad (t \leq \tau_W^\alpha,\ P\text{-a.s.}) \tag{1.41}$$

which satisfies Eq. (1.1) with $\xi_0^1 = 0$, and by the construction, the process $(\xi^1_{t\wedge\tau_W^\alpha})$ is $\mathcal{F}^W_{t\wedge\tau_W^\alpha}$-adapted.

Therefore, we construct a τ_W^α-local strong solution ξ^1 of Eq. (1.1) such that

$$\alpha_t \leq \xi_t^1 \leq \beta_t \quad (t \leq \tau_W^\alpha,\ P\text{-a.s.}).$$

Starting from Eq. (1.39), one can construct a τ_W^α-local strong solution ξ^2 of Eq. (1.1) with

$$-\alpha_t \geq \xi_t^2 \geq -\beta_t \quad (t \leq \tau_W^\alpha,\ P\text{-a.s.}).$$

Finally, note that there is no function $\beta^* = (\beta_t^*)$ such that $\beta_t^* < \beta_t$ and conditions (1.24) and (1.27) are satisfied. Indeed, if β^* satisfies condition (1.24), then $\beta^* \geq \alpha$. Thus, by (1.37), condition (1.27) does not hold.

5. (a) The example below shows that condition (1.34) is not necessary for the existence of a strong solution of Eq. (1.1).

Consider the linear equation

$$d\xi_t = a_t \xi_t dt + dW_t, \quad 0 \leq t \leq T, \quad \xi_0 = 0, \tag{1.42}$$

where the function $a = (a_t)$ is such that

$$\int_\varepsilon^T |a_t|\, dt < \infty \quad \text{for any} \quad \varepsilon > 0.$$

Along with Eq. (1.42) we consider the equation

$$d\xi_t = a_t\xi_t dt + dW_t, \quad \varepsilon \le t \le T, \quad \xi_\varepsilon = C \in \mathbb{R}_1 \tag{1.43}$$

with the same coefficient a_t, $\varepsilon \le t \le T$, as in Eq. (1.42).

Obviously, (1.43) has a unique strong solution given by the formula

$$\xi_t^C = \exp\biggl(\int_\varepsilon^t a_u du\biggr)(C - W_\varepsilon) + \int_\varepsilon^t \exp\biggl(\int_s^t a_u du\biggr) a_s W_s ds + W_t,$$
$$\xi_\varepsilon = C, \quad \varepsilon \le t \le T.$$

If, in addition,

$$\int_0^{T^0} |a_t|\, dt < \infty$$

for some small number $T^0 > 0$, then taking $\varepsilon = 0$, we obtain that Eq. (1.42) has a unique strong solution, (ξ_t^0), $0 \le t \le T$. Condition (1.24) here takes the form

$$\int_0^{T^0} \exp\biggl(\int_s^{T^0} |a_u|\, du\biggr)|a_s|\alpha_s ds < \infty,$$

where α_t is defined by (1.18).

Under such a condition, the general solution of Eq. (1.42) has, obviously, the form

$$\xi_t^C = C\exp\biggl(\int_{t_0}^t a_u du\biggr) + \int_{t_0}^t \exp\biggl(\int_s^t a_u du\biggr) a_s W_s ds + W_t, \qquad 0 \le t, t_0 \le T, \tag{1.44}$$

where C is a function of ω, $\omega \in \Omega$.

If

$$\int_0^{T^0} a_t dt = +\infty,$$

then ξ_t^C is a solution of Eq. (1.42) for any C.

In this case, the functional $\Phi(t, \omega, \omega_1)$ from Theorem 1.5 takes the form

$$\Phi(t, \omega, \omega_1) = \xi_t^{C(\omega_1)}(\omega), \quad \omega, \omega_1 \in \Omega.$$

Thus it is seen that the inclusion of the solution of Eq. (1.42) in different classes of solutions (strong, anticipating) depends on the choice of $C(\omega)$.

If

$$\int_0^{T^0} a_t dt = -\infty,$$

then we must take $C = 0$ in (1.44).

(b) It is well known ([126]) that if the stochastic process $\xi = (\xi_t)$, $0 \le t \le T$, is adapted to a family of σ-algebras $(\mathcal{F}_t)$, $0 \le t \le T$, to which, in turn, a Wiener process $W = (W_t)$, $0 \le t \le T$, is adapted, then the pathwise uniqueness of the weak solution of the general equation (1.3) implies that the solution ξ is, actually, a strong one.

Now, with the example of Eq. (1.42) we show that with a specific choice of the function a_t one can construct a (pathwise) unique anticipating (but not a strong) solution of Eq. (1.42) on the whole interval $[0, T]$. Thus, the condition that the solution is adapted to $(\mathcal{F}_t)$, $0 \le t \le T$, in [126] cannot be omitted.

Indeed, set, e.g.,

$$a_t = \begin{cases} 0 \quad \text{if} \ \ t = 0, \\ \dfrac{c_1}{t}, \quad 0 < c_1 < \dfrac{1}{2} \quad \text{if} \ \ 0 \le t \le \dfrac{T}{2}, \\ \dfrac{c_2}{t - \frac{T}{2}}, \quad c_2 < 0 \quad \text{if} \ \ \dfrac{T}{2} < t \le T. \end{cases}$$

It can be easily seen that there is no solution of Eq. (1.43) with the initial condition $\xi_{T/2} \neq 0$ in the interval $\left[\dfrac{T}{2}, T\right]$, but there exists a unique solution with the initial condition $\xi_{T/2} = 0$. The last condition, in turn, defines $C = C(\omega)$ uniquely in formula (1.44) of the general solution of Eq. (1.42) on the interval $\left[0, \dfrac{T}{2}\right]$. Namely,

$$C(\omega) = -\exp\biggl(-\int_{t_0}^{T/2} a_u du\biggr)\biggl[\int_{t_0}^{T/2} \exp\biggl(\int_s^{T/2} a_u du\biggr) a_s W_s ds + W_{T/2}\biggr] \in \mathcal{F}_{T/2}^W.$$

Thus, the solution just constructed is a pathwise unique anticipating solution of Eq. (1.42) on the whole interval $[0, T]$ but, obviously, it is not strong.

1.5. Innovation Problem for Nonlinear Filtering

Consider a stochastic basis $(\Omega, \mathcal{F}, F = (\mathcal{F}_t), t \ge 0, P)$ with a Wiener process $W = (W_t)$, $t \ge 0$, defined on it, and let θ be a random variable independent of W with a distribution function $F(a) = P(\theta \le a)$, $a \in \mathbb{R}_1$.

Further, we consider the Itô process $\xi = (\xi_t)$, $t \ge 0$, with the differential

$$d\xi_t = \theta dt + dW_t, \quad t \ge 0, \quad \xi_0 = 0. \tag{1.45}$$

We want to construct an innovation process $\overline{W}$, i.e., represent the process ξ in the form of the diffusion-type process:

$$d\xi_t = A_t dt + d\overline{W}_t, \quad t \ge 0, \quad \xi_0 = 0,$$

where the process A_t is $\mathcal{F}_t^\xi$-measurable for every t and $\overline{W} = (\overline{W}_t)$, $t \ge 0$, is a Wiener process with

$$\mathcal{F}_t^{\overline{W}} = \mathcal{F}_t^\xi, \quad t \ge 0 \quad (\text{mod } P).$$

Note that both σ-algebras are augmented with respect to the measure P for each $t \ge 0$.

In order to do this, we introduce the function $A(t, x)$: $[0, \infty) \times \mathbb{R}_1 \to \mathbb{R}_1$ by the formula

$$A(t, x) = \begin{cases} \dfrac{\displaystyle\int_{-\infty}^{+\infty} a \exp\left(ax - \frac{a^2 t}{2}\right) dF(a)}{\displaystyle\int_{-\infty}^{+\infty} \exp\left(ax - \frac{a^2 t}{2}\right) dF(a)} & \text{if} \quad t > 0, \\ 0 & \text{if} \quad t = 0. \end{cases} \tag{1.46}$$

From the Bayes formula we have

$$A(t, \xi_t) = E(\theta \mid \mathcal{F}_t^\xi) \quad (dt \times P\text{-a.s.}).$$

Thus, by virtue of [77], $\overline{W} = (\overline{W}_t)$, $t \ge 0$, where

$$\overline{W}_t = \xi_t - \int_0^t A(s, \xi_s)\, ds \tag{1.47}$$

is a Wiener process adapted to $(\mathcal{F}_t^\xi)$, $t \ge 0$.

Under rather strong restrictions on the distribution F of the random variable θ it has been shown that the Wiener process constructed in (1.47) is an innovation process. In [14], θ was assumed to be bounded, in [66] the pair (θ, W) was assumed to be Gaussian, and in [96], $E|\theta| < \infty$.

If we write Eq. (1.47) in the form

$$d\xi_t = A(t, \xi_t)\, dt + d\overline{W}_t, \quad t \geq 0, \quad \xi_0 = 0 \tag{1.48}$$

and consider Eq. (1.48) as a stochastic differential equation, then we note that the problem of the construction of an innovation process can be solved by proving the existence of the unique strong solution of Eq. (1.48).

a. Direct Probabilistic Proof

Application of Theorem 1.1 makes it possible to reduce the problem of construction of the unique solution of Eq. (1.48) to a simpler problem: that is, to the proof of the weak uniqueness of the solution of Eq. (1.48).

Note that the existence of a weak solution of Eq. (1.48) follows directly from (1.47).

We show first that for any $s > 0$ and $z \in \mathbb{R}_1$, Eq. (1.48), considered in the interval $t \geq s > 0$ with the initial condition $\xi_s = z$, is strongly regular.

With this aim in view, we fix $s > 0$ and $z \in \mathbb{R}_1$ and consider the following Itô process:

$$\xi_t(s, z) = z + \theta_{s,z} \cdot (t - s) + W_t - W_s, \quad t > s, \tag{1.49}$$

where $\theta_{s,z}$ is a random variable independent of the future increments $W_t - W_s$, $t \geq s > 0$, of a Wiener process W, and $P(\theta_{s,z} \leq a) = F_{s,z}(a)$, $a \in \mathbb{R}_1$, with

$$F_{s,z}(a) = \frac{\displaystyle\int_{-\infty}^{a} e^{bz - \frac{b^2 s}{2}}\, dF(b)}{\displaystyle\int_{-\infty}^{+\infty} e^{bz - \frac{b^2 s}{2}}\, dF(b)}. \tag{1.50}$$

We introduce the function

$$A_{s,z}(t, x) = \begin{cases} \dfrac{\displaystyle\int_{-\infty}^{+\infty} a \exp\left(ax - \dfrac{a^2(t-s)}{2}\right) dF_{s,z}(a)}{\displaystyle\int_{-\infty}^{+\infty} \exp\left(ax - \dfrac{a^2(t-s)}{2}\right) dF_{s,z}(a)} & \text{for} \quad t > s; \\ 0 \quad \text{for} \quad t = s. \end{cases} \tag{1.51}$$

Note that for any $s > 0$ and $z \in \mathbb{R}_1$,

$$A_{s,z}(t, \xi_t(s, z)) = E(\theta_{s,z} \mid \mathcal{F}_t^{\xi(s,z)}) \qquad (dt \times P\text{-a.s.}).$$

Thus, by virtue of [77], in this case we have again that

$$\overline{W}_t(s, z) = \xi_t(s, z) - z - \int_s^t A_{s,z}(u, \xi_u(s, z))\, du \tag{1.52}$$

is a Wiener process with respect to the filtration $(\mathcal{F}_t^{\xi(s,z)})$, $t \geq s > 0$.

Substituting Eq. (1.50) into (1.51), we obtain

$$A_{s,z}(t, x) = A(t, x + z).$$

Hence, by virtue of Eq. (1.52), we obtain

$$d\xi_t(s, z) = A(t, \xi_t(s, z))\, dt + d\overline{W}_t(s, z), \quad \xi_s(s, z) = z, \quad t \geq s > 0. \tag{1.53}$$

Consequently, the process $\xi_t(s, z)$, $t \geq s$, $z \in \mathbb{R}_1$, is a weak solution of Eq. (1.53) and hence, of Eq. (1.48), on the whole interval $[0, \infty)$ with the initial condition $\xi_s(s, z) = z$.

It can be easily seen now that the function $A(t,x)$ is bounded in the domain

$$\{(t,x) : 0 < s \le t \le T,\ |x| \le H,\ 0 < T < \infty,\ 0 < H < \infty\}.$$

Indeed,

$$|A(t,x)| \le \frac{\displaystyle\int_{-\infty}^{+\infty} |a| \exp\left(|a|H - \frac{a^2 s}{2}\right) dF(a)}{\displaystyle\int_{-\infty}^{+\infty} \exp\left(-|a|H - \frac{a^2 T}{2}\right) dF(a)} < \infty.$$

Hence, by virtue of Corollary 1.2, Eq. (1.48), considered on the interval $[s,\infty)$, $\xi_s = z$, $z \in \mathbb{R}_1$, for any $s > 0$ is strongly regular until the explosion time (see the definition in the statement of Corollary 1.2).

As we can see, the weak solution which was constructed earlier (see Eq. (1.47)) is finite on the whole interval $[s,\infty)$, and we obtain that Eq. (1.48) is strongly regular on the whole interval $[s,\infty)$, $s > 0$, for any initial condition $\xi_s = z \in \mathbb{R}_1$.

Thus, there exists a non-anticipating functional $\Phi_t(s, z, (u_\tau - u_s, \tau \ge s))$, $s > 0$, $t \ge s$, $z \in \mathbb{R}_1$, $u \in C_{[s,\infty)}$, such that for any weak solution (ξ_t, W_t), $t \ge s > 0$, of Eq. (1.48), with the initial condition $\xi_s = \eta$, where η is a random variable independent of the σ-algebra $\mathcal{F}^W_{[s,\infty)}$, the following relation holds:

$$\xi_t = \Phi_t(s, \eta, (W_\tau - W_s,\ \tau \ge s)) \quad (P\text{-a.s.}).$$

Now we introduce a measure $Q_{s,z}(\cdot)$ on the space $(C_{[s,\infty)}, \mathcal{B}_{[s,\infty)})$. For any $s > 0$, $z \in \mathbb{R}_1$, we set

$$Q_{s,z}(B) = P(\Phi_\cdot(s, z, (W_\tau - W_s,\ \tau \ge s)) \in B),$$

where $B \in \mathcal{B}_{[s,\infty)}$.

Obviously, $Q_{s,z}(B) = P(\xi(s,z) \in B)$.

One can easily verify the following properties of the measure $Q_{s,z}$: for all $s > 0$ and $z \in \mathbb{R}_1$ we have

(1) $Q_{s,z}\big(\widehat{\theta} := \lim\limits_{t\to\infty} \dfrac{u_t}{t} \text{ exists}\big) = 1$;

(2) $Q_{s,z}\big((\overline{W}_t(s,z) := u_t - z - (t-s)\widehat{\theta},\ t \ge s) \in B\big) = P^s(B)$, where P^s is a Wiener measure on $(C_{[s,\infty)} \mathcal{B}_{[s,\infty)})$;

(3) $Q_{s,z}(\widehat{\theta} \le a) = F_{s,z}(a)$, $a \in \mathbb{R}_1$;

(4) $Q_{s,z}\big(\widehat{\theta} \le a, (\overline{W}_t(s,z),\ t \ge s) \in B\big) = F_{s,z} \cdot P^s(B)$;

(5) if (ξ_t) is a weak solution of Eq. (1.48) on the whole interval $[0,\infty)$, then

$$P((\xi_t,\ t \ge s) \in B \mid \xi_s = z) = Q_{s,z}(B).$$

Now let $\widetilde{\xi}_t$, $t \ge 0$, be a weak solution of Eq. (1.48) considered on the interval $t \ge 0$ with the initial condition $\widetilde{\xi}_0 = 0$. Let $P_{\widetilde{\xi}}(\cdot)$ be a distribution corresponding to the process $\widetilde{\xi}$, $P(\cdot)$ be a Wiener measure on the space $(C_{[0,\infty)}, \mathcal{B}_{[0,\infty)})$, and $P_{\widetilde{\xi}_s}(\cdot)$ be a distribution of the random variable $\widetilde{\xi}_s$, where $s > 0$ is a fixed point. It is obvious that the following equalities are true:

(1′)

$$\begin{aligned} P_{\widetilde{\xi}}\left(\lim_{t\to\infty} \frac{u_t}{t} \quad \text{exists}\right) &= \int_{\mathbb{R}_1} P_{\widetilde{\xi}}\left(\lim_{t\to\infty} \frac{u_t}{t} \quad \text{exists} \mid u_s = z\right) P_{\widetilde{\xi}_s}(dz) \\ &= \int_{\mathbb{R}_1} Q_{s,z}\left(\lim_{t\to\infty} \frac{u_t}{t} \quad \text{exists}\right) P_{\widetilde{\xi}_s}(dz) = 1. \end{aligned}$$

Let $\lim\limits_{t\to\infty} \dfrac{\widetilde{\xi}_t}{t} = \widetilde{\theta}$.

For any $s > 0$, $z \in \mathbb{R}_1$, we have

(2′)

$$
\begin{aligned}
&P_{\widetilde{\xi}}\big((u_t - u_s - \widetilde{\theta}(t-s),\, t \ge s) \in B\big) \\
&\quad = \int_{\mathbb{R}_1} P_{\widetilde{\xi}}\big((u_t - u_s - \widetilde{\theta}(t-s),\, t \ge s) \in B \mid u_s = z\big) P_{\widetilde{\xi}_s}(dz) \\
&\quad = \int_{\mathbb{R}_1} Q_{s,z}(B) \cdot P_{\widetilde{\xi}_s}(dz) = \int_{\mathbb{R}_1} P^s(B) P_{\widetilde{\xi}_s}(dz) = P^s(B),
\end{aligned}
$$

where $B \in \mathcal{B}_{[s,\infty)}$.

(3′)

$$
\begin{aligned}
&P_{\widetilde{\xi}}\big((u_t - u_s - \widetilde{\theta}(t-s),\, t \ge s) \in B, \widetilde{\theta} \le a\big) \\
&\quad = \int_{\mathbb{R}_1} P_{\widetilde{\xi}}\big((u_t - u_s - \widetilde{\theta}(t-s),\, t \ge s) \in B,\, \widetilde{\theta} \le a \mid u_s = z\big) P_{\widetilde{\xi}_s}(dz) \\
&\quad = \int_{\mathbb{R}_1} Q_{s,z}(B, \widetilde{\theta} \le a) P_{\widetilde{\xi}_s}(dz) = \int_{\mathbb{R}_1} P^s(B) F_{s,z}(a) P_{\widetilde{\xi}_s}(dz) = P^s(B)\widetilde{F}(a),
\end{aligned}
$$

where $\widetilde{F}(a)$ is the distribution of the random variable $\widetilde{\theta}$.

(4′) $P(\widetilde{\theta} \le a \mid \widetilde{\xi}_s = z) = F_{s,z}(a)$.

Thus, if $\widetilde{W}_t = \widetilde{\xi}_t - \widetilde{\theta}t$, $t \ge 0$, then by virtue of properties (2′) and (3′), we easily obtain that the σ-algebra $\sigma(\widetilde{\theta})$ is independent of the σ-algebra $\mathcal{F}^{\widetilde{W}}_{[0,+\infty)} = \sigma\big(\cup_{s>0}\sigma(\widetilde{W}_\tau - \widetilde{W}_s,\, \tau \ge s)\big)$. But $\mathcal{F}^{\widetilde{W}}_{[0,\infty)} = \sigma(\widetilde{W}_t)$, $t \ge 0 = \mathcal{F}^{\widetilde{W}}_{0+} \vee \mathcal{F}^{\widetilde{W}}_{[0+,\infty)} = \mathcal{F}^{\widetilde{W}}_{[0+,\infty)}$, since $\mathcal{F}^{\widetilde{W}}_{0+} = \mathcal{F}^{\widetilde{W}}_{0} = (\varnothing, \Omega)$ (mod P). Hence $\widetilde{W}_t$, $t \ge 0$, is a Wiener process independent of the random variable $\widetilde{\theta}$. In order to show that the distribution of $\widetilde{\xi}$ coincides with that of ξ and, therefore, that Eq. (1.48) has a unique weak solution, it is sufficient to show that $\widetilde{F}(a) = F(a)$, $a \in \mathbb{R}_1$.

It can be easily seen that

$$
P(\widetilde{\theta} \le a \mid \mathcal{F}^{\widetilde{\xi}}_s) = P(\widetilde{\theta} \le a \mid \widetilde{\xi}_s) = F_{s,\widetilde{\xi}_s}(a), \quad a \in \mathbb{R}_1, \quad s > 0 \quad (P_{\widetilde{\xi}}\text{-a.s.}). \tag{1.54}
$$

But

$$
\lim_{s \to 0+} P(\widetilde{\theta} \le a \mid \mathcal{F}^{\widetilde{\xi}}_s) = P(\widetilde{\theta} \le a \mid \mathcal{F}^{\widetilde{\xi}}_{0+})
$$

by virtue of the properties of the reverse martingale.

Further, we note that $P_{\widetilde{\xi}} \ll P_W$ and $\mathcal{F}^W_{0+} = \mathcal{F}^W_0$. Thus, by the zero–one law,

$$
P(\widetilde{\theta} \le a \mid \mathcal{F}^{\widetilde{\xi}}_{0+}) = P(\widetilde{\theta} \le a) = \widetilde{F}(a) \quad (P_{\widetilde{\xi}}\text{-a.s.}).
$$

Finally, by virtue of (1.54),

$$
\widetilde{F}(a) = \lim_{s \to 0+} F_{s,\widetilde{\xi}_s}(a) = F(a).
$$

b. Proof Based on the Extension of Theorem 1.1

Let $(C_{[0,T]}, \mathcal{B}_{[0,T]})$ be a measure space of continuous functions. Let, further, $Y \in \mathcal{B}_{[0,T]}$.

We call the continuous process $\xi = (\xi_t)$, $0 \le t \le T$, a *Y-solution* (see [126]) (weak or strong) of Eq. (1.1) if the process ξ satisfies Eq. (1.1) and is such that

$$
P\{((\xi_t),\, 0 \le t \le T) \in Y\} = 1.
$$

Similarly the notions of *Y-weakly* and *Y-strongly* regular equations can be introduced.

Consider the class

$$
K = \{Y : u, v \in Y \Longrightarrow u \vee v, u \wedge v \in Y\},
$$

where $u \vee v = \max(u_t, v_t)$, $0 \le t \le T$, and $u \wedge v = \min(u_t, v_t)$, $0 \le t \le T$.

It can be easily seen that an analogue of Theorem 1.1 holds also for Y-solutions.

Namely, the following theorem takes place.

Theorem 1.1′. *For any $Y \in K$, Eq.* (1.1) *is Y-strongly regular iff Eq.* (1.1) *is Y-weakly regular.*

The proof coincides with that of Theorem 1.1. The only difference is that here we can apply the Yamada–Watanabe theorem formulated for Y-solutions.

Namely, if Eq. (1.1) has a Y-weak solution and if any two Y-weak solutions given on the same (arbitrary) probability space with the same initial distributions coincide pathwise, then Eq. (1.1) has a pathwise unique Y-strong solution.

Consider the class

$$Y_0 = \left\{ u \in C_{[0,T]} : \int_0^T A^2(t, u_t)\, dt < \infty \right\}.$$

It is obvious that $Y_0 \in K$.

The solution (weak or strong) of Eq. (1.1) is called an AC-*solution* (absolutely continuous solution) if the measure corresponding to this solution is absolutely continuous with respect to the measure of the process $\xi_0 + W$.

It follows from the criterion of absolute continuity of the measure of the diffusion process with respect to the Wiener measure (see [66]) that any Y_0-solution is an AC-solution and vice versa.

On the other hand (as follows from the form of the Radon–Nikodym derivative), the AC-solutions are equivalent in distributions, which implies the following corollary.

Corollary 1.3. *If there exists a weak AC-solution of Eq.* (1.1)*, then it is a pathwise unique strong AC-solution.*

Returning to the innovation problem, note that Eq. (1.48) has a weak solution, namely, the initial process ξ (see (1.47)). But the measure of the Itô process ξ is absolutely continuous with respect to the Wiener measure (see (1.45)). The problem is solved.

Chapter 2

PARTIALLY OBSERVABLE DIFFUSION-TYPE PROCESSES. CONSTRUCTION OF AN INNOVATION PROCESS

2.1. A Stochastic Version of the Gronwall–Bellman Lemma

The following lemma is a stochastic version (multidimensional) of the well-known Gronwall–Bellman lemma.

Lemma 2.1. *Let, on a stochastic basis $(\Omega, \mathcal{F}, F = (\mathcal{F}_t),\ 0 \le t \le T,\ P)$, the following objects be given:*

(1) *a multidimensional continuous process $X_t = (X_t^1, ..., X_t^n)$ with $X_t \ge 0$ (P-a.s.), $0 \le t \le T$;*

(2) *a matrix process*

$$K_t = (K_t^{ij}), \quad i, j = 1, ..., n, \quad 0 \le t \le T, \quad K(0) = 0,$$

which is increasing and continuous;

(3) *a multidimensional continuous local martingale*

$$M_t = (M_t^1, ..., M_t^n), \quad M_0^i = 0, \quad i = 1, ..., n.$$

Let

$$0 \le X_t \le \int_0^t X_s dK_s + M_t, \quad 0 \le t \le T \quad (P\text{-a.s.}).$$

Then

$$P\big(\sup_{0 \le t \le T} \|X_t\| = 0\big) = 1,$$

where $\|X_t\| = \sum\limits_{j=1}^{n} |X_t^j|$.

Proof. Note that if B is an increasing process, $M \in \mathcal{M}^c_{\mathrm{loc}}$, and $0 \le M_t + B_t$, then if $EB_\tau < \infty$, where τ is a stopping time, $M_{t\wedge\tau}$ is a supermartingale.

Indeed,

$$M_{t\wedge\tau} \ge -B_{t\wedge\tau} \ge -B_\tau.$$

Hence, $M_{t\wedge\tau}$ is bounded by an integrable random variable and, therefore, is a supermartingale.

It is obvious that

$$\|X_t\| \le \int_0^t \|X_s\|\, d\widetilde{K}_s + \widetilde{M}_t,$$

where

$$\widetilde{K}_t := t + \sup_{0\le s\le t} \|X_s\| + \max_{1\le i,j\le n} K_t^{ij}, \quad \widetilde{M}_t = \sum_{i=1}^{n} M_t^i.$$

Consider the stopping time τ_c defined by the equality

$$\widetilde{K}_{\tau_c} = c.$$

We show that for

$$B_t = \int_0^t \|X_s\|\, d\widetilde{K}_s,$$

$EB_{\tau_c} < \infty$ is true.

Indeed,

$$B_t \le \sup_{0\le s\le t} \|X_s\| \widetilde{K}_t \le (\widetilde{K}_t)^2.$$

Hence,

$$B_{\tau_c} \le (\widetilde{K}_{\tau_c})^2 = c^2.$$

Thus, $\widetilde{M}_{t\wedge\tau_c}$ is a supermartingale and $E\widetilde{M}_{t\wedge\tau_c} \le 0$.

Changing the time $X_{\tau_c} = Y(c)$, we obtain

$$\|Y(c)\| \le \int_0^\tau \|X_s\|\, d\widetilde{K}_s + \widetilde{M}_{\tau_c}.$$

Since $\tau_c = \widetilde{K}^{-1}(c)$, we have

$$\|Y(c)\| \le \int_0^c \|Y(l)\|\, dl + \widetilde{M}_{\tau_c}.$$

Hence,

$$E\|Y(c)\| \le \int_0^t E\|Y(l)\|\, dl.$$

The latter, by the Gronwall–Bellman lemma, leads to the equality $E\|Y(c)\| = 0$, which implies $Y(c) = 0$ (P-a.s.). But this means that $\sup\limits_c \|X_{\tau_c}\| = 0$ and since $\tau_c \to \infty$ as $c \to \infty$, we have $\sup\limits_{0\le t\le T} \|X_t\| = 0$ (P-a.s.). □

2.2. Innovation Problem for a Component of the Diffusion-Type Process

Consider a diffusion-type process (η, ξ) which is a solution of the following SDE:

$$\begin{aligned} d\eta_t &= a_t(\xi, \eta)dt + b_t(\eta)dv_t, \quad \eta_0 = 0, \\ d\xi_t &= A_t(\xi, \eta)dt + dW_t, \quad \xi_0 = 0, \quad 0 \le t \le T, \end{aligned} \tag{2.1}$$

where v and W are independent Wiener processes and a, b, and A are nonanticipating functionals.

Let $g_t(x, y)$, $0 \le t \le T$, $x, y \in C_{[0,T]}$ denote any of the functionals $a_t(x, y)$, $A_t(x, y)$ and assume that the following conditions are satisfied for any $x, \overline{x}, y \in C_{[0,T]}$ and $t \in [0, T]$:

(1) the Lipschitz condition with respect to the variable x,

$$|g_t(x, y) - g_t(\overline{x}, y)|^2 \le \mathrm{const}\left((x_t - \overline{x}_t)^2 + \int_0^t (x_s - \overline{x}_s)^2 dK_s\right);$$

(2) the linear growth condition

$$|g_t(x, y) \le \mathrm{const}\left(1 + x_t^2 + y_t^2 + \int_0^t (x_s^2 + y_s^2)\, dK_s\right),$$

where $K = (K_t)$, $K_0 = 0$, is an increasing continuous function;

(3) the coefficient $b_t(y)$ is such that the equation

$$d\zeta_t = b_t(\zeta)dv_t, \quad \zeta_0 = 0$$

has a unique strong solution.

Theorem 2.1. *Let $(\eta, \xi) = (\eta_t, \xi_t)$, $0 \le t \le T$, be a strong solution of SDE* (2.1)*. If conditions* (1), (2), *and* (3) *are satisfied, then there exists an innovation Wiener process $\overline{W} = (\overline{W}, F^\xi)$ for the process ξ, and*

$$F^{\overline{W}} = F^\xi \pmod P.$$

Remark 2.1. The process $\overline{W} = (\overline{W}, F^\xi)$ and the filtrations $F^{\overline{W}}$ and F^ξ are defined in the Introduction, (0.6), (0.7), and (0.8).

Proof. For simplicity we consider the case where $b_t(y) = 1$ (in the general case it is sufficient to consider a distribution of the solution of the equation

$$d\zeta = b\, dv$$

instead of the Wiener distribution). We suppose also that $dK_t = dt$.

Both assertions of the theorem will be proved if we show that the process ξ is represented in the form of a diffusion-type process

$$d\xi_t = m_t(\xi)dt + d\overline{W}_t, \quad \xi_0 = 0, \quad 0 \le t \le T, \tag{2.2}$$

where the process $m = (m_t(\xi))$ is F^ξ-adapted, $\overline{W} = (\overline{W}, F^\xi)$ is a Wiener process, and Eq. (2.2) has a unique strong solution.

As m_t we take

$$m_t(\xi) = E(A_t \mid \mathcal{F}_t^\xi),$$

where $A_t = A_t(\xi, \eta(\xi, v))$, $\eta(\xi, v) = (\eta_t(\xi, v))$, $0 \le t \le T$, is a strong solution of the first equation of system (2.1) with given ξ.

The condition of linear growth, as is well known [66], leads to the existence of moments of all orders for ξ and η and, surely, to the existence of m_t and to its square integrability, i.e.,

$$\int_0^T m_t^2(\xi)\, dt < \infty \quad (P\text{-a.s.})$$

(moreover, this condition implies $E\int\limits_0^T m_t^2(\xi)\, dt < \infty$).

Thus, the process $\overline{W} = (\overline{W}, F^{\xi})$ with

$$\overline{W}_t = \xi_t - \int_0^t m_s(\xi)\, ds$$

is a Wiener process and, hence, we have to prove the strong solvability of Eq. (2.2).

Since the weak solution of (2.1) has just been constructed, we prove the strong uniqueness of Eq. (2.2).

We use the generalized Bayes formula (see, e.g., [66]) to obtain an explicit expression for $m_t(\xi)$.

Introduce the following distributions on the measure space $(C_{[0,T]}, \mathcal{B}_{[0,T]})$ of continuous functions: $\forall B \in \mathcal{B}_{[0,T]}$ let $Q_v(B) = P(v \in B)$ and $Q_\eta(B) = \int\limits_{C_{[0,T]}} I_{\{y:\eta(\xi,y)\in B\}} Q_v(dy)$.

Note that the process v is independent of the Wiener process W and, again, by the condition of linear growth, the following Bayes' formula is true (P-a.s.):

$$\begin{aligned} m_t(\xi) &= E\big(A_t(\xi, \eta(\xi, v)) \mid \mathcal{F}_t^{\xi}\big) \\ &= \int_{C_{[0,T]}} A_t(\xi, \eta(\xi, y)) \exp\bigg[\int_0^t A_s(\xi, \eta(\xi, y)) \\ &\quad - m_s(\xi))\, d\overline{W}_s - \frac{1}{2}\int_0^t (A_s(\xi, \eta(\xi, y)) - m_s(\xi))^2 ds\bigg] Q_v(dy). \end{aligned}$$

Change the integration variable and pass to the distribution of the process η with a given ξ, $Q_\eta(\cdot)$. We obtain

$$\begin{aligned} m_t(\xi) &= \int_{C_{[0,T]}} A_t(\xi, y) \exp\bigg[\int_0^t (A_s(\xi, y) - m_s(\xi))\, d\overline{W} \\ &\quad - \frac{1}{2}\int_0^t (A_s(\xi, y) - m_s(\xi))^2 ds\bigg] Q_\eta(dy) \\ &= \int_{C_{[0,T]}} A_t(\xi, y) \exp\bigg[\int_0^t (A_s(\xi, y) - m_s(\xi))\, d\overline{W}_s \\ &\quad - \frac{1}{2}\int_0^t (A_s(\xi, y) - m_s(\xi))^2 ds\bigg] \exp\bigg[\int_0^t a_s(\xi, y)\, dv_s - \frac{1}{2}\int_0^t a_s^2(\xi, y)\, ds\bigg] Q_v(dy). \end{aligned} \tag{2.3}$$

The theorem will be proved if we establish the following: if the processes ξ^1 and ξ^2 defined on the same probability space are the solutions of the equation

$$d\xi_t^i = m_s^i dt + d\overline{W}_t \tag{2.4}$$

such that

$$P\bigg(\int_0^T (m_t^i)^2 dt < \infty\bigg) = 1, \quad i = 1, 2, \tag{2.5}$$

where $m_t^i = m_t(\xi^i)$, then

$$P\big(\sup_{0\le t\le T} |\xi_t^1 - \xi_t^2| = 0\big) = 1.$$

Let ξ^1 and ξ^2 satisfy Eq. (2.4) and condition (2.5) with the initial notation for both the probability space and Wiener process (for convenience).

Condition (2.5) implies that the distributions of ξ^1 and ξ^2 are absolutely continuous w.r.t. a Wiener measure and, moreover, the distributions Q_{ξ_1}, Q_{ξ_2}, and Q_ξ of ξ^1, ξ^2, and ξ, respectively, coincide, i.e., $Q_{\xi_1}(\cdot) = Q_{\xi_2}(\cdot) = Q_\xi(\cdot)$, $\cdot \in \mathcal{B}_{[0,T]}$.

Note that

$$P\big(\sup_{0\le t\le T} |\xi_t^1 - \xi_t^2| = 0\big) \ge P\bigg(\int_0^T (m_t^1 - m_t^2)^2 dt = 0\bigg).$$

We prove that the last probability is equal to 1.

Introduce the following notation and simplification: for $i = 1, 2$, $t \in [0, T]$,

$$A_t^i(y) = A_t(\xi^i, y), \quad m_t^i = m_t(\xi^i), \quad a_t^i(y) = a_t(\xi^i, y),$$

$$g_t^i(y) = \exp\left[\int_0^t (A_s^i(y) - m_s^i)d\overline{W}_s - \frac{1}{2}\int_0^t (A_s^i(y) - m_s^i)^2 ds\right] \exp\left[\int_0^t a_s^i(y)dv_s - \frac{1}{2}\int_0^t (a_s^i(y))^2 ds\right],$$

$$G_t(y) = \frac{1}{2}(g_t^1(y) + g_t^2(y)), \quad Q_v(dy) = Q(dy),$$

$$\int_{C_{[0,T]}} f(y)Q(dy) = \int f(y)Q(dy),$$

where f is an integrable function.

Note that

$$\int G_t(y)Q(dy) = 1,$$

$(Q_{\xi_1} = Q_{\xi_2} = Q_\xi)$.

Further, we denote

$$z_t = \int_0^t (m_s^1 - m_s^2)^2 ds, \quad 0 \le t \le T.$$

The square for any expression f will be written by using the brackets $(f)^2$ (instead of $f^2 = (f(\xi^2))$. The exact values of the constants appearing during the estimation does not matter and it is not necessary to observe their change.

Fix t and consider the difference $m_t^1 - m_t^2$. It can be easily seen that

$$\begin{aligned} m_t^1 - m_t^2 &= \int (A_t^1(y)g_t^1(y) - A_t^2(y)g_t^2(y))\, Q(dy) \\ &= \frac{1}{2}\left[\int (A_t^1(y) - A_t^2(y))G_t(y)Q(dy)\right. \\ &\quad \left. + \int (A_t^1(y) + A_t^2(y))(g_t^1(y) - g_t^2(y))Q(dy)\right]. \end{aligned}$$

Using the Schwartz inequality and the simple inequality

$$|e^x - e^y| \le \frac{e^x + e^y}{2}|x - y|,$$

we obtain the estimate

$$\begin{aligned} (m_t^1 - m_t^2)^2 \le \mathrm{const}\Bigg\{ &\int (A_t^1(y) - A_t^2(y))^2 G_t(y)Q(dy) \\ &+ \int (A_t^1(y) + A_t^2(y))^2 G_t(y)Q(dy)\left[\int \left(\int_0^t A_s^1(y) - A_s^2(y))d\overline{W}_s\right)^2 G_t(y)Q(dy)\right. \\ &+ \int \left(\int_0^t (m_s^1 - m_s^2)d\overline{W}_s\right)^2 G_t(y)Q(dy) + \int \left(\int_0^t (a_s^1(y) - a_s^2(y))dv_s\right)^2 G_t(y)Q(dy) \end{aligned}$$

$$+\int\left(\int_0^t(m_s^1-m_s^2)^2ds\right)\left(\int_0^t(m_s^1+m_s^2)^2ds\right)G_t(y)Q(dy)$$
$$+\int\left(\int_0^t(A_s^1(y)-A_s^2(y))^2ds\right)\left(\int_0^t(m_s^1+m_s^2)^2ds\right)G_t(y)Q(dy)$$
$$+\int\left(\int_0^t(m_s^1-m_s^2)^2ds\right)\left(\int_0^t(A_s^1(y)+A_s^2(y))^2ds\right)G_t(y)Q(dy)$$
$$+\int\left(\int_0^t(a_s^1(y)-a_s^2(y))^2ds\right)\left(\int_0^t(a_s^1(y)-a_s^2(y))^2ds\right)G_t(y)Q(dy)$$
$$+\int\left(\int_0^t(A_s^1(y)-A_s^2(y))^2ds\right)\left(\int_0^t(A_s^1(y)+A_s^2(y))^2ds\right)G_t(y)Q(dy)\Bigg]\Bigg\}.$$

Each integral should be estimated separately. We have:

1. By condition (1) of the theorem, from Eq. (2.2) and definition of m^i, $i=1,2$, we have

$$(A_t^1(y)-A_t^2(y))\le \text{const}\left[(\xi_t^1-\xi_t^2)^2+\int_0^t(\xi_s^1-\xi_s^2)^2ds\right]$$
$$\le \text{const}\int_0^t(m_s^1-m_s^2)^2ds=\text{const}\cdot z_t.$$

Hence

$$\int(A_t^1(y)-A_t^2(y))^2G_t(y)Q(dy)\le\text{const}\cdot z_t. \tag{2.6}$$

2. It can be easily seen that

$$\int(A_t^1(y)+A_t^2(y))^2G_t(y)Q(dy)\le\text{const}\sum_{i=1}^2\int(A_t^i(y))^2g_t^i(y)Q(dy)=\text{const}\sum_{i=1}^2\mu_t^i, \tag{2.7}$$

where

$$\mu_t^i=\int(A_t^i(y))^2g_t^i(y)Q(dy).$$

Now we note that

$$P\left(\int_0^T\mu_t^i dt<\infty\right)=1,\quad i=1,2.$$

Indeed, denote $\eta^i=\eta(\xi^i,v)$ and recall that $Q_{\xi_1}=Q_{\xi_2}=Q_\xi$. We obtain

$$\int(A_t^i(y))^2g_t^i(y)Q(dy)=E\big(A_t^2(\xi^i,\eta^i)\mid\mathcal{F}_t^{\xi^i}\big).$$

But it is well known (see [66]) that $\exists C>0$,

$$E[(\xi_t^i)^2+(\eta_t^i)^2]\le e^{Ct}-1.$$

Thus, by virtue of condition (2) of the theorem,

$$\int_0^T EA_t^2(\xi^i,\eta^i)dt<\infty.$$

The following three terms are denoted by I^1, I^2, and I^3 and studied later on:

3.

$$I_t^1=\int\left(\int_0^t(A_s^1(y)-A_s^2(y))d\overline{W}_s\right)^2G_t(y)Q(dy).$$

4.

$$I_t^2=\int\left(\int_0^t(a_s^1(y)-a_s^2(y))dy_s\right)^2G_t(y)Q(dy).$$

5.
$$I_t^3 = \int \left(\int_0^t (m_s^1 - m_s^2) d\overline{W}_s \right)^2 G_t(y) Q(dy) = \left(\int_0^t (m_s^1 - m_s^2) d\overline{W}_s \right)^2.$$

Applying inequality (2.6) and similar estimates with the functions $a_t^i(y)$, $i = 1, 2$, instead of the functions $A_t^i(y)$, $i = 1, 2$, and denoting $\gamma_t = \sum\limits_{i=1}^{2} \int\limits_0^t (m_s^i)^2 ds$, we obtain

6.
$$\int \left(\int_0^t (m_s^1 - m_s^2)^2 ds \right) \left(\int_0^t (m_s^1 + m_s^2)^2 ds \right) G_t(y) Q(dy) \le \text{const} \cdot \gamma_t z_t.$$

7.
$$\int \left(\int_0^t (A_s^1(y) - A_s^2(y))^2 ds \right) \left(\int_0^t (m_s^1 + m_s^2)^2 ds \right) G_t(y) Q(dy) \le \text{const} \cdot \gamma_t z_t.$$

8.
$$\int \left(\int_0^t (m_s^1 - m_s^2)^2 ds \right) \left(\int_0^t (A_s^1(y) + A_s^2(y))^2 ds \right) G_t(y) Q(dy)$$
$$\le \text{const} \cdot z_t \left(\sum_{i=1}^{2} \int \left(\int_0^t (A_s^i(y))^2 ds \right) g_t^i(y)\, Q(dy) \right) = \text{const} \cdot z_t I_t^4,$$

where
$$I_t^4 = \sum_{i=1}^{2} \int \left(\int_0^t (A_s^i(y))^2 ds \right) g_t^i(y)\, Q(dy).$$

9.
$$\int \left(\int_0^t (a_s^1(y) - a_s^2(y))^2 ds \right) \left(\int_0^t (a_s^1(y) + a_s^2(y))^2 ds \right) G_t(y) Q(dy) \le \text{const} \cdot z_t I_t^5,$$

where
$$I_t^5 = \sum_{i=1}^{2} \int \left(\int_0^t (a_s^i(y))^2 ds \right) g_t^i(y)\, Q(dy).$$

10.
$$\int \left(\int_0^t (A_s^1(y) - A_s^2(y))^2 ds \right) \left(\int_0^t (A_s^1(y) + A_s^2(y))^2 ds \right) G_t(y) Q(dy) \le \text{const} \cdot z_t I_t^4.$$

Combining the estimates obtained in 1–10, denoting
$$\delta_t = \max\left(\gamma_t, \sum_{i=1}^{2} \mu_t^i \right),$$
and taking, for simplicity, const $\equiv 1$, we obtain
$$(m_t^1 - m_t^2)^2 \le z_t (1 + \delta_t + I_t^4 + I_t^5) + \delta_t (I_t^1 + I_t^2 + I_t^3).$$

Our next aim is to estimate the integrals $I^1, ..., I^5$. By application of the Itô formula to the integrands of $I^1, ..., I^5$, we obtain linear stochastic inequalities whose solution gives the desired estimates.

We perform this procedure taking as an example the integral I^1. Introduce the product stochastic basis
$$(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{F} = (\widetilde{\mathcal{F}}_t), 0 \le t \le T, \widetilde{P})$$
$$= (\Omega, \mathcal{F}, F = (\mathcal{F}_t), 0 \le t \le T, P) \times (C_{[0,T]}, \mathcal{B}_{[0,T]}, B = (\mathcal{B}_{[0,t]}), 0 \le t \le T, Q),$$
where Q is a Wiener measure, and let
$$\overline{W}_t = \overline{W}_t(\omega, y) = \overline{W}_t(\omega), \quad \xi_t^i = \xi_t^i(\omega, y) = \xi_t^i(\omega), \quad i = 1, 2,$$
$$v_t = v_t(\omega, y) = v_t(y) = y_t.$$

Then a Wiener process v is independent of processes $\overline{W}$, ξ^1, and ξ^2. Moreover,

$$I_t^1 = I_t^1(\xi^1, \xi^2) = \widetilde{E}(i_t^1(\xi^1, \xi^2, v) \mid \mathcal{F}_t^{\xi^1, \xi^2})$$

with

$$i_t^1(\xi^1, \xi^2, v) = i_t^1(y),$$

where the process $i_t^1(y) = i_t^1(\omega, y)$ defined on the stochastic basis $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{F}, \widetilde{P})$ is given by the formula

$$i_t^1(y) = \left(\int_0^t (A_s^1(y) - A_s^2(y))\, d\overline{W}_s\right)^2 G_t(y),$$

or, in more detail,

$$i_t^1(\omega, y) = \left(\int_0^t (A_s^1(\omega, y) - A_s^2(\omega, y))\, d\overline{W}_s\right)^2 G_t(\omega, y),$$

where

$$G_t(\omega, y) = \frac{1}{2}(g_t^1(\omega, y) + g_t^2(\omega, y)),$$

$$g_t^i(\omega, y) = \mathcal{E}_t\left(\int_0^{\cdot} (A_s^i(\omega, y) - m_s^i)\, d\overline{W}_s\right) \mathcal{E}_t\left(\int_0^{\cdot} a_s^i(\omega, y)\, dv_s\right),$$

where $i = 1, 2$, and $\mathcal{E}_t(M)$ is the Dolean exponential of the martingale M (see [67]).

Applying the Itô formula to each summand (recall that $G = \frac{1}{2}(g^1 + g^2)$), after addition and some simple calculations we obtain

$$\begin{aligned}
i_t^1(y) = \frac{1}{2} \int_0^t \sum_{i=1}^2 \Bigg[&\left(\int_0^s (A_u^1(y) - A_u^2(y))\, d\overline{W}_u\right)^2 (A_s^i(y) - m_s^i) g_s^i(y)\, d\overline{W}_s \\
&+ 2\left(\int_0^s (A_u^1(y) - A_u^2(y))\, d\overline{W}_u\right) (A_s^1(y) - A_s^2(y)) g_s^i(y)\, d\overline{W}_s \\
&+ \left(\int_0^t (A_u^1(y) - A_u^2(y))\, d\overline{W}_u\right)^2 a_s^i(y) g_s^i(y)\, dv_s \Bigg] + \int_0^t (A_s^1(y) - A_s^2(y))^2 G_s(y)\, ds \\
&+ \int_0^t \left(\int_0^s (A_u^1(y) - A_u^2(y))\, d\overline{W}_u\right) (A_s^1(y) - A_s^2(y)) \left(\sum_{i=1}^2 (A_s^i(y) - m_s^i) g_s^i(y)\right) ds.
\end{aligned} \tag{2.8}$$

Estimate the integrand of the last summand (using the simple inequality $ab \leq \frac{1}{2}(a^2 + b^2)$), we easily obtain that this summand is less than or equal to

$$\begin{aligned}
&\int_0^t \left(\int_0^s (A_u^1(y) - A_u^2(y))\, d\overline{W}_u\right)^2 G_s(y)\, ds + \text{const} \int_0^t (A_s^1(y) - A_s^2(y))^2 G_s(y) l_1(s, y)\, ds \\
&\qquad \leq \int_0^t i_s^1(y)\, ds + \text{const} \int_0^t \left(\int_0^s (m_u^1 - m_u^2)^2 du\right) l_1(s, y)\, ds,
\end{aligned} \tag{2.9}$$

where $l_1(s, y) = \sum_{i=1}^2 ((A_s^i(y))^2 + (m_s^i)^2)$.

Hence, from (2.8) we obtain a linear stochastic inequality for i_t^1 of the form

$$\begin{aligned}
i_t^1(y) \leq &\int_0^t i_s^1(y)\, ds + \text{const} \int_0^t \left(\int_0^s (m_u^1 - m_u^2)^2 du\right) G_s(y) l_1(s, y)\, ds \\
&+ \int_0^t K_1(s, y)\, d\overline{W}_s + \int_0^t p_1(s, y)\, dv_s, \quad 0 \leq t \leq T.
\end{aligned} \tag{2.10}$$

Solving the last inequality, we obtain the estimate

$$i_t^1(y) \le \int_0^t \overline{K}_1(s,y)\,d\overline{W}_s + \int_0^t \overline{p}_1(s,y)\,dv_s + \int_0^t \left(\int_0^s (m_u^1 - m_u^2)^2 du\right)\overline{l}_1(s,y)\,ds, \tag{2.11}$$

where $\overline{K}_1$, $\overline{p}_1$, and $\overline{l}_1$ ($\overline{l}_1 \ge 0$) are random processes with

$$\begin{gathered} P\left(\int_0^T (\overline{K}_1(s,y))^2 ds < \infty\right) = P\left(\int_0^T (\overline{p}_1(s,y))^2 ds < \infty\right) = 1 \ (Q\text{-a.s.}), \\ P\left(\int_0^T \left(\int \overline{l}_1(s,y)\,Q(dy)\right) ds < \infty\right) = 1. \end{gathered} \tag{2.12}$$

Quite similarly we estimate i_t^2 in the expression

$$I_t^2 = \int i_t^2(y)\,Q(dy).$$

A direct application of the Itô formula to I^3 yields

$$I_t^3 = \int_0^t (m_s^1 - m_s^2)^2 ds + \int_0^t K_3(s)\,d\overline{W}_s$$

with

$$P\left(\int_0^T (K_3(s))^2 ds < \infty\right) = 1.$$

By a simple application of the Itô formula, we obtain for the integrands i_t^4 and i_t^5 of integrals I_t^4 and I_t^5, respectively:

$$i_t^i(y) = \int_0^t \overline{K}_i(s,y)\,d\overline{W}_s + \int_0^t \overline{p}_i(s,y)\,dv_s + \int_0^t \overline{l}_i(s,y)ds, \quad i = 4,5,$$

where

$$P\left(\int_0^T (\overline{K}_i(s,y))^2 ds < \infty\right) = P\left(\int_0^T (\overline{p}_i(s,y))^2 ds < \infty\right) = 1 \quad (Q\text{-a.s.})$$

and $\overline{l}_i \ge 0$, $i = 4,5$, with

$$P\left(\int_0^T \left(\int \overline{l}_i(s,y)\,Q(dy)\right) ds < \infty\right) = 1.$$

For the further consideration we need the following lemma.

Lemma 2.2. *Let $X(t,y) = X(t,\omega,y)$ be an Itô process (defined on the above-mentioned product space) with*

$$0 \le X(t,y) = \int_0^t n(s,y)\,d\overline{W}_s + \int_0^t m(s,y)\,dv_s + \int_0^t u(s,y)\,ds,$$

where $\overline{W}_t = \overline{W}_t(\omega)$ and $v_t = v_t(y)$ are Wiener processes and $u(t,y) \ge 0$ (recall that $n(s,y) = n(s,\omega,y)$, $m(s,y) = m(s,\omega,y)$, and $u(s,y) = u(s,\omega,y)$).

Assume that

$$P\left(\int_0^T \left(\int u(s,y)\,Q(dy)\right) ds < \infty\right) = 1$$

and

$$P\left(\int_0^T (n(s,y))^2 ds < \infty\right) = P\left(\int_0^T (m(s,y))^2 ds < \infty\right) = 1 \quad (Q\text{-a.s.}).$$

Denote

$$\zeta(t,y)=\int_0^t n(s,y)\,d\overline{W}_s+\int_0^t m(s,y)\,dv_s,$$
$$\eta(t,y)=\int_0^t u(s,y)\,ds.$$

Then if $\tau=\tau(\omega,y)=\tau(\omega)$ is an $(\mathcal{F}_t)$-stopping time (and, therefore, $(\widetilde{\mathcal{F}}_t)$-stopping time) such that $\tau\le T$ and

$$E\int\eta(\tau,y)\,Q(dy)<\infty,$$

then the process

$$\zeta(t\wedge\tau)=\int\zeta(t\wedge\tau,y)\,Q(dy)$$

is an $(\mathcal{F}_t)$-supermartingale with

$$\sup_{0\le t\le T}E|\zeta(t\wedge\tau)|<\infty,$$

and, in particular,

$$P\bigg(\sup_{0\le t\le T}\int X(t,y)\,Q(dy)<\infty\bigg)=1.$$

Proof. Consider the $(\widetilde{\mathcal{F}}_t)$-stopping time

$$\tau_c=\tau_c(y)=\inf\{t>0:\zeta(t,y)\ge c\}\wedge T,\quad c>0.$$

Note that for each y, τ_c is an $(\mathcal{F}_t)$-stopping time.

It is obvious that for each y (Q-a.s.), $\zeta(t\wedge\tau\wedge\tau_c,y)$ is a uniformly integrable $\mathcal{F}_t$-martingale with $\zeta^-(t\wedge\tau,y)$, $0\le t\le T$, where $\zeta^-=\max(0,-\zeta)$.

The conditions of the lemma result in

$$\sup_{0\le t\le T}E\zeta^-(t\wedge\tau)\le\int E\sup_{0\le t\le T}\zeta^-(t\wedge\tau,y)\,Q(dy)<\infty.$$

By the Fatou lemma, for each y (Q-a.s.),

$$E\zeta(t\wedge\tau,y)\le\varliminf_{c\to\infty}E\zeta(t\wedge\tau\wedge\tau_c,y)=0.$$

This leads to the relation

$$E\zeta^+(t\wedge\tau)=\int(E\zeta^+(t\wedge\tau,y))\,Q(dy)$$
$$=\int E\zeta(t\wedge\tau,y)\,Q(dy)+E\zeta^-(t\wedge\tau)\le\sup_{0\le t\le T}E\zeta^-(t\wedge\tau).$$

Hence,

$$\sup_{0\le t\le T}E|\zeta(t\wedge\tau)|<\infty.\tag{2.13}$$

Further, let $s\le t$. Then, again, by the Fatou lemma, we obtain

$$E(\zeta(t\wedge\tau,y)\mid\mathcal{F}_s)\le\varliminf_{c\to\infty}E(\zeta(t\wedge\tau\wedge\tau_c,y)\mid\mathcal{F}_s)=\varliminf_{c\to\infty}\zeta(s\wedge\tau\wedge\tau_c,y)=\zeta(s\wedge\tau,y)$$

for each y (Q-a.s.) and by an average with respect to the measure $Q(dy)$ we obtain

$$E\bigg(\int\zeta(t\wedge\tau,y)\,Q(dy)\mid\mathcal{F}_s\bigg)\le\int\zeta(s\wedge\tau,y)\,Q(dy),\quad s\le t.$$

Hence, $\zeta(t\wedge\tau)$ is an $(\mathcal{F}_t)$-supermartingale.

Now we consider the $(\mathcal{F}_t)$-stopping time

$$\sigma_c = \inf\Big\{t > 0 : \int \eta(t,y)\,Q(dy) \geq c\Big\} \wedge T, \quad c > 0.$$

Obviously,

$$\lim_{c\to\infty} P(\sigma_c = T) = 1 \tag{2.14}$$

and from the martingale inequalities (see [66]) it follows that

$$P\Big(\sup_{0\leq t\leq \sigma_c} \zeta(t) < \infty\Big) = 0. \tag{2.15}$$

To prove the last assertion of the lemma, it is enough to show that

$$P\Big(\sup_{0\leq t\leq T} \zeta(t) < \infty\Big) = 1.$$

For all $a > 0$ we have

$$P\Big(\sup_{0\leq t\leq T} \zeta(t) > a\Big) \leq P\Big(\sup_{0\leq t\leq \sigma_c} \zeta(t) > a, \sigma_c = T\Big)$$
$$+P\Big(\sup_{0\leq t\leq T} \zeta(t) > a, \sigma_c < T\Big) \leq P\Big(\sup_{0\leq t\leq \sigma_c} \zeta(t) > a\Big) + P(\sigma_c < T).$$

Passing to the limit first as $a \to \infty$ and then as $c \to \infty$, we obtain from inequalities (2.14) and (2.13) the desirable result. $\square$

Applying this lemma to I^4 and I^5, we obtain

$$P\Big(\sup_{0\leq t\leq T} I_t^i < \infty\Big) = 1, \quad i = 4, 5.$$

Put

$$\Gamma_t = \max\bigg(1 + \delta_t + I_t^4 + I_t^5; \int \bar{l}_i(t,y)\,Q(dy), \quad i = 1, 2; \quad (K_t^3)^2\bigg).$$

We have

$$P\bigg(\int_0^T \Gamma_t dt < \infty\bigg) = 1.$$

Finally, combining the obtained inequalities, we obtain

$$z_t \leq \int_0^t \Gamma_s z_s ds + \int_0^t \Gamma_s (I_s^1 + I_s^2 + I_s^3)\,ds,$$
$$I_t^1 \leq \int_0^t \Gamma_s z_s ds + \zeta_1(t), \quad I_t^2 \leq \int_0^t \Gamma_s z_s ds + \zeta_2(t),$$
$$I_t^3 \leq z_t + \zeta_3(t) \leq \int_0^t \Gamma_s z_s ds + \int_0^t \Gamma_s (I_s^1 + I_s^2 + I_s^3)\,ds + \zeta_3(t),$$

where ζ_i, $i = 1, 2, 3$, are local martingales.

To complete the proof, it is sufficient to refer to Lemma 2.1. Indeed, if for each t we set

$$X_t = (z_t, I_t^1, I_t^2, I_t^3), \quad M_t = (0, \zeta_1(t), \zeta_2(t), \zeta_3(t)), \quad dK_t = \alpha\Gamma_t dt,$$

where α is the matrix

$$\alpha = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix},$$

then we obtain the desirable result. $\square$

2.3. Coincidence of σ-Algebras in a Filtering Problem of a Multidimensional Partially Observable Diffusion-Type Process

Let $\mathbb{R}_l$, $l = m, n, k$, be Euclidean spaces with a fixed orthonormal basis and the usual Euclidean norm $|\cdot|$.

Fix $T > 0$.

Let $(C^l_{[0,T]}, \mathcal{B}^l_{[0,T]})$, $l = m, n, k$, be measure spaces of continuous l-dimensional functions with the uniform metric, and let $a(t,x,y)$ and $A(t,x,y)$ be, respectively, m- and n-dimensional vectors, $\sigma_1(t,x,y)$ and $\sigma_2(t,x,y)$ be, respectively, $(m \times k)$- and $(n \times k)$-matrices defined for $t \in [0,T]$, $x \in C^m_{[0,T]}$, $y \in C^n_{[0,T]}$ (nonrandom).

Let a k-dimensional Wiener process $W = (W, F)$ be given on a stochastic basis $(\Omega, \mathcal{F}, F = (\mathcal{F}_t), 0 \le t \le T, P)$.

Consider the following system of SDEs:

$$\begin{aligned} d\eta(t) &= a(t,\eta,\xi)dt + \sigma_1(t,\eta,\xi)dW_t, \quad \eta(0) = 0, \\ d\xi(t) &= A(t,\eta,\xi)dt + \sigma_2(t,\xi)dW_t, \qquad \xi(0) = 0. \end{aligned} \tag{2.16}$$

Denote by $g(t,x,y)$ each of the coefficients a, A, σ_1, and σ_2 of (2.16). Assume that $g(t,x,y)$ is a non-anticipating functional and

(1) $|g(t,x,y)| \le \text{const}$ for all t, x, y; const > 0;

(2) the functional $g(t,x,y)$ satisfies the Lipschitz condition with respect to the pair (x,y):

$$|g(t,x_1,y_1) - g(t,x_2,y_2)|^2 \le \text{const}\bigg(|x_1(t) - x_2(t)| + |y_1(t) - y_2(t)| + \int_0^t \left(|x_1(s) - x_2(s)|^2 + |y_1(s) - y_2(s)|^2\right) dK_s\bigg),$$

where $K(t) \ge 0$, $K(0) = 0$ is an increasing right-continuous nonrandom function, and $|\cdot|$ is a norm on a suitable Euclidean space.

We set

$$\sigma(t,y) = \sigma_2(t,y)\sigma_2^*(t,y),$$

where $*$ denotes transposition;

(3) assume that there exists a constant $\lambda > 0$ such that for all $t \in [0,T]$, $y \in C^n_{[0,T]}$, and $u = (u_1, ..., u_n) \in \mathbb{R}_n$,

$$\sum_{i,j=1}^n \sigma_{ij}(t,y)u_i u_j \ge \lambda |u|^2.$$

If conditions (1) and (2) hold, then as is well known, there exists a pathwise unique strong solution of Eq. (2.16).

Further, if $\sigma^{1/2}$ is a positive symmetric square root of the matrix σ, then, thanks to condition (3), there exists the matrix $\sigma^{-1/2}$, which is a bounded function of $(t,y) \in [0,T] \times C^n_{[0,T]}$.

Denote by

$$F^\xi = (\mathcal{F}^\xi_t), \quad 0 \le t \le T,$$

the P-augmented filtration generated by the process ξ and consider the process

$$\overline{W}(t) = \int_0^t \sigma^{-1/2}(s,\xi)\, d\xi(s) - \int_0^t \sigma^{-1/2}(s,\xi) E(A(s,\eta,\xi) \mid \mathcal{F}^\xi_s)\, ds, \tag{2.17}$$

where $E(\cdot \mid \mathcal{F}^\xi_s) = E(\cdot \mid \mathcal{F}^\xi_s)(\omega, s)$ is a (ω, s)-measurable modification of a conditional expectation.

Further, we denote the P-augmented filtration generated by the process $\overline{W}$ by

$$F^{\overline{W}} = (\mathcal{F}^{\overline{W}}_t), \quad 0 \le t \le T.$$

Theorem 2.2. *In scheme* (2.16), *under conditions* (1), (2), *and* (3), *the process* $\overline{W} = (\overline{W}, F^{\xi})$, *defined by Eq.* (2.17), *is an innovation process, i.e., a Wiener process with*

$$F^{\overline{W}} = F^{\xi} \quad (\text{mod } P).$$

Proof. Let us show that system (2.16) can be reduced to the following triangular system:

$$\begin{aligned} d\eta(t) &= a(t,\eta,\xi)dt + c(t,\eta,\xi)d\widetilde{W}(t) + b(t,\eta,\xi)dv(t), \quad \eta(0) = 0, \\ d\xi(t) &= A(t,\eta,\xi)dt + \widetilde{\sigma}(t,\xi)d\widetilde{W}(t), \quad \xi(0) = 0, \end{aligned} \tag{2.18}$$

where $\widetilde{\sigma} = \sigma^{1/2}$, b and c are some matrices satisfying conditions (1) and (2), and v and $\widetilde{W}$ are independent Wiener processes (the dimensions of the just-described objects are given in the proof below).

Indeed, let

$$d\widetilde{W} = \widetilde{\sigma}^{-1}\sigma_2 dW$$

(for simplicity, here and below, the arguments of the functions and processes will be omitted). Obviously, $\widetilde{W}$ is an n-dimensional Wiener process with

$$\widetilde{\sigma} d\widetilde{W} = \sigma_2 dW.$$

Denote

$$L = I - \sigma_2^*(\sigma^{-1})\sigma_2, \quad c = \sigma_1\sigma_2^*\sigma^{-1/2},$$

where I is the identity matrix, and consider the process x with

$$dx = LdW.$$

It is easy to see that the processes x and $\widetilde{W}$ are independent and

$$\sigma_1 dW = cd\widetilde{W} + \sigma_1 LdW.$$

It is well known that the matrix L can be represented in the form

$$L = MBD,$$

where the matrix B is nondegenerate and matrices M, B, and D have dimensions $m \times p$, $p \times p$, and $p \times k$, respectively, where $p \le k - n$ (obviously, it is always possible to take $k - n > 0$).

Further, the matrix

$$E = BDD^*B^*$$

is nondegenerate. Hence

$$LdW = ME^{1/2}(E^{-1/2}BDdW).$$

Now, if we take

$$dv = E^{-1/2}BDdW,$$

then v is a Wiener process independent of $\widetilde{W}$, and we obtain the desirable system (2.18).

In scheme (2.18), without loss of generality (due to condition (3)), we can set

$$\widetilde{\sigma}(t,y) = I,$$

where I is the identity matrix.

Below we omit the sign "$\sim$" over the process $\widetilde{W}$ and simply write W.

The process $\overline{W}$ with

$$d\overline{W}(t) = d\xi(t) - m(t,\xi)\,dt,$$

where

$$m(t,\xi) = E(A(t,\eta,\xi) \mid \mathcal{F}_t^{\xi}),$$

is, as is well known, a Wiener process.

Thus, from the second equation of system (2.18) we derive

$$dW(t) = d\xi(t) - A(t,\eta,\xi)dt = d\overline{W}(t) + [m(t,\xi) - A(t,\eta,\xi)]dt.$$

Hence,

$$d\eta(t) = a(t,\eta,\xi)dt + b(t,\eta,\xi)dv(t) + c(t,\eta,\xi)d\overline{W}(t) + c(t,\eta,\xi)[m(t,\xi) - A(t,\eta,\xi)]dt.$$

It follows from conditions (1) and (2) that the last equation has a strong solution, i.e.,

$$\eta(t) = F(t,v,\xi),$$

where $F(t,x,y)$ is a nonanticipating functional.

Therefore, we obtain

$$d\xi(t) = \phi(t,v,\xi)dt + dW_t \quad (= m(t,\xi)dt + d\overline{W}(t)), \tag{2.19}$$

where

$$\phi(t,x,y) = A(t,F(t,x,y),y)$$

is a superposition of nonanticipating functionals and hence itself is nonanticipating.

Obviously, $\mathcal{F}^{\overline{W}} \subset \mathcal{F}^{\xi}$ (mod P), by the construction of the process $\overline{W}$. Thus, it remains to prove that $\mathcal{F}^{\xi} \subset \mathcal{F}^{\overline{W}}$ (mod P). To this end, it suffices to show the pathwise uniqueness of the solution of Eq. (2.19). The proof then follows from the Yamada–Watanabe theorem.

Denote by $\mu(dx)$ a distribution of the Wiener process v on a measure space $(C^k_{[0,T]}, \mathcal{B}^k_{[0,T]})$. Recall that the processes v and W are independent. Hence, from the Bayes formula we have

$$m(t,\xi) = E(\phi(t,v,\xi) \mid \mathcal{F}^{\xi}_t) = \int_{C^k_{[0,T]}} \phi(t,x,\xi)\rho(t,x,\xi)\,\mu(dx), \tag{2.20}$$

where

$$\rho(t,x,\xi) = \exp(f(t,x,\xi)),$$

$$f(t,x,\xi) := \sum_{p=1}^{n} \int_0^t (\phi_p(s,x,\xi) - m_p(s,\xi))\,d\overline{W}_p(s) - \frac{1}{2}\int_0^t |\phi(s,x,\xi) - m(s,\xi)|^2 ds. \tag{2.21}$$

Introduce the probability space

$$(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P}) = (\Omega \times C^k_{[0,T]}, \mathcal{F} \times \mathcal{B}^k_{[0,T]}, P \times \mu),$$

$\widetilde{\omega} = (\omega, x)$, and let $\widetilde{\xi}(\widetilde{\omega}) = \xi(\omega)$, $\widetilde{v}(\widetilde{\omega}) = x$. Then $\widetilde{v}$ is a Wiener process independent of the process ξ.

Denoting expectations with respect to measures $\widetilde{P}$ and μ by $\widetilde{E}$ and E^{μ}, respectively, we obtain from (2.20)

$$\begin{aligned} m(t,\widetilde{\xi}) &= \int_{C^k_{[0,T]}} \phi(t,x,\widetilde{\xi})\rho(t,x,\widetilde{\xi})\,\mu(dx) \\ &:= E^{\mu}\big(\phi(t,\widetilde{v},\widetilde{\xi})\rho(t,\widetilde{v},\widetilde{\xi})\big) = \widetilde{E}\big(\phi(t,\widetilde{v},\widetilde{\xi})\rho(t,\widetilde{v},\widetilde{\xi}) \mid \mathcal{F}^{\widetilde{\xi}}_t\big) \quad (P\text{-a.s.}), \end{aligned} \tag{2.22}$$

where $\widetilde{v}$ and $\widetilde{\xi}$ are independent.

For simplicity, introduce the notation:

(a) in Eq. (2.22) we omit the sign "$\sim$";

(b) each constant is denoted by const;

(c) in condition (2) we set $dK(t) = dt$;

(d)

$$G(t) = \frac{1}{2}(\rho^1(t) + \rho^2(t)), \quad \rho^i(t) = \rho(t,v,\xi^i),$$
$$m^i(\xi) = m(t,\xi^i), \quad \phi^i(t) = \phi(t,v,\xi^i), \quad f^i(t) = f(t,v,\xi^i),$$

where ξ^i $(i = 1,2)$ are two solutions of Eq. (2.19) defined on the space introduced above.

It obviously follows from condition (1) of the theorem that the distributions of ξ^1 and ξ^2 coincide.

In particular,

$$E^{\mu}G(t) = 1, \quad 0 \le t \le T.$$

According to (2.22), for each $i = 1, \dots, n$ we have

$$m_i^1(t) - m_i^2(t) = E^{\mu}(\phi_i^1(t) - \phi_i^2(t))G(t) + E^{\mu}(\rho^1(t) - \rho^2(t))\frac{1}{2}(\phi_i^1(t) + \phi_i^2(t)) \tag{2.23}$$

for each t.

Hence, if we use the inequality $|e^x - e^y| \le \frac{1}{2}(e^x + e^y)|x - y|$, then we obtain

$$|\rho^1(t) - \rho^2(t)| \le G(t)|f^1(t) - f^2(t)|.$$

Note that, according to condition (1),

$$|\phi^i(t)| \le \text{const}, \quad i = 1, 2, \quad \text{for all } t.$$

Thus, from Eq. (2.23) we obtain

$$\begin{aligned} |m_i^1(t) - m_i^2(t)| &\le \text{const}\big[E^{\mu}\big(|\phi_i^1(t) - \phi_i^2(t)|\sqrt{G(t)}\sqrt{G(t)}\big) \\ &+ E^{\mu}\big(|f^1(t) - f^2(t)|\sqrt{G(t)}\sqrt{G(t)}\big)\big], \quad i = 1, 2, \dots, n. \end{aligned}$$

Squaring each part of the last inequalities, applying the elementary relation $(a + b)^2 \le 2a^2 + 2b^2$ and the Schwartz inequality, and then averaging and summing up, we obtain

$$E|m^1(t) - m^2(t)|^2 \le \text{const}[E|\phi^1(t) - \phi^2(t)|^2 G(t) + E(f^1(t) - f^2(t))^2 G(t)]. \tag{2.24}$$

Denote

$$z(t) = E\int_0^t |m^1(s) - m^2(s)|^2 ds \tag{2.25}$$

and show that each summand in inequality (2.24) is less than or equal to const $\cdot z(t)$.

Then we obtain from (2.24)

$$z(t) \le \text{const}\int_0^t z(s)\, ds, \quad 0 \le t \le T,$$

whence, according to the Gronwall–Bellman lemma, we have

$$P(z(T) = 0) = 1.$$

Now the assertion of the theorem follows from the inequality

$$P\Big(\sup_{0 \le t \le T} |\xi^1(t) - \xi^2(t)| = 0\Big) \ge P(z(T) = 0).$$

Thus, we have to prove that for each t,

$$\begin{aligned} E|\phi^1(t) - \phi^2(t)|^2 G(t) &\le \text{const} \cdot z(t), \\ E(f^1(t) - f^2(t))^2 G(t) &\le \text{const} \cdot z(t). \end{aligned} \tag{2.26}$$

The proofs of each of the above inequalities are quite similar (simple application of the Lipschitz condition and the Itô formula easily shows that $|\phi^1(t) - \phi^2(t)|^2$ and $(f^1 - f^2)^2$ are estimated from above by the same expressions).

Let us prove, for example, the first inequality.

Introduce the following notation: for each $t \in [0,T]$ we have

$$
\begin{aligned}
x(t) &= |\eta^1(t) - \eta^2(t)|^2 G(t),\\
h(t) &= |\xi^1(t) - \xi^2(t)|^2 G(t),\\
d(t) &= G(t)\int_0^t |m^1(s) - m^2(s)|^2 ds,\\
p(t) &= G(t)\int_0^t |\eta^1(s) - \eta^2(s)|^2 ds,\\
u(t) &= G(t)\int_0^t |\xi^1(s) - \xi^2(s)|^2 ds,\\
M_1(t) &= \sum_{i=1}^m \sum_{j=1}^k \left(\int_0^t (b_{ij}(s,\eta^1,\xi^1) - b_{ij}(s,\eta^2,\xi^2))\, dv_j(s)\right)^2,\\
M_2(t) &= \sum_{i=1}^m \sum_{j=1}^n \left(\int_0^t (c_{ij}(s,\eta^1,\xi^1) - c_{ij}(s,\eta^2,\xi^2))\, d\overline{W}_j(s)\right)^2,\\
l(t) &= M_1(t)G(t),\\
y(t) &= M_2(t)G(t).
\end{aligned}
$$

It is easy to see that

$$
E(|\phi^1(t) - \phi^2(t)|^2 G(t)) \le \text{const} \cdot E(h(t) + u(t) + x(t) + p(t)). \tag{2.27}
$$

Let us estimate each term on the right-hand side of inequality (2.27).

We have

$$
\begin{aligned}
Eh(t) = EE^\mu h(t) &= E(|\xi^1(t) - \xi^2(t)|^2 E^\mu G(t))\\
&= E|\xi^1(t) - \xi^2(t)|^2 \le \text{const} \cdot z(t).
\end{aligned} \tag{2.28}
$$

In a complete analogy, we write

$$
Eu(t) \le \text{const} \cdot z(t). \tag{2.29}
$$

Further, it is easy to see that

$$
\begin{aligned}
|\eta^1(t) - \eta^2(t)|^2 \le & \int_0^t |a(s,\eta^1,\xi^1) - a(s,\eta^2,\xi^2)|^2 ds\\
&+ \int_0^t |c(s,\eta^1,\xi^1)m^1(s) - c(s,\eta^2,\xi^2)m^2(s)|^2 ds\\
&+ \int_0^t |c(s,\eta^1,\xi^1)A(s,\eta^1,\xi^1) - c(s,\eta^2,\xi^2)A(s,\eta^2,\xi^2)|^2 ds\\
&+ M_1(t) + M_2(t).
\end{aligned} \tag{2.30}
$$

According to conditions (1) and (2) of the theorem, we obtain from (2.30)

$$
x(t) \le \text{const}(p(t) + l(t) + y(t) + d(t) + u(t)). \tag{2.31}
$$

Note that in our calculations below there arise stochastic integrals with respect to the Wiener processes v and $\overline{W}$ possessing cumbersome integrands. But according to condition (1) of the theorem, each of them is a martingale equal to zero at the point $t = 0$ (indeed, it is sufficient to note that each coefficient is bounded and $E(G(t))^2 \le \text{const}$). Denote these martingales by the symbol “mart.”

Using the Itô formula and condition (2), we obtain the following:

(e)

$$p(t) = \text{mart} + \int_0^t x(s)\,ds;$$

(f)

$$\begin{aligned} l(t) &= \text{mart} + \sum_{i=1}^m \sum_{j=1}^n \int_0^t (b_{ij}(s,\eta^1,\xi^1) - b_{ij}(s,\eta^2,\xi^2))^2 G(s)\,ds \\ &\le \text{mart} + \text{const}\left(\int_0^t x(s)\,ds + \int_0^t h(s)\,ds\right); \end{aligned}$$

(g)

$$\begin{aligned} y(t) &= \text{mart} + \sum_{i=1}^m \sum_{j=1}^n \int_0^t (c_{ij}(s,\eta^1,\xi^1) - c_{ij}(s,\eta^2,\xi^2))^2 G(s)\,ds \\ &+ \sum_{i=1}^m \sum_{j=1}^n \int_0^t \left(\int_0^s (c_{ij}(u,\eta^1,\xi^1) - c_{ij}(u,\eta^2,\xi^2))\,d\overline{W}_j(u)\right)(c_{ij}(s,\eta^1,\xi^1) - c_{ij}(s,\eta^2,\xi^2)) \\ &\times \left(\sum_{i=1}^n [(\phi_i^1(s) - m_i^1(s))\rho^1(s) + (\phi_i^2(s) - m_i^2(s))\rho^2(s)]\right) ds. \end{aligned} \tag{2.32}$$

Now we note that just as in (j),

$$\sum_{i=1}^m \sum_{j=1}^n \int_0^t (c_{ij}(s,\eta^1,\xi^1) - c_{ij}(s,\eta^2,\xi^2))^2 G(s) ds \le \text{const}\left(\int_0^t x(s) ds + \int_0^t h(s) ds\right).$$

Recall that by condition (1) of the theorem,

$$|\phi^i| \le \text{const}, \quad |m^i| \le \text{const}, \quad i = 1, 2.$$

Thus if we apply the elementary inequality

$$ab \le \frac{1}{2}(a^2 + b^2)$$

and condition (2) of the theorem to the third term on the right-hand side of equality (2.32), we can easily see that it is overestimated by the expression

$$\text{const}\left(\int_0^t x(s)\,ds + \int_0^t h(s)\,ds + \int_0^t y(s)\,ds\right).$$

Therefore, we obtain

$$y(t) \le \text{mart} + \text{const}\left(\int_0^t x(s)\,ds + \int_0^t h(s)\,ds + \int_0^t y(s)\,ds\right). \tag{2.33}$$

Now, from inequality (2.31), according to (i), (j), and (k), we find that

$$\begin{aligned} x(t) &\le \text{mart} + \text{const}\left(\int_0^t x(s)\,ds + \int_0^t h(s)\,ds + y(t) + d(t) + u(t)\right) \\ &\le \text{mart} + \text{const}\left(\int_0^t x(s)\,ds + \int_0^t h(s)\,ds + \int_0^t y(s)\,ds + u(t) + d(t)\right). \end{aligned} \tag{2.34}$$

From (2.28) we obtain

$$E\int_0^t h(s)\,ds = \int_0^t Eh(s)\,ds \le \text{const}\int_0^t z(s)\,ds.$$

But $z(t)$ is a positive increasing function. Hence

$$\int_0^t z(s)\,ds \le \int_0^t z(t)\,ds \le \text{const} \cdot z(t)$$

and, therefore,

$$E\int_0^t h(s)\,ds \le \text{const} \cdot z(t). \tag{2.35}$$

Further,

$$Ed(t) = EE^{\mu}d(t) = E\int_0^t |m^1(s) - m^2(s)|^2 ds \cdot E^{\mu}G(t) = z(t). \tag{2.36}$$

Now, averaging inequalities (2.33) and (2.34) and then adding them, we obtain

$$E(x(t) + y(t)) \le \text{const}\bigg(\int_0^t E(x(s) + y(s))ds + E\int_0^t h(s)ds + Eu(t) + Ed(t)\bigg). \tag{2.37}$$

Hence, by (2.36), (2.29), and (2.37), we have

$$E(x(t) + y(t)) \le \text{const}\bigg(\int_0^t E(x(s) + y(s))\,ds + z(t)\bigg).$$

Solving this inequality and taking into account that the process $y(t) \ge 0$, we obtain

$$Ex(t) \le \text{const} \cdot z(t). \tag{2.38}$$

Item (i) and inequality (2.35) yield

$$Ep(t) = E\int_0^t x(s)\,ds \le \text{const} \cdot z(t). \tag{2.39}$$

Finally, the desired inequality (2.26) follows from inequality (2.27), according to relations (2.28), (2.29), (2.38), and (2.39). ☐

Chapter 3

ESTIMATIONAL STOCHASTIC EQUATIONS AND ROBUST ESTIMATORS IN STATISTICAL MODELS ASSOCIATED WITH SEMIMARTINGALES. CONTIGUOUS ALTERNATIVES

3.1. The Limiting Behavior of Roots of the Estimational Stochastic Equations

A key role in robust estimation theory is played by the Huber M-estimators. ([32,35,109,112–115]). In general, M-estimators can be considered as follows.

Consider a sequence of filtered statistical models

$$\mathcal{E} = \{(\Omega^n, \mathcal{F}^n, F^n = (\mathcal{F}_t^n), 0 \le t \le T, (Q_\theta^n, \theta \in \Theta \subset \mathbb{R}_1))\}_{n\ge1}, \tag{3.1}$$

where for each $n \ge 1$ and $\theta \ne \theta'$, the probability measures Q_θ^n and $Q_{\theta'}^n$ are equivalent, $Q_\theta^n \sim Q_{\theta'}^n$, $\mathcal{F}^n = \mathcal{F}_T^n$, and $T > 0$ is a number, the σ-algebra $\mathcal{F}^n$ is complete, and the filtration F^n satisfies the usual conditions w.r.t. Q_θ^n for some and, hence, for each θ (see Subsection 3.3.1, (a), (d) and Remark 3.16 below).

Let for each $\theta \in \Theta$ and $n \ge 1$ the process $(L_n(\theta, t), 0 \le t \le T)$ be a local (square-integrable) Q_θ^n-martingale.

Denote $L_n(\theta) = L_n(\theta, t)|_{t=T}$ and consider the stochastic equation (with respect to the parameter θ)

$$L_n(\theta) = L_n(\theta, \omega) = 0, \quad n \ge 1. \tag{3.2}$$

A sequence $\{T_n(\omega),\ \omega \in \Omega^n\}_{n\geq 1}$ of $\mathcal{F}^n$-measurable roots of these equations (i.e., for each $n \geq 1$, $T_n(\omega)$ is a random variable defined on $(\Omega^n, \mathcal{F}^n)$ with values in Θ, and such that

$$L_n(T_n(\omega), \omega) = 0, \quad n \geq 1) \tag{3.3}$$

is called a generalized M-estimator.

Note that equality (3.3) can hold only asymptotically (in some sense, see, e.g., Theorem 3.1 below).

The proof of assertions concerning the asymptotic behavior of M-estimators as solutions of Eq. (3.2) is carried out in two steps: first, the asymptotic properties are established for the left-hand side of Eq. (3.2); second, the asymptotic properties of estimators (considered as implicit functions) are obtained by linearization. In this way one can construct the so-called CLAN (consistent, linear, asymptotically normal) estimators, which are asymptotically equivalent to M-estimators (see, e.g., (3.15) below). The class of CLAN estimators is a basic class of estimators in robust estimation theory, developed below in this chapter.

3.1.1. Local limiting behavior of roots. The Dugue–Cramer–Le Breton method. Given a sequence of statistical models (3.1), let $\{c_n(\theta)\}_{n\geq 1}$, $c_n(\theta) > 0$, $\theta \in \Theta$, be a normalizing deterministic sequence.

Consider the sequence of random variables $\{L_n(\theta)\}_{n\geq 1} = \{L_n(\theta, \omega),\ \omega \in \Omega^n\}_{n\geq 1}$, depending on the parameter $\theta \in \Theta$.

Remark 3.1. We will use the following abbreviation:

$$Q^n_\theta\text{-}\lim_{n\to\infty} \xi^n = K,$$

where $\xi = \{\xi_n\}_{n\geq 1}$ is a sequence of random variables defined for each n on Ω^n and K is a real number.

This equality means that $\forall \rho > 0$,

$$\lim_{n\to\infty} Q^n_\theta\{\omega \in \Omega^n : |\xi_n(\omega) - K| > \rho\} = 0.$$

Theorem 3.1. *Let the following conditions hold*:

(a) *for each* $\theta \in \Theta$, $\lim\limits_{n\to\infty} c_n(\theta) = 0$;

(b) *for each* $n \geq 1$, *the mapping* $\theta \rightsquigarrow L_n(\theta)$ *is continuously differentiable in* θ Q^n_θ*-a.s.* $(\dot{L}_n(\theta) = \dfrac{\partial}{\partial\theta} L_n(\theta))$;

(c) *for each* $\theta \in \Theta$, *there exists a function* $\Delta_Q(\theta, y)$, $\theta, y \in \Theta$, *such that*

$$Q^n_\theta\text{-}\lim_{n\to\infty} c_n^2(\theta) L_n(y) = \Delta_Q(\theta, y) \tag{3.4}$$

and the equation

$$\Delta_Q(\theta, y) = 0$$

with respect to the variable y *has a unique solution* $\theta^* = b^Q(\theta)$;

(d) $Q^n_\theta\text{-}\lim\limits_{n\to\infty} c_n^2(\theta)\dot{L}_n(\theta^*) = -\gamma_Q(\theta)$, *where* $\gamma_Q(\theta)$ *is a positive number for each* $\theta \in \Theta$;

(e) $\lim\limits_{r\to 0} \limsup\limits_{n\to\infty} Q^n_\theta\{\sup\limits_{\{y:|y-\theta^*|\leq r\}} c_n^2(\theta)|\dot{L}_n(y) - \dot{L}_n(\theta^*)| > \rho\} = 0$ *for each* $\rho > 0$.

Then for each $\theta \in \Theta$ *there exists a sequence of random variables* $T = \{T_n\}_{n\geq 1}$ *taking the values in* Θ *such that*

(I) $\lim\limits_{n\to\infty} Q^n_\theta\{L_n(T_n) = 0\} = 1$;

(II) $Q^n_\theta\text{-}\lim\limits_{n\to\infty} T_n = \theta^*$;

(III) *if* $\{\widetilde{T}_n\}_{n\geq 1}$ *is another sequence with properties* (I) *and* (II), *then*

$$\lim_{n\to\infty} Q^n_\theta\{T_n = \widetilde{T}_n\} = 1.$$

If, in addition,

(f) *the sequence of distributions* $\{\mathcal{L}\{c_n(\theta)L_n(\theta^*) \mid Q_\theta^n\}\}_{n\geq 1}$ *weakly converges to a certain distribution* Φ, *then*

(IV)

$$\text{(i)} \qquad \mathcal{L}\{\gamma_Q(\theta)c_n^{-1}(\theta)(T_n-\theta^*) \mid Q_\theta^n\} \overset{w}{\to} \Phi,$$

$$\text{(ii)} \qquad c_n^{-1}(\theta)(T_n-\theta^*) = \frac{c_n^{-1}(\theta)L_n(\theta^*)}{\gamma_Q(\theta)} + R_n(\theta), \quad R_n(\theta) \overset{Q_\theta^n}{\to} 0.$$

Proof. 1. By the Taylor formula we have

$$L_n(y) = L_n(\theta^*) + \dot{L}_n(\theta^*)(y-\theta^*) + [\dot{L}_n(\overline{\theta}) - \dot{L}_n(\theta^*)](y-\theta^*),$$

where $\overline{\theta} = \theta^* + \alpha(\theta^*)(y-\theta^*)$, $\alpha(\theta^*) \in [0,1]$, and the point $\overline{\theta}$ is chosen so that $\overline{\theta} \in \mathcal{F}^n$ ($\xi \in \mathcal{F}$ means that the r.v. ξ is $\mathcal{F}$-measurable).

From this we obtain

$$c_n^2(\theta)L_n(y) = c_n^2(\theta)L_n(\theta^*) - \gamma_Q(\theta)(y-\theta^*) + \varepsilon_n(\overline{\theta},\theta^*)(y-\theta^*), \tag{3.5}$$

where $\varepsilon_n(y,\theta^*) \in \mathcal{F}^n$,

$$\varepsilon_n(y,\theta^*) = c_n^2(\theta)[\dot{L}_n(y) - \dot{L}_n(\theta^*)] + [c_n^2(\theta)\dot{L}_n(\theta^*) + \gamma_Q(\theta)], \quad y \in \Theta.$$

Obviously, conditions (d) and (e) ensure that

$$\lim_{r\to 0}\limsup_{n\to\infty} Q_\theta^n\big\{\sup_{\{y:|y-\theta^*|\leq r\}} |\varepsilon_n(y,\theta^*)| > \rho\big\} = 0 \tag{3.6}$$

for each $\rho > 0$.

2. We now show that there exists a family $\{\Omega_\theta(n,r)\colon\ n \geq 1,\ r \geq 0,\ \theta \in \Theta\}$ with properties

$$\text{(1)} \qquad \Omega_\theta(n,r) \in \mathcal{F}^n,$$

$$\text{(2)} \qquad \lim_{r\to 0}\liminf_{n\to\infty} Q_\theta^n\{\Omega_\theta(n,r)\} = 1,$$

and for any $r > 0$, $n \geq 1$, and $\omega \in \Omega_\theta(n,r)$ the equation

$$L_n(y) = 0$$

has a unique solution T_n in the segment $|y-\theta^*| \leq r$.

Expansion (3.5) yields

$$c_n^2(\theta)L_n(\theta^*+u)u = c_n^2(\theta)L_n(\theta^*)u - u^2\gamma_Q(\theta) + u^2\varepsilon_n(\overline{\theta},\theta^*). \tag{3.7}$$

For any $\theta \in \Theta$, $n \geq 1$, and $r > 0$, we define

$$\Omega_\theta(n,r) = \left\{\omega \in \Omega^n : |c_n^2(\theta)L_n(\theta^*) < \frac{\gamma_Q(\theta)r}{2},\ \sup_{\{y:|y-\theta^*|\leq r\}} |\varepsilon_n(y,\theta^*)| < \frac{\gamma_Q(\theta)}{2}\right\}.$$

Obviously, $\Omega_\theta(n,r) \in \mathcal{F}^n$. Hence, if $\omega \in \Omega_\theta(r,n)$, then from Eq. (3.7) we obtain $L_n(\theta^*+u)u < 0$ for $|u| = r$.

Since the mapping $u \rightsquigarrow L_n(\theta^*+u)$ is continuous with respect to u, the equation $L_n(\theta^*+u) = 0$ for $|u| \leq r$ has at least one solution $u_n(\theta^*)$ with $|u_n(\theta^*)| \leq r$.

It can be easily seen that if $\omega \in \Omega_\theta(n,r)$ and $|u| \leq r$, then $\dot{L}_n(\theta^*+u) < 0$.

On the other hand, for $\omega \in \Omega_\theta(n,r)$ and $|u| \leq r$,

$$L_n(\theta^*+u,\omega) - L_n(\theta^*+u_n(\theta^*),\omega)$$
$$= \int_0^1 \frac{\partial}{\partial\alpha}[L_n((\theta^*+u_n(\theta^*)) + \alpha(u-u_n(\theta^*)),\omega)]\,d\alpha.$$

Consequently,

$$L_n(\theta^* + u, \omega) = \int_0^1 \dot{L}(\theta^* + u_n(\theta^*) + \alpha(u - u_n(\theta^*)), \omega)(u - u_n(\theta^*))\, d\alpha$$

and

$$\begin{aligned} &L_n(\theta^* + u, \omega)(u - u_n(\theta^*)) \\ &= \int_0^1 \dot{L}_n(\theta^* + u_n(\theta^*) + \alpha(u - u_n(\theta^*)), \omega)(u - u_n(\theta^*))^2 d\alpha < 0, \end{aligned}$$

provided that $u \neq u_n(\theta^*)$. Hence $L_n(\theta^* + u, \omega) \neq 0$ for $|u| \leq r$, $u \neq u_n(\theta^*)$. By the construction of the set $\Omega_\theta(n, r)$ and due to conditions (c), (d), and (e) it is easily seen that (2) is true as well.

3. Now we construct the sequence $T = \{T_n\}_{n \geq 1}$ with properties (I), (II), and (III). Define

$$\Omega_n^\theta := \underset{k>0}{\cup} \Omega_\theta(n, k^{-1}).$$

Obviously, $\Omega_n^\theta \in \mathcal{F}^n$. Let $\omega \in \Omega_n^\theta$. Then it follows from the previous statement that there exists a number $k(\omega) > 0$ such that the equation $L_n(y) = 0$ has a unique solution $\widetilde{T}_n(\omega)$ in the segment $|y - \theta^*| \leq (k(\omega))^{-1}$ with the mapping $\omega \rightsquigarrow \widetilde{T}_n(\omega)$ which is $\Omega_n^\theta \cap \mathcal{F}^n$-measurable (see, e.g., [60]).

We set

$$T_n(\omega) = \begin{cases} \widetilde{T}_n(\omega) & \text{if } \omega \in \Omega_n^\theta, \\ \theta_0 & \text{if } \omega \notin \Omega_n^\theta, \end{cases}$$

where θ_0 is a point in Θ.

It is easily seen that, by construction, T_n possesses properties (I), (II), and (III).

4. Finally, we prove assertion (IV). By expansion (3.5) we have

$$\begin{aligned} &|c_n(\theta) L_n(T_n) - c_n(\theta) L_n(\theta^*) - \gamma_Q(\theta) c_n^{-1}(\theta)(T_n - \theta^*)| \\ &\leq |\varepsilon(\overline{T}_n, \theta^*) \gamma_Q^{-1}(\theta)| |\gamma_Q(\theta) c_n^{-1}(\theta)(T_n - \theta^*)| \end{aligned} \tag{3.8}$$

and $\limsup\limits_{n\to\infty} Q_\theta^n\{|\varepsilon_n(\overline{T}_n, \theta^*)| \geq \rho\} = 0\ \forall \rho > 0$, which follows directly from the relation

$$\{|\overline{T}_n - \theta^*| \leq r\} \cap \{\sup_{\{y: |y - \theta^*| \leq r\}} |\varepsilon_n(y, \theta^*)| < \rho\} \subset \{|\varepsilon_n(\overline{T}_n, \theta^*)| < \rho\}.$$

Denote $X_n := c_n(\theta)(L_n(T_n) - L_n(\theta^*))$, $Y_n := \gamma_Q(\theta) c_n^{-1}(\theta)(T_n - \theta^*)$, and $Z_n := |\varepsilon_n(\overline{T}_n, \theta^*) \gamma_Q^{-1}(\theta)|$. Then inequality (3.8) takes the form

$$|X_n - Y_n| \leq Z_n |Y_n|.$$

It is well known (see [8], Problem 2, Sec. 1.4) that if X_n converges weakly to X ($X_n \overset{w}{\to} X$) and $Z_n \overset{P}{\to} 0$, then $Y_n \overset{w}{\to} X$. Thus, we obtain

$$\lim_{n\to\infty} \mathcal{L}\{\gamma_Q(\theta) c_n^{-1}(\theta)(T_n - \theta^*) \mid Q_\theta^n\} = \lim_{n\to\infty} \mathcal{L}\{c_n(\theta) L_n(\theta^*) \mid Q_\theta^n\}.$$

Assertion (i) is proved. The proof of assertion (ii) easily follows from (i) and inequality (3.8). □

3.1.2. Global limiting behavior of roots. The Perlman-type conditions. We use the objects introduced in Subsection 3.1.1. Assume that $\Theta = [a, b]$. Furthermore, for convenience, we set $a = -\infty$ and $b = +\infty$.

For every θ we consider the set

$$S_\theta = \big\{\widehat{T} = \{\widehat{T}_n\}_{n \geq 1} :\ \text{for each } n \geq 1, \widehat{T}_n \in \mathcal{F}^n,\ \text{and } Q_\theta^n\text{-}\lim_{n\to\infty} c_n^2(\theta) L_n(\widehat{T}_n) = 0\big\}.$$

Theorem 3.2. *Let the following condition* (sup c) *hold*:

(sup c)$_1$ *the function* $\Delta_Q(\theta, y)$ *is* y*-continuous for every* θ;

(sup c)$_2$ *for any* K, $0 < K < \infty$, *and* $\rho > 0$,

$$\lim_{n\to\infty} Q_\theta^n\big\{\sup_{|y|\le K} |c_n^2(\theta)L_n(y) - \Delta_Q(\theta, y)| > \rho\big\} = 0.$$

Then

I. *The following alternative holds*: *if* $\widehat{T} \in S_\theta$, *then either*

$$Q_\theta^n\text{-}\lim_{n\to\infty} \widehat{T}_n = \theta^* = b^Q(\theta) \tag{3.9}$$

or

$$\overline{\lim_{n\to\infty}}\, Q_\theta^n\{|\widehat{T}_n| > K\} > 0 \tag{3.10}$$

for any K, $0 < K < \infty$.

II. *If, in addition, the condition*

$$(c^+) \qquad\qquad \lim_{|y|\to\infty} |\Delta_Q(\theta, y)| = K(\theta) > 0$$

holds and

$$\lim_{n\to\infty} Q_\theta^n\big\{\sup_{-\infty<y<+\infty} |c_n^2(\theta)L_n(y) - \Delta_Q(\theta, y)| > \rho\big\} = 0$$

for any $\rho > 0$, *then* (3.9) *holds.*

Proof. Let $\widehat{T} = \{\widehat{T}_n\}_{n\ge1} \in S_\theta$ and suppose that inequality (3.10) is not satisfied. Then there exists a number $K_0 > 0$ such that

$$\lim_{n\to\infty} Q_\theta^n\{|\widehat{T}_n| > K_0\} = 0.$$

Therefore,

$$\begin{aligned}
Q_\theta^n\big\{|c_n^2(\theta)L_n(\widehat{T}_n) - \Delta_Q(\theta, \widehat{T}_n)| > \rho\big\} &\le Q_\theta^n\{|\widehat{T}_n| > K_0\} \\
+Q_\theta^n\big\{|c_n^2(\theta)L_n(\widehat{T}_n) - \Delta_Q(\theta, \widehat{T}_n)| > \rho,\ |\widehat{T}_n| \le K_0\big\} &\le Q_\theta^n\{|\widehat{T}_n| > K_0\} \\
+Q_\theta^n\big\{\sup_{|y|\le K_0} |c_n^2(\theta)L_n(y) - \Delta(\theta, y)| > \rho\big\} &\to 0 \quad \text{as} \quad n\to\infty.
\end{aligned}$$

On the other hand,

$$Q_\theta^n\text{-}\lim_{n\to\infty} c_n^2(\theta)L_n(\widehat{T}_n) = 0$$

and, hence,

$$Q_\theta^n\text{-}\lim_{n\to\infty} \Delta_Q(\theta, \widehat{T}_n) = 0. \tag{3.11}$$

Now we assume that Eq. (3.9) also fails. Then one can choose $\varepsilon > 0$ such that

$$\overline{\lim_{n\to\infty}}\, Q_\theta^n\big\{|\widehat{T}_n - b^Q(\theta)| > \varepsilon\big\} > 0.$$

By condition (sup c)$_1$,

$$\Delta(\varepsilon) = \inf_{\{y:|y-b^Q(\theta)>\varepsilon,|y|\le K_0\}} |\Delta_Q(\theta, y)| > 0,$$

whence

$$\begin{aligned}
&\overline{\lim_{n\to\infty}}\, Q_\theta^n\{|\Delta_Q(\theta, \widehat{T}_n)| > \Delta(\varepsilon)\} \\
&\quad\ge \overline{\lim_{n\to\infty}}\, Q_\theta^n\{|\Delta_Q(\theta, \widehat{T}_n)| > \Delta(\varepsilon),\ |\widehat{T}_n| \le K_0\} \\
&\quad\ge \overline{\lim_{n\to\infty}}\, Q_\theta^n\{|\widehat{T}_n - b^Q(\theta)| > \varepsilon,\ |\widehat{T}_n| \le K_0\} > 0,
\end{aligned}$$

which contradicts Eq. (3.11).

In order to prove the second assertion of the theorem, it is sufficient to note that under condition (c^+),

$$\inf_{\{y:|y-b^Q(\theta)|\geq\varepsilon\}} |\Delta_Q(\theta,y)| > 0$$

and to repeat the previous arguments. $\square$

Suppose that the conditions of Theorem 3.1 are satisfied.

For every $n \geq 1$, consider the set

$$A_n = \{\omega \in \Omega^n : \text{ the equation } L_n(y,\omega) = 0 \text{ has at least one solution}\}.$$

Note that $A_n \in \mathcal{F}^n$. Indeed, recall that the σ-algebra $\mathcal{F}^n$ is complete, $L_n(y,\cdot) \in \mathcal{F}^n$ for each fixed y, and $L_n(\cdot,\omega)$ is a.s. continuous. Hence, the mapping $(y,\omega) \rightsquigarrow L_n(y,\omega)$ is measurable and $B_n := \{(y,\omega): L_n(y,\omega) = 0\} \in \mathcal{B}(\mathbb{R}_1) \times \mathcal{F}^n$. But $A_n = \Pi_{\Omega^n}(B_n)$, where $\Pi_{\Omega^n}(\cdot)$ is a projection operator. Thus, $A_n \in \mathcal{F}^n$.

Obviously, for any θ, we have $\Omega_n^\theta \subset A_n$, where the set Ω_n^θ is defined in item 3 of the proof of Theorem 3.1.

Since, under the conditions of Theorem 3.1, $Q_\theta^n\{\Omega_n^\theta\} \to 1$ for any θ, we have

$$\lim_{n\to\infty} Q_\theta^n\{A_n\} = 1.$$

For each $n \geq 1$, we introduce the sets

$$S_n = \bigl\{\widetilde{T}_n : \widetilde{T}_n \text{ is } \mathcal{F}^n\text{-measurable; } L_n(\widetilde{T}_n) = 0 \text{ if } \omega \in A_n; \\ \widetilde{T}_n = \theta_0 \text{ if } \omega \notin A_n\bigr\},$$

where θ_0 is a real number.

We consider now the set of estimators

$$S_{\text{sol}} = \{\widetilde{T} = \{\widetilde{T}_n\}_{n\geq1} : \forall n \geq 1,\ \widetilde{T}_n \in S_n\}.$$

Corollary 3.1. *If, along with the conditions of Theorem* 3.1, *conditions* (sup c) *are satisfied for any θ, then there exists an estimator $T^* = \{T_n^*\}_{n\geq1} \in S_{\text{sol}}$ such that*

$$Q_\theta^n\text{-}\lim_{n\to\infty} T_n^* = b^Q(\theta) \tag{3.12}$$

for any θ.

If, moreover, for any θ condition (c^+) is satisfied, then any estimator $\widetilde{T} \in S_{\text{sol}}$ has property (3.12).

Proof. It is sufficient to construct an estimator $T^* = \{T_n^*\}_{n\geq1}$ for which (3.10) fails for each θ.

For any $n \geq 1$ and $\varepsilon > 0$, there exists $T_n^* \in S_n$ such that

$$|T_n^*| \leq \operatorname*{ess\,inf}_{\widetilde{T}_n\in S_n} |\widetilde{T}_n| + \varepsilon.$$

By virtue of Theorem 3.1, for any θ there exists a sequence $\widehat{T}(\theta) = \{\widehat{T}_n(\theta)\}_{n\geq1}$ such that

$$\lim_{n\to\infty} Q_\theta^n\{L_n(\widehat{T}_n(\theta)) = 0\} = 1 \tag{3.13}$$

and

$$Q_\theta^n\text{-}\lim_{n\to\infty} \widehat{T}_n(\theta) = b^Q(\theta). \tag{3.14}$$

Thus, we have

$$\begin{aligned}\overline{\lim_{n\to\infty}}\, Q_\theta^n\{|T_n^*| > K\} &\leq \overline{\lim_{n\to\infty}}\, Q_\theta^n\{|T_n^*| > K, L_n(\widehat{T}_n(\theta)) \neq 0\} + \overline{\lim_{n\to\infty}}\, Q_\theta^n\{|T_n^*| > K, L_n(\widehat{T}_n(\theta)) = 0\} \\ &\leq \overline{\lim_{n\to\infty}}\, Q_\theta^n\{L_n(\widehat{T}_n(\theta)) \neq 0\} + \overline{\lim_{n\to\infty}}\, Q_\theta^n\{|\widehat{T}_n(\theta)| + \varepsilon > K\}.\end{aligned}$$

The first and second terms on the right-hand side converge to zero by virtue of Eqs. (3.13) and (3.14). $\square$

Remark 3.2. We call conditions (sup c) and (c^+) the Perlman-type conditions. In [79], Perlman investigates analogous problem in the i.i.d. case.

Remark 3.3. If the conditions of Corollary 3.1 are satisfied, then by virtue of Theorem 3.1, IV (ii) there exists an estimator $T = \{T_n\}_{n\geq 1}$ such that

$$\begin{gathered} T_n = \theta^* + \frac{L_n(\theta^*)}{\gamma_Q(\theta)} + R_n(\theta), \\ c_n^{-1}(\theta) R_n(\theta) \overset{Q_\theta^n}{\to} 0. \end{gathered} \tag{3.15}$$

If $\theta^* = b^Q(\theta) = \theta$ and the distribution Φ from Theorem 3.1, (f) is Gaussian, then we obtain a CLAN estimator.

3.2. Robust Estimators in Discrete-Time Statistical Models

3.2.1. The statement of the problem. Section 3.2 deals with the robust estimation of a one-dimensional parameter for the contaminated models described in terms of shrinking contamination neighborhoods of nominal conditional densities.

Let a sequence of statistical models

$$\mathcal{E} = \{\mathcal{E}_n\}_{n\geq 1} := (\Omega^n, \mathcal{F}^n, P_\theta^n, \theta \in \Theta \subset \mathbb{R}^1)\}_{n\geq 1}$$

be given, where

$$\Omega^n = \mathcal{X}_0 \times \prod_1^n \mathcal{X}, \quad \mathcal{F}^n = \mathcal{B}_0 \times \prod_1^n \mathcal{B},$$

$(\mathcal{X}_0, \mathcal{B}_0)$ and $(\mathcal{X}, \mathcal{B})$ are some measure Blackwell spaces (see [45]), $x_0 \in \mathcal{X}_0$, $x_i \in \mathcal{X}$, $i \geq 1$. We assume that $P_\theta^n \sim P^n$, where P^n is some probability on $(\Omega^n, \mathcal{F}^n)$. Let $\{f^n := (f_i^n(\theta) := f_i^n(x_i, \theta \mid x_{i-1}, ..., x_0))_{1\leq i\leq n}\}_{n\geq 1}$ be the corresponding system of regular conditional densities, i.e.,

$$P_{\theta,i}^n(dz \mid x_{i-1}, ..., x_0) = f_i^n(z, \theta \mid x_{i-1}, ..., x_0) P^n(dz \mid x_{i-1}, ..., x_0), \quad i \geq 1.$$

In addition,

$$f_0^n(\theta) := f_0^n(x_0, \theta) = dP_{\theta,0}^n / dP_0^n, \quad P_{\theta,0}^n = P_\theta^n \mid \mathcal{B}_0, \ P_0^n = P^n \mid \mathcal{B}_0.$$

This system is referred to as a nominal system.

Suppose that the function $f_i^n(\theta)$ is continuously differentiable with respect to θ for all $0 \leq i \leq n$, $n \geq 1$, P^n-a.s.

We denote

$$l^n = \left(l_i^n(\theta) := \frac{\partial}{\partial\theta} \ln f_i^n(\theta), \quad 0 \leq i \leq n \right),$$

$$I_n(\theta) = E_\theta^n \sum_{i=0}^n (l_i^n(\theta))^2.$$

Assume that for each $n \geq 1$ and $\theta \in \Theta$,

$$0 < I_n(\theta) < \infty, \ n^{-1} I_n(\theta) \to I(\theta) \ \text{ as } \ n \to \infty, \ 0 < I(\theta) < \infty.$$

Introduce the following abbreviations.

If $W = W(i, x_i, ..., x_0)$ and $U = U(i, x_i, ..., x_0)$, $i \geq 1$, are real functions, then we write

$$W_i(z \mid x) := W(i, z, x_{i-1}, ..., x_0), \ \ U_i(x) := U(i, x_{i-1}, ..., x_0).$$

Further, we introduce the measures $\widehat{\mu}_n = \widehat{\mu}_n(dz, di, x; \theta)$, $\mu_n = \mu_n(dz, di, dx; \theta)$, and $\nu_n = \nu_n(di, dx; \theta)$ defined by the relations

$$\int W\, d\widehat{\mu}_n := \int W_i(z \mid x)\widehat{\mu}_n(dz, di, x; \theta) := n^{-1}\sum_{i=1}^n \int W_i(z \mid x) f_i^n(z, \theta \mid x) P_i^n(dz \mid x),$$

$$\int W\, d\mu_n := E_\theta^n \int W\, d\widehat{\mu}_n,$$

$$\int U\, d\nu_n := \int U_i(x)\, \nu_n(di, dx, \theta) := n^{-1} E_\theta^n \sum_{i=1}^n U_i.$$

We consider the class of CLAN estimators.

The estimator $T^\psi = \{T_n^\psi\}_{n\ge 1}$ is said to be CLAN if there exists a sequence of score functions $\psi = \{\psi^n := (\psi_i^n(z, \theta \mid x)_{1\le i\le n}\}_{n\ge 1}$, such that for each θ the following conditions are satisfied:

(c.1) $\psi^n \in L_2(\mu_n)$,

$$\int \psi_i^n(z, \theta \mid x) f_i^n(z, \theta \mid x) P_i^n(dz \mid x) = 0, \quad 0 \le i \le n, \quad \text{for each } \ n \ge 1;$$

(c.2) the Lindeberg condition: for each $a \in (0, 1]$,

$$\int (\psi^n)^2 I_{\{|\psi^n| > an^{1/2}\}} d\widehat{\mu}_n \overset{P_\theta^n}{\to} 0 \ \text{ as } \ n \to \infty;$$

(c.3)

$$\int (\psi^n)^2 d\widehat{\mu}_n \overset{P_\theta^n}{\to} \Gamma^\psi(\theta),$$

$$\int \psi^n l^n d\widehat{\mu}_n \overset{P_\theta^n}{\to} \gamma^\psi(\theta) \ \text{ as } \ n \to \infty,$$

where $0 < \Gamma^\psi(\theta) < \infty$, $0 < \gamma^\psi(\theta) < \infty$, $\Gamma^\psi(\theta)$ and $\gamma^\psi(\theta)$ are deterministic;

(c.4)

$$T_n^\psi = \theta + \frac{n^{-1}\sum\limits_{i=1}^n \psi_i^n(x_i, \theta \mid x)}{\int \psi^n l^n d\widehat{\mu}_n} + R_n(\theta),$$

where

$$n^{1/2} R_n(\theta) \overset{P_\theta^n}{\to} 0.$$

According to the central limit theorem for martingales (see [45]), we obtain from (c.1)–(c.4)

$$\mathcal{L}\{n^{1/2}(T_n^\psi - \theta) \mid P_\theta^n)\} \overset{w}{\to} \mathcal{N}(0, \Gamma^\psi(\theta)/(\gamma^\psi(\theta))^2).$$

It should be noted that condition (c.2) and the ergodicity condition (c.3) are automatically satisfied in "good" ergodic situations with a suitably chosen sequence ψ.

We introduce shrinking contamination neighborhoods for nominal systems of conditional densities.

For each $R > 0$ and $n \ge 1$, we consider the following sets of functions:

$$\Lambda_R^n := \Big\{\lambda^n : \lambda^n = (\lambda_i^n(x, \theta)), \quad 1 \le i \le n, \quad \lambda_i^n(x, \theta) \ge 0, \quad \int \lambda^n d\nu_n \le R\Big\},$$

$$\mathcal{H}_R^n := \Big\{H^n \in L_2(\mu_n) : H^n = (H_i^n(z, \theta \mid x)), \quad 1 \le i \le n,$$
$$\int H_i^n(z, \theta \mid x) f_i^n(z, \theta \mid x) P_i^n(dz \mid x) = 0,$$
$$H_i^n(z, \theta \mid x) \ge -\lambda_i^n(x, \theta) \quad \mu_n\text{-a.s.}, \quad \lambda^n \in \Lambda_R^n\Big\}.$$

Now we define the neighborhood of conditional density $f^n = (f_i^n(z, \theta \mid x))_{1 \le i \le n}$,

$$F_R^n = \{f^{n,H} : f^{n,H} = (f_i^{n,H}(z,\theta \mid x)),\ 1 \le i \le n : f^{n,H} = (1 + n^{-1/2} H^n) f^n : H^n \in \mathcal{H}_R^n\}. \tag{3.16}$$

Assume that for each θ

$$\sup_{i,x,n} \lambda_i^n(x,\theta) < \infty. \tag{3.17}$$

Remark 3.4. Obviously, by virtue of (3.17), $f^{n,H}$ is a conditional density for a sufficiently large n. This property is a sufficient condition for the asymptotic theory developed below. For convenience, without loss of generality, assume that $f^{n,H}$ is a density for all $n \ge 1$.

Remark 3.5. Consider the set of functions

$$\begin{aligned}\Phi_R^n = \Big\{ & f^{n,\lambda,h} : f^{n,\lambda,h} = \big(f_i^{n\lambda,h}(z,\theta \mid x)\big), \quad i \le i \le n, \\ & f^{n,\lambda,h} = \left(1 - \frac{\lambda^n}{\sqrt{n}}\right) f^n + \frac{\lambda^n}{\sqrt{n}} h^n, \\ & \lambda^n \in \Lambda_R^n, \quad h^n = (h_i^n(z,\theta \mid x)), \quad 1 \le i \le n, \\ & \int h_i^n(z,\theta \mid x) P_i^n(dz \mid x) = 1, \quad h_i^n(z,\theta \mid x) \ge 0 \Big\}\end{aligned}$$

(the generalized Huber's "gross error" model).

It is easy to see that there exists a one-to-one correspondence between the sets F_R^n and Φ_R^n given by the following relations:

$$\begin{aligned} & H^n = \lambda^n \frac{h^n}{f^n} - \lambda^n \quad \text{on the set } \{(i,x) : \lambda_i^n(x,\theta) > 0\}; \\ & H^n = 0 \quad \text{and } h^n \text{ is any density on the set } \{(i,x) : \lambda_i^n(x,\theta) = 0\}. \end{aligned}$$

Recall that θ and n are fixed.

Let $\mathcal{H}_{\text{seq}}$ denote a class of sequences $H = \{H^n\}_{n \ge 1}$ with the following properties:
(1) $H^n \in \mathcal{H}_R^n$ for each $n \ge 1$;
(2) $\lim\limits_{N \to \infty} \lim\limits_{n \to \infty} P_\theta^{n,H} \Big\{ \int (H^n)^2 d\widehat{\mu}_n > N \Big\} = 0$,
where $P_\theta^{n,H}$ is the probability measure corresponding to the set of conditional densities $(f_i^{n,H})_{1 \le i \le n} \in F_R^n$ and to the initial density $f_0^n(\theta)$.

The sequence $\{P_\theta^{n,H}\}_{n \ge 1}$ is referred to as the sequence of alternative measures or, briefly, alternatives.

Proposition 3.1. *Let $H = \{H^n\}_{n \ge 1} \in \mathcal{H}_{\text{seq}}$. Then:*
(i) $(P_\theta^{n,H}) \lhd (P_\theta^n)$;
(ii) *if T^ψ is CLAN, then*

$$\mathcal{L}\{n^{1/2}(T_n^\psi - \theta) - b_n^{\psi,H}(\theta)/\gamma^\psi(\theta) \mid P_\theta^{n,H}\} \xrightarrow{w} \mathcal{N}(0, \Gamma^\psi(\theta)/(\gamma^\psi(\theta))^2),$$

where

$$b_n^{\psi,H}(\theta) = \int \psi^n H^n d\widehat{\mu}_n;$$

(iii) *if*

$$b_n^{\psi,H}(\theta) \xrightarrow{P_\theta^n} \beta^{\psi,H}(\theta),$$

where $\beta^{\psi,H}(\theta)$ is deterministic, then

$$\mathcal{L}\{n^{1/2}(T_n^\psi - \theta) \mid P_\theta^{n,H}\} \xrightarrow{w} \mathcal{N}(\beta^{\psi,H}(\theta)/\gamma^\psi(\theta), \Gamma^\psi(\theta)/(\gamma^\psi(\theta))^2).$$

Below, in Sec. 3.3, we prove this proposition in the general case.

Proposition 3.1 enables us to derive an explicit expression for the asymptotic mean-square error under the alternatives $\{P_\theta^{n,H}\}_{n\geq 1}$ (the risk functional). Namely, we have

$$\lim_{a\to\infty}\lim_{n\to\infty} E_\theta^{n,H}[n(T_n^\psi-\theta)^2\wedge a] = D(\psi,H;\theta),$$

where

$$D(\psi,H;\theta) = \frac{[\beta^{\psi,H}(\theta)]^2+\Gamma^\psi(\theta)}{(\gamma^\psi(\theta))^2}. \tag{3.18}$$

In what follows, we assume that all ergodicity conditions expressed in terms of the convergence of integrals with respect to the measure $(\widehat{\mu}_n)$ and with respect to the measure (μ_n) are equivalent, i.e., if $\phi=\{\phi^n\}_{n\geq 1}$ and $\chi=\{\chi^n\}_{n\geq 1}$ are some sequences of functions with ϕ^n, $\chi^n\in L_2(\mu_n)$ $\forall n\geq 1$, then

$$\int \phi^n\chi^n d\widehat{\mu}_n \overset{P_\theta^n}{\to} C \iff \int \phi^n\chi^n d\mu_n \to C \ \text{ as } \ n\to\infty.$$

Then we have

$$D(\psi,H;\theta) = \lim_{n\to\infty} D_n(\psi^n,H^n;\theta), \tag{3.19}$$

where

$$D_n(\psi,H;\theta) = \frac{\left[\left(\int \psi H\, d\mu_n\right)^2+\int \psi^2 d\mu_n\right]}{\left(\int \psi l^n d\mu_n\right)^2}. \tag{3.20}$$

We introduce the optimization criterion and the optimization problem.

Let Ψ and $\mathcal{H}_\Psi$ be some classes of sequences $\psi=\{\psi^n\}_{n\geq 1}$ and $H=\{H^n\}_{n\geq 1}$ such that for each $\psi\in\Psi$ and $H\in\mathcal{H}_\Psi$ the conditions of Proposition 3.1 are satisfied.

The sequence of score functions $\psi^*=\{\psi^{*,n}\}_{n\geq 1}\in\Psi$ is said to be $(\Psi,\mathcal{H}_\Psi)$-optimal in the minimax sense with respect to the risk functional $D(\psi,H;\theta)$ if

$$\sup_{H\in\mathcal{H}_\Psi} D(\psi^*,H;\theta) = \inf_{\psi\in\Psi}\sup_{H\in\mathcal{H}_\Psi} D(\psi,H;\theta) \tag{3.21}$$

or, equivalently, if for each $\varepsilon>0$ and $\psi\in\Psi$, there exists $H^{\varepsilon,\psi}\in\mathcal{H}_\Psi$ such that for each $H\in\mathcal{H}_\Psi$, the inequality

$$D(\psi^*,H;\theta)\leq D(\psi,H^{\varepsilon,\psi};\theta)(1+\varepsilon) \tag{3.22}$$

is satisfied.

Remark 3.6. (3.21) $\iff$ (3.22).

(1) (3.21) $\Longrightarrow$ (3.22). Indeed, from (3.21) $\forall\psi\in\Psi$, $\forall H\in\mathcal{H}_\Psi$ we have

$$D(\psi^*,H;\theta)\leq \sup_{H\in\mathcal{H}_\Psi} D(\psi,H;\theta).$$

By the definition of the supremum, $\forall\varepsilon>0$ $\exists H^{\varepsilon,\psi}\in\mathcal{H}_\Psi$:

$$\sup_{H\in\mathcal{H}_\Psi} D(\psi,H;\theta)\leq D(\psi,H^{\varepsilon,\psi};\theta)+\varepsilon = D(\psi,H^{\varepsilon,\psi};\theta)\left(1+\frac{\varepsilon}{D(\psi,H^{\varepsilon,\psi};\theta)}\right).$$

(2) (3.22) $\Longrightarrow$ (3.21). Indeed, (3.22) is true for each ψ and H. Therefore,

$$\sup_{H\in\mathcal{H}_\Psi} D(\psi^*,H;\theta)\leq \sup_{H\in\mathcal{H}_\Psi} D(\psi,H;\theta)+\varepsilon.$$

The last inequality yields

$$\sup_{H\in\mathcal{H}_\Psi} D(\psi^*, H;\theta) \le \inf_{\psi\in\Psi}\sup_{H\in\mathcal{H}_\Psi} D(\psi, H;\theta) + \varepsilon \quad \forall\varepsilon > 0.$$

Hence,

$$\sup_{H\in\mathcal{H}_\Psi} D(\psi^*, H;\theta) \le \inf_{\psi\in\Psi}\sup_{H\in\mathcal{H}_\Psi} D(\psi, H;\theta).$$

The inverse inequality is trivial.

The CLAN estimator $T^* = T^{\psi^*}$, which corresponds to the optimal sequence of score functions ψ^*, is called the $(\Psi, \mathcal{H}_\Psi)$-minimax estimator.

The problem is to construct an optimal sequence of score functions ψ^*.

The solution of this problem is directly associated with the analytic form of the risk functional and the classes Ψ and $\mathcal{H}_\Psi$ with respect to which the minimax operation is taken.

In general, $\beta = \beta^{\psi,H}$, $\Gamma = \Gamma^\psi$, and $\gamma = \gamma^\psi$ are functionals of the sequences ψ, H, and l. In particular, they are limits of certain scalar products (see previous definitions).

Two alternatives are possible:

(1) β, Γ, and γ preserve the form of the scalar product in an appropriate $L_2(\mu)$-space with a finite or σ-finite measure μ;

(2) β, Γ, and γ do not possess this property.

In the first case, by a suitable choice of the classes Ψ and $\mathcal{H}_\Psi$ we arrive at the standard minimax problem

$$\inf_{\psi\in\Psi}\sup_{H\ge 0, \int H\le R} \frac{\langle\psi, H\rangle^2 + \|\psi\|^2}{\langle\psi, l\rangle^2}, \tag{3.23}$$

with "rich" classes of alternatives. The solution of this problem is the Huber function. This type of contamination will be called a *full contamination.*

In the second case, in order to solve the above minimax problem, we develop the "approximation technique" which consists in the following: construct the optimal score function $\psi^{*,n}$ for the fixed step n; describe sufficiently wide classes Ψ and $\mathcal{H}_\Psi$ such that the sequence $\psi^* = \{\psi^{*,n}\}_{\ge 1}$ is $(\Psi, \mathcal{H}_\Psi)$-optimal.

3.2.2. Fixed-step optimization problem. In this subsection, for each fixed $n \ge 1$, we construct the optimal score function $\psi^{*,n}$ with respect to the risk functional $D_n(\psi, H;\theta)$ (see (3.19), (3.20)). More precisely, $\psi^{*,n} \in \Psi_n^0$ is said to be optimal at the nth step if

$$\sup_{H\in\mathcal{H}_R^n} D_n(\psi^{*,n}, H;\theta) = \inf_{\psi\in\Psi_n^0}\sup_{H\in\mathcal{H}_R^n} D_n(\psi, H;\theta),$$

where

$$\Psi_n^0 = \Big\{\psi^n \in L_2(\mu) : \int \psi_i^n(z,\theta \mid x) f_i^n(z,\theta \mid x) P_i^n(dz \mid x) = 0, 1 \le i \le n\Big\}. \tag{3.24}$$

Denote by

$$Q_i^n(\cdot,\theta \mid x) := \int I_{\{z: l_i^n(z,\theta|x)\in\cdot\}} f_i^n(z,\theta \mid x) P_i^n(dz \mid x) \tag{3.25}$$

the conditional distribution of $l_i^n(z,\theta \mid x)$ with a given (i,x) and consider the equation (with respect to β)

$$\int [y-\beta]_{-m}^m Q_i^n(dy,\theta \mid x) = 0, \tag{3.26}$$

where $[x]_a^b = (x\wedge b)\vee a$, $m > 0$ is a number.

We denote by

$$\beta^n := \beta_i^n(x,m,\theta) := \beta(Q_i^n(\cdot,\theta \mid x), m)$$

the solution of Eq. (3.26).

Remark 3.7. The question of solvability of Eq. (3.26) is considered below, in Lemma 3.2.

Assume that the distribution $Q_i^n(\cdot, \theta \mid x)$ satisfies the following conditions: for each fixed n and θ
(a) $Q_i^n(\cdot, \theta \mid x)$ has a unique median;
(b) $Q_i^n(\{\beta_i^n(x, 0, \theta)\}, \theta \mid x) = 0$, where

$$\beta_i^n(x, 0, \theta) = \lim_{m \to 0} \beta_i^n(x, m, \theta).$$

Note that the assertions of Theorem 3.3 below are true without condition (a). But the proof is technically more complicated (see [119]).

Theorem 3.3. (i) *There exists an optimal* $\psi^{*,n}$ *which is equal to*

$$\psi^{*,n} = \big([l_i^n(z, \theta \mid x) - \beta_i^n(x, m_n^*(\theta), \theta)]_{-m_n^*(\theta)}^{m_n^*(\theta)}\big)_{1 \le i \le n}, \tag{3.27}$$

where $m_n^*(\theta)$ *is the unique solution of the equation*

$$\begin{aligned} R^2 m^2 = \int\int &([y - \beta_i^n(x, m, \theta)]_{-m}^m y \\ &- ([y - \beta_i^n(x, m, \theta)]_{-m}^m)^2) Q_i^n(dy, \theta \mid x) \nu_n(di, dx, \theta). \end{aligned} \tag{3.28}$$

(ii) *This* $\psi^{*,n}$ *is* μ_n*-a.s. unique (up to a constant factor).*

Proof. First we prove three lemmas. In the sequel, the parameter θ is fixed and omitted.
Let for each $n \ge 1$

$$\Psi_n = \left\{\psi^n \in L_2(\mu_n) : \int \psi_i^n(z \mid x) f_i^n(z \mid x) P_i^n(dz \mid x) = 0,\ 1 \le i \le n,\ \int \psi^n l^n d\mu_n = 1\right\},$$
$$\Psi_{\gamma,n} = \{\psi^n : |\psi_i^n(z \mid x)| \le \gamma \quad \mu_n\text{-a.s.}\}, \quad \gamma > 0,$$
$$\Gamma_n = \{\gamma : \Psi_n \cap \Psi_{\gamma,n} \ne \varnothing\}.$$

Below we omit the index n as well.

Lemma 3.1. $\Gamma = \{\gamma : \gamma \ge d^{-1}\}$, *where*

$$d := \int\int |y - \mathrm{med}_i(x)| Q_i(dy \mid x) \nu(di, dx)$$

and $\mathrm{med}_i(x)$ *is the median of the distribution* $Q_i(\cdot \mid x)$.

Proof. First let us prove the inclusion $\Gamma \subset \{\gamma\colon \gamma \ge d^{-1}\}$.
Let $\phi_i(x)$ be a measurable function with $|\phi_i(x)| \le E_{i,x}|l|$, where $E_{i,x}$ is the sign of conditional expectation (see (3.29) and Remark 3.10 below).
For each $\gamma \in \Gamma$, $\psi \in \Psi \cap \Psi_\gamma$,

$$\begin{aligned} 1 &= \left|\int \psi_i(z \mid x)(l_i(z \mid x) - \phi_i(x)) \mu(dt, di, dx)\right| \\ &\le \gamma \int |l_i(z \mid x) - \phi_i(x)|\, \mu(dt, di, dx). \end{aligned}$$

Hence

$$\gamma \ge \left(\inf_\phi \int |l_i(z \mid x) - \phi_i(x)|\, \mu(dz, di, dx)\right)^{-1} = d^{-1}$$

by the well-known minimization property of the median.
The inverse inclusion follows from Lemma 3.3. ∎

Everywhere below, by the derivative we mean the right derivative.

Denote

$$E_{i,x}|l| := \int |y| Q_i(dy \mid x). \tag{3.29}$$

Lemma 3.2. (1) *For each $m > 0$, the equation*

$$\int [y-\beta]_{-m}^{m} Q_i(dy \mid x) = 0 \tag{3.30}$$

has a unique measurable solution $\beta_i(x,m)$;

(2) *there exists a constant $C \geq 1$ such that*

$$|\beta_i(x,m)| \leq C E_{i,x}|l| \quad (\nu\text{-a.s.}); \tag{3.31}$$

(3) *for each (i,x), the function $m \rightsquigarrow \beta_i(x,m)$ has a derivative $\beta_i'(x,m)$ and $|\beta_i'(x,m)| \leq 1$.*

Proof. (1) We consider the function

$$f(m,\beta) = \int [y-\beta]_{-m}^{m} Q_i(dy \mid x) \quad (:= f_{(i,x)}(m,\beta)).$$

It is easy to verify that

$$|[y-\beta_1]_{-m_1}^{m_1} - [y-\beta_2]_{-m_2}^{m_2}| \leq |m_1 - m_2| \vee |\beta_1 - \beta_2|$$

and hence $f(m,\beta)$ is continuous with respect to (m,β). Further, we observe that

$$f_\beta'(m,\beta) = -\int I_{\{-m+\beta<y\leq m+\beta\}} Q_i(dy \mid x) \leq 0, \tag{3.32}$$

and $\lim\limits_{\beta\to\pm\infty} f(m,\beta) = \mp m$. Therefore, for each fixed $m > 0$, Eq. (3.30) has a solution.

Recall that $Q_i(\cdot \mid x)$ has a unique median. This implies that $f_\beta'(m,\beta) < 0$ in a neighborhood of each point (m,β) such that $f(m,\beta) = 0$ and, according to the implicit function theorem, $\beta_i(x,m)$ is continuous in m.

Thus, there exists a solution $\beta_i(x,m)$ of Eq. (3.30), which is unique and m-continuous. Finally, we note that for fixed m, $\beta_i(x,m)$ is (i,x)-measurable, since it can be considered as a hitting time of level zero by the continuous process $f_{(i,x)}(m,\beta)$. □

Remark 3.8. It is easy to verify that $\lim\limits_{m\to\infty} \beta_i(x,m) = 0$ and if $Q_i(\{\beta_i(x,0)\} \mid x) = 0$, then $\beta_i(x,0) = \lim\limits_{m\to 0} \beta_i(x,m) = \mathrm{med}_i(x)$.

(2) Recall that the function $\beta \to f(m,\beta)$ is continuous, decreases from m to $-m$ (see (3.32)), and $f(m,\beta_i(x,m)) = 0$. Now the statement is equivalent to the existence of a constant $C \geq 1$ such that

$$f(m, CE_{i,x}|l|) \leq 0, \quad f(m, -CE_{i,x}|l|) \geq 0$$

for each $m > 0$.

Let us prove, e.g., the first inequality. Show that for each $m > 0$,

$$\beta_i(x,m) \leq E_{i,x}|l| + m.$$

It suffices to prove that

$$f(m, E_{i,x}|l| + m) \leq 0 \quad \forall m > 0.$$

This inequality follows from the following calculations:

$$\begin{aligned}
f(m, E_{i,x}|l| + m) = {} & \int y I_{\{E_{i,x}|l| \le y \le E_{i,x}|l|+2m\}} Q_i(dy \mid x) \\
& - E_{i,x}|l| Q_i((E_{i,x}|l|, E_{i,x}|l| + 2m) \mid x) \\
& - m Q_i((E_{i,x}|l|, E_{i,x}|l| + 2m) \mid x) \\
& - m Q_i((-\infty, E_{i,x}|l|] \mid x) \\
& + m Q_i([E_{i,x}|l| + 2m, +\infty) \mid x) \\
& \le E_{i,x}|l| Q_i((E_{i,x}|l|, E_{i,x}|l| + 2m) \mid x) \\
& + 2m Q_i((E_{i,x}|l|, E_{i,x}|l| + 2m) \mid x) \\
& - E_{i,x}|l| Q_i((E_{i,x}|l|, E_{i,x}|l| + 2m) \mid x) \\
& - m\{Q_i(E_{i,x}|l|, E_{i,x}|l| + 2m) \mid x) + Q_i((-\infty, E_{i,x}|l|] \mid x) \\
& - Q_i([E_{i,x}|l| + 2m, +\infty) \mid x)\} = 0.
\end{aligned}$$

Now we show that $\exists \alpha > 2$ such that

$$\beta_i(x, m) \le \alpha E_{i,x}|l|$$

for each $m > \alpha E_{i,x}|l|$.

Indeed, fix ε_0, $0 < \varepsilon_0 < \dfrac{1}{2}$, and take $\alpha \ge \left(\dfrac{2}{1 - 2\varepsilon_0} \vee \dfrac{1}{\varepsilon_0}\right)$.

We have

$$E_{i,x}|l| \ge E_{i,x}(|l| I_{\{l \ge \alpha E_{i,x}|l|\}}) \ge \alpha E_{i,x}|l| Q_i([\alpha E_{i,x}|l|, +\infty) \mid x).$$

Thus,

$$Q_i([\alpha E_{i,x}|l|, +\infty) \mid x) \le \frac{1}{\alpha} \le \varepsilon_0.$$

Denote

$$\begin{aligned}
\varepsilon_+ &= Q_i([\alpha E_{i,x}|l| + m, +\infty) \mid x), \\
\varepsilon_- &= Q_i((-\infty, \alpha E_{i,x}|l| - m] \mid x).
\end{aligned}$$

In these notations,

$$\varepsilon_+ \le \varepsilon_0, \quad m\varepsilon_+ \le E_{i,x}|l| \quad \text{for each } \ m > 0$$

(indeed, $m\varepsilon_+ \le m \dfrac{E_{i,x}|l|}{\alpha E_{i,x}|l| + m} \le E_{i,x}|l| \ \forall m > 0$), and for each $m > \alpha E_{i,x}|l|$,

$$\begin{aligned}
f(m, \alpha E_{i,x}|l|) = {} & \int y I_{\{-m+\alpha E_{i,x}|l| \le y \le m + \alpha E_{i,x}|l|\}} Q_i(dy \mid x) \\
& - \alpha E_{i,x}|l|(1 - \varepsilon_+ - \varepsilon_-) - m(\varepsilon_- - \varepsilon_+) \\
& \le \begin{cases} E_{i,x}|l| - \alpha E_{i,x}|l| + \alpha\varepsilon_+ E_{i,x}|l| + \alpha\varepsilon_- E_{i,x}|l| \\ \qquad - \alpha\varepsilon_- E_{i,x}|l| + \alpha\varepsilon_+ E_{i,x}|l| & \text{if } \ \varepsilon_- \ge \varepsilon_+, \\ 2E_{i,x}|l| - \alpha E_{i,x}|l| + 2\alpha\varepsilon_+ E_{i,x}|l| & \text{if } \ \varepsilon_- < \varepsilon_+ \end{cases} \\
& = \left.\begin{cases} E_{i,x}|l|(1 + 2\alpha\varepsilon_+ - \alpha) & \text{if } \ \varepsilon_- \ge \varepsilon_+ \\ E_{i,x}|l|(2 + 2\alpha\varepsilon_+ - \alpha) & \text{if } \ \varepsilon_- < \varepsilon_+ \end{cases}\right\} \le 0.
\end{aligned}$$

Finally, we obtain the following: $\exists \alpha > 2$,

$$\begin{aligned}
\beta_i(x, m) &\le \alpha E_{i,x}|l| \le (\alpha + 1) E_{i,x}|l| \quad \text{if } \ m > \alpha E_{i,x}|l|, \\
\beta_i(x, m) &\le E_{i,x}|l| + m \le (\alpha + 1) E_{i,x}|l| \quad \text{if } \ m \le \alpha E_{i,x}|l|,
\end{aligned}$$

whence the desirable result follows with $C = 1 + \alpha$. $\square$

Remark 3.9. If we take $\varepsilon_0 = \dfrac{1}{4}$, then we obtain $\alpha = 4$ and hence $|\beta_i(x,m)| \le 5E_{i,x}|l|$.

Remark 3.10. $|\operatorname{med}_i(x)| \le E_{i,x}|l|$.

Indeed, from the conditionally centering property of a random variable l, we have

$$E_{i,x}|l| = 2E_{i,x}l^+, \quad l^+ = \max(0,l).$$

Hence

$$E_{i,x}l^+ \ge E_{i,x}(l^+ I_{\{l \ge E_{i,x}|l|\}}) \ge 2E_{i,x}l^+ Q_i([E_{i,x}|l|, +\infty) \mid x)$$

and, therefore,

$$Q_i([E_{i,x}|l|, +\infty) \mid x) \le \frac{1}{2}.$$

Remark 3.11. Inequality (3.31) holds for any conditionally centered random variable (not only for l).

(3) Suppose that $(\Delta m > 0) \Rightarrow (\Delta\beta(m) \ge 0)$, where Δ denotes an increment.

Recall that in the neighborhood of the point (m,β) such that $f(m,\beta) = 0$, we have $f'_\beta(m,\beta) < 0$. From the implicit function theorem we obtain

$$\beta'(m) = -\frac{f'_m(m,\beta+0)|_{\beta=\beta(m)}}{f'_\beta(m,\beta)|_{\beta=\beta(m)}}.$$

Thus we obtain

$$\beta'_i(x,m) = \frac{\displaystyle\int (I_{\{y-\beta_i(x,m)>m\}} - I_{\{y-\beta_i(x,m)\le -m\}}) Q_i(dy \mid x)}{\displaystyle\int I_{\{-m<y-\beta_i(x,m)\le m\}} Q_i(dy \mid x)}. \tag{3.33}$$

Using Eq. (3.33), it is easy to verify that $|\beta'_i(x,m)| \le 1$.

The consideration of the case $(\Delta m > 0) \Rightarrow (\Delta\beta(m) < 0)$ is quite similar. In this case,

$$\beta'_i(x,m) = \frac{\displaystyle\int (I_{\{y-\beta_i(x,m)\ge m\}} - I_{\{y-\beta_i(x,m)< -m\}}) Q_i(dy \mid x)}{\displaystyle\int I_{\{-m\le y-\beta_i(x,m)< m\}} Q_i(dy \mid x)}.$$

Lemma 3.3. *If* $\gamma \in \Gamma$*, then* (i)

$$\inf_{\psi \in \Psi_n \cap \Psi_{\gamma,n}} \int \psi^2 d\mu_n = \int (\psi^*)^2 d\mu_n,$$

where

$$\psi_i^*(z,m \mid x) = [l_i(z \mid x) - \beta_i(x,m)]_{-m}^m \left(\int\int [y - \beta_i(x,m)]_{-m}^m y Q_i(dy \mid x) \nu(di,dx) \right)^{-1}, \tag{3.34}$$

and if $\gamma > d^{-1}$*, then* m *is a solution of the equation*

$$m \left(\int\int [y - \beta_i(x,m)]_{-m}^m y Q_i(dy \mid x) \nu(di,dx) \right)^{-1} = \gamma; \tag{3.35}$$

if $\gamma = d^{-1}$*, then*

$$\psi_i^*(z,m \mid x) = d^{-1} \operatorname{sign}(l_i(z \mid x) - \operatorname{med}_i(x));$$

(ii) $\psi^* = (\psi_i^*(z,m \mid x))_{1\le i\le n}$ *is* μ_n*-a.s. unique.*

Remark 3.12. It is obvious that

$$\int\int [y-\beta_i(x,m)]_{-m}^{m} y Q_i(dy \mid x)\nu(di,dx)$$
$$= \int\int [|y-\beta_i(x,m)|]_0^m |y-\beta_i(x,m)| Q_i(dy \mid x)\nu(di,dx) \geq 0$$

and this relation is equal to zero iff a random variable $l_i(z \mid x) = \beta_i(x,m)$ (μ_n-a.s.). In this case, $\Psi_n = \varnothing$.

Proof. Note that if $m > 0$, then

$$|m^{-1}[y-\beta]_{-m}^{m} y| = |[y/m - \beta/m]_{-1}^{1} y| \leq |y|$$

and by the Lebesgue theorem, the function

$$F(m) := m^{-1} \int\int [y-\beta_i(x,m)]_{-m}^{m} y Q_i(dy \mid x)\nu(di,dx)$$

is continuous and, moreover, $\lim\limits_{m\to\infty} F(m) = 0$, $\lim\limits_{m\to 0} F(m) = d$. Thus, the equation $\gamma F(m) = 1$ has a solution.

In what follows, for simplicity, we use the following notation:

$$I^+ := I^+(y,i,x) = \begin{cases} I_{\{y-\beta_i(x,m)>m\}} & \text{if } (\Delta m > 0) \Rightarrow (\Delta\beta_i(m,x) \geq 0), \\ I_{\{y-\beta_i(x,m)\geq m\}} & \text{if } (\Delta m > 0) \Rightarrow (\Delta\beta_i(m,x) < 0), \end{cases}$$
$$I^- := I^-(y,i,x) = \begin{cases} I_{\{y-\beta_i(x,m)\leq -m\}} & \text{if } (\Delta m > 0) \Rightarrow (\Delta\beta_i(m,x) \geq 0), \\ I_{\{y-\beta_i(x,m)< -m\}} & \text{if } (\Delta m > 0) \Rightarrow (\Delta\beta_i(m,x) < 0), \end{cases}$$
$$I^0 = 1 - I^+ - I^-.$$

If $\phi = \phi(y,i,x)$ is some real function, then

$$\begin{gathered} \int \phi := \int \phi(y,i,x)\, Q_i(dy \mid x), \\ \int\int \phi = \int\Big(\int \phi\Big) := \int\Big(\int \phi(y,i,x)\, Q_i(dy \mid x)\Big)\nu(di,dx). \end{gathered} \tag{3.36}$$

We show that $F'(m) < 0$.

We rewrite $F(m)$ as

$$F(m) = m^{-1} \int\int [y-\beta(m)]_{-m}^{m} (y-\beta(m)).$$

Then, by virtue of statements (2) and (3) of Lemma 3.2, we have

$$\begin{aligned} F'(m) = m^{-2}\Big\{ \int\int & [(y-\beta(m))I^+ - m\beta'(m)I^+ - (y-\beta(m))I^- \\ & + m\beta'(m)I^- - 2\beta'(m)(y-\beta(m))I^0] - \int\int [y-\beta(m)]_{-m}^{m}(y-\beta(m)) \Big\} \\ = m^{-2}\Big\{ m \int\int & [(y-\beta(m))I^+ - (y-\beta(m))I^- - \beta'(m)(y-\beta(m))I^0] \\ & - \int\int [y-\beta(m)]_{-m}^{m}(y-\beta(m)) \Big\}. \end{aligned}$$

The last equality follows from the equality

$$m\int I^+ - m\int I^- + \int (y-\beta(m))I^0 = 0. \tag{3.37}$$

Further,

$$F'(m) = m^{-2}\int\left[-m\beta'(m)\int(y-\beta(m))I^0 - \int(y-\beta(m))^2 I^0\right]$$
$$= m^{-2}\int\left[-m\frac{\int I^+ - \int I^-}{\int I^0}\int(y-\beta(m))I^0 - \int(y-\beta(m))^2 I^0\right]$$
$$= m^{-2}\int\left[\frac{(\int(y-\beta(m))I^0)^2}{\int I^0} - \int(y-\beta(m))^2 I^0\right] < 0,$$

by virtue of Eq. (3.37) and the Schwartz inequality.

Hence, in this case, Eq. (3.35) has a unique solution.

It is well known ([32]) that in this case and in the limit case as $m \to 0$, the optimal score function ψ^* has the form given by (3.34). □

Proof of Theorem 3.3. From Lemmas 3.1 and 3.3 we have

$$\inf_{\psi\in\Psi_n^0}\sup_{H\in\mathcal{H}_R^n} D_n(\psi,H) = \inf_{\gamma\in\Gamma}\left(R^2\gamma^2 + \inf_{m:m\ge 0,\frac{m}{v(m)}\le\gamma}\frac{u(m)}{v(m)}\right) = \inf_{m\ge 0}\Phi(m),$$

where $\Phi(m) = (R^2m^2 + u(m))v^{-2}(m)$,

$$u(m) = \int\int([y-\beta(m)]_{-m}^{m})^2, \tag{3.38}$$

$$v(m) = \int\int[y-\beta(m)]_{-m}^{m}y \tag{3.39}$$

(for the notation see (3.36)).

Consider the behavior of the function $\Phi(m)$. From Lemma 3.2 it immediately follows that there exists $\Phi'(m)$ and $\Phi'(m) = 2v^{-3}(m)p(m)q(m)$, where

$$p(m) = m(R^2 - \varphi(m)), \tag{3.40}$$

$$\varphi(m) = m^{-2}[v(m) - u(m)], \tag{3.41}$$

$$q(m) = \int\int(y-\beta(m))(y-\beta(m)+m\beta'(m))I^0.$$

It is easy to verify that

$$\varphi(m)\ge 0,\quad \varphi(m)\to_0^\infty \quad\text{as}\quad m\to_\infty^0 .$$

We show that

$$\varphi'(m) < 0. \tag{3.42}$$

Indeed, we have

$$\varphi'(m) = -m^{-2}\int\int(I^+ - I^- - \beta'(m)I^0)(y-\beta(m))$$

and

$$\int\int(I^+ - I^- - \beta'(m)I^0)(y-\beta(m)) > m > 0.$$

Thus, there exists a unique point m^* such that $R^2 - \varphi(m^*) = 0$.

Consider the function $q(m)$. It is easy to verify that $q(m) > 0$ for each m, as immediately follows from the assumption that $Q_i(\cdot \mid x)$ has a unique median and the definition of I^0.

Hence there exists a unique minimum point m^* of the function $\Phi(m)$ given by Eq. (3.28).

Finally, the optimal ψ^* has the form (3.27) and is unique (up to a constant factor). □

3.2.3. Approximation by a fixed-step solution. Let for each $n \geq 1$ the system $Q^n = (Q_i^n(\cdot \mid x))_{1 \leq i \leq n}$ of regular conditional distributions on $(\mathbb{R}_d, \mathcal{B}(\mathbb{R}_d))$, $d = 1, 2$, be given. Further, let $(\mathcal{M}_d, \mathcal{B}(\mathcal{M}_d))$ be a measure space of probability measures on $(\mathbb{R}_d, \mathcal{B}(\mathbb{R}_d))$, where $\mathcal{B}(\mathcal{M}_d)$ is a Borel σ-algebra generated by the open sets with respect to the Prokhorov metric. Define

$$\mathcal{L}^{n,Q}(\cdot) = \int I_{\{Q_i^n(\cdot|x) \in \cdot\}} \nu_n(di, dx).$$

Definition. The sequence $\{Q^n\}_{n \geq 1}$ is said to be generalized weakly convergent to the random element (r.e.) Q (this is denoted by $Q^n \Rightarrow Q$), if

$$\mathcal{L}^{n,Q} \overset{w}{\to} \mathcal{L}^Q \quad \text{as} \quad n \to \infty,$$

where $\mathcal{L}^Q$ is the distribution of the random element Q.

In other words, $Q^n \Rightarrow Q$ means that

$$\int F(Q_i^n(\cdot \mid x)) \nu_n(di, dx) = \int F(\nu) \mathcal{L}^{n,Q}(d\nu) \to \int F(\nu) \mathcal{L}^Q(d\nu) \ \text{ as } \ n \to \infty$$

for each continuous bounded functional $F : \mathcal{M}_d \to \mathbb{R}_1$.

Note that if

$$\overline{Q}^n(\cdot) := \int Q_i^n(\cdot \mid x) \nu_n(di, dx) = \int \nu(\cdot) \mathcal{L}^{n,Q}(d\nu)$$

and

$$\overline{Q}(\cdot) = \int \nu(\cdot) \mathcal{L}^Q(\nu),$$

then

$$(Q^n \Rightarrow Q) \Rightarrow (\overline{Q}^n \overset{w}{\to} \overline{Q}).$$

Indeed, it is sufficient to take

$$F(\nu) = \int f(y) \nu(dy)$$

for $f \in C_b$ (the class of continuous bounded functions).

We fix the parameter θ and omit it.

Let Ψ^0 be the class of sequences $\psi = \{\psi^n\}_{n \geq 1}$ with $\psi^n \in \Psi_n^0$ for each $n \geq 1$ (see (3.24)).

For every $A \subset \Psi^0$, we introduce the set

$$\mathcal{H}(A) = \Big\{ \{H^n\}_{n \geq 1} \in \mathcal{H}_{\text{seq}} : \sup_n \int (H^n)^2 d\mu_n < \infty, \\ \int \psi^n H^n d\mu_n \to, \ \forall \psi \in A \Big\}.$$

The symbol "$a_n \to$" means that the sequence $\{a_n\}_{n \geq 1}$ has a finite limit.

Everywhere below we use the following abbreviations.

If $\{Q^{n,\psi}\}$ or $\{Q^{n,\psi_1,\psi_2}\}$ generalized weakly converges to some random element, we write, for definiteness,

$$Q^{n,\psi} \Rightarrow Q^\psi, \quad Q^{n,\psi_1,\psi_2} \Rightarrow Q^{\psi_1,\psi_2}.$$

Analogously, we write

$$\overline{Q}^{n,\psi} \overset{w}{\to} \overline{Q}^\psi, \quad \overline{Q}^{n,\psi_1,\psi_2} \overset{w}{\to} \overline{Q}^{\psi_1,\psi_2}.$$

Finally, we denote

$$\mathcal{L}^{n,Q^\psi} := \mathcal{L}^{n,\psi}, \quad \mathcal{L}^{Q^\psi} := \mathcal{L}^\psi.$$

Now we introduce the classes Ψ and $\mathcal{H}_\Psi$. Denote

$$Q_i^{n,\psi}(\cdot \mid x) = \int I_{\{z:\psi_i(z|x)\in\cdot\}} f_i^n(z \mid x) P_i^n(dz \mid x),$$

$$Q_i^{n,\psi^1,\psi^2}(\cdot \mid x) = \int I_{\{z:(\psi_i^{1,n}(z|x),\psi_i^{2,n}(z|x))\in\cdot\}} f_i^n(z \mid x) P_i^n(dz \mid x).$$

Let Ψ be a subset of Ψ^0 with the following properties:
(1) the sequence $\{(\psi^n)^2\}_{n\ge1}$ is uniformly integrable with respect to a sequence of measures $\{\mu_n\}_{n\ge1}$;
(2) $(\psi^1,\psi^2\in\Psi) \Rightarrow (\psi^1\in\Phi^{\psi^2})$, where $\Phi^{\psi^2} = \{\psi\in\Psi^0\colon\ Q^{n,\psi^2,\psi} \Rightarrow Q^{\psi^2,\psi}\}$;
(3) $l = \{l^n\}_{n\ge1}\in\Psi$;
(4) if $Q^{n,\psi}\Rightarrow Q$, then

$$\mathcal{L}^Q\Big\{\nu : \nu \text{ has a unique median}, \int y\,\nu(dy) = 0\Big\} = 1;$$

(5) $(\widetilde{\psi} \in \bigcap_{\psi\in\Psi} \Phi^\psi) \Rightarrow (\widetilde{\psi}\in\Psi)$.

Further, we suppose that

$$\mathcal{L}^l\{\nu : \nu \text{ is nondegenerate}, \ \nu(\{\mathrm{med}\}) = 0\} = 1.$$

Finally, define $\mathcal{H}_\Psi = \mathcal{H}(\Psi)$.

Remark 3.13. Show, for example, that the set

$$C = \{\nu : \nu \text{ has a unique median}\}$$

is measurable. Indeed, the functions

$$a(\nu) = \inf\{a : \nu(-\infty,a] \ge 1/2\}, \quad b(\nu) = \sup\{b : \nu[b,\infty) \ge 1/2\}$$

are measurable and $C = \{\nu : a(\nu) = b(\nu)\}$.

Remark 3.14. If $Q_i^{n,l}(\cdot \mid x)$ is symmetric with respect to zero, then $\beta(Q_i^{n,l}(\cdot \mid x), m) = 0$. In this case, in the definition of the class Ψ, the property $Q^{n,\psi^2,\psi} \Rightarrow Q^{\psi^2,\psi}$ can be replaced by

$$\overline{Q}^{n,\psi^2,\psi} \xrightarrow{w} \overline{Q}^{\psi^2,\psi}.$$

A similar remark is valid for property (4) as well.

Theorem 3.4. *Let, for each $n\ge1$, $\psi^{*,n}$ be a score function, constructed in Theorem 3.3. Then the sequence $\psi^* = \{\psi^{*,n}\}_{n\ge1}$ is $(\Psi,\mathcal{H}_\Psi)$-optimal.*

Proof. We prove some lemmas.

Lemma 3.4. *The sequence $\widetilde{\psi} := \{[l^n - \beta^n]_{-m}^m\}_{n\ge1}\in\Psi$.*

Proof. First, we prove that $\widetilde{\psi}\in\bigcap_{\psi\in\Psi}\Phi^\psi$. It is easy to verify that if ν^n and ν are some measures on $(\mathbb{R}_2,\mathcal{B}(\mathbb{R}_2))$ such that $\nu^n \xrightarrow{w} \nu$ and the first marginal ν_1 of ν has a unique median, then $\beta(\nu^n)\to\beta(\nu)$ as $n\to\infty$, where the functional $\beta(\cdot)$ is defined as a solution of the equation

$$\int [u-\beta]_{-m}^m \nu(du,dv) = 0. \tag{3.43}$$

Indeed, $\beta(\nu) = \beta(\nu_1)$, where $\beta(\nu_1)$ is the unique solution of Eq. (3.43) with ν_1 instead of ν.
If now

$$\underline{\beta} = \varliminf_{n\to\infty}\beta(\nu^n), \quad \overline{\beta} = \varlimsup_{n\to\infty}\beta(\nu^n),$$

then there exist subsequences, say $\{n'\}$ and $\{n''\}$, such that

$$\beta(\nu^{n'})\to\underline{\beta}, \quad \beta(\nu^{n''})\to\overline{\beta}.$$

Now, if we pass to the limit as $n' \to \infty$ (or $n'' \to \infty$) in Eq. (3.43), where we substitute $\nu^{n'}$ (or $\nu^{n''}$) for ν, then from the uniqueness of the solution of Eq. (3.43) we obtain $\underline{\beta} = \overline{\beta} = \beta(\nu_1) = \beta(\nu)$.

Introduce the mapping $h : \nu \to h(\nu)$ by the relation

$$h(\nu)(\cdot) = \int I_{\{([u-\beta(\nu)]_{-m}^{m}, v\in\cdot\}}\nu(du, dv),$$

and show that the mapping h is continuous (with respect to a weak convergence) at the point ν, with marginal ν_1, having a unique median. This immediately follows from the continuity of the functional $\beta(\cdot)$.

Indeed, for each continuous bounded function f on $\mathbb{R}_2$ we have

$$\int f(u, v)h(\nu^n)(du, dv) = \int f([u - \beta(\nu^n)]_{-m}^{m}, v)\nu^n(du, dv)$$
$$\to \int f([u - \beta(\nu)]_{-m}^{m}, v)\nu(du, dv) = \int f(u, v)h(\nu)(du, dv).$$

Let $\psi \in \Psi$. Obviously,

$$\beta^n = (\beta(Q_i^{n,l}(\cdot \mid x), m))_{1\le i\le n} = (\beta(Q_i^{n,l,\psi}(\cdot \mid x), m)_{1\le i\le n}$$

and $Q^{n,\widetilde{\psi},\psi} = h(Q^{n,l,\psi})$. Now the convergence

$$Q^{n,\widetilde{\psi},\psi} \Rightarrow Q^{\widetilde{\psi},\psi}$$

follows from the convergence

$$Q^{n,l,\psi} \Rightarrow Q^{l,\psi}.$$

Indeed, for any bounded continuous functional F on $\mathcal{M}_2$, the superposition $F(h)$ has the same properties.

It remains to verify property (4) from the definition of the class Ψ. For each $F \in C_b(\mathcal{M}_1)$, we have

$$\int F(Q_i^{n,\widetilde{\psi}}(\cdot \mid x))\nu_n(di, dx) = \int F(\widetilde{h}(Q_i^{n,l}(\cdot \mid x)))\nu_n(di, dx), \tag{3.44}$$

where the mapping $\widetilde{h} : \nu \to \widetilde{h}(\nu)$, $\nu \in \mathcal{M}_1$, given by

$$\widetilde{h}(\nu)(\cdot) = \int I_{\{[y-\beta(\nu)]_{-m}^{m}\in\cdot\}}\nu(dy)$$

is continuous. Note, in addition, that if ν has a unique median or if $\int y\nu(dy) = 0$, then $\widetilde{h}(\nu)$ has the same property. Now we have

$$\int F(Q_i^{n,\widetilde{\psi}}(\cdot \mid x))\nu_n(di, dx) = \int F(v)\mathcal{L}^{n,\widetilde{\psi}}(dv)$$
$$= \int F(\widetilde{h}(\nu))\mathcal{L}^{n,l}(d\nu) \to \int F(\widetilde{h}(v))\mathcal{L}^{1}(dv) = \int F(\nu)\mathcal{L}^{\widetilde{\psi}}(d\nu),$$

where $\mathcal{L}^{\widetilde{\psi}} = \mathcal{L}\{\widetilde{h} \mid \mathcal{L}^1\}$. Hence

$$\mathcal{L}^{\widetilde{\psi}}\Big\{\nu : \nu \text{ has a unique median, } \int y\nu(dy) = 0\Big\} = 1,$$

since $l \in \Psi$ and, therefore, $\mathcal{L}^1$ has the latter property. □

Corollary 3.2. *If* $m_n \to m > 0$, *then* $\psi = \{[l^n - \beta^n]_{-m_n}^{m_n}\}_{n\ge1} \in \Psi$.

Proof. It is sufficient to note that the function $\beta = \beta(\nu, m)$ is continuous in both arguments. □

Lemma 3.5. *The sequence* $\psi^* = \{\psi^{*,n}\}_{n\ge1} \in \Psi$ (*see* (3.27)).

Proof. It is sufficient to show that $m_n^* \to m^* > 0$. From Eq. (3.28) we have

$$1 = \frac{R^2(m_n^*)^2 + u(m_n^*)}{v(m_n^*)} \geq \frac{R^2(m_n^*)^2}{v(m_n^*)}$$

(the functions $u(m)$ and $v(m)$ are defined in (3.38) and (3.39)).

Hence

$$(m_n^*)^2 \leq \frac{v(m_n^*)}{R^2} \leq \frac{\int x^2 \overline{Q}^{n,l}(dx)}{R^2} \to \frac{\int x^2 \overline{Q}^{l}(dx)}{R^2} < \infty.$$

Therefore, the sequence $\{m_n^*\}$ is bounded. Denote $m_1 = \liminf\limits_{n\to\infty} m_n^*$, $m_2 = \limsup\limits_{n\to\infty} m_n^*$, and let $\{n'\}$ and $\{n''\}$ be subsequences such that $m_{n'}^* \to m_1$ and $m_{n''}^* \to m_2$. Rewrite Eq. (3.28) as follows:

$$R^2(m_n^*)^2 = \int\int F(m_n^*, y, z)\overline{Q}^{n,l,\beta}(dy, dz), \tag{3.45}$$

where

$$F(m, y, z) = [y - z]_{-m}^{m} y - ([y - z]_{-m}^{m})^2.$$

It is easy to see that in Eq. (3.45) we may pass to the limit as $n \to \infty$. Indeed, it is sufficient to show that

$$Q^{n,l,\beta} \Rightarrow Q^{l,\beta}. \tag{3.46}$$

For each continuous bounded functional F on $\mathcal{M}_2$, we have

$$\int F(\nu)\mathcal{L}^{n,l,\beta}(d\nu) = \int F(h(\nu))\mathcal{L}^{n,l}(d\nu) \to \int F(h(\nu))\mathcal{L}^{1}(d\nu) = \int F(\nu)\mathcal{L}^{1,\beta}(d\nu),$$

where the mapping $h : \mathcal{M}_1 \to \mathcal{M}_2$ is given by the relation

$$h(\nu)(du, dv) = \delta_{\beta(\nu)}(du)\nu(dv),$$

where $\delta_{\{a\}}$ is the Dirac measure at the point a, and hence $\mathcal{L}^{l,\beta} = \mathcal{L}(h \mid \mathcal{L}^l)$. Note that the mapping h is $\mathcal{L}^l$-a.s. continuous. Thus relation (3.46) is proved.

Passing to the limit in Eq. (3.45) first as $n' \to \infty$ and then as $n'' \to \infty$, we find that m_1 and m_2 are solutions of the equation

$$R^2 m^2 = \int\int F(m, y, z)\overline{Q}^{l,\beta}(dy, dz). \tag{3.47}$$

But this equation has a unique solution. Hence, $m_1 = m_2$. □

In the sequel, we need the following sets:

$$O_R^n = \Big\{H^n \in L_2(\mu_n) : H_i^n(z \mid x) \geq 0 \ \ (\mu_n\text{-a.s.}),\ \ \int H^n d\mu_n \leq R\Big\},$$
$$\overline{\mathcal{H}}_\Psi = \{\{\overline{H}^n\}_{n\geq1} : \overline{H}^n \in O_R^n \ \ \forall n \geq 1,\ \ \{H^n\}_{n\geq1} \in \mathcal{H}_\Psi\},$$

where

$$H_i^n(z \mid x) = \overline{H}_i^n(z \mid x) - \int \overline{H}_i^n(z \mid x) f_i^n(z \mid x) P_i^n(dz \mid x). \tag{3.48}$$

It is easy to see that there exists a one-to-one correspondence between $\mathcal{H}_\Psi$ and $\overline{\mathcal{H}}_\Psi$: for each $H \in \mathcal{H}_\Psi$, there exists $\overline{H} \in \overline{\mathcal{H}}_\Psi$ such that $D(\psi, H) = D(\psi, \overline{H})$ for each $\psi \in \Psi$, and vice versa. We write $H \sim \overline{H}$.

Consider the sequences $\psi = \{\psi^n\}_{n\geq1}$ for which $\sup\limits_{H\in\mathcal{H}_\Psi} D(\psi, H) < \infty$. We call such ψ *admissible.*

Lemma 3.6. *The sequence $\psi \in \Psi$ is admissible iff there exists a constant $c > 0$ such that*

$$\int I_{\{|y|>c\}} \overline{Q}^{\psi}(dy) = 0,$$

where

$$\overline{Q}^{n,\psi} \xrightarrow{w} \overline{Q}^{\psi}.$$

Proof. Assume that for each $c > 0$,

$$\lim_{n\to\infty} \int I_{\{\psi_i^n(z|x)>c\}} \mu_n(dz, di, dx) = \int I_{\{y>c\}} \overline{Q}^{\psi}(dy) > 0.$$

Consider the function

$$H^{n,c} = (H_i^{n,c}(z \mid x))_{1\le i\le n} := \left(R I_{\{\psi_i^n(z|x)>c\}} \left(\int I_{\{\psi_i^n(z|x)>c\}} d\mu^n \right)^{-1} \right)_{1\le i\le n}.$$

It will be proved in Lemma 3.9 that $H^c = \{H^{n,c}\}_{n\ge1} \in \overline{H}_{\Psi}$. Now we have

$$D(\psi, H^c) \ge \lim_{n\to\infty} \left[c^2 R^2 \left(\int I_{\{\psi^n>c\}} d\mu_n \right)^{-1} + \int (\psi^n)^2 d\mu_n \right] \left[\int \psi^n l^n d\mu_n \right]^{-2},$$

and the last expression tends to $+\infty$ as $c \to \infty$. □

Let $\psi \in \Psi$ be admissible and denote

$$\eta := \operatorname*{ess\,sup}_{\overline{Q}^{\psi}(dy)} |y|.$$

It follows from Lemma 3.6 that $\eta < \infty$. *Connect* with ψ the sequence $\widetilde{\psi}$,

$$\widetilde{\psi} = \{\widetilde{\psi}^n\}_{n\ge1} = \{[\psi^n - \beta^n]_{-\eta}^{\eta}\}_{n\ge1},$$

where $\beta^n = (\beta(Q_i^{n,\psi}(\cdot \mid x), \eta))_{i\le n}$.

Lemma 3.7. *$\widetilde{\psi} \in \Psi$, $\widetilde{\psi}$ is admissible, and $\psi \sim \widetilde{\psi}$, i.e.,*

$$D(\widetilde{\psi}, H) = D(\psi, H) \quad \forall H \in \mathcal{H}_{\Psi}.$$

Proof. It is sufficient to prove the last relation. We show, e.g., that

$$\left(\int \psi^n H^n d\mu_n \right)^2 - \left(\int \widetilde{\psi}^n H^n d\mu_n \right)^2 \to 0 \quad \text{as} \quad n \to \infty$$

(see (3.19) and (3.20)).

We have

$$\begin{aligned} \overline{\lim_{n\to\infty}} \left| \left(\int \psi^n H^n d\mu_n \right)^2 - \left(\int \widetilde{\psi}^n H^n d\mu_n \right)^2 \right| &\le \overline{\lim_{n\to\infty}} \left| \int (\psi^n + \widetilde{\psi}^n) H^n d\mu_n \right| \\ &\times \left| \int \{(\psi^n - [\psi^n]_{-\eta}^{\eta}) + ([\psi^n]_{-\eta}^{\eta} - [\psi^n - \beta^n]_{-\eta}^{\eta})\} H^n d\mu_n \right|. \end{aligned} \tag{3.49}$$

From the simple inequality $(a+b)^2 \le 2(a^2+b^2)$, the Schwartz inequality, and the definitions of Ψ and $\mathcal{H}_{\Psi}$, we have

$$\left(\int (\psi^n + \widetilde{\psi}^n) H^n d\mu_n \right)^2 \le \text{const.}$$

Hence, the right-hand side of inequality (3.49)

$$\le \text{const} \varlimsup_{n\to\infty} \Bigg[\bigg(\int (\psi^n - [\psi^n]_{-\eta}^{\eta})^2 \bigg)^{1/2}$$
$$+ \bigg(\int ([\psi^n]_{-\eta}^{\eta} - [\psi^n - \beta^n]_{-\eta}^{\eta})^2 d\mu_n \bigg)^{1/2} \Bigg] := \text{const} \varlimsup_{n\to\infty} (I_1^n + I_2^n).$$

From the uniform integrability of $\{(\psi^n)^2\}_{n\ge1}$ and the definition of η it immediately follows that $I_1^n \to 0$ as $n \to \infty$. Indeed,

$$(I_1^n)^2 = \int (y - [y]_{-\eta}^{\eta})^2 \overline{Q}^{n,\psi}(dy) \to \int (y - [y]_{-\eta}^{\eta})^2 \overline{Q}^{\psi}(dy)$$
$$= \int \int (y - [y]_{-\eta}^{\eta})^2 \nu(dy) \mathcal{L}^{\psi}(d\nu) = 0,$$

since

$$\mathcal{L}^{\psi}\{\nu : \nu([-\eta,\eta]^c) = 0\} = 1,$$

where A^c denotes the complement of the set A.

In fact,

$$0 = \overline{Q}^{\psi}([-\eta,\eta]^c) = \int \nu([-\eta,\eta]^c) \mathcal{L}^{\psi}(d\nu).$$

Further,

$$(I_2^n)^2 = \int \int ([y]_{-\eta}^{\delta} - [y-z]_{-\eta}^{\eta})^2 \overline{Q}^{n,\psi,\beta}(dy,dz)$$
$$\to \int \int ([y]_{-\eta}^{\eta} - [y - \beta(\nu)]_{-\eta}^{\eta})^2 \nu(dy) \mathcal{L}^{\psi}(d\nu) = 0,$$

since

$$\mathcal{L}^{\psi}\{\nu : \beta(\nu) = 0\} = \mathcal{L}^{\psi}\{\nu : \beta(\nu) = 0,\ \nu([-\eta,\eta]) = 1\}$$
$$= \mathcal{L}^{\psi}\Big\{\nu : \int [y]_{-\eta}^{\eta} \nu(dy) = 0,\ \nu([-\eta,\eta]) = 1\Big\}$$
$$= \mathcal{L}^{\psi}\Big\{\nu : \int y\,\nu(dy) = 0,\ \nu([-\eta,\eta]) = 1\Big\} = \mathcal{L}^{\psi}\Big\{\nu : \int y\,\nu(dy) = 0\Big\} = 1,$$

thanks to property (4) of the definition of the class Ψ. □

Lemma 3.8. *If $\psi \in \Psi$ and $\overline{Q}^{\psi}(\{y\}) = 0$, then $\psi^1 := I_{\{\psi > y\}} \in \overline{H}_{\Psi}$.*

Proof. For any $\psi^0 \in \Psi$, we have

$$\int \psi_i^{n,0}(z \mid x) \psi_i^{n,1}(z \mid x)\, \mu_n(dz, di, dx) = \int u I_{\{\nu > y\}} \overline{Q}^{n,\psi^0,\psi}(du, dv) \to,$$

as follows from the uniform integrability of $\{(\psi^{n,0})^2\}_{n\ge1}$ and the ($\overline{Q}^{\psi^0,\psi}$-a.s.) continuity of the function $uI_{\{v>y\}}$. □

Let $\psi \in \Psi$ and let $\widetilde{\psi}$ be connected with ψ.

Lemma 3.9. *For $\forall \varepsilon > 0$ there exists $\overline{H}^{\varepsilon} = \{\overline{H}^{n,\varepsilon}\}_{n\ge1}$ such that:*

(i) $\overline{H}^{\varepsilon} \in \overline{\mathcal{H}}_{\Psi}$;

(ii) $\varlimsup_{n\to\infty} \sup_{H \in O_R^n} D_n(\widetilde{\psi}^n, H)(D_n(\widetilde{\psi}^n, \overline{H}^{n,\varepsilon}))^{-1} \le 1 + \text{const} \cdot \varepsilon.$

Proof. Without loss of generality, we assume that

$$\overline{Q}^{\widetilde{\psi}}(\{\eta(1-\varepsilon)\}) = 0$$

and

$$\overline{Q}^{\widetilde{\psi}}((\eta(1-\varepsilon), +\infty)) > 0,$$

where $\eta = \operatorname*{ess\,sup}_{\overline{Q}^{\widetilde{\psi}}} |y|$. We set

$$\overline{H}^{n,\varepsilon} = RI_{\{\widetilde{\psi}^n > \eta(1-\varepsilon)\}} \left(\int I_{\{\widetilde{\psi}^n > \eta(1-\varepsilon)\}} d\mu_n \right)^{-1} := RI_{A_n} \left(\int I_{A_n} d\mu_n \right)^{-1},$$

where $A_n = \{(z,i,x) : \widetilde{\psi}_i^n(z \mid x) > \eta(1-\varepsilon)\}$. Define $H^{n,\varepsilon}$ by formula (3.48). Then

$$\begin{aligned} \overline{\lim_{n\to\infty}} \int (H^{n,\varepsilon})^2 d\widehat{\mu}_n &= \overline{\lim_{n\to\infty}}\, R^2 n^{-1} \sum_{i=1}^n \left[\int I_{A_n} f_i^n - \left(\int I_{A_n} f_i^n \right)^2 \right] \left(\int I_{A_n} d\mu_n \right)^{-2} \\ &\le \overline{\lim_{n\to\infty}}\, R^2 \left(\int I_{A_n} d\mu_n \right)^{-2} < \infty, \end{aligned}$$

since

$$\lim_{n\to\infty} \int I_{A_n} d\mu_n = \int I_{\{y > \eta(1-\varepsilon)\}} \overline{Q}^{\widetilde{\psi}}(dy) > 0.$$

Now (i) follows from Lemma 3.8.

To prove (ii), it suffices to verify that

$$\liminf_{n\to\infty} \frac{\left| \int \widetilde{\psi}^n \overline{H}^{n,\varepsilon} d\mu_n \right|}{\sup_{H\in O_R^n} \left| \int \widetilde{\psi}^n H\, d\mu_n \right|} \ge 1 - \varepsilon. \tag{3.50}$$

Indeed,

$$\begin{aligned} \sup_{H\in O_R^n} D_n(\widetilde{\psi}^n, H) &= D_n(\widetilde{\psi}^n, \overline{H}^{n,\varepsilon}) \\ &+ \frac{R^2 \left(\int \widetilde{\psi}^n \overline{H}^{n,\varepsilon} d\mu_n \right)}{\left(\int \widetilde{\psi}^n l^n d\mu_n \right)^2} \cdot \left(\frac{\sup_{H\in O_R^n} \left(\int \widetilde{\psi}^n H\, d\mu_n \right)^2}{\left(\int \widetilde{\psi}^n \overline{H}^{n,\varepsilon} d\mu_n \right)^2} - 1 \right), \end{aligned}$$

and (ii) follows from the definition of the classes Ψ and $\mathcal{H}_\Psi$.

But the left-hand side of expression (3.50)

$$\ge \lim_{n\to\infty} (\eta(1-\varepsilon))(\eta_n)^{-1} = (\eta(1-\varepsilon))\eta^{-1} = 1 - \varepsilon,$$

where $\eta_n = \operatorname*{ess\,sup}_{\overline{Q}^{n,\widetilde{\psi}}} |y|$ and, obviously, $\eta_n \to \eta$. $\square$

Remark 3.15. The essential point in the proof is that the relation $\eta_n \to \eta$ holds for $\widetilde{\psi}$, whereas, in general, it is not true for ψ.

Proof of Theorem 3.4. For each $\psi \in \Psi$ and $H \in \mathcal{H}_\Psi$, from the fixed-step optimization problem we have

$$D_n(\psi^{*,n}, H^n) \le \sup_{H\in\mathcal{H}_R^n} D_n(\psi^n, H) = \sup_{H\in O_R^n} D_n(\psi^n, H).$$

Let $\widetilde{\psi}$ be connected with ψ. Then from item (ii) of Lemma 3.9 we obtain

$$\varlimsup_{n\to\infty} \frac{D_n(\psi^{*,n}, H^n)}{D_n(\widetilde{\psi}^n, \overline{H}^{n,\varepsilon})} \le \varlimsup_{n\to\infty} \sup_{H\in O_R^n} \frac{D_n(\widetilde{\psi}^n, H)}{D_n(\widetilde{\psi}^n, \overline{H}^{n,\varepsilon})} \le 1 + \mathrm{const}\cdot\varepsilon,$$

and, therefore,

$$D(\psi^*, H) \le D(\widetilde{\psi}, \overline{H}^{\varepsilon})(1 + \mathrm{const}\cdot\varepsilon).$$

It remains to note that $\widetilde{\psi} \sim \psi$ and $\overline{H}^{\varepsilon} \sim H^{\varepsilon}$. $\square$

We denote $\overline{\beta}^n = \beta(Q_i^n(\cdot \mid x), m^*)$, where m^* is a solution of Eq. (3.47).

Corollary 3.3. *The sequence*

$$\overline{\psi}^* = \{\overline{\psi}^{*,n}\}_{n\ge 1} := \{[l^n - \overline{\beta}^n]_{-m^*}^{m^*}\}_{n\ge 1}$$

is optimal.

Corollary 3.4. *If there exists a CLAN estimator $T^* = T^{\psi^*}$, then it is $(\Psi, \mathcal{H}_\Psi)$-minimax.*

The proofs of the corollaries are obvious and we omit them.

3.2.4. Special models.

I.1. Independent identically distributed (i.i.d.) observations. Let $f_i^n(z,\theta \mid x) \equiv f(z,\theta)$ be a density (with respect to the Lebesgue measure), $\lambda_i^n(x,\theta) \equiv R$, $h_i^n(z,\theta \mid x) \equiv h(z,\theta)$ be a density with $\int h^2(z,\theta) f(z,\theta)\,dz < \infty$, and $H_i^n(z,\theta \mid x) \equiv h(z,\theta) - 1$, $l_i^n(z,\theta \mid x) \equiv l(z,\theta) = \dfrac{\partial}{\partial\theta} \ln f(z,\theta)$, $\psi_i^n(z,\theta \mid x) \equiv \psi(z,\theta)$ with

$$\int \psi(z,\theta) f(z,\theta)\,dz = 0,$$

$$\int \psi^2(z,\theta) f(z,\theta)\,dz < \infty, \quad I^n(\theta) \equiv nI(\theta),$$

$$I(\theta) = \int l^2(z,\theta) f(z,\theta)\,dz < \infty.$$

Then

$$D(\psi, H; \theta) = \frac{R^2\Big(\displaystyle\int \psi(z,\theta) H(z,\theta) f(z,\theta)\,dz\Big)^2 + \displaystyle\int \psi^2(z,\theta) f(z,\theta)\,dz}{\Big(\displaystyle\int \psi(z,\theta) l(z,\theta) f(z,\theta)\,dz\Big)^2},$$

$\psi^*(z,\theta) = [l(z,\theta) - \beta(\theta, m^*)]_{-m^*}^{m^*}$, where $\beta(\theta, m^*)$ is the solution of the equation

$$\int [l(z,\theta) - \beta]_{-m^*}^{m^*} f(z,\theta)\,dz = 0$$

and m^* satisfies the equation

$$R^2 m^2 = \int \Big\{ [l(z,\theta) - \beta(\theta, m)]_{-m}^{m} l(z,\theta) - ([l(z,\theta) - \beta(\theta,m)]_{-m}^{m})^2 \Big\} f(z,\theta)\,dz.$$

I.2. Independent, nonidentically distributed observations. For $i \geq 1$, let

$$X_i = \theta + \sigma_i \varepsilon_i, \quad X_0 = 0,$$

where $\theta \in \mathbb{R}_1$ is an unknown parameter, $\{\varepsilon_i\}_{i \geq 1}$ is an i.i.d. sequence of standard normal random variables (r.v.), and the sequence $\{\sigma_i\}_{i \geq 1}$ of numbers is such that $\sigma_i > 0$ and for each $y \in \mathbb{R}_1$,

$$\chi_n(y) := n^{-1} \sum_{i=1}^{n} I_{\{\sigma_i^{-2} < y\}} \to \chi(y) \quad \text{as} \quad n \to \infty.$$

We note that

$$l_i(X_i, \theta) = (X_i - \theta)/\sigma_i^2 \sim \mathcal{N}(0, \sigma_i^{-2}).$$

Assume that the function $f(y) = y$ is uniformly integrable with respect to χ_n. Then

$$n^{-1} I_n(\theta) = n^{-1} \sum_{i=1}^{n} \sigma_i^{-2} \to \int y \chi(dy).$$

(a) Let the contamination model be

$$Y_i^n = \theta + \sigma_i \varepsilon_i^n, \quad i \geq 1, \quad Y_0 = 0$$

with

$$\varepsilon_i^n = (1 - Z_i^n)\varepsilon_i + Z_i^n W_i, \quad 1 \leq i \leq n, \quad n \geq 1,$$

where $(\varepsilon_i, Z_i^n, W_i)_{1 \leq i \leq n}$ is the i.i.d. sequence of random vectors with mutually independent components,

$$P(Z_i^n = 1) = \lambda/n^{1/2}, \quad \lambda > 0 \text{ is a number,}$$

the density of the r.v. W_i is $h(z)$. Then for the density $f_i^{n,H}(z, \theta)$ of the r.v. Y_i^n we have

$$f_i^{n,H}(z, \theta)/\phi(\theta, \sigma_i, z) = 1 + n^{-1/2} H_i(z, \theta), \tag{3.51}$$

where

$$H_i(z, \theta) = \lambda \cdot \left(\frac{h((z - \theta)/\sigma_i)}{\phi((z - \theta)/\sigma_i)} - 1 \right).$$

Here $\phi(a, b, \cdot)$ is the density of the function of the normal distribution $\mathcal{N}(a, b^2, \cdot)$ with parameters (a, b), $\phi(u) = \phi(0, 1, u)$. Consider the class Ψ of all score functions of the form

$$\psi_i(z, \theta) = \psi((z - \theta)/\sigma_i^2),$$

where $\psi(u)$ is a continuous function,

$$|\psi(u)| \leq K|u|, \quad \int \psi(u/\sigma_i)\phi(u)\, du = 0,$$

$$\int \psi^2(u/\sigma_i)\phi(u)\, du < \infty, \quad K > 0 \text{ is a constant.}$$

Obviously, $(l_i)_{i \geq 1} \in \Psi$, where $l_i = \dfrac{\partial}{\partial \theta} \ln \phi \left(\dfrac{z - \theta}{\sigma_i^2} \right)$.

Let us show that all ergodicity conditions are satisfied. In particular,

$$\int (\psi)^2 d\widehat{\mu}_n = n^{-1} \sum_{i=1}^{n} \int (\psi((z - \theta)/\sigma_i^2))^2 \phi(\theta, \sigma_i, z)\, dz$$

$$= \int \int (\psi(u y^{1/2}))^2 \phi(u)\, du \chi_n(dy) \to \int \int (\psi(u y^{1/2}))^2 \phi(u)\, du \chi(dy) = \Gamma^{\psi}, \tag{3.52}$$

$$\int \psi l\, d\widehat{\mu}_n = \int \int \psi(u y^{1/2}) u y^{1/2} \phi(u)\, du \chi_n(dy) \to \int \psi(u y^{1/2}) u y^{1/2} \phi(u)\, du \chi(dy) = \gamma^{\psi}.$$

Let

$$\int (h(u)/\phi(u))^2 \phi(u)\, du < \infty.$$

Then

$$H = \{(H_i(z,\theta))_{1 \le i \le n}\}_{n \ge 1} \in \mathcal{H}_\Psi.$$

Obviously,

$$b_n^{\psi,H} = \int\int \psi(uy^{1/2}) H(u)\phi(u)\, du\chi_n(dy) \to \int\int \psi(uy^{1/2})H(u)\phi(u)\, du\chi(dy) = \beta^{\psi,H}, \tag{3.53}$$

where

$$H(u) = \lambda \cdot \left(\frac{h(u)}{\phi(u)} - 1\right).$$

From expressions (3.52) and (3.53) it is obvious that such a contamination is not full.

(b) Let the contamination model be as in (a), but $(Z_i^n)_{1 \le i \le n}$ and $(W_i)_{1 \le i \le n}$ be mutually independent sequences of independent nonidentically distributed r.v. with

$$P(Z_i^n = 1) = \lambda(\sigma_i^{-1})/n^{1/2},$$

and the density of the r.v. W_i be $h_i(z) = h(z, \sigma_i^{-1})$. Then $f_i^{n,H}$ has the form (3.51) with

$$H_i(z,\theta) = \lambda(\sigma_i^{-1})\left(\frac{h((z-\theta)/\sigma_i, \sigma_i^{-1})}{\phi((z-\theta)/\sigma_i)} - 1\right).$$

Denote by Ψ the set of all score functions of the form

$$\psi_i(z,\theta) = \psi((z-\theta)/\sigma_i, \sigma_i^{-1}), \quad i \le n,$$

where $\psi(u_1,u_2)$ is continuous, $|\psi(u_1,u_2)| \le K|u_1||u_2|$, and $K > 0$ is a constant.

It is easy to verify that

$$\overline{Q}^{n,\psi^1,\psi^2} \xrightarrow{w} \overline{Q}^{\psi^1,\psi^2}.$$

Indeed,

$$\begin{aligned}\overline{Q}^{n,\psi^1,\psi^2}(\cdot) &= n^{-1}\sum_{i=1}^n \int I_{\{(\psi^1(u,\sigma_i^{-1}),\psi^2(u,\sigma_i^{-1}))\in\cdot\}}\phi(u)\, du \\ &= \int\int I_{\{(\psi^1(u,y^{1/2}),\psi^2(u,y^{1/2}))\in\cdot\}}\, du\chi_n(dy),\end{aligned}$$

and for each $f \in C^b_{\mathbb{R}_2}$, we have

$$\begin{aligned}\int\int f(z_1,z_2)\overline{Q}^{n,\psi^1,\psi^2}(dz^1,dz^2) &= \int\int f(\psi^1(u,y^{1/2}),\psi^2(u,y^{1/2}))\phi(u)\, du\chi_n(dy) \\ &\to \int\int f(\psi^1(u,y^{1/2}),\psi^2(u,y^{1/2}))\phi(u)\, du\chi(dy) \\ &= \int\int f(z_1,z_2)\overline{Q}^{\psi^1,\psi^2}(dz_1,dz_2).\end{aligned}$$

A straightforward calculation shows that all ergodicity conditions are satisfied. Namely,

$$\begin{aligned}\Gamma^\psi &= \int\int (\psi(u,y^{1/2}))^2\phi(u)\, du\chi(dy), \\ \gamma^\psi &= \int\int (\psi(u,y^{1/2}))^2 uy^{1/2}\phi(u)\, du\chi(dy).\end{aligned}$$

Assume that the functions $\lambda(u_1)$ and $h(u_1, u_2)$ are continuous and

$$|\lambda(u_1)| \le \lambda |u_1|, \quad |h(u, u_1)| \le h(u)(1 + |u_1|),$$
$$\int h^2(u)/\phi(u)\, du < \infty, \quad \lambda > 0 \text{ is a constant.}$$

Then

$$H = \{(H_i(z,\theta)_{i\le n})\}_{n\ge 1} \in \mathcal{H}_\Psi.$$

Indeed,

$$\int \psi H d\widehat{\mu}_n = \int\int \psi(u, y^{1/2})\lambda(y^{1/2})h(u, y^{1/2})\, du\chi_n(dy)$$
$$\to \int\int \psi(u, y^{1/2})\lambda(y^{1/2})h(u, y^{1/2})\, du\chi(dy) = \beta^{\psi,H}.$$

We have obtained the full contamination and, therefore, we can conclude that the optimal score function ψ^* is

$$\psi_i^* = [(X_i - \theta)/(\sigma_i^2)]_{-m^*}^{m^*}, \quad i \ge 1,$$

where m^* is the solution of the equation

$$R^2 m^2 = \int\int ([uy^{1/2}]_{-m}^{m} uy^{1/2} - ([uy^{1/2}]_{-m}^{m})^2)\phi(u)\, du\chi(dy).$$

If the distribution χ is unknown, then, by virtue of Theorem 3.4, the optimal sequence is $\psi^* = \{(\psi_i^{*,n})_{1\le i\le n}\}_{n\ge 1}$, where

$$\psi_i^{*,n} = [(X_i - \theta)/(\sigma_i^2)]_{-m_n^*}^{m_n^*}, \quad 1 \le i \le n,$$

and m_n^* is the solution of the above equation with χ_n instead of χ. Note that $m_n^* \to m^*$ as $n \to \infty$.

II. The Markov Chain

1. Stationary ergodic Markov chain. Let $Y_1, ..., Y_k, Y_{k+1}, ..., Y_n$, $k \ge 1$, $n \ge k$, be observations of a stationary homogeneous ergodic Markov chain defined by the initial density $f_0(y_k, ..., y_1, \theta)$ and the transition density $f(z, \theta \mid y_{i-1}, ..., y_{i-k})$, $i > k$, where $\theta \in \Theta \subset \mathbb{R}_1$ is an unknown parameter.

Denote $x_0 = (y_k, ..., y_1)$, $x_i = y_{k+i}$, $i \ge 1$, and let

$$f_0^n(x_0, \theta) = f(x_0, \theta), \quad f_1^n(z, \theta \mid x) = f(z, \theta \mid x_0),$$
$$f_i^n(z, \theta \mid x) = f(z, \theta \mid x_{i-1}, ..., x_0^{(i)}) \quad \text{if } 1 \le i \le k,$$

where $x_0^{(i)} = (y_k, ..., y_i)$ and $f_i(z, \theta \mid x) = f(z, \theta \mid x_{i-1}, ..., x_{i-k})$, $i > k$.

Further, let

$$\Psi = \bigcup_{q \ge k} \Psi^q,$$

where Ψ^q is the class of sequences $\psi = \{(\psi_i^n)_{i\le n}\}_{n\ge 1}$ such that $\psi_i^n(z, \theta \mid x) = \psi(z, \theta \mid x_{i-1}, ..., x_{i-q})$, $i > q$, where ψ satisfies the usual conditions of integrability. Note that

$$l_i^n \equiv l_i(z, \theta \mid x) = l_i(z, \theta \mid x_{i-1}, ..., x_{i-k}), \quad i > k,$$

and, hence, $l \in \Psi^k$.

It is obvious that

$$|\langle \psi, H\rangle| \le R \sup |\psi| = \sup_{H \in \mathcal{H}^{q_1}} |\langle \psi, H\rangle|$$

for each $\psi \in \Psi^{q_1}$ and $H \in \mathcal{H}^{q_2}$, $q_1, q_2 \ge k$. Thus, we have

$$\sup_{H \in \mathcal{H}} |\langle \psi, H\rangle| = R \sup |\psi|$$

and

$$\inf_{\psi\in\Psi}\sup_{H\in\mathcal{H}} D(\psi,H;\theta)=\inf_{q\ge k}\inf_{\psi\in\Psi^q}\frac{\sup|\psi|^2+\|\psi\|^2}{(\langle\psi,l\rangle)^2},$$

where the inner product and the norm are considered in $L^2(\mathbb{R}_q,\mu)$, where

$$d\mu=f_{q+1}(x_{q+1},...,x_1)dx_{q+1}...dx_1$$

and f_{q+1} is the $(q+1)$-length stationary density.

But the optimal score function ψ^* for the standard optimization problem

$$\frac{\sup|\psi|^2+\|\psi\|^2}{(\langle\psi,l\rangle)^2}\underset{\psi\in\Psi^q}{\Rightarrow}\min$$

has the form

$$\psi^*=[l-\beta^*]_{-m^*}^{m^*},$$

where the pair (β^*,m^*) is the solution of Eqs. (3.26), (3.28), which does not depend on the index q.

This result involves the case of the stationary ergodic $AR(k)$ model (see Künsch, [57]):

$$Y_i=\sum_{l=1}^{k}\alpha_l(\theta)Y_{i-l}+\varepsilon_i,\quad i>k,$$

where $\theta\in\mathbb{R}_1$, $\alpha_l(\theta)$, $1\le l\le k$, is a known function, differentiable in θ, and $\{\varepsilon_i\}_{i\ge1}$ is the i.i.d. sequence with the density function $g(\cdot)$, $g>0$.

In this case we have

$$f_i(z;\theta\mid x_{i-1},...,x_{i-k})=g\biggl(z-\sum_{l=1}^{k}\alpha_l(\theta)x_{i-1}\biggr)$$

for $i>k$ and

$$l_i=-\lambda(\widetilde{\varepsilon}_i)\sum_{l=1}^{k}\dot{\alpha}_l(\theta)X_{i-1},$$

where

$$\dot{\alpha}_l=\frac{\partial}{\partial\theta}\alpha_l(\theta),\quad \lambda=\frac{\partial}{\partial z}\ln g(z),\quad \widetilde{\varepsilon}_1=\varepsilon_{k+i}.$$

Then Eqs. (3.26) and (3.28) take the form

$$\int\biggl[-\lambda(z)\sum_{l=1}^{k}\dot{\alpha}_l(\theta)x_{i-l}-\beta\biggr]_{-m}^{m}g(z)\,dz=0,$$

$$R^2m^2=\int\biggl(\int\biggl\{\biggl[-\lambda(z)\sum_{l=1}^{k}\dot{\alpha}_l(\theta)x_{k+1-l}-\beta(x_k,...,x_1,m,\theta)\biggr]_{-m}^{m}\biggl(-\lambda(z)\sum_{l=1}^{k}\dot{\alpha}_l(\theta)x_{k+1-l}\biggr)$$
$$-\biggl(\biggl[-\lambda(z)\sum_{l=1}^{k}\dot{\alpha}_l(\theta)x_{k+1-l}-\beta(x_k,...,x_1,m,\theta)\biggr]_{-m}^{m}\biggr)^2\biggr\}g(z)\,dz\biggr)f_k(x_k,...,x_1)\,dx_k...dx_1,$$

where $f_k(x_k,...,x_1)$ is a k-length stationary density.

2. Ergodic Markov chain. The consideration of this case is quite similar to that of the previous one. Note only that from ergodicity we have

$$\int F(Q_i^{n,\psi^1,\psi^2}(\cdot \mid x)(di, dx) = n^{-1}E\sum_{i=1}^{n} F(Q^{\psi^1,\psi^2}(\cdot \mid X_{i-1}, ..., X_{i-q}))$$
$$\to \int F(Q^{\psi^1,\psi^2}(\cdot \mid x_q, ..., x_1) f_q(x_q, ..., x_1)\, dx_q...dx_1,$$

where f_q is a q-length invariant density.

III. Stationary Ergodic $MA(1)$ Model

Let $..., Y_{-1}, Y_0, Y_1, ...$ be a stationary ergodic $MA(1)$ process, i.e.,

$$Y_i = v_i - \theta v_{i-1}, \quad |\theta| < 1,$$

where (v_i) is a double infinite sequence of i.i.d. r.v. with common density g,

$$Ev_i = 0, \quad Ev_i^2 < \infty.$$

Let $\Lambda := \dot{g}/g$ and

$$I_g = \int \Lambda^2 g < \infty.$$

Denote

$$X_0 = \overline{v}_0 = (v_0, v_{-1}, ...), \quad X_i = Y_i, \quad i \geq 1,$$

and let for $1 \leq i \leq n$, $P_\theta^{n,i}(\cdot \mid x_{i-1}, ..., x_0)$ be a regular conditional distribution of X_i with a given $\mathcal{F}_{i-1} = \sigma(X_{i-1}, ..., X_0)$. Then the density of $P_\theta^{n,i}$ with respect to the Lebesgue measure is

$$f_i(z, \theta \mid x) = g\left(z + \sum_{j=1}^{i-1} \theta^i x_{i-j} + \theta^i v_0\right). \tag{3.54}$$

Staab [92] considered a class of CLAN estimators $T_n(X_n, ..., X_0)$ such that

$$n^{1/2}(T_n - \theta) = n^{-1/2} \sum_{j=1}^{n} \psi_{\theta,j} + o_{P_\theta^n}(1)$$

with

$$P_\theta^n = \mathcal{L}\{X_n, ..., X_0\}, \quad \psi_{\theta,j} = \psi_\theta(\overline{v}_j),$$

where ψ_θ is a measurable mapping from $\prod_{-\infty}^{1} \mathbb{R}_1$ to $\mathbb{R}_1$ such that $\psi_\theta(\overline{v}_1)$ is a square-integrable r.v.,

$$E_\theta(\psi_\theta(\overline{v}_1 \mid \overline{v}_0) = 0, \quad E_\theta[\psi_\theta(\overline{v}_1)\overline{c}_1 \Lambda(v_1)] = 1, \quad \overline{v}_i = (v_i, v_{i-1}, ...),$$
$$\overline{c}_i = \sum_{j=0}^{\infty} -\theta^j v_{i-j-1} = \sum_{j=0}^{\infty} j\theta^{j-1} Y_{i-j}.$$

It is obvious that

$$\mathcal{L}\{n^{1/2}(T_n - \theta) \mid P_\theta^n\} \xrightarrow{\omega} \mathcal{N}(0, c(\psi_\theta)),$$

where $c(\psi_\theta) = E_\theta \psi_{\theta,1}^2$.

Now we briefly describe a contamination model proposed by Staab [92], the so-called submodel. Within this model, the shrinking contamination neighborhood of the basic measure P_θ^n contains all measures Q_θ^n such that $Q_\theta^{(n,0)} = P_\theta^{(n,0)}$, where $P_\theta^{(n,0)} = \mathcal{L}(\overline{v}_0)$, $Q_\theta^n = \prod_{i=0}^{n} Q_\theta^{(n,i)}$,

$$Q_\theta^{(n,i)} = (1 + n^{-1/2}\varepsilon_i f_i) P_\theta^{(n,i)}, \quad \varepsilon_i = \varepsilon(\overline{v}_{i-1}), \quad f_i = f(\overline{v}_i), \quad 1 \leq i \leq n,$$

where

$$\varepsilon(\cdot) \geq 0, \quad E_\theta \varepsilon(\overline{v}_0) \leq R, \quad R > 0, \quad \varepsilon(\overline{v}_0) \in L_\infty(P_\theta),$$
$$f(\cdot) \geq -1, \quad f(\overline{v}_0) \in L_\infty(P_\theta), \quad E_\theta(f(\overline{v}_1) \mid \overline{v}_0) = 0.$$

Obviously, $(Q_\theta^n) \lhd (P_\theta^n)$, and by the Le Cam third lemma,

$$\mathcal{L}\{n^{1/2}(T_n - \theta) \mid Q_\theta^n(f,\varepsilon)\} \xrightarrow{w} \mathcal{N}(b(\psi_\theta, f, \varepsilon), c(\psi_\theta)),$$

where

$$b(\psi_\theta, f, \varepsilon) = E_\theta \psi_{\theta,1} \cdot f_1 \varepsilon_1.$$

As usual, an asymptotic mean-square error is taken for the risk functional.

First of all, note that this scheme is a special case of the model given in the present work, if we restrict the consideration to the classes Ψ and $\mathcal{H}_\Psi$ of sequences $\psi = \{(\psi_i^n)_{i \leq n}\}_{n \geq 1}$ and $H = \{(H_i^n)_{i \leq n}\}_{n \geq 1}$ such that $\psi_i^n(X_i, ..., X_0) = \psi(\overline{v}_i)$ and $H_i^n(X_i, ..., X_0) = \varepsilon(\overline{v}_{i-1}) f(\overline{v}_i)$. Obviously, all ergodicity conditions are satisfied. Note only that in this case the functional $D(\psi, H; \theta)$ can be written in an explicit integral form and hence we obtain standard minimax problem (3.23) with the resulting function

$$\overline{\psi}_{i,\theta}^* = [\overline{c}_i \Lambda(v_i) - \beta_i^*]_{-m^*}^{m^*},$$

where m^* satisfies Eq. (3.28).

It remains to construct an estimator $T_n = T_n(X_n, ..., X_1)$ with $\{(\overline{\psi}_{i,\theta}^*)_{i \leq n}\}_{n \geq 1}$ in its asymptotic expansion. Staab proposed an approach based on the assumption of the approximation $\overline{\psi}_{i,\theta}^*$ by sufficiently smooth $\widehat{\psi}_{i,\theta}^n(X_i, ..., X_1)$ in such a way that: (1) the CLAN estimator T_n corresponding to $\widehat{\psi}_{i,\theta}^n$ can be constructed; (2) $\widehat{\psi}_{i,\theta}^n$ can be replaced by $\overline{\psi}_{i,\theta}^*$ in an asymptotic expansion of T_n. To illustrate this approach, Staab considered the case of standard normal innovations v_i. Note that in this case,

$$\beta_i^* = 0, \;\; \widehat{\psi}_i^n = \widehat{\psi}_i = [c_i \Lambda(\widehat{v}_i)]_{-m^*}^{m^*},$$

where

$$\widehat{c}_i = \sum_{j=0}^{i-1} j \theta^{j-1} X_{i-j}, \quad \widehat{v}_i = \sum_{j=0}^{i-1} \theta^{j-1} X_{i-j}, \quad i \geq 1.$$

Let us go back to the general model considered in this paper. Assume again that $X_0 = \overline{v}_0$, $X_i = Y_i$, $i \geq 1$, $f_i(z, \theta \mid x)$ are defined by (3.54), and all the objects are introduced in a standard way (cf. full model of [92]). For a correct definition of the class Ψ it is sufficient to show that $l = \{l^n\}_{n \geq 1} \in \Psi$, i.e., for any bounded Lipschitz function $F : \mathcal{M} \to \mathbb{R}_1$,

$$\int F(\nu) \mathcal{L}^{n,l}(d\nu) \to \int F(\nu) \mathcal{L}^1(d\nu), \tag{3.55}$$

where $\mathcal{L}^1$ is defined by relation (3.58) below. Recall that $l_i = \Lambda(v_i) c_i(\theta)$,

$$c_i(\theta) = \sum_{j=0}^{i-1} j \theta^{j-1} X_{i-j} + i \theta^i v_0.$$

Denote $\overline{l}_i = \Lambda(v_i) \overline{c}_i(\theta)$ and show that

$$R_n := n^{-1} \sum_{i=1}^{n} E|F(Q_i^l(\cdot \mid x)) - F(Q_i^{\overline{l}}(\cdot \mid x))| \to 0 \text{ as } n \to \infty, \tag{3.56}$$

where

$$Q_i^{\overline{l}}(\cdot \mid x) = \int I_{\{\Lambda(v)\overline{c}_i \in \cdot\}} g(v)\, dv.$$

Indeed,

$$R_n \le \text{const} \cdot n^{-1} \sum_{i=1}^{n} E d_{BL}(Q_i^{\overline{l}}(\cdot \mid x), Q_i^{l}(\cdot \mid x)) \le \text{const} \cdot n^{-1} \sum_{i=1}^{n} E|c_i - \overline{c}_i| \int |\Lambda(v)| g(v)\, dv \to 0.$$

Here $d_{BL}(\rho_1, \rho_2)$ denotes a bounded Lipschitz metric on $\mathcal{M}$ (see, e.g., [88]). From ergodicity we have

$$\lim_{n\to\infty} n^{-1} E_\theta \sum_{i=1}^{n} F(Q_i^{\overline{l}}(\cdot \mid x)) = EF(Q_1^{\overline{l}}(\cdot \mid \overline{Y}_0)) = \int F(\nu)\mathcal{L}^1(d\nu), \tag{3.57}$$

where $\overline{Y}_0 = (Y_0, Y_{-1}, ...)$ and

$$\mathcal{L}^l\{\cdot\} = \int I_{\{Q_1^{\overline{l}}(\cdot|\overline{Y}_0)\in\cdot\}} P_\theta(d\overline{Y}_0). \tag{3.58}$$

Now (3.55) follows from (3.56) and (3.57). By Theorem 3.4, the optimal sequence $\psi^* = \{\psi^{*,n}\}_{n\ge1}$ is defined by the relation $\psi_i^{*,n} = [l_i - \beta_i^*(m_n^*)]_{-m_n^*}^{m_n^*}$, where m_n^* satisfies Eq. (3.28). Moreover, $m_n^* \to m^*$, where m^* satisfies Eq. (3.47). Hence (see Corollary 3.3), $\widetilde{\psi}_i^* = [l_i - \beta_i^*(m^*)]_{-m^*}^{m^*}$ is also optimal. Since $\widetilde{\psi}_i^*$ contains an unobservable variable v_0, it becomes necessary to construct a sufficiently smooth function $\widehat{\psi}_i$ depending only on the real observations $(Y_1, ..., Y_n)$ such that

$$n^{-1/2} \sum_{i=1}^{n} |\widetilde{\psi}_i^* - \widehat{\psi}_i| = o_{P_\theta^n}(1).$$

Let $\widehat{\psi}_i = [\Lambda(\widehat{v}_i)\widehat{c}_i - \widehat{\beta}_i]_{-m^*}^{m^*}$, where $\widehat{\beta}_i$ is a solution of the equation

$$\int [\widehat{c}_i \Lambda(v) - \beta]_{-m^*}^{m^*} g(v)\, dv = 0.$$

We have

$$|\psi_i^* - \widehat{\psi}_i| \le |\Lambda(\widehat{v}_i)\widehat{c}_i - \Lambda(v_i)c_i| + |\beta_i^* - \widehat{\beta}_i|.$$

Assume that the function Λ is such that

$$n^{-1/2} \sum_{i=1}^{n} |\Lambda(\widehat{v}_i)\widehat{c}_i - \Lambda(v_i)c_i| = o_{P_\theta^n}(1)$$

(e.g., satisfies the Lipschitz condition).

Now we have to show that

$$n^{-1/2} \sum_{i=1}^{n} |\beta_i^* - \widehat{\beta}_i| = o_{P_\theta^n}(1).$$

It suffices to verify that if $\beta(c)$ is a solution of the equation

$$\int [c\Lambda(v) - \beta]_{-m}^{m} g(v)\, dv = 0,$$

then the function $\beta(c)$ has a Lipschitz property. With the latter in view, we calculate the derivative $\beta'(c)$ and obtain

$$\beta'(c) = \frac{\displaystyle\int \Lambda(v) I_{\{|c\Lambda(v)-\beta|\le m\}} g(v)\, dv}{\displaystyle\int I_{\{|c\Lambda(v)-\beta|\le m\}} g(v)\, dv}.$$

It is easy to verify that $\beta'(c)$ is a continuous bounded function. Consequently,

$$|\beta_i^* - \widehat{\beta}_i| \le \text{const}\, |c_i - \widehat{c}_i|$$

and the desirable convergence holds.

Let $\widehat{\theta}_n$ be a $\sqrt{n}$-consistent estimator. For example, we may take the least-square estimator defined by the equation

$$\sum_{i=1}^{n} \widehat{v}_i(\theta)\widehat{c}_i(\theta) = 0.$$

Define the estimator T_n^* by the relation

$$T_n^* = \widehat{\theta}_n + \frac{\widehat{\psi}_i(\widehat{\theta}_n)}{\gamma(\widehat{\theta}_n)},$$

where $\gamma(\theta) = P_\theta^n\text{-}\lim\limits_{n\to\infty} n^{-1} \sum\limits_{i=1}^{n} \psi_i^* l_i$ (see Subsection 3.2.5).

Then Λ is a sufficiently smooth function with respect to θ and the following asymptotic expansion holds:

$$\begin{aligned}\sqrt{n}(T_n^* - \theta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\widehat{\psi}_i}{\gamma(\theta)} + o_{P_\theta^n}(1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\widetilde{\psi}_i^*}{\gamma(\theta)} + o_{P_\theta^n}(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\overline{\psi}_i^*}{\gamma(\theta)} + o_{P_\theta^n}(1)\end{aligned}$$

(see also Subsection 3.2.5 below).

IV. Innovation Contamination

One possible scheme for the realization of the contamination

$$f_i^{n,H} = f_i^n(1 + n^{-1/2} H_i^n) \tag{3.59}$$

is as follows.

Let $\rho = \{\rho_i\}_{i\geq 1}$, $\tau = \{\tau_i\}_{i\geq 1}$, and $\sigma = \{\sigma_i\}_{i\geq 1}$ be mutually independent sequences of i.i.d. r.v. Further, let the sequences of measurable functions

$$\{(u_i^n(x_{i-1}, ..., x_0, y, \theta),\ \ w_i^n(x_{i-1}, ..., x_0, y, \theta),\ \ d_i^n(x_{i-1}, ..., x_0, y))_{i\leq n}\}_{n\geq 1}$$

be such that the distributions of the random variables $u_i^n(x_{i-1}, ..., x_0, \rho_i, \theta)$ and $w_i^n(x_{i-1}, ..., x_0, \tau_i, \theta)$ possess the densities $f_i^n(z, \theta \mid x)$ and $h_i^n(z, \theta \mid x)$, respectively, and the function d_i^n takes only values 0 and 1, and

$$P\{(d_i^n(x_{i-1}, ..., x_0, \sigma_i) = 1\} = n^{-1/2} \lambda_i^n(x_{i-1}, ..., x_0).$$

Let

$$\begin{aligned}Y_i^n &= (1 - d_i^n(Y_{i-1}^n, ..., Y_0, \sigma)) u_i^n(Y_{i-1}^n, ..., Y_0^n, \rho_i, \theta) \\ &\quad + d_i^n(Y_{i-1}^n, ..., Y_0^n, \rho_i) w_i^n(Y_{i-1}^n, ..., Y_0^n, \tau_i, \theta), \quad 1 \leq i \leq n.\end{aligned} \tag{3.60}$$

Then the density $f_i^{n,H}$ of the conditional distribution Y_i^n with a given $Y_{i-1}^n = x_{i-1}, ..., Y_0^n = x_0$ has the form (3.59), where

$$H_i^n(z, \theta \mid x) = \lambda_i^n(x)(h_i^n(z, \theta \mid x) - f_i^n(z, \theta \mid x))/f_i^n(z, \theta \mid x).$$

Obviously, Y_i^n can be written as

$$Y_i^n = (1 - Z_i^n) X_i^n + Z_i^n W_i^n, \quad 1 \leq i \leq n. \tag{3.61}$$

Relation (3.61) coincides in its form with the replacement model in the sense of Martin and Yohai [70], but it has a somewhat different meaning, since the triple (X, W, Z) of the process cannot be defined a priori as is assumed in the definition of the replacement model. It includes the innovation contamination in many models, in particular, in ARMA models.

1. $AR(1)$ **model.** We assume in (3.60) that $u_i^n = \theta x_{i-1} + y$, $|\theta| < 1$.

(a) Let

$$x_0 = 0, \quad P\{d_i^n = 1\} = \lambda n^{-1/2}, \quad \lambda > 0,$$
$$w_i^n := \theta x_{i-1} + y.$$

Then

$$Y_i^n = \theta Y_{i-1}^n + (1 - d_i^n)\rho_i + d_i^m \tau_i = \theta Y_{i-1}^n + \widehat{\rho}_i^n, \quad Y_0 = 0,$$

corresponds to the innovation contamination for the $AR(1)$ process defined by the relation

$$X_i = \theta X_{i-1} + \rho_i, \quad X_0 = 0.$$

Let g and h be densities of r.v. ρ_i and τ_i, respectively. Then

$$f_i^{n,H}(z, \theta \mid x_{i-1}, ..., x_1) = g(z - \theta x_{i-1})(1 + n^{-1/2} H(z, \theta \mid x_{i-1})),$$

where

$$H(z, \theta \mid x_{i-1}) = \lambda \frac{h - g}{g}(z - \theta x_{i-1}).$$

Obviously, this form of contamination is not full for score functions of the form $\psi(x_i - \theta x_{i-1}, x_{i-1})$, to which the maximum likelihood function

$$l(z, \theta \mid x_{i-1}) = \Lambda(z - \theta x_{i-1}) x_{i-1}, \quad \Lambda := \dot{g}/g$$

belongs.

(b) Let

$$x_0 = 0, \quad P\{d_i^n = 1\} = \lambda(x_{i-1}) n^{-1/2},$$
$$w_i^n(x_{i-1}, ..., y, \theta) := \theta x_{i-1} + \overline{w}(x_{i-1}, y).$$

In this case,

$$Y_i^n = \theta Y_{i-1}^n + \widetilde{\rho}_i^n, \quad Y_0 = 0,$$

where $\widetilde{\rho}_i^n = (1 - d_i^n)\rho_i + d_i^n \overline{w}_i$.

Here

$$H(z, \theta \mid x_{i-1}) = \lambda(x_{i-1}) \frac{h(z - \theta x_{i-1} \mid x_{i-1}) - g(z - \theta x_{i-1}}{g(z - \theta x_{i-1})}$$

and $h(\cdot \mid x_{i-1})$ is the density of the r.v. $\overline{w}(x_{i-1}, \tau_i)$. This form of contamination is full for the score function of type $\psi(x_i - \theta x_{i-1}, x_{i-1})$.

(c) Let

$$x_0 = 0, \quad P\{d_i^n = 1\} = \lambda(x_{i-1}, ..., x_{(i-q)\wedge 0}) n^{-1/2},$$
$$w_i^n = \theta x_{i-1} + \overline{w}(x_{i-1}, ..., x_{(i-q)\wedge 0}, y), \quad q \geq 1.$$

In this case, we again obtain the full contamination for score functions of the type $\psi(x_i - \theta x_{i-1}, x_{i-1}, ..., x_{(i-q)\wedge 0})$.

(d) If d_i^n and w_i^n are of the general form, then we obtain the full contamination with respect to the general form of score functions.

2. $ARMA(p,q)$ **model.** For $1 \le i \le n$, let

$$u_i^n(x_{i-1},...,x_0,y,\theta) := -\sum_{k=1}^{p}\sum_{j=1}^{i-1}\beta_k(\theta)\gamma_i(\theta)\widetilde{x}_{i-j-k} - \sum_{r=0}^{q-1}\chi_{r_i}(\theta)\overline{x}_{-r} + y,$$

where

$$x_0 = (\overline{x}_0, \overline{x}_{-1}, ...), \quad \widetilde{x}_m = x_m, \quad m \ge 1,$$

and

$$\widetilde{x}_m = \sum_{l=0}^{\infty}\sum_{k=0}^{q}\delta_l(\theta)\alpha_k(\theta)\overline{x}_{m-l-k}, \quad m \le 0,$$

where

$$\beta_0(\theta) = \alpha_0(\theta) = 1, \quad \chi_{r_i}(\theta) = \sum_{k=0}^{r}\gamma_{i+r-k}(\theta)\alpha_k(\theta).$$

The sequences $(\gamma_i(\theta))_{i\ge0}$ and $(\delta_i(\theta))_{i\ge0}$ are uniquely defined in a standard way by the vectors $(\alpha_i(\theta), 1 \le i \le q)$ and $(\beta_i(\theta), 1 \le i \le p)$, respectively (see [92], p. 1.5, (2.8), p. 1.6 (2.10)). Then with arbitrary d_i^n and w_i^n and with $\rho_i = v_i$, $i \ge 1$, $Y_0 = (v_0, v_{-1}, ...)$, Eq. (3.60) corresponds to the full contamination of the innovation $(v_j, j \ge 1)$ for the $ARMA(p,q)$ process defined by the relation

$$\sum_{j=0}^{p}\beta_j(\theta)X_{i-j} = \sum_{j=0}^{q}\alpha_i(\theta)v_{i-j}, \quad i \ge 1,$$

or an equivalent relation

$$X_i = -\sum_{k=1}^{p}\sum_{j=1}^{i-1}\beta_k(\theta)\gamma_i(\theta)X_{i-j-k} - \sum_{r=0}^{q-1}\chi_{r_i}(\theta)v_{-r} + v_i, \quad i \ge 1,$$

where

$$X_m = \sum_{l=0}^{\infty}\sum_{k=0}^{q}\delta_l(\theta)\alpha_k(\theta)v_{m-l-k}, \quad m \le 0.$$

Indeed, in this case for $i \ge 1$ we have

$$\begin{aligned} Y_i^n = &-\sum_{k=1}^{p}\sum_{j=1}^{i-1}I_{\{i-j-k\ge1\}}\beta_k(\theta)\gamma_i(\theta)Y_{i-j-k}^n \\ &-\sum_{k=1}^{p}\sum_{j=1}^{i-1}I_{\{i-j-k\le0\}}\beta_k(\theta)\gamma_i(\theta)X_{i-j-k} - \sum_{r=0}^{q-1}\chi_{r_i}(\theta)v_{-r} + \widetilde{v}_i^n = \widetilde{u}_i^n + \widetilde{v}_i^n, \end{aligned}$$

where $\widetilde{v}_i^n = (1-d_i^n)v_i + d_i^n(w_i - \widetilde{u}_i^n)$.

For the $AR(1)$ case, we illustrate the scope of application of the theory developed in this work.

Suppose that $H(z,\theta \mid x_{i-1})$ are such that they define contiguous alternatives, i.e., $(P_\theta^{n,Y}) \lhd (P_\theta^n)$. Then in case (a), the asymptotic bias is

$$\beta^{\psi,H} = \int\int \psi(z,y)H(z)g(z)\,dz\pi(dy),$$

where $\pi(\cdot)$ is the invariant measure corresponding to the process X, which leads to a minimax problem whose solution, obviously, cannot be the Huber function. But in case (b),

$$\beta^{\psi,H} = \int\int \psi(z,y)H(z \mid y)g(z)\,dz\pi(dy);$$

therefore, this case provides the full contamination and hence the Huber function is optimal in the given class Ψ.

Since in cases (c) and (d) we have full contamination for the given classes of score functions, the general theory allows one to conclude that in these cases the Huber function is also optimal.

It should be noted that while in cases (a), (b), and (c) the risk functional can be written in an explicit form involving the inner products and norm in an appropriately chosen L_2-space, in case (d) the nominal conditional distribution defines the stationary ergodic $AR(1)$ process, $D(\psi, H; \theta)$ cannot be written in an explicit form, and the developed theory should be applied in its full capacity.

3.2.5. A method of constructing the optimal CLAN estimators. Let $\psi = \{\psi^n\}_{n\geq 1} \in \Psi$ be a sequence of score functions. Let $\{\overline{T}_n\}_{n\geq 1}$ be some $\sqrt{n}$-consistent estimator of the unknown parameter θ, i.e.,

$$n^{1/2}(\overline{T}_n - \theta) = O_{P_\theta^n}(1). \tag{3.62}$$

Consider the one-step approximation procedure

$$T_n^\psi = \overline{T}_n + \frac{n^{-1}L_n(\overline{T}_n)}{\gamma^\psi(\overline{T}_n)}, \tag{3.63}$$

where $L_n(\theta) = \sum\limits_{i=1}^n \psi_i^n(x_i, \theta \mid x)$ and $\gamma^\psi(\theta)$ is defined in (c.3) of the definition of the CLAN estimator.

Assume that for each $n \geq 1$ the function $\theta \mapsto \psi_i^n(z, \theta \mid x)$ is θ-continuously differentiable $\mu_n(dz, di, dx)$-a.s. with the derivatives $\dot{\psi}_i^n(z, \theta \mid x)$ and $\dot{\psi} = \{\dot{\psi}^n\}_{n\geq 1} \in \Psi$.

Everywhere below, for any function $\varphi(x, \theta)$ we denote $\dot{\varphi}(x, \theta) = \dfrac{\partial}{\partial\theta}\varphi(x, \theta)$.

Further, denote $\dot{L}_n(\theta) = \sum\limits_{i=1}^n \dot{\psi}_i^n(x_i, \theta \mid x)$, $B(n, u, \theta) := \{y\colon\ |y - \theta| \leq n^{-1/2}u\}$ and suppose that the following conditions are satisfied: for each $\theta \in \Theta$, $\rho > 0$, $0 < u \leq K$, $K > 0$,

$$\varlimsup_{n\to\infty} P_\theta^n\Big\{\sup_{B(n,u,\theta)} |n^{-1}\dot{L}_n(y) + \gamma^\psi(\theta)| > \rho\Big\} = 0, \tag{3.64}$$

$$\sup_{B(n,u,\theta)} |\gamma^\psi(y) - \gamma^\psi(\theta)| \to 0 \quad \text{as } n \to \infty. \tag{3.65}$$

Proposition 3.2. *If the above conditions are satisfied, then the estimator* $\{T_n^\psi\}_{n\geq 1}$ *constructed in* (3.63) *is CLAN.*

Proof. It is sufficient to show that

$$T_n^\psi = \theta + \frac{n^{-1}\sum\limits_{i=1}^n \psi_i^n(x_i, \theta \mid x)}{\gamma^\psi(\theta)} + o_{P_\theta^n}(n^{-1/2}). \tag{3.66}$$

Now we rewrite Eq. (3.63) as follows:

$$\begin{aligned} n^{1/2}(T_n^\psi - \theta) &= \frac{n^{1/2}(\overline{T}_n - \theta)(\gamma^\psi(\overline{T}_n) - \gamma^\psi(\theta))}{\gamma^\psi(\overline{T}_n)} \\ &+ \frac{n^{1/2}(\overline{T}_n - \theta)[n^{-1}(L_n(\overline{T}_n) - L_n(\theta))(\overline{T}_n - \theta)^{-1} + \gamma^\psi(\theta)]}{\gamma^\psi(\overline{T}_n)} + \frac{n^{-1}L_n(\theta)}{\gamma^\psi(\overline{T}_n)}. \end{aligned} \tag{3.67}$$

The first term on the right-hand side of Eq. (3.67) tends to zero by (3.62) and (3.65) while the last term – by the ergodicity condition. Further, if $L_n(\theta)$ is asymptotically continuously differentiable with a derivative $\gamma^\psi(\theta)$ (this means that $\forall\theta \in \Theta$, $\forall\rho > 0$, $\forall u$, $0 < u \leq K$, $K > 0$,

$$\varlimsup_{n\to\infty} P_\theta^n\Big\{\sup_{B(n,u,\theta)} \Big|\frac{1}{n}\frac{L_n(y) - L_n(\theta)}{y - \theta} + \gamma^\psi(\theta)\Big| > \rho\Big\} = 0), \tag{3.68}$$

then the second term in Eq. (3.67) tends to zero and (3.63) is satisfied. It remains to note that (3.64) is a simple sufficient condition for (3.68). $\square$

Now we give another sufficient condition for $T^{\psi^*} = \{T_n^{\psi^*}\}_{n\geq 1}$ to be a CLAN estimator, where $\psi^* = \{\psi^{*,n}\}_{n\geq 1}$ is the optimal score sequence.

Suppose that $\forall\theta \in \Theta$, $\forall\rho > 0$, $0 < u \leq K$, $K > 0$,

$$\overline{\lim_{n\to\infty}}\, P_\theta^n\Big\{\sup_{B(n,u,\theta)} n^{-1}|\dot{L}_n^*(y) - \dot{L}_n^*(\theta)| > \rho\Big\} = 0, \tag{3.69}$$

where $\dot{L}_n^*(u) = \sum_{i=1}^n \dot{\psi}_i^{*,n}(x_i, u \mid x)$, $\forall u \in \Theta$.

Proposition 3.3. *If conditions* (3.64), (3.65), *and* (3.69) *are satisfied, then the estimator* $T^{\psi^*} = \{T_n^{\psi^*}\}_{n\geq 1}$ *constructed in* (3.63) *with* $L_n^*(\theta)$ *instead of* $L_n(\theta)$ *and* γ^{ψ^*} *instead of* γ^ψ *is CLAN.*

Proof. We have

$$\begin{aligned}\dot{L}_n^*(\theta) &= \sum_{i=1}^n\Big[\dot{\psi}_i^{*,n}(x_i,\theta \mid x) - \int \dot{\psi}_i^{*,n}(z,\theta \mid x) f_i^n(z,\theta \mid x) P_i^n(dz \mid x)\Big] \\ &\quad + \sum_{i=1}^n \int \dot{\psi}_i^{*,n}(z,\theta \mid x) f_i^n(z,\theta \mid x) P_i^n(dz \mid x) := \sum_{i=1}^n m_i^n + \sum_{i=1}^n a_i^n \\ &:= M_1^n(\theta) + A_1^n(\theta),\end{aligned} \tag{3.70}$$

where for each t, $0 \leq t \leq 1$,

$$M_t^n(\theta) = \sum_{i=1}^{[nt]} m_i^n, \quad A_t^n(\theta) = \sum_{i=1}^{[nt]} a_i^n, \quad M_1^n(\theta) = M_t^n(\theta)|_{t=1}, \quad A_1^n = A_1^n(\theta)|_{t=1}.$$

Introduce the abbreviation

$$\int \varphi_i^n(z,\theta \mid x) f_i^n(z,\theta \mid x) P_i^n(dz \mid x) = \int \varphi^n f^n$$

for any function $\varphi^n = \varphi_i^n(z,\theta \mid x)$.

Now we observe that $\int \psi^{*,n} f^n = -\int \psi^{*,n} l^n f^n$ and

$$A_1^n(\theta) = -\sum_{i=1}^n \int \psi_i^{*,n} l_i^n f_i^n. \tag{3.71}$$

Further, note that $(M_t^n(\theta))_{0\leq t\leq 1}$ is a P_θ^n-martingale with the square characteristic

$$\langle M^n(\theta)\rangle_1 = \sum_{i=1}^n \Big[\int (\dot{\psi}_i^{*,n})^2 f_i^n - \Big(\int \psi_i^{*,n} l_i^n f_i^n\Big)^2\Big].$$

But $n^{-1}\sum_{i=1}^n \int (\dot{\psi}_i^{*,n})^2 f_i^n$ converges, since $\dot{\psi}^* \in \Psi$ and

$$n^{-1}\sum_{i=1}^n \Big(\int \psi_i^{*,n} l_i^n f_i^n\Big)^2 \leq \text{const}\cdot n^{-1}\sum_{i=1}^n \int (l_i^n)^2 f_i^n \to,$$

since $|\psi^{*,n}| \leq m_n^* \to m^* < \infty$ by the construction. Hence

$$\lim_{N\to\infty}\overline{\lim_{n\to\infty}}\, P_\theta^n\{n^{-1}\langle M^n(\theta)\rangle_1 \geq N\} = 0,$$

and we conclude that $n^{-1}M_1^n(\theta) \overset{P_\theta^n}{\to} 0$.

Now (3.70) and (3.71) imply that for each $\theta \in \Theta$ and $\rho > 0$,

$$\overline{\lim_{n\to\infty}} P_\theta^n\Big\{|n^{-1}\dot{L}_n^*(\theta) + \gamma^{\psi^*}(\theta)| > \rho\Big\} = 0, \tag{3.72}$$

which leads to the relation $\forall\theta \in \Theta$, $\forall\rho > 0$, $\forall u$, $0 < u \le K$, $K > 0$,

$$P_\theta^n\Big\{\sup_{B(n,u,\theta)} |n^{-1}\dot{L}_n^*(y) + \gamma^{\psi^*}(\theta)| > \rho\Big\} \le P_\theta^n\Big\{\sup_{B(n,u,\theta)} n^{-1}|\dot{L}_n^*(y) - \dot{L}_n^*(\theta)| > \frac{\rho}{2}\Big\}$$
$$+P_\theta^n\Big\{|n^{-1}\dot{L}_n^*(\theta) + \gamma^{\psi^*}(\theta)| > \frac{\rho}{2}\Big\} \to 0 \quad \text{as} \;\; n \to \infty,$$

and the desirable follows from Proposition 3.2. □

As we have seen above, it is necessary to study the question on the θ-differentiability of the function $\psi^{*,n}(z, \theta \mid x)$. Let us investigate this question based on the implicit function theorem and Eqs. (3.26) and (3.28).

Denote $\Phi(i, x, \beta, m, \theta) := \int [y-\beta]_{-m}^m Q_i(dy, \theta \mid x)$ (index n is fixed and omitted here) and let $\beta_i(x, m, \theta)$ be a solution of the equation

$$\Phi(i, x, \beta, m, \theta) = 0. \tag{3.73}$$

Using the implicit function theorem, we find that there exist β'_m and $\dot{\beta}(= \beta'_\theta)$, and

$$\begin{aligned} \beta'_m(i, x, m, \theta) &= -\frac{\Phi'_m(i, x, \beta, m, \theta)}{\Phi'_\beta(i, x, \beta, m, \theta)}\bigg|_{\beta=\beta_i(x,m,\theta)}, \\ \dot{\beta}(i, x, \theta) &= -\frac{\Phi'_\theta(i, x, \beta, m, \theta)}{\Phi'_\beta(i, x, \beta, m, \theta)}\bigg|_{\beta=\beta_i(x,m,\theta)} \end{aligned} \tag{3.74}$$

(note that $\Phi'_\beta(i, x, \beta, m, \theta) = -\int I_{\{|y-\beta|\le m\}} Q_i(dy; \theta \mid x) < 0$ in a neighborhood of the solution of Eq. (3.73), see (3.32) and the subsequent text).

Now we denote

$$F(\theta, m) = \varphi(\theta, m) - R^2$$

(see (3.40) and (3.41)).

Then again by the implicit function theorem, we have that there exists $\dot{m}(\theta)$, where $m(\theta)$ is a solution of the equation $F(\theta, m) = 0$, and

$$\dot{m} = -\frac{F'_\theta(\theta, m)}{F'_m(\theta, m)}\bigg|_{m=m(\theta)} \tag{3.75}$$

($F'_m(\theta, m) = \varphi'_m(\theta, m) < 0$, see (3.42)).

Of course, one can obtain explicit formulas for $\dot{\beta}$ in terms of the basic system of conditional densities $\{(f_i^n(z, \theta \mid x))_{i\le n}\}_{n\ge1}$ and their derivatives by means of the calculation of Φ'_θ, Φ'_β, F'_θ, and F'_m based on the above formulas.

It is now obvious that if the basic model is sufficiently smooth, then $\psi^{*,n}$ is also smooth, and, e.g.,

$$\dot{\psi}_i^{*,n}(z, \theta \mid x) = (\dot{l}_i^n(z, \theta \mid x) - \dot{\beta}_i^{*,n}(x, \theta)) I_{\{|l_i^n(z,\theta|x) - \beta_i^{*,n}(x,\theta)| \le m_n^*(\theta)\}}$$
$$+\dot{m}_n^*(\theta)(I_{\{l_i^n(z,\theta|x)-\beta_i^{*,n}(x,\theta)\ge m_n^*\}} - I_{\{l_i^n(z,\theta|x)-\beta_i^{*,n}(x,\theta)\le -m_n^*\}}),$$

where $\dot{\beta}^{*,n}$ and $\dot{m}_n^*$ are given by relations (3.74) and (3.75).

The high-order derivatives are calculated analogously.

Finally, assume that (3.65) holds with γ^{ψ^*} instead of γ^ψ and there exists a sequence of functions

$$c_k = \{c^{n,k}\}_{n\ge1} = \{(c_i^{n,k}(z, \theta \mid x))_{i\le n}\}_{n\ge1}$$

such that $\forall\theta \in \Theta$, $\forall u$, $0 < u \le K$, $K > 0$,

$$\sup_{B(n,u,\theta)} \max(|\ddot{l}_i^n(z, y \mid x)|, |\ddot{\beta}_i^{*,n}(x, y)|, |\ddot{m}_n^*(y)|) \le c_i^{n,k}(z, \theta \mid x)$$

and

$$n^{-3/2} \sum_{i=1}^{n} c_i^{n,k}(x_i, \theta \mid x) \xrightarrow{P_\theta^n} 0 \quad \text{as } \ n \to \infty.$$

Proposition 3.4. *If the above conditions are satisfied, then the assertion of Proposition* 3.3 *holds.*

Proof. Indeed, for $0 < u \le K$, $K > 0$,

$$\begin{aligned}
\sup_{B(n,u,\theta)} n^{-1}|\dot{L}_n^*(y) - \dot{L}_n^*(\theta)| &\le n^{-1} \sum_{i=1}^{n} \sup_{B(n,u,\theta)} |\dot{\psi}_i^{*,n}(x_i, y \mid x) - \dot{\psi}_i^{*,n}(x_i, \theta \mid x)| \\
&\le n^{-1} \sum_{i=1}^{n} \sup_{B(n,u,\theta)} |\ddot{\psi}^{*,n}(x_i, y \mid x)||y - \theta| \\
&\le \text{const} \cdot n^{-3/2} \sum_{i=1}^{n} c_i^{n,k}(x_i, \theta \mid x) \xrightarrow{P_\theta^n} 0 \quad \text{as } \ n \to \infty.
\end{aligned}$$

□

3.3. Robust Estimators in General Statistical Models with Filtration

3.3.1. Specification of the model. Regularity. Ergodicity. (a) Let

$$\mathcal{E} = (\Omega, \mathcal{F}, F = (\mathcal{F}_t)_{t\ge0}, \{P_\theta, \theta \in \Theta\}, P) \tag{3.76}$$

be a general statistical model with filtration. This means that $(\Omega, \mathcal{F}, F, P)$ is a stochastic basis, i.e., a complete probability space with filtration $F = (\mathcal{F}_t)_{t\ge0}$ satisfying the usual conditions, P_θ is a probability measure depending on the parameter θ to be estimated, and Θ is an open subset of $\mathbb{R}_1$. It is assumed that $P_\theta \overset{\text{loc}}{\sim} P$ $\forall\theta \in \Theta$.

Remark 3.16. Consider a statistical model

$$(\Omega, \mathcal{F}, F, \{P_\theta, \theta \in \Theta\})$$

and assume that the set of measures

$$\{P_\theta, \theta \in \Theta\}$$

is such that $P_{\theta'} \overset{\text{loc}}{\sim} P_\theta$ for each $\theta' \neq \theta$.

Fix some value of the parameter θ, say, θ_0, and denote $P = P_{\theta_0}$. Then $P_\theta \sim P$ $\forall\theta \in \Theta$. Now we assume that $(\Omega, \mathcal{F}, F, P)$ is a stochastic basis. Thus we obtain the previous model with reference measure P.

Remark 3.17. All the notation concerning the martingale theory that is used below can be found in [45, 67].

Let $P(t) = P \mid \mathcal{F}_t$, $P_\theta(t) = P_\theta \mid \mathcal{F}_t$ be the restrictions of the measures P and P_θ to the σ-algebra $\mathcal{F}_t$, and let $\rho_\theta = (\rho_\theta(t))_{t\ge0}$ be the likelihood ratio process with cadlag trajectories. For simplicity we assume that $\rho_\theta(0) = 1$. As is well known (see [48]),

$$\rho_\theta := \frac{dP_\theta}{dP} = \mathcal{E}(M_\theta) := \exp\Big\{M_\theta - \frac{1}{2}\langle M_\theta^c\rangle\Big\} \prod (1 + \Delta M_\theta) e^{-\Delta M_\theta},$$

where $M \in \mathcal{M}_{\text{loc}}(P)$ is a local P-martingale.

(b) Let Q be some other probability measure on $(\Omega, \mathcal{F}, F)$ such that $Q \overset{\text{loc}}{\ll} P$ and $\dfrac{dQ}{dP} = \mathcal{E}(M)$, M being a local P-martingale.

If m is a local P-martingale, then the process (Girsanov transform)

$$L(m, M) := m - \frac{1}{1+\Delta M} \cdot [m, M] = m - \langle M^c, m^c\rangle - \sum \frac{\Delta m \Delta M}{1+\Delta M} \tag{3.77}$$

is a local Q-martingale.

(c) An experiment $\mathcal{E}$ is said to be *regular* (see also [82]) if:

1. for each $t \geq 0$ (P-a.s.), the function $\theta \to M_\theta(t, \omega)$ is continuously differentiable and the derivative $\dot{M}_\theta := \dfrac{\partial}{\partial\theta} M_\theta$ is a local P-martingale $\forall\theta$;
2. for all $t \geq 0$ (P-a.s.), there exists $\dfrac{\partial}{\partial\theta} \ln \rho_\theta = L(\dot{M}_\theta, M_\theta) \in \mathcal{M}^2(P_\theta)$, the class of square-integrable P_θ-martingales;
3. the Fisher information $I(\theta) = E_\theta \langle L(\dot{M}_\theta, M_\theta)\rangle$ is finite and positive.

(d) Consider the sequence of regular statistical models

$$\mathcal{E} = \{\mathcal{E}_n\}_{n\geq 1} = \{(\Omega^n, \mathcal{F}^n, F^n = (\mathcal{F}^n_t)_{0\leq t\leq T}, \{P^n_\theta, \theta \in \Theta \subset \mathbb{R}_1\}, P^n)\}_{n\geq 1},$$

where $T > 0$ is a number.

Denote $c_n(\theta) = (I^n_T(\theta))^{-1/2} = (E^n_\theta \langle L(\dot{M}^n_\theta, M^n_\theta)\rangle_T)^{-1/2}$. Then, if

$$(1) \quad \lim_{n\to\infty} c_n(\theta) = 0,$$

$$(2) \quad c_n^2(\theta) \widehat{I}^n(\theta) \overset{P^n_\theta}{\to} 1 \quad \text{as } n \to \infty,$$

where $\widehat{I}^n_\theta = \langle L(\dot{M}^n_\theta, M^n_\theta)\rangle_T$, then we call the sequence $\mathcal{E} = \{\mathcal{E}_n\}_{n\geq 1}$ *ergodic.*

3.3.2. CLAN estimators. Denote by $\mathcal{M} = \mathcal{M}(\{P^n_\theta\}_{n\geq 1})$ the class of sequences of processes $L_\theta = \{L^n_\theta\}_{n\geq 1}$, $L^n_\theta = (L^n_\theta(t), 0 \leq t \leq T)$ with the following properties:

(1) for each $n \geq 1$, $L^n_\theta \in \mathcal{M}^2(P^n_\theta)$;

(2) the sequence $\{L^n_\theta\}_{n\geq 1}$ satisfies the Lindeberg condition

$$\int_0^T \int_{|x|>\varepsilon} x^2 \nu_n(dt, dx) \overset{P^n_\theta}{\to} 0 \quad \text{as } n \to \infty \ \ \forall \varepsilon \in (0, 1], \tag{3.78}$$

where ν_n is the compensator (with respect to the measure P^n_θ) of the jump measure of the process $c_n(\theta) L^n_\theta$;

(3)

$$c_n^2(\theta) \langle L^n_\theta\rangle_T \overset{P^n_\theta}{\to} \Gamma_L(\theta) \quad \text{as } n \to \infty, \ \ 0 < \Gamma_L(\theta) < \infty; \tag{3.79}$$

(4)

$$c_n^2(\theta) \langle L^n_\theta, L(\dot{M}^n_\theta, M^n_\theta)\rangle_T \overset{P^n_\theta}{\to} \gamma_L(\theta) \quad \text{as } n \to \infty, \ \ 0 < \gamma_L(\theta) < \infty, \tag{3.80}$$

$\Gamma_L(\theta)$ and $\gamma_L(\theta)$ are deterministic functions.

We assume that $\{L(\dot{M}^n_\theta, M^n_\theta)\}_{n\geq 1} \in \mathcal{M}(\{P^n_\theta\}_{n\geq 1})$. Note that if $L_\theta \in \mathcal{M}(\{P^n_\theta\}_{n\geq 1})$, then

$$\mathcal{L}(c_n(\theta) L^n_\theta(T) \mid P^n_\theta) \overset{\omega}{\to} \mathcal{N}(0, \Gamma_L(\theta)) \tag{3.81}$$

as simply follows from the central limit theorem (CLT) for martingales (see, e.g., [45]).

The sequence $T^L = \{T^L_n\}_{n\geq 1}$ of $\mathcal{F}^n_T$-measurable random variables with values in Θ is called a CLAN estimator if for each $\theta \in \Theta$ there exists $L_\theta \in \mathcal{M}(\{P^n_\theta\}_{n\geq 1})$ such that

$$T^L_n = \theta + \frac{L^n_\theta(T)}{\langle L^n_\theta, L(\dot{M}^n_\theta, M^n_\theta)\rangle_T} + R_n(\theta), \tag{3.82}$$

where $c_n^{-1}(\theta)R_n(\theta) \xrightarrow{P_\theta^n} 0$ as $n \to \infty$.

It is obvious that

$$\mathcal{L}(c_n^{-1}(\theta)(T_n^L - \theta \mid P_\theta^n) \xrightarrow{\omega} \mathcal{N}\left(0, \frac{\Gamma_L(\theta)}{\gamma_L^2(\theta)}\right).$$

3.3.3. Shrinking neighborhoods. For each $n \geq 1$, denote by $\mathcal{P}_\theta^n := \{\widetilde{P}_\theta^n \colon \widetilde{P}_\theta^n \sim P_\theta^n\}$ some neighborhood of the basic (core) probability measure P_θ^n. We know that for each $\widetilde{P}_\theta^n \in \mathcal{P}_\theta^n$, there exists $\widetilde{N}_\theta^n \in \mathcal{M}_{\text{loc}}(P_\theta^n)$ such that

$$\frac{d\widetilde{P}_\theta^n}{dP_\theta^n} = \mathcal{E}(\widetilde{N}_\theta^n). \tag{3.83}$$

If the martingale $\widetilde{N}_\theta^n$ has the form

$$\widetilde{N}_\theta^n = c_n(\theta)N_\theta^n \tag{3.84}$$

and

$$\{\widetilde{P}_\theta^n\} \lhd \{P_\theta^n\}, \tag{3.85}$$

then we say that the sequence $\{\widetilde{P}_\theta^n\}_{n\geq 1}$ belongs to the set of shrinking neighborhoods of the core sequence $\{P_\theta^n\}_{n\geq 1}$.

For definiteness, denote such a set by $\mathcal{P}_\theta$ and each element of $\mathcal{P}_\theta$ by $\{\widetilde{P}_\theta^n\}_{n\geq 1}$ or by $\{P_\theta^{n,N}\}_{n\geq 1}$.

Proposition 3.5. *Let $\{P_\theta^{n,N}\}_{n\geq 1} \in \mathcal{P}_\theta$ and $T^L = \{T_n^L\}_{n\geq 1}$ be the CLAN estimator with asymptotic expansion* (3.82)*. Then*

$$\mathcal{L}\left(c_n^{-1}(\theta)(T_n^L - \theta) - \frac{\widetilde{B}_T^n}{c_n^2(\theta)\langle L_\theta^n, L(\dot{M}_\theta^n, M_\theta^n)\rangle_T} \mid P_\theta^{n,N}\right) \xrightarrow{\omega} \mathcal{N}\left(0, \frac{\Gamma_L(\theta)}{\gamma_L^2(\theta)}\right), \tag{3.86}$$

where $\widetilde{B}^n = (\widetilde{B}_t^n,\ 0 \leq t \leq T)$ is the first characteristic of the process $c_n(\theta)L_\theta^n$ with respect to the measure $P_\theta^{n,N}$.

Proof. Let $W = (W_t)$, $0 \leq t \leq T$, be a standard Wiener process defined on a stochastic basis $(\Omega, \mathcal{F}, F, P)$. Denote $M := \sqrt{\frac{\Gamma_L(\theta)}{T}}W$. Then $C_T = E^P(M_T^2) = \Gamma_L(\theta)$. Further, denote $X^n = c_n(\theta)L_\theta^n$, and let

$$C_T^{1,n} = C_T + \int_0^T \int_{|x|\leq 1} x^2\nu^n(dt, dx) - \sum_{0<t\leq T}\left(\int_{|x|\leq 1} x\nu^n(\{t\}, dx)\right)^2$$

be the second modified characteristic of the semimartingale X^n with respect to the truncation function $h(x) = xI_{\{|x|\leq 1\}}$.

Then by conditions (3.78) and (3.79), $I_{\{|x|\geq\varepsilon\}} \cdot \nu_T^n \xrightarrow{P_\theta^n} 0$ as $n \to \infty$ $\forall \varepsilon \in (0, 1]$ and $C_T^{1,n} \xrightarrow{P_\theta^n} C_T = \Gamma_L(\theta)$. Hence (see [45])

$$X_T^n - \widetilde{B}_T^n \xrightarrow{\mathcal{L}(T|P_\theta^{n,N})} M_T,$$

where $\widetilde{B}_T^n = -\int_0^T \int_{|x|\geq 1} x\widetilde{\nu}^n(dt, dx)$ is a first characteristic of the process X^n with respect to the measure $P_\theta^{n,N}$.

Now the desirable follows from conditions (3.82) and (3.83). □

Proposition 3.6. *Let the sequence $\{\widetilde{P}_\theta^n\}$ be such that for each $n \geq 1$, $\widetilde{P}_\theta^n \in \mathcal{P}_\theta^n$, (3.83)–(3.84) are satisfied, $N_\theta^n \in \mathcal{M}^2(P_\theta^n)$, and*

$$\lim_{d\to\infty} \overline{\lim_{n\to\infty}}\, \widetilde{P}_\theta^n\{c_n^2(\theta)\langle N_\theta^n\rangle_T > d\} = 0. \tag{3.87}$$

Let $T^L = \{T_n^L\}_{n\geq 1}$ be a CLAN estimator with asymptotic expansion (3.82). Then:

(1) *$\{\widetilde{P}_\theta^n\} \lhd \{P_\theta^n\}$, and, therefore, $\{\widetilde{P}_\theta^n\} \in \mathcal{P}_\theta$;*

(2) *relation (3.86) remains true with $c_n^2(\theta)\langle L_\theta^n, N_\theta^n\rangle_T$ instead of $\widetilde{B}_T^n$;*

(3) *if, in addition, there exists a deterministic limit*

$$\beta_{L,N}(\theta) = \widetilde{P}_\theta^n\text{-}\lim_{n\to\infty} c_n^2(\theta)\langle L_\theta^n, N_\theta^n\rangle_T, \tag{3.88}$$

($\widetilde{P}_\theta^n$-$\lim\limits_{n\to\infty}$ denotes the limit in probability), then

$$\mathcal{L}\Big(c_n^{-1}(\theta)(T_n^L - \theta) \mid \widetilde{P}_\theta^n\Big) \xrightarrow{\omega} \mathcal{N}\left(\frac{\beta_{L,N}(\theta)}{\gamma_L(\theta)}, \frac{\Gamma_L(\theta)}{\gamma_L^2(\theta)}\right). \tag{3.89}$$

Proof. Denote $X^n = c_n(\theta)L_\theta^n$. For each $n \geq 1$, the process X^n is a semimartingale with a triplet of predictable characteristics $(-xI_{\{|x|\geq 1\}} \cdot \nu^n,\ c_n^2\langle L_\theta^{n,c}\rangle,\ \nu^n)$ (with respect to the measure P_θ^n).

(1) The following necessary and sufficient condition for $\{\widetilde{P}_\theta^n\} \lhd \{P_\theta^n\}$ is well known (see, e.g., [45]):

$$\lim_{\eta\to\infty}\lim_{n\to\infty} \widetilde{P}_\theta^n\left(h_T\left(\frac{1}{2}; \widetilde{P}_\theta^n, P_\theta^n\right) \geq \eta\right) = 0, \tag{3.90}$$

$$\lim_{\eta\to\infty}\lim_{n\to\infty} \widetilde{P}_\theta^n\left(\sup_{t\leq T}\alpha_n(t) \geq \eta\right) = 0, \tag{3.91}$$

where $\left(h_t\left(\frac{1}{2}; \widetilde{P}_\theta^n, P_\theta^n\right)\right)$, $0 \leq t \leq T$, is the Hellinger process of order $\frac{1}{2}$, and $\alpha_n(t) = \widetilde{\rho}_\theta^n(t)/\rho_\theta^n(t-)$, where

$$\widetilde{\rho}_\theta^n = \frac{d\widetilde{P}_\theta^n}{dP_\theta^n} = \mathcal{E}(c_n(\theta)N_\theta^n).$$

It can be easily seen that

$$h\left(\frac{1}{2}; \widetilde{P}_\theta^n, P_\theta^n\right) \leq \frac{1}{2}c_n^2(\theta)\langle N_\theta^n\rangle. \tag{3.92}$$

Indeed,

$$h\left(\frac{1}{2}; \widetilde{P}_\theta^n, P_\theta^n\right) = \frac{1}{8}c_n^2(\theta)\langle N_\theta^{n,c}\rangle + \frac{1}{2}\Big(\sum\Big(1 - \sqrt{1 + c_n(\theta)\Delta N_\theta^n}\Big)^2\Big)^{p,P_\theta^n}.$$

But, since $(1 - \sqrt{1+x})^2 \leq \dfrac{x^2}{(1+\sqrt{1+x})^2} \leq x^2$ for $x \geq -1$, we have

$$h\left(\frac{1}{2}; \widetilde{P}_\theta^n, P_\theta^n\right) \leq \frac{1}{8}c_n^2(\theta)\langle N_\theta^{n,c}\rangle + \frac{1}{2}\Big(\sum c_n^2(\theta)(\Delta N_\theta^n)^2\Big)^{p,P_\theta^n} \leq \frac{1}{2}c_n^2(\theta)\langle N_\theta^n\rangle. \tag{3.93}$$

Further,

$$\begin{aligned}\widetilde{P}_\theta^n\Big(\sup_{t\leq T}\alpha_n(t) \geq \eta\Big) &= \widetilde{P}_\theta^n\Big(\sup_{t\leq T}(1 + c_n(\theta)\Delta N_\theta^n(t)) \geq \eta\Big)\\ \leq \widetilde{P}_\theta^n\Big(\sup_{t\leq T}c_n^2(\theta)(\Delta N_\theta^n(t))^2 \geq \eta - 1\Big) &\leq \widetilde{P}_\theta^n\Big(\sum_{t\leq T}c_n^2(\theta)(\Delta N_\theta^n(t))^2 \geq \eta - 1\Big).\end{aligned} \tag{3.94}$$

By virtue of the Lenglart inequality, for any $\delta > 0$ we have

$$\widetilde{P}_\theta^n\left(\sum_{t\le T} c_n^2(\theta)(\Delta N_\theta^n(t))^2 \ge \eta\right) \le \frac{d}{\eta} + \widetilde{P}_\theta^n\left(\left(\sum_{t\le T} c_n^2(\theta)(\Delta N_\theta^n(t))^2\right)^{p,P_\theta^n} \ge d\right). \tag{3.95}$$

Now assertion (1) follows from (3.90)–(3.95) and (3.87).

(2) Now we show that

$$\left|\widetilde{B}_T^n - c_n^2(\theta)\langle L_\theta^n, N_\theta^n\rangle_T\right| \overset{\widetilde{P}_\theta^n}{\to} 0 \quad \text{as } n\to\infty. \tag{3.96}$$

Denote $\overline{X}^n = X^n - \sum \Delta X^n I_{\{|\Delta X^n|\ge 1\}}$. Since $\overline{X}^n$ is a special semimartingale, the unique decomposition

$$\overline{X}^n = \overline{M}^n + \overline{A}^n$$

takes place with a predictable $\overline{A}^n$.

Rewrite $\overline{X}^n$ as follows:

$$\overline{X}^n = X^n - xI_{\{|x|\ge 1\}} * (\mu^n - \nu^n) - xI_{\{|x|\ge 1\}} * \nu^n.$$

We have

$$\overline{M}^n = X^n - xI_{\{|x|\ge 1\}} * (\mu^n - \nu^n).$$

Further, applying the triplet transformation formulas under an absolutely continuous change of measure, we obtain

$$\widetilde{B}^n = B^n + c_n(\theta)\langle \overline{M}^n, N_\theta^n\rangle.$$

Hence

$$\widetilde{B}^n - c_n^2(\theta)\langle L_\theta^n, N_\theta^n\rangle = B^n - c_n(\theta)\langle X^n - \overline{M}^n, N_\theta^n\rangle.$$

By the Lindeberg condition and the contiguity $\{\widetilde{P}_\theta^n\} \lhd \{P_\theta^n\}$, we have

$$B_T^n \overset{\widetilde{P}_\theta^n}{\to} 0.$$

Further,

$$\langle X^n - \overline{M}^n\rangle = x^2 I_{\{|x|\ge 1\}} * \nu^n - \sum_{t\le\cdot}\left(\int xI_{\{|x|\ge 1\}}\cdot \nu^n(\{t\}, dx)\right)^2.$$

Again, the Lindeberg condition and the contiguity yield

$$\langle X^n - \overline{M}^n\rangle_T \overset{\widetilde{P}_\theta^n}{\to} 0. \tag{3.97}$$

But by the Kunita–Watanabe inequality, (3.97), and (3.87), we obtain

$$c_n^2(\theta)\langle X^n - \overline{M}^n, N_\theta^n\rangle_T^2 \le \langle X^n - \overline{M}^n\rangle_T c_n^2(\theta)\langle N_\theta^n\rangle_T \overset{\widetilde{P}_\theta^n}{\to} 0 \quad \text{as } n\to\infty.$$

Assertion (2) is proved.

Assertion (3) is an easy consequence of assertion (2). □

Denote the measure $\widetilde{P}_\theta^n := P_\theta^{n,N}$. For the validity of (3.88), the existence of the deterministic limit with respect to $\{P_\theta^n\}_{n\ge 1}$ is sufficient.

Now (3.89) implies simply that

$$\lim_{a\to\infty}\lim_{n\to\infty} E_\theta^{n,N}\left\{(c_n^{-1}(\theta)(T_n^L - \theta))^2 \wedge a\right\} = D(L,N,\theta), \tag{3.98}$$

where $E_\theta^{n,N}$ is the expectation with respect to the measure $P_\theta^{n,N}$ and

$$D(L,N,\theta) = \frac{\beta_{L,N}^2(\theta) + \Gamma_L(\theta)}{\gamma_L^2(\theta)}. \tag{3.99}$$

3.3.4. Optimization criteria. Denote $\mathcal{N} = \mathcal{N}(\{P_\theta^n\}_{n\geq 1})$ the class of sequences of processes $\widetilde{N}_\theta = \{\widetilde{N}_\theta^n\}_{n\geq 1}$ satisfying conditions (3.84), (3.85), and (3.88), and, moreover, such that the process $\mathcal{E}(\widetilde{N}_\theta^n)$ from (3.83) is a uniformly integrable martingale with $\inf_t \mathcal{E}_t(\widetilde{N}_\theta^n) > 0$.

Further, suppose that for each $n \geq 1$, $\widetilde{N}_\theta^n \in \mathcal{N}_R^n$, $R > 0$, is some domain in $\mathcal{M}^2(P_\theta^n)$. Consideration of such a set $\mathcal{N}_R^n$ is needed if we want to obtain a nontrivial optimization problem.

Remark 3.18. Under the above-mentioned conditions we obtain that: (1) $\widetilde{P}_\theta^n$ defined by (3.83) is a probability measure, equivalent to P_θ^n; (2) $\{\widetilde{P}_\theta^n\} \lhd \{P_\theta^n\}$ and, therefore, $D(L, N, \theta)$, given by (3.99), has the statistical meaning of a risk functional (see (3.98)).

The CLAN estimator $T^* = \{T_n^*\}_{n\geq 1}$ is called $(\mathcal{M},\mathcal{N})$-optimal in the minimax sense over the class of CLAN estimators $\{T^L,\, L \in \mathcal{M}(\{P_\theta^n\}_{n\geq 1})\}$ for each $\theta \in \Theta$,

$$\forall \varepsilon > 0,\ \ \forall L \in \mathcal{M},\ \ \exists N^{\varepsilon,L} \in \mathcal{N},\ \ \forall N \in \mathcal{N}$$

$$\lim_{a\to\infty}\lim_{n\to\infty} \frac{E_\theta^{n,N}((c_n^{-1}(\theta)(T_n^* - \theta))^2 \wedge a)}{E_\theta^{n,N^{\varepsilon,L}}((c_n^{-1}(\theta)(T_n^* - \theta))^2 \wedge a)} \leq 1 + \varepsilon. \tag{3.100}$$

The score sequence $L_\theta^* = \{L_\theta^{*,n}\}_{n\geq 1} \in \mathcal{M}(\{P_\theta^n\}_{n\geq 1})$ is said to be $(\mathcal{M},\mathcal{N})$-optimal in the minimax sense if for each $\theta \in \Theta$

$$\sup_{N\in\mathcal{N}} D(L^*, N, \theta) = \inf_{L\in\mathcal{M}} \sup_{N\in\mathcal{N}} D(L, N, \theta). \tag{3.101}$$

It is obvious that if a score sequence L_θ^* is optimal and the corresponding CLAN estimator exists, then $T^* = T^{L^*}$ is also optimal.

Introduce the strong ergodicity condition: let for $\forall K_\theta^n$ and $R_\theta^n \in \mathcal{M}^2(P_\theta^n)$

$$P_\theta^n\text{-}\lim_{n\to\infty} c_n^2(\theta)\langle K_\theta^n, R_\theta^n\rangle_T = C \Leftrightarrow \lim_{n\to\infty} E_\theta^n c_n^2(\theta)\langle K_\theta^n, R_\theta^n\rangle_T = C \tag{3.102}$$

($C \geq 0$ is some constant).

Denote

$$D_n(L^n, N^n, \theta) = \frac{(E_\theta^n\langle L_\theta^n, N_\theta^n\rangle_T)^2 + E_\theta^n\langle L_\theta^n\rangle_T}{(E_\theta^n\langle L_\theta^n, L(\dot{M}_\theta^n, M_\theta^n)\rangle_T)^2}. \tag{3.103}$$

Obviously, under condition (3.102),

$$D(L, N, \theta) = \lim_{n\to\infty} D_n(L^n, N^n, \theta). \tag{3.104}$$

3.3.5. Calculation of the explicit form of the risk functional $D_n(L^n, N^n, \theta)$. At the beginning of this section the index n is fixed and omitted.

Consider the statistical model (3.76) (see also Remark 3.16) associated with the one-dimensional F-adapted cadlag process $X = (X_t)$, $0 \leq t \leq T$, in the following way: for each $\theta \in \Theta$, P_θ is the unique measure on $(\Omega, \mathcal{F})$ such that the process X is a (P_θ, F)-semimartingale with predictable characteristics $(B(\theta), C(\theta), \nu_\theta)$ (with respect to the truncation function $h(x) = xI_{\{|x|\leq 1\}}$). Assume for convenience that all measures P_θ coincide on $\mathcal{F}_0$. Also, we assume that under the measure P, X is a semimartingale with triplet $B = B(0)$, $C = C(0)$, $\nu = \nu_0$.

We know ([45], Ch. III) that in this situation there exist a $\widetilde{\mathcal{P}}$-measurable positive function

$$Y_\theta = \{Y_\theta(\omega, t, x), (\omega, t, x) \in \Omega \times \mathbb{R}_+ \times \mathbb{R}_1\}$$

and a predictable process $(\beta_\theta(t))$, $0 \leq t \leq T$, with

$$|h(Y_\theta - 1)| * \nu \in \mathcal{A}_{\mathrm{loc}}(P),$$

$$\beta_\theta^2 \cdot C \in \mathcal{A}_{\mathrm{loc}}^+(P),$$

such that the following is true:

$$\begin{aligned} &(1) \quad B(\theta) = B + \beta_\theta \cdot C + h(Y_\theta - 1) * \nu, \\ &(2) \quad C(\theta) = C, \\ &(3) \quad \nu_\theta = Y_\theta * \nu, \quad Y_\theta > 0. \end{aligned} \tag{3.105}$$

In addition, the function Y_θ can be chosen in such a way that

$$a(t) := \nu(\{t\}, \mathbb{R}_1) = 1 \Leftrightarrow a_\theta(t) := \nu_\theta(\{t\}, \mathbb{R}_1) = \widehat{Y}_\theta(t) = \int Y_\theta(t,x)\nu(\{t\}, dx) = 1.$$

If the measure P is such that any (P, F)-local martingale admits the integral representation property with respect to X, then the likelihood-ratio process can be given by the explicit formula

$$\rho_\theta = \frac{dP_\theta}{dP} = \mathcal{E}(M_\theta),$$

where

$$M_\theta = \beta_\theta \cdot X^c + \left(Y_\theta - 1 + \frac{\widehat{Y}_\theta - a}{1-a} \right) * (\mu - \nu) \in \mathcal{M}_{\text{loc}}(P) \tag{3.106}$$

(with the usual convention $\dfrac{0}{0} = 0$).

Assume that our statistical model is regular. Thus we assume that for almost all (ω, t, x) (with respect to the corresponding Dolean's measure), the functions β_θ: $\theta \rightsquigarrow \beta_\theta(\omega, t)$ and Y_θ: $\theta \rightsquigarrow Y_\theta(\omega, t, x)$ are continuously differentiable (we denote $\dot{\beta}_\theta = \dfrac{\partial}{\partial\theta}\beta_\theta$ and $\dot{Y}_\theta = \dfrac{\partial}{\partial\theta}Y_\theta$) and the differentiation under the integral sign is possible.

Let us calculate

$$\frac{\partial}{\partial\theta} \ln \rho_\theta = \dot{M}_\theta - \langle \dot{M}_\theta^c, M_\theta^c \rangle - \sum \frac{\Delta \dot{M}_\theta \Delta M_\theta}{1 + \Delta M_\theta} = L(\dot{M}_\theta, M_\theta).$$

Below we use the following proposition.

Proposition 3.7. *Let P-martingales m and M admit integral representations*

$$\begin{aligned} m &= \beta \cdot n + \psi * (\mu - \nu), \\ M &= \gamma \cdot n + \chi * (\mu - \nu), \end{aligned} \tag{3.107}$$

where n is a continuous P-martingale, μ is an integer-valued measure on $[0, T] \times E$, $E = \mathbb{R}_1 \setminus \{0\}$, and ν is its P-compensator.

Let $\widetilde{P}$ be a measure, $\widetilde{P} \sim P$ with

$$\frac{d\widetilde{P}}{dP} = \mathcal{E}(M).$$

Then for the Girsanov transform $L(m, M)$ ($\in \mathcal{M}_{\text{loc}}(\widetilde{P})$, see (3.77)) we have

$$L(m, M) = \beta \cdot (n - \gamma \cdot \langle n \rangle) + \Phi * (\mu - \widetilde{\nu}), \tag{3.108}$$

where $\widetilde{\nu}$ is the $\widetilde{P}$-compensator of the measure μ and

$$\Phi = \frac{\psi - \widehat{\psi}}{1 + \chi - \widehat{\chi}} + \frac{\widehat{\psi}}{1 - \widehat{\chi}} \cdot I_{\{\widehat{\chi} < 1\}} \tag{3.109}$$

with

$$\widehat{\psi}(t) = \int_E \psi(t,x)\nu(\{t\}, dx), \quad \widehat{\chi}(t) = \int_E \chi(t,x)\nu(\{t\}, dx).$$

Proof. First, we note that

$$1 + \chi - \widehat{\chi} > 0,$$

since M is a density-determining martingale and $\widetilde{P} \sim P$. Further, from the definition of the Girsanov transform (3.77) we easily obtain

$$\Delta_t L^d(m, M) = \frac{\psi(t, \beta_t) I_D - \widehat{\psi}(t)}{1 + \chi(t, \beta_t) I_D - \widehat{\chi}(t)}, \tag{3.110}$$

where $D = \{(\omega, t) : \mu(\{t\}, E) = 1\}$ with the Dirac measure $\mu(\{t\}, dx) = \delta_{\beta_t}(dx)$.

On the other hand,

$$\Delta_t \Phi * (\mu - \widetilde{\nu}) = \Phi(t, \beta_t) I_D - \widehat{\widetilde{\Phi}}(t), \tag{3.111}$$

where

$$\widehat{\widetilde{\Phi}}(t) = \int_E \Phi(t, x) \widetilde{\nu}(\{t\}, dx) = \frac{\widehat{\psi}(t)}{1 - \widehat{\chi}(t)} \cdot I_{\{\widehat{\chi}(t) < 1\}}. \tag{3.112}$$

Now, substituting (3.109) and (3.112) into (3.111), we obtain

$$\Delta_t L^d(m, M) = \Delta_t \Phi * (\mu - \nu)$$

and, hence, the purely discontinuous part of the $\widetilde{P}$-martingale $L(m, M)$ is equal to $\Phi * (\mu - \widetilde{\nu})$. The continuous part

$$L^c(m, M) = m^c - \langle m^c, M^c \rangle = \beta \cdot n - \beta\gamma \cdot \langle n \rangle.$$

$\square$

From (3.106) we have

$$\dot{M}_\theta = \dot{\beta}_\theta \cdot X^c + \left(\dot{Y}_\theta + \frac{\dot{\widehat{Y}}_\theta}{1 - a} \right) * (\mu - \nu). \tag{3.113}$$

Note that $\widehat{Y}_\theta(t) = \int Y_\theta(t, x) \nu(\{t\}, dx)$ $(:= a_\theta(t))$. Hence

$$\dot{\widehat{Y}}_\theta(t) = \widehat{\dot{Y}}_\theta(t) = \dot{a}_\theta(t).$$

Now from (3.106), (3.113), and (3.108) we obtain

$$L(\dot{M}_\theta, M_\theta) = \dot{\beta}_\theta (X^c - \beta_\theta \cdot C) + \Phi_\theta * (\mu - \nu_\theta), \tag{3.114}$$

where

$$\Phi_\theta = \frac{\dot{Y}_\theta}{Y_\theta} + \frac{\dot{a}_\theta}{1 - a_\theta} \cdot I_{\{a_\theta < 1\}}$$

with $I_{\{a_\theta = 1\}} \dot{a}_\theta = 0$.

We give a more detailed description of the function Φ_θ.

For this purpose, recall (see [45]) that one can choose a version of characteristics $C(\theta)$ and ν_θ such that

$$\begin{aligned}
C_t(\theta) &= (c^\theta \cdot A^{\theta, c})_t, \\
\nu_\theta(\omega, dt, dx) &= dA_t^\theta(\omega) B_{\omega,t}^\theta(dx) \ \ (= \nu_\theta^c(\omega, dt, dx) + \nu_\theta^d(\omega, dt, dx)) \\
&= \nu_\theta^c(\omega, dt, dx) + dA_t^{\theta, d}(\omega) B_{\omega,t}^\theta(dx),
\end{aligned}$$

where $\nu_\theta^c = I_{\{a_\theta = 0\}} \nu_\theta$ is a continuous part of ν_θ, the process $A^\theta = (A_t^\theta)_{0 \le t \le T} \in \mathcal{A}_{\text{loc}}^+$, $c^\theta = (c_t^\theta)$, $0 \le t \le T$, is a nonnegative predictable process, and $B_{\omega,t}^\theta(dx)$ is a transition kernel from $(\Omega \times \mathbb{R}_1, \mathcal{P})$ into $(\mathbb{R}_1, \mathcal{B}(\mathbb{R}_1))$ with $B_{\omega,t}^\theta(\{0\}) = 0$ and $\Delta A_t^\theta B_{\omega,t}^\theta(\mathbb{R}_1) \le 1$.

Now for each integrable (with respect to ν_θ) function $\psi = (\psi(\omega, t, x))$ we have

$$
\begin{aligned}
(\psi * \nu_\theta)_t &= (\psi * \nu_\theta^c)_t + \int_0^t \int_{\mathbb{R}_1} \psi(\omega, s, x) B_{\omega,s}^\theta(dx) dA_s^{\theta,d} \\
&= (\psi * \nu_\theta^c)_t + \sum_{s \le t} B_{\omega,s}^\theta(\mathbb{R}_1) \left(\int_{\mathbb{R}_1} \psi(\omega, s, x) q_{\omega,s}^\theta(dx) \right) \Delta A_s^\theta \\
&= (\psi * \nu_\theta^c)_t + \sum_{s \le t} a_\theta(s) \int_{\mathbb{R}_1} \psi(\omega, s, x) q_{\omega,s}^\theta(dx), \qquad (3.115)
\end{aligned}
$$

where

$$
a_\theta(t) = \Delta A_t^\theta B_{\omega,t}^\theta(\mathbb{R}_1) \qquad (3.116)
$$

and

$$
q_{\omega,t}^\theta(dx) I_{\{a_\theta(t)>0\}} = \frac{B_{\omega,t}^\theta(dx)}{B_{\omega,t}^\theta(\mathbb{R}_1)} I_{\{a_\theta(t)>0\}}. \qquad (3.117)
$$

Thus, $q_{\omega,t}^\theta(dx)$ is a probability measure:

$$
I_{\{a_\theta(t)>0\}} \int_{\mathbb{R}_1} q_{\omega,t}^\theta(dx) = I_{\{a_\theta(t)>0\}}.
$$

Denote

$$
\frac{d\nu_\theta^c}{d\nu^c} := F_\theta, \quad \frac{q_{\omega,t}^\theta(dx)}{q_{\omega,t}(dx)} = f_\theta(\omega, t, x) \quad (\text{simply } \ f_\theta).
$$

Then we have

$$
\begin{aligned}
Y_\theta &= F_\theta I_{\{a=0\}} + \frac{a_\theta}{a} f_\theta I_{\{a>0\}}, \\
\dot{Y}_\theta &= \dot{F}_\theta I_{\{a=0\}} + \left(\frac{\dot{a}_\theta}{a} f_\theta + \frac{a_\theta}{a} \dot{f}_\theta \right) I_{\{a>0\}}.
\end{aligned}
$$

Therefore (recall that $Y_\theta > 0$),

$$
\Phi_\theta = \frac{\dot{F}_\theta}{F_\theta} I_{\{a_\theta=0\}} + \left(\frac{\dot{f}_\theta}{f_\theta} + + \frac{\dot{a}_\theta}{a_\theta(1-a_\theta)} \cdot I_{\{a_\theta<1\}} \right) I_{\{a_\theta>0\}}. \qquad (3.118)
$$

Denote

$$
l_c^\theta = \dot{\beta}_\theta, \quad l_\pi^\theta = \frac{\dot{F}_\theta}{F_\theta}, \quad l_\delta^\theta = \frac{\dot{f}_\theta}{f_\theta}, \quad l_b^\theta = \frac{\dot{a}_\theta}{a_\theta(1-a_\theta)}.
$$

Then

$$
L(\dot{M}_\theta, M_\theta) = l_c^\theta (X^c - \beta_\theta \cdot C) + (l_\pi^\theta I_{\{a_\theta=0\}} + l_\delta^\theta I_{\{a_\theta>0\}} + l_b^\theta I_{\{0<a_\theta<1\}}) * (\mu - \nu_\theta). \qquad (3.119)
$$

Further, let $\widetilde{P}_\theta \sim P$ and

$$
\frac{d\widetilde{P}_\theta}{dP} = \mathcal{E}(\widetilde{M}_\theta)
$$

with

$$
\widetilde{M}_\theta = \widetilde{\beta}_\theta \cdot X^c + \left(\widetilde{Y}_\theta - 1 + \frac{\widehat{\widetilde{Y}}_\theta - a}{1-a} \right) (\mu - \nu)
$$

(cf. (3.106)).

Then it is easy to see that

$$
\frac{d\widetilde{P}_\theta}{dP_\theta} = \mathcal{E}(L(\widetilde{M}_\theta - M_\theta, M_\theta)) = \mathcal{E}(\widetilde{N}_\theta)
$$

with $\widetilde{N}_\theta \in \mathcal{M}_{\rm loc}(P_\theta)$, and, therefore,

$$\widetilde{N}_\theta = (\widetilde{\beta}_\theta - \beta_\theta)(X^c - \beta_\theta \cdot C) + \Bigg[\left(\frac{\widetilde{F}_\theta}{F_\theta} - 1\right) I_{\{a_\theta = 1\}} \\ + \left(\frac{\widetilde{a}_\theta \widetilde{f}_\theta}{a_\theta f_\theta} - 1\right) I_{\{a_\theta > 0\}} + \frac{\widetilde{a}_\theta - a_\theta}{1 - a_\theta} I_{\{0 < a_\theta < 1\}}\Bigg] * (\mu - \nu_\theta). \tag{3.120}$$

Starting from Eqs. (3.119) and (3.120), it is natural to represent the core martingale L_θ and martingale N_θ (for definitions see previous Subsections 3.3.2 and 3.3.3) as follows:

$$L_\theta = \psi_c^\theta (X^c - \beta_\theta \cdot C) + (\psi_\pi^\theta I_{\{a_\theta = 0\}} + \psi_\delta^\theta I_{\{a_\theta > 0\}} + \psi_b^\theta I_{\{0 < a_\theta < 1\}}) * (\mu - \nu_\theta), \tag{3.121}$$

where $\psi_\alpha^\theta = \psi_\alpha^\theta(\omega, t)$ if $\alpha = c, b$; $\psi_\alpha^\theta = \psi_\alpha^\theta(\omega, t, x)$ if $\alpha = \pi, \delta$, and $\widehat{\psi}_\delta^\theta = 0$;

$$N_\theta = H_c^\theta (X^c - \beta_\theta \cdot C) + (H_\pi^\theta I_{\{a_\theta = 0\}} + H_\delta^\theta I_{\{a_\theta > 0\}} + H_b^\theta I_{\{0 < a_\theta < 1\}}) * (\mu - \nu_\theta), \tag{3.122}$$

where $H_\alpha^\theta = H_\alpha^\theta(\omega, t)$ if $\alpha = c, b$; $H_\alpha^\theta = H_\alpha^\theta(\omega, t, x)$ if $\alpha = \pi, \delta$, and $\widehat{H}_\delta^\theta = 0$.

Let us again endow all the objects with the index n and introduce the Dolean measures

$$\begin{aligned} \mu_c^{\theta,n}(dt, d\omega) &= c_n^2(\theta)\, dC_t^n P_\theta^n(d\omega), \\ \mu_\pi^{\theta,n}(dt, dx, d\omega) &= c_n^2(\theta) I_{\{a_\theta^n = 0\}} \nu_\theta^n(dt, dx) P_\theta^n(d\omega), \\ \mu_\delta^{\theta,n}(dt, dx, d\omega) &= c_n^2(\theta) I_{\{a_\theta^n = 0\}} \nu_\theta^n(dt, dx) P_\theta^n(d\omega), \\ \mu_b^{\theta,n}(dt, d\omega) &= c_n^2(\theta) p_\theta^n(\omega, dt) P_\theta^n(d\omega), \end{aligned} \tag{3.123}$$

where the measure $p_\theta^n(\omega, dt)$ is defined by the relation

$$p_\theta^n(\omega, B) = \sum_{t \in B} I_{\{0 < a_\theta^n(\omega,t) < 1\}} a_\theta^n(\omega, t)(1 - a_\theta^n(\omega, t)) \quad \forall B \in \mathcal{B}(\mathbb{R}_+).$$

Assume that for each $\alpha = c, \pi, \delta, b$ and $n \geq 1$, $l_\alpha^{\theta,n}$, $\psi_\alpha^{\theta,n}$, $H_\alpha^{\theta,n} \in L_2(\mu_\alpha^{\theta,n})$.

Denote

$$\begin{aligned} l^n &= (l_\alpha^{\theta,n},\ \alpha = c, \pi, \delta, b), \\ \psi^n &= (\psi_\alpha^{\theta,n},\ \alpha = c, \pi, \delta, b), \\ H^n &= (H_\alpha^{\theta,n},\ \alpha = c, \pi, \delta, b), \\ \mu^n &= (\mu_\alpha^{\theta,n},\ \alpha = c, \pi, \delta, b). \end{aligned} \tag{3.124}$$

Then a simple calculation results in

$$D_n(L^n, N^n; \theta) := D_n(\psi^n, H^n; \theta) = \frac{\left(\sum\limits_\alpha \psi_\alpha^{\theta,n} H_\alpha^{\theta,n} * \mu_\alpha^{\theta,n}\right)^2 + \sum\limits_\alpha (\psi_\alpha^{\theta,n})^2 * \mu_\alpha^{\theta,n}}{\left(\sum\limits_\alpha \psi_\alpha^{\theta,n} l_\alpha^{\theta,n} * \mu_\alpha^{\theta,n}\right)^2}, \tag{3.125}$$

where the sign "$*$" denotes the integral, $\alpha = c, \pi, \delta, b$.

3.3.6. Fixed-step optimization problem for statistical models associated with semimartingales. Fix the index $n \geq 1$, the real number $R > 0$, and consider the following sets of functions:

$$\Psi_n^0 = \Big\{ \psi^n = (\psi_\alpha^{\theta,n}, \alpha = c, \pi, \delta, b) : \psi_\alpha^{\theta,n} \in L_2(\mu_\alpha^{\theta,n}), \alpha = c, \pi, \delta, b, \widehat{\psi}_\delta^{\theta,n} = 0 \Big\}. \tag{3.126}$$

$$\begin{aligned} \mathcal{H}_R^n = \Big\{ & H^n = (H_\alpha^{\theta,n}, \alpha = c, \pi, \delta, b) : H_\alpha^{\theta,n} \in L_2(\mu_\alpha^{\theta,n}), \alpha = c, \pi, \delta, b, \\ & \widehat{H}_\delta^{\theta,n} = 0, H_\delta^{\theta,n}(\omega, t, x) \geq -\lambda^{\theta,n}(\omega, t), \ \lambda^{\theta,n}(\omega, t) \geq 0, \\ & \sum_{\alpha \neq \delta} |H_\alpha^{\theta,n}| * \mu_\alpha^{\theta,n} + \lambda^{\theta,n} * \mu_\delta^{\theta,n} \leq R \Big\}. \end{aligned} \tag{3.127}$$

The score function $\psi^{*,n} = (\psi_\alpha^{*,n}, \alpha = c, \pi, \delta, b) \in \Psi_n^0$ is said to be (Ψ_n^0, H_R^n)-optimal in the minimax sense if for each $\theta \in \Theta$,

$$\sup_{H^n \in \mathcal{H}_R^n} D_n(\psi^{*,n}, H^n; \theta) = \inf_{\psi \in \Psi_n^0} \sup_{H^n \in \mathcal{H}_R^n} D_n(\psi^n, H^n; \theta), \tag{3.128}$$

where $D_n(\psi^n, H^n; \theta)$ is given by (3.125).

Remark 3.19. Consider the following simple construction. Let the Hilbert spaces

$$L_\alpha^2(\Omega_\alpha, \mathcal{F}_\alpha, \mu_\alpha), \ \ \alpha = 1, ..., 4; \ \Omega_\alpha \cap \Omega_j = \varnothing, \quad \alpha \neq j,$$

be given.

We denote

$$\Omega = \cup \Omega_\alpha, \quad \mathcal{F} = \cup \mathcal{F}_\alpha = \{ \underset{\alpha}{\cup} A_\alpha : A_\alpha \in \mathcal{F}_\alpha, \quad \alpha = 1, ..., 4 \},$$
$$\mu(A) = \sum_\alpha \mu_\alpha(A \cap \Omega_\alpha) \quad \forall A \in \mathcal{F}$$

and consider a new Hilbert space $L^2(\Omega, \mathcal{F}, \mu)$ with the inner product $\langle \cdot, \cdot \rangle$. If $X, Y \in L_2(\Omega, \mathcal{F}, \mu)$, then, obviously,

$$\langle X, Y \rangle = \sum_\alpha \langle X_\alpha, Y_\alpha \rangle_\alpha,$$

where $X_\alpha = X I_{\{\Omega_\alpha\}}$, $Y_\alpha = Y I_{\{\Omega_\alpha\}}$, and $\langle \cdot, \cdot \rangle_\alpha$ is the inner product in L_α^2.

Now for each $n \geq 1$ and θ let

$$\begin{gathered} \Omega_1 = \Omega^n \times [0, T]; \\ \Omega_2 = \{\Omega^n \times [0, T]\} \cap \{(\omega, t) : a_\theta^n = 0\} \times \mathbb{R}_1, \\ \Omega_3 = \{\Omega^n \times [0, T]\} \cap \{(\omega, t) : a_\theta^n > 0\} \times \mathbb{R}_1, \\ \Omega_4 = \{\Omega^n \times [0, T]\} \cap \{(\omega, t) : 0 < a_\theta^n < 1\} \times \{\chi\}, \quad \chi \notin \mathbb{R}_1, \end{gathered}$$

with the corresponding σ-algebras and let

$$\mu = (\mu_\alpha, \alpha = 1, ..., 4) = (\mu_\alpha^{\theta,n}, \alpha = c, \pi, \delta; \ \ \mu_b^{\theta,n} \times \delta_{\{\chi\}});$$

$\delta_{(\cdot)}$ is the corresponding Dirac measure.

Further, let

$$\begin{aligned} X &= (X_\alpha, \alpha = 1, ..., 4) = (l_\alpha^{\theta,n}, \alpha = c, \pi, \delta, b), \\ Y &= (Y_\alpha, \alpha = 1, ..., 4) = (\psi_\alpha^{\theta,n}, \alpha = c, \pi, \delta, b), \\ Z &= (Z_\alpha, \alpha = 1, ..., 4) = (H_\alpha^{\theta,n}, \alpha = c, \pi, \delta, b). \end{aligned}$$

Then the right-hand side of (3.128) can be written as follows:

$$A = \inf_{Y \in \mathcal{Y}} \sup_{|Z| * \mu \leq R} \frac{(YZ * \mu)^2 + Y^2 * \mu}{(YX * \mu)^2}, \tag{3.129}$$

where

$$\mathcal{Y} = \{Y : VY * \mu = 0 \text{ for all } V = V(\omega, t) I_{\{\Omega_3\}}(\omega, t, x) \text{ with } |VY| * \mu < \infty\}.$$

But

$$A = \inf_{Y \in \mathcal{Y}} \frac{R^2 (\operatorname*{ess\,sup}_\mu |Y|)^2 + Y^2 * \mu}{(YX * \mu)^2}. \tag{3.130}$$

Indeed,

$$|YZ * \mu| \leq |Y||Z| * \mu \leq \operatorname*{ess\,sup}_\mu |Y| (|Z| * \mu) \leq R \operatorname*{ess\,sup}_\mu |Y|.$$

For each $\varepsilon > 0$ consider the sets

$$A_\varepsilon^1 = \Big\{Y > 0,\ Y \geq \operatorname*{ess\,sup}_\mu |Y| - \varepsilon\Big\},$$
$$A_\varepsilon^2 = \Big\{Y < 0,\ -Y \geq \operatorname*{ess\,sup}_\mu |Y| - \varepsilon\Big\}.$$

From the definition of ess sup it follows that

$$\mu(A_\varepsilon^1 \cup A_\varepsilon^2) > 0.$$

Suppose, e.g., that $\mu(A_\varepsilon^1) > 0$. Then we have

$$\mu(A_\varepsilon^1) = \sum_\alpha \mu_\alpha (A_\varepsilon^1 \cap \Omega_\alpha) > 0.$$

Hence $\mu_\alpha(A_\varepsilon^1 \cap \Omega_\alpha) > 0$ for some $\alpha = 1, 2, 3, 4$. Let, for definiteness,

$$\mu_1(A_\varepsilon^1 \cap \Omega_1) > 0.$$

Consider the function $Z^\varepsilon = (Z_1^\varepsilon, 0, 0, 0)$ with

$$Z_1^\varepsilon = \frac{R I_{\{A_\varepsilon^1 \cap \Omega_1\}}}{\mu_1(A_\varepsilon^1 \cap \Omega_1)}.$$

Now we obtain

$$|Z^\varepsilon| * \mu = Z^\varepsilon * \mu = Z_1^\varepsilon I_{\{\Omega_1\}} * \mu = Z_1^\varepsilon * \mu_1 = R$$

and

$$\begin{aligned} |YZ^\varepsilon * \mu| = |YZ^\varepsilon I_{\{\Omega_1\}} * \mu| = |YZ_1^\varepsilon * \mu_1| &= |YZ_1^\varepsilon \cdot I_{\{A_\varepsilon^1\}} * \mu_1 + YZ_1^\varepsilon \cdot I_{\{(A_\varepsilon^1)^c\}} * \mu_1| \\ &= YZ_1^\varepsilon I_{\{A_\varepsilon^1\}} * \mu_1 \geq \Big(\operatorname*{ess\,sup}_\mu |Y| - \varepsilon\Big) R \end{aligned}$$

for each $\varepsilon > 0$, where A^c is a complement of the set A.

Now from (3.130) we obtain that the optimal Y has the form

$$Y^* = \text{const}[X - \beta]_{-m}^m, \quad m > 0,$$
$$\beta = \beta(\omega, t) I_{\{\Omega_3\}}(\omega, t, x).$$

Our problem is to find equations for the pair (β, m) (compare with Sec. 3.2).

Remark 3.20. The above-mentioned optimization problem is an analytic problem and does not have a statistical meaning. More exact specification of sets of score functions and alternatives are needed.

As in the case of discrete time, denote

$$Q^{\delta,n}_{\omega,t}(\cdot,\theta) = \int I_{\{x: l^{\theta,n}_{\delta}\in\cdot\}} q^{\theta,n}_{\omega,t}(dx), \tag{3.131}$$

where the probability $q^{\delta,n}_{\omega,t}$ is given by (3.117) (see also (3.116)), and consider the equation with respect to β ($m > 0$ is a number, $[x]^b_a = (x \wedge b) \vee a$, $a < b$),

$$\int [y-\beta]^m_{-m} Q^{\delta,n}_{\omega,t}(dy,\theta) = 0. \tag{3.132}$$

Denote by

$$\beta^{\delta,n} = \beta^{\delta,n}(m,\theta) = \beta(Q^{\delta,n}_{\omega,t}(\cdot,\theta), m) \tag{3.133}$$

the solution of Eq. (3.132).

Assume that the distribution $Q^{\delta,n}_{\omega,t}(\cdot,\theta)$ satisfies conditions (a) and (b) after Remark 3.7 and $\overline{Q}^{n,l}_{\alpha}([-a,a]\setminus\{0\},\theta) > 0$ for each $a > 0$, $\alpha = c,\pi,b$. The distribution $\overline{Q}^{n,l}_{\alpha}(\cdot,\theta)$ is defined below, before Eq. (3.152).

Theorem 3.5. (i) *There exists optimal* $\psi^{*,n}$ *equal to*

$$\psi^{*,n} = \begin{cases} [l^{\theta,n}_{\alpha}]^{m^*_n(\theta)}_{-m^*_n(\theta)} & \text{if } \alpha \neq \delta, \\ [l^{\theta,n}_{\delta} - \beta^{\delta,n}(m^*_n(\theta),\theta)]^{m^*_n(\theta)}_{-m^*_n(\theta)} & \text{if } \alpha = \delta, \end{cases} \tag{3.134}$$

where $m^*_n(\theta)$ *is the unique solution of the equation*

$$\begin{aligned} R^2m^2 = {} & \sum_{\alpha\neq\delta}\Big\{[l^{\theta,n}_{\alpha}]^m_{-m} l^{\theta,n}_{\alpha} * \mu^{\theta,n}_{\alpha} - ([l^{\theta,n}_{\alpha}]^m_{-m})^2 * \mu^{\theta,n}_{\alpha}\Big\} \\ & + \Big\{[l^{\theta,n}_{\delta} - \beta^{\delta,n}(m,\theta)]^m_{-m} l^{\theta,n}_{\delta} * \mu^{\theta,n}_{\delta} - ([l^{\theta,n}_{\delta} - \beta^{\delta,n}(m,\theta)]^m_{-m})^2 * \mu^{\theta,n}_{\delta}\Big\}. \end{aligned} \tag{3.135}$$

(ii) *This* $\psi^{*,n}$ *is unique (up to a constant factor).*

The proof is quite similar to that of Theorem 3.3 and we omit it here.

3.3.7. Comments and special models.

1. To make clear the notion of optimal (with respect to the risk functional $D(L,N;\theta)$) estimators in the spirit of robust statistics, let us recall the asymptotic behavior of the estimational equation.

Consider the estimational stochastic equation

$$L_n(\theta) = L_n(\theta,\omega) = 0, \quad \theta\in\Theta, \quad n \geq 1,$$

where for each θ and n the random variable $L_n(\theta,\omega)$ is defined on the stochastic basis $(\Omega^n, \mathcal{F}^n, P^n)$. Further, we consider a family $Q = \{Q^n_\theta\}_{n\geq1}$ of measures, where for each θ and n the measure Q^n_θ is defined on the σ-algebra $\mathcal{F}^n$.

We know that if the conditions of Corollary 3.1 are satisfied, then there exists the CLAN estimator

$$T^L_n = b^Q(\theta) - \frac{L_n(b^Q(\theta))}{\dot{L}_n(b^Q(\theta))} + R_n(\theta) \tag{3.136}$$

with Q^n_θ-$\lim c^{-1}_n(\theta)R_n(\theta) = 0$, where $b^Q(\theta)$ is the unique solution of the equation

$$\Delta_Q(\theta,y) = 0$$

with

$$\Delta_Q(\theta,y) = Q^n_\theta\text{-}\lim_{n\to\infty} c^2_n(\theta)L_n(y)$$

and $\{c_n(\theta)\}_{n\geq 1}$ is some normalizing sequence. Thus we have

$$T_n^L \xrightarrow{Q_\theta^n} b^Q(\theta).$$

The quantity $b^Q(\theta)$ is called an *asymptotic version* of the estimator $T^L = \{T_n^L\}_{n\geq 1}$.

Moreover, if the measure Q_θ^n depends on the parameter γ in such a way that

$$Q_\theta^{\gamma,n} \to Q_\theta^{0,n} \quad \text{as} \ \ \gamma \to 0,$$

in certain sense, then one can define the so-called *influence functional*

$$IF(\{Q_\theta^{0,n}\}, \{T_n^L\}, \{Q_\theta^{\gamma,n}\}) = \lim_{\gamma\to 0} \frac{b^{Q^\gamma}(\theta) - b^{Q^0}(\theta)}{\gamma} \tag{3.137}$$

(see [70]).

It should be mentioned that if for each $n \geq 1$, $L_n(\theta) = L_t^n(\theta)|_{t=T}$, where $\{L_t^n(\theta),\ 0 \leq t \leq T\}_{n\geq 1} \in \mathcal{M}(\{P_\theta^n\})$ and $Q_\theta^{0,n} = P_\theta^n$, then expansions (3.136) and (3.82) are equivalent and $b^{Q^0}(\theta) = \theta$.

Now let $\gamma = \gamma(n) = c_n(\theta)$ (recall that θ is fixed) and $\widetilde{P}_\theta^n = Q_\theta^{\gamma_n,n}$ be such as in Proposition 3.6. In this case expansion (3.136) also remains true with respect to $\{\widetilde{P}_\theta^n\}_{n\geq 1}$, since $\{\widetilde{P}_\theta^n\} \lhd \{P_\theta^n\}$, i.e.,

$$T_n^L \xrightarrow{\widetilde{P}_\theta^n} \theta \quad \text{as} \ \ n \to \infty.$$

Hence $b^{\widetilde{P}}(\theta) = \theta$ and the asymptotic versions of $T^L = \{T_n^L\}_{n\geq 1}$ for $P_\theta = \{P_\theta^n\}_{n\geq 1}$ and $\widetilde{P}_\theta = \{\widetilde{P}_\theta^n\}_{n\geq 1}$ coincide (and are equal to θ).

Therefore, the direct transference of the notion of the influence functional is impossible.

Nevertheless, relations (3.86) and (3.88) allow us to define an analogous characteristic.

In view of these relations, the expression

$$\frac{c_n(\theta)\langle L_\theta^n, N_\theta^n\rangle_T}{\langle L_\theta^n, L(\dot{M}_\theta^n, M_\theta^n)\rangle_T}$$

can be regarded as a bias at a fixed "nth step" and the expression

$$\widetilde{P}_\theta^n\text{-}\lim_{n\to\infty} c_n^{-1}(\theta)\left(\theta + \frac{c_n(\theta)\langle L_\theta^n, N_\theta^n\rangle_T}{\langle L_\theta^n, L(\dot{M}_\theta^n, M_\theta^n)\rangle_T} - \theta\right) = \frac{\beta_{L,N}(\theta)}{\gamma_L(\theta)}$$

can be regarded as a "bias variation rate" (cf. (3.137)).

Hence, $\dfrac{\beta_{L,N}(\theta)}{\gamma_L(\theta)}$ can be interpreted as an influence functional. At the same time, $\dfrac{\Gamma_L(\theta)}{\gamma_L^2(\theta)}$ is an asymptotic variance of the estimator T^L under the basic (core) sequence of measures $P_\theta = \{P_\theta^n\}_{n\geq 1}$ and, therefore, the solution of the optimization problem based on the risk functional $D(L, N; \theta)$ is equivalent to the construction of the optimal B-robust estimator (see [32]).

2. It is obvious from decomposition (3.119) that the score martingale $L(\dot{M}_\theta^n, M_\theta^n)$ is fully specified by the function $l = (l_c, l_\pi, l_\delta, l_b)$, where the subscripts c, π, δ, and b have the following sense: c corresponds to the continuous part, π to the Poisson-type part, δ to the jumps at predictable moments (including the basic special case, the discrete-time case), and b to the binomial-type part of the score martingale.

3. Consider the case where

$$\begin{aligned} \nu^c(dt, dx) &= \nu^c(dt) P_t(dx), \\ \nu_\theta^c(dt, dx) &= \nu_\theta^c(dt) P_t^\theta(dx), \end{aligned} \tag{3.138}$$

where $\int P_t(dx) = \int P_t^\theta(dx) = 1$ (here the index n is omitted).

In this case,

$$\frac{\dot{F}_\theta}{F_\theta} = \frac{\dot{f}^c_\theta}{f^c_\theta} + \frac{\dot{a}^c_\theta}{a^c_\theta},$$

where

$$f^c_\theta = \frac{dP^\theta}{dP}, \quad a^c_\theta = \frac{d\nu^c_\theta}{d\nu^c}.$$

Obviously,

$$\int \frac{\dot{f}^c_\theta}{f^c_\theta} P^\theta_t(dx) = 0 \quad \forall t.$$

Further, if we denote

$$\psi_{\pi_1} = \psi_\pi - \int \psi_\pi P^\theta_t(dx)$$

and

$$\psi_{\pi_2} = \int \psi_\pi P^\theta_t(dx),$$

then

$$\int \psi_{\pi_1} P^\theta_t(dx) = 0.$$

In this case it is convenient to represent

$$H_\pi = (H_{\pi_1}, H_{\pi_2})$$

with

$$\int H_{\pi_1} P^\theta_t(dx) = 0.$$

Correspondingly, if we again endow the objects under consideration with the index n and introduce the measures

$$\mu^{\theta,n}_{\pi_1}(ds, dx, d\omega) = c^2_n(\theta)\nu^{c,n}_\theta(ds, dx)P^n_\theta(d\omega)$$

and

$$\mu^{\theta,n}_{\pi_2}(ds, d\omega) = c^2_n(\theta)\nu^{c,n}_\theta(ds)P^n_\theta(d\omega),$$

then everywhere there arise new objects with indices π_1 and π_2 instead of the objects with index π, for example,

$$l^n = (l^{\theta,n}_c, l^{\theta,n}_{\pi_1}, l^{\theta,n}_{\pi_2}, l^{\theta,n}_\delta, l^{\theta,n}_b),$$
$$\psi^n = (\psi^{\theta,n}_c, \psi^{\theta,n}_{\pi_1}, \psi^{\theta,n}_{\pi_2}, \psi^{\theta,n}_\delta, \psi^{\theta,n}_b),$$

etc.

4. To make the sense of contamination models clear, let us consider some special cases.

(i) Diffusion-type process. We consider this case in detail.

Let, for each $n \geq 1$, $\xi_n = (\xi_n(t))$, $0 \leq t \leq T$, be a diffusion-type process with the differential

$$d\xi_n(t) = \beta_n(t, \xi_n; \theta)dt + dW_n(t), \quad \xi_n(0) = 0, \tag{3.139}$$

defined on the stochastic basis $(\Omega, \mathcal{F}, F = (\mathcal{F}_t), 0 \leq t \leq T, P)$ with a Wiener process $W_n = (W_n(t), \mathcal{F}_t)$, $0 \leq t \leq T$, given on it, $\theta \in \Theta \subset \mathbb{R}_1$ be an unknown parameter, $\beta_n(t, x, \theta)$ be a nonanticipating functional for each $n \geq 1$, $\theta \in \Theta$.

This case is covered by the general scheme of statistical models $\mathcal{E} = \{\mathcal{E}_n\}_{n\geq 1}$ in the following way.

We set $\Omega^n = C_{[0,T]}$, the space of continuous functions (x_t), $0 \leq t \leq T$, where $x_0 = 0$, $\mathcal{F}^n = \mathcal{B}_T = \sigma(x\colon x_t,\ t \leq T)$, $F^n = (\mathcal{F}^n_t = \sigma(x\colon\ x_s,\ s \leq t))$, $0 \leq t \leq T$, P^n is a Wiener measure, and P^n_θ is the distribution of the process ξ_n (with given θ). In other words, the coordinate process $x = \{x_t(\omega), \omega \in C_{[0,T]},$

$0 \le t \le T\}$, with $x_t(\omega) = \omega_t$, is a (P^n, F)-semimartingale with a triplet $(0, t, 0)$, $0 \le t \le T$, and a (P^n_θ, F)-semimartingale with a triplet

$$\left(\int \beta_n(s, x, \theta)\, ds,\ t, 0\right), \quad 0 \le t \le T.$$

Assume that for each $n \ge 1$,

$$P^n\left(\int_0^T \beta_n^2(t, x, \theta)\, dt < \infty\right) = P^n_\theta\left(\int_0^T \beta_n^2(t, x, \theta)\, dt < \infty\right) = 1.$$

Under these conditions there exists a unique weak solution of Eq. (3.139), $P^n_\theta \sim P^n$, and the likelihood-ratio process has the form

$$\rho^n_\theta = \exp\left(\int_0^t \beta_n(s, x, \theta)\, dx_s - \frac{1}{2}\int_0^t \beta_n^2(s, x, \theta)\, ds\right) = \mathcal{E}_t(M^n_\theta),$$

where

$$M^n_\theta(t) = \int_0^t \beta_n(s, x, \theta)\, dx_s, \quad 0 \le t \le T,$$

is a local (P^n, F)-martingale.

Further, let for each $n \ge 1$, $x \in C_{[0,T]}$, and $t \in [0, T]$, the mapping $\theta \rightsquigarrow \beta_n(t, x, \theta)$ be continuously differentiable $\left(\dfrac{\partial}{\partial\theta}\beta := \dot{\beta}\right)$, and

$$\frac{\partial}{\partial\theta}\int_0^t \beta_n(s, x, \theta)\, dx_s = \int_0^t \dot{\beta}_n(s, x, \theta)\, dx_s,$$
$$\frac{\partial}{\partial\theta}\int_0^t \beta_n(s, x, \theta)\, ds = \int_0^t \dot{\beta}_n(s, x, \theta)\, ds,$$
$$0 < I^n_T(\theta) := E_\theta \int_0^T (\dot{\beta}_n(t, x, \theta))^2 dt < \infty.$$

Then the regularity conditions are satisfied, and

$$L_t(\dot{M}^n_\theta, M^n_\theta) = \frac{\partial}{\partial\theta}\ln \rho^n_\theta(t) = \int_0^t \dot{\beta}_n(s, x, \theta)(dx_s - \beta_n(s, x, \theta) ds) \in \mathcal{M}^2(P^n_\theta),$$
$$I^n_T(\theta) = E^n_\theta \langle L(\dot{M}^n_\theta, M^n_\theta)\rangle_T.$$

Thus,

$$l^{\theta,n}_c(t, x) = \dot{\beta}_n(t, x, \theta),$$
$$c_n^{-2}(\theta) = E^n_\theta \int_0^T (l^{\theta,n}_c(t, x))^2 dt.$$

The ergodicity means that $c_n(\theta) \to 0$ and

$$\int_0^T (\dot{\beta}_n(t, x, \theta))^2 dt \cdot \left(E^n_\theta \int_0^T (\dot{\beta}_n(t, x, \theta))^2 dt\right)^{-1} \xrightarrow{P^n_\theta} 1 \quad \text{as } n \to \infty.$$

The score martingale $L^n_\theta \in \mathcal{M}^2(P^n_\theta)$ is given by the formula

$$L^n_\theta(t) = \int_0^t \psi^{\theta,n}_c(s, x)(dx_s - \beta_n(s, x, \theta) ds), \quad 0 \le t \le T.$$

The "contamination model" means that $\widetilde{P}^n_\theta$ is the distribution of the process $\widetilde{\xi}_n = (\widetilde{\xi}_n(t))$, $0 \le t \le T$, with the differential

$$d\widetilde{\xi}_n(t) = (\beta_n(t, \widetilde{\xi}_n, \theta) + c_n(\theta) H^{\theta,n}_c(t, \widetilde{\xi}_n))dt + dW_n(t), \quad \widetilde{\xi}_n(0) = 0.$$

Hence the main objects are given by the following equalities:

$$N_\theta^n(t) = \int_0^t H_c^{\theta,n}(s,x)(dx_s - \beta_n(s,x,\theta)ds), \quad 0 \le t \le T,$$

$$\mu_c^{\theta,n}(dt,dx) := c_n^2(\theta)dtP_\theta^n(dx), \quad x \in C_{[0,T]},$$

$$D_n(\psi^n, H^n;\theta) = \frac{\left(\int_{C_{[0,T]}}\int_0^T \psi_c^{\theta,n}(t,x)H_c^{\theta,n}(t,x)dtP_\theta^n(dx)\right)^2 + \int_{C_{[0,T]}}\int_0^T (\psi_c^{\theta,n}(t,x))^2 dtP_\theta^n(dx)}{\left(\int_{C_{[0,T]}}\int_0^T \psi_c^{\theta,n}(t,x)l_c^{\theta,n}(t,x)dtP_\theta^n(dx)\right)^2},$$

$$H_R^n = \left\{H_c^{\theta,n} : \int_{C_{[0,T]}}\int_0^T |H_c^{\theta,n}(t,x)|dtP_\theta^n(dx) \le R\right\}.$$

Finally, the optimal score martingale is

$$L_\theta^{*,n}(t) = \int_0^t [\dot\beta_n(s,x,\theta)]_{-m_n^*(\theta)}^{m_n^*(\theta)}(dx_s - \beta_n(s,x,\theta)ds), \quad 0 \le t \le T,$$

where $m_n^*(\theta)$ is the unique solution of the equation

$$R^2m^2 = \int_{C_{[0,T]}}\int_0^T [\dot\beta_n(t,x,\theta)]_{-m}^{m}\beta_n(t,x,\theta)dtP_\theta^n(dx) - \int_{C_{[0,T]}}\int_0^T \left([\dot\beta_n(t,x,\theta)]_{-m}^{m}\right)^2 dtP_\theta^n(dx).$$

In the following special cases, we briefly describe only main objects.

(ii) Poisson-type point process. For each $n \ge 1$, let P_θ^n be the distribution of the point process $\xi_n = (\xi_n(t))$, $0 \le t \le T$, with compensator $\nu_\theta^n(t) = \int_0^t a_\theta^n(s)\,d\alpha_s^n$, and let P^n be the distribution of the point process with compensator α^n.

Then

$$L_t(\dot M_\theta^n, M_\theta^n) = \int_0^t \frac{\dot a_\theta^n(s)}{a_\theta^n(s)}(dx_s - a_\theta^n(s)d\alpha_s^n),$$

$$L_\theta^n(t) = \int_0^t \psi_{\pi_2}^{\theta,n}(dx_s - a_\theta^n(s)d\alpha_s^n),$$

$$I_T^n(\theta) = E_\theta^n \int_0^T \left(\frac{\dot a_\theta^n(t)}{a_\theta^n(t)}\right)^2 a_\theta^n(t)d\alpha_t^n, \quad c_n^{-2}(\theta) = I_T^n(\theta).$$

The "contaminated" measure $\widetilde{P}_\theta^n$ in this case is the distribution of the point process with the compensator

$$\widetilde{\nu}_\theta^n(t) = \int_0^t (a_\theta^n(s) + c_n(\theta)H_{\pi_2}^{\theta,n}(s))d\alpha_s^n.$$

Thus,

$$N_\theta^n(t) = \int_0^t H_{\pi_2}^{\theta,n}(s)(dx_s - a_\theta^n(s)d\alpha_s^n).$$

(iii) Processes with jumps at predictable moments. Consider the case where $X^c = 0$, $\nu^c = 0$. Then from Eqs. (3.120) and (3.122) we find that

$$\left(\frac{\widetilde{a}^n_\theta \widetilde{f}^n_\theta}{a^n_\theta f^n_\theta} - 1\right) I_{\{a^n_\theta > 0\}} + \frac{\widetilde{a}^n_\theta - a^n_\theta}{1 - a^n_\theta} I_{\{0 < a^n_\theta < 1\}} = c_n(\theta)(H^{\theta,n}_\delta I_{\{a^n_\theta > 0\}} + H^{\theta,n}_b I_{\{0 < a^n_\theta < 1\}}). \tag{3.140}$$

To illustrate the role of terms with indices δ and b, let us consider two special cases. Let for each $n \geq 1$:

(1)
$$a^n(t) = a^n_\theta(t) = \widetilde{a}^n_\theta(t) = \begin{cases} 1 & \text{if } t = \dfrac{i}{n} \cdot T, \quad 1 \leq i \leq n, \\ 0, & \text{otherwise.} \end{cases}$$

Then from (3.140) we obtain

$$\frac{\widetilde{f}^n_\theta}{f^n_\theta} - 1 = c_n(\theta) H^{\theta,n}_\delta,$$

i.e., a discrete-time model.

(2) P^n_θ is the distribution of the point process $\xi_n = (\xi_n(t))$, $0 \leq t \leq T$, with the compensator $\nu^n_\theta(dt, dx) = \delta_{\{1\}}(dx)\nu^n_\theta(dt)$, where $\nu^{c,\theta}_\theta(dt) = 0$. Then $a^n_\theta = \Delta\nu^n_\theta(t)$.

Further, let $\widetilde{P}^n_\theta$ be the distribution of the point process $\widetilde{\xi}_n$ with the compensator $\widetilde{\nu}^n_\theta$ possessing the same properties.

Then (3.140) has the form

$$\frac{\widetilde{a}^n_\theta - a^n_\theta}{a^n_\theta(1 - a^n_\theta)} I_{\{0 < a^n_\theta < 1\}} = c_n(\theta) H^{\theta,n}_b I_{\{0 < a^n_\theta < 1\}}.$$

Hence,

$$N^n_\theta(t) = \int_0^t H^{\theta,n}_b I_{\{0 < a^n_\theta < 1\}}(dx_s - \nu^n_\theta(ds)), \quad 0 \leq t \leq T.$$

The simplest case of such a model is the binomial model with a random probability of success, i.e., when the observation is a sequence of indicators $I^n_1, \dots, I^n_n$, $a^n_\theta = P^n_\theta\{I^n_i = 1 \mid \mathcal{F}^n_{i-1}\}$, $1 \leq i \leq n$, $\mathcal{F}^n_i = \sigma(I^n_j,$ $j \leq i)$.

The latter fact explains the meaning of the index b.

3.3.8. Construction of the sequence of optimal score functions. We need some auxiliary notions concerning the weak convergence of σ-finite distributions.

1. Let $\overline{Q}^n(\cdot)$, $n \geq 1$, and $\overline{Q}(\cdot)$ be σ-finite distributions on $(\mathbb{R}_d, \mathcal{B}(\mathbb{R}_d))$, $d = 1, 2$, satisfying the conditions

$$\begin{aligned} &\int |x|^2 \overline{Q}^n(dx) < \infty, \quad n \geq 1, \\ &\int |x|^2 \overline{Q}(dx) < \infty, \end{aligned} \tag{3.141}$$

where $x \in \mathbb{R}_d$, $|\cdot|$ is the usual norm in $\mathbb{R}_d$, $d = 1, 2$.

Suppose that

$$\int |x|^2 \overline{Q}^n(dx) \to \int |x|^2 \overline{Q}(dx) \quad \text{as } n \to \infty. \tag{3.142}$$

Define the sets

$$\overline{C}_d := \Big\{ f : f \text{ is a continuous function on } \mathbb{R}_d,$$
$$f(0) = 0 \text{ and } \frac{f(x)}{|x|^2} \text{ is bounded} \Big\}, \quad d = 1, 2.$$

We say that the sequence of distributions $\{\overline{Q}^n(\cdot)\}_{n\geq 1}$ weakly converges to the distribution $\overline{Q}(\cdot)$ and write

$$\overline{Q}^n \xrightarrow{W} \overline{Q} \quad \text{as} \quad n \to \infty \tag{3.143}$$

if relations (3.141) and (3.142) are satisfied, and

$$\int f(x)\overline{Q}^n(dx) \to \int f(x)\overline{Q}(dx) \quad \text{as} \quad n \to \infty \quad \forall f \in \overline{C}_d, \quad d = 1, 2.$$

2. Let $Q^n = Q^n_{\omega,t}(\cdot)$, $n \geq 1$, and $Q = Q_{\omega,t}(\cdot)$ be regular conditional probability measures on $(\mathbb{R}_d, \mathcal{B}(\mathbb{R}_d))$, $d = 1, 2$. For each $B \in \mathcal{B}(\mathcal{M}_d)$, $d = 1, 2$, let

$$\mathcal{L}^{n,Q}(B) = \int I_{\{Q^n_{\omega,t}(\cdot)\in B\}}\nu^n(d\omega, dt)$$

and

$$\mathcal{L}^{Q}(B) = \int I_{\{Q_{\omega,t}(\cdot)\in B\}}\nu(d\omega, dt)$$

be σ-finite distributions on $(\mathcal{M}_d, \mathcal{B}(\mathcal{M}_d))$ induced by some σ-finite measures $\nu^n(d\omega, dt)$ and $\nu(d\omega, dt)$ defined on $(\Omega \times [0, T],\ \mathcal{F} \times \mathcal{B}([0, T]))$.

Here $(\mathcal{M}_d, \mathcal{B}(\mathcal{M}_d))$, $d = 1, 2$, are measure spaces of the probability measures on $(\mathbb{R}_d, \mathcal{B}(\mathbb{R}_d))$, $d = 1, 2$, and $\mathcal{B}(\mathcal{M}_d)$ is a Borel σ-algebra generated by open (with respect to the Prokhorov metric ρ) sets.

Introduce on $(\mathbb{R}_d, \mathcal{B}(\mathbb{R}_d))$, $d = 1, 2$, the measures

$$\overline{Q}^n(\cdot) = \int \nu(\cdot)\mathcal{L}^{n,Q}(d\nu), \quad n \geq 1, \tag{3.144}$$

and

$$\overline{Q}(\cdot) = \int \nu(\cdot)\mathcal{L}^{Q}(d\nu), \tag{3.145}$$

where $\nu(\cdot) \in \mathcal{M}_d$, and suppose that these measures satisfy conditions (3.141) and (3.142).

Introduce the functional

$$\overline{F}(\nu) = \int |x|^2\nu(dx)$$

defined on the space $(\mathcal{M}_d, \rho)$, $d = 1, 2$.

Then by conditions (3.141) and (3.142) we have

$$\int \overline{F}(\nu)\mathcal{L}^{n,Q}(d\nu) = \int |x|^2\nu(dx)\mathcal{L}^{n,Q}(d\nu) = \int |x|^2\overline{Q}^n(dx) < \infty, \tag{3.146}$$

$$\int \overline{F}(\nu)\mathcal{L}^{Q}(d\nu) = \int |x|^2\overline{Q}(dx) < \infty, \tag{3.147}$$

and

$$\int \overline{F}(\nu)\mathcal{L}^{n,Q}(d\nu) \to \int \overline{F}(\nu)\mathcal{L}^{Q}(d\nu) \quad \text{as} \quad n \to \infty. \tag{3.148}$$

Denote

$$\begin{aligned} C_d := \Big\{ & F : F \ \text{ is a continuous functional on } \ (\mathcal{M}_d, \rho), \quad d = 1, 2; \\ & \text{if } \ \overline{\nu}(\cdot) \ \text{ is a distribution degenerated at } \ 0, \ \text{ then } \ F(\overline{\nu}) = 0; \\ & \frac{F(\nu)}{\overline{F}(\nu)} \ \text{ is bounded} \Big\}, \quad d = 1, 2. \end{aligned}$$

We say that the sequence of random measures $\{Q^n\}_{n\geq 1}$ generalized weakly converges to the random measure Q and write

$$Q^n \Rightarrow Q \quad \text{as} \quad n \to \infty$$

if

$$\mathcal{L}^{n,Q} \xrightarrow{W} \mathcal{L}^Q \quad \text{as } n \to \infty.$$

The latter convergence means that conditions (3.146), (3.147), and (3.148) are satisfied, and

$$\int F(\nu)\mathcal{L}^{n,Q}(d\nu) \to \int F(\nu)\mathcal{L}^Q(d\nu) \quad \text{as } n \to \infty, \quad \forall F \in C_d, \quad d = 1, 2.$$

From definitions, it easily follows that

$$(Q^n \Rightarrow Q) \Rightarrow (\overline{Q}^n \xrightarrow{W} \overline{Q}).$$

Indeed, it is sufficient to take

$$F(\nu) = \int f(x)\nu(dx), \quad f \in C_d, \quad x \in \mathbb{R}_d. \tag{3.149}$$

We consider the case described in Subsection 3.3.5. As follows from (3.125), for each $n \geq 1$,

$$D_n(L^n, N^n; \theta) = D_n(\psi^n, H^n; \theta).$$

Now we define the classes of sequences of functions Ψ and $\mathcal{H}_\Psi$ such that if

$$\psi = \{\psi^n\}_{n\geq 1} \in \Psi \quad \text{and} \quad H = \{H^n\}_{n\geq 1} \in \mathcal{H}_\Psi,$$

then

$$D(\psi, H; \theta) \; (= D(L, N; \theta)) \; = \lim_{n\to\infty} D_n(\psi^n, H^n; \theta)$$

(see (3.98), (3.99), (3.103), and (3.104)).

Then we construct the sequence $\psi^* = \{\psi^{*,n}\}_{n\geq 1}$ of score functions $(\Psi, \mathcal{H}_\Psi)$-optimal in the minimax sense, i.e., the sequence $\psi^* \in \Psi$ such that

$$\sup_{H\in\mathcal{H}_\Psi} D(\psi^*, H; \theta) = \inf_{\psi\in\Psi} \sup_{H\in\mathcal{H}_\Psi} D(\psi, H; \theta)$$

for each $\theta \in \Theta$.

Below the parameter θ is fixed and omitted.

(i) Definition of the class Ψ. Denote by Ψ_α^0, $\alpha = c, \pi, \delta, b$, the class of sequences $\psi_\alpha = \{\psi_\alpha^n\}_{n\geq 1}$ such that

(a)

$$\psi_\alpha^n \in L_2(\mu_\alpha^n) \tag{3.150}$$

for each $n \geq 1$;

(b) for each $n \geq 1$ and $\eta > 0$ there exists a constant $r_\alpha^n = r_\alpha^n(\eta) > 0$ such that

$$P_\theta^n\left(\int I_{\{|u|>\eta\}}\widetilde{Q}_\alpha^{n,\psi}(\omega, du) \leq r_\alpha^n\right) = 1, \tag{3.151}$$

and the sequence $\{r_\alpha^n\}_{n\geq 1}$ is bounded (for each fixed η).

Here

$$\widetilde{Q}_\alpha^{n,\psi}(\omega, \cdot) = \int I_{\{\psi_\alpha^n \in \cdot\}} d\widetilde{\mu}_{\alpha,\omega}^n,$$

where

$$\begin{aligned}
\widetilde{\mu}_{c,\omega}^n(dt) &= c_n^2(\theta)dC_t^n, \\
\widetilde{\mu}_{\pi,\omega}^n(dt, dx) &= c_n^2(\theta)I_{\{a_\theta^n=0\}}\nu_\theta^n(dt, dx), \\
\widetilde{\mu}_{\delta,\omega}^n(dt, dx) &= c_n^2(\theta)I_{\{a_\theta^n>0\}}\nu_\theta^n(dt, dx), \\
\widetilde{\mu}_{b,\omega}^n(dt) &= c_n^2(\theta)p_\theta^n(\omega, dt).
\end{aligned}$$

First we define the classes Ψ_α for $\alpha = c, \pi, b$.

For each $n \geq 1$, we denote

$$\overline{Q}_\alpha^{n,\psi}(\cdot) = \int I_{\{\psi_\alpha^n \in \cdot\}} d\mu_\alpha^n, \quad \overline{Q}_\alpha^{n,\psi^1,\psi^2}(\cdot) = \int I_{\{(\psi_\alpha^{1,n}, \psi_\alpha^{2,n}) \in \cdot\}} d\mu_\alpha^n,$$

where the measures μ_α^n are defined by (3.123).

Fix the sequence $\psi_\alpha^0 \in \Psi_\alpha^0$ and introduce the set

$$\Phi^{\psi_\alpha^0} = \Big\{ \psi \in \Psi_\alpha^0 : \overline{Q}_\alpha^{n,\psi,\psi^0} \xrightarrow{W} \overline{Q}_\alpha^{\psi,\psi^0} \Big\}. \tag{3.152}$$

Let $\Psi_\alpha \subset \Psi_\alpha^0$ be a set of sequences with the following properties:

(1)

$$l_\alpha \in \Psi_\alpha; \tag{3.153}$$

(2)

$$((\psi_\alpha^1, \psi_\alpha^2) \in \Psi_\alpha) \Rightarrow (\psi_\alpha^1 \in \Phi^{\psi_\alpha^2}); \tag{3.154}$$

(3)

$$\Big(\widetilde{\psi}_\alpha \in \bigcap_{\psi_\alpha \in \Psi_\alpha} \Phi^{\psi_\alpha} \Big) \Rightarrow (\widetilde{\psi}_\alpha \in \Psi_\alpha); \tag{3.155}$$

(4) the sequence $\{(\psi_\alpha^n)^2\}_{n \geq 1}$ is uniformly integrable with respect to the sequence of measures $\{\mu_\alpha^n\}_{n \geq 1}$.

Assume, at last, that $\overline{Q}_\alpha^l([-a, a] \setminus \{0\}) > 0$ for each $a > 0$.

Let $\alpha = \delta$.

For each $n \geq 1$, we denote

$$Q_\delta^{n,\psi}(\cdot \mid \omega, t) = \int I_{\{\psi_\delta^n \in \cdot\}} q_{\omega,t}^n(dx),$$
$$Q_\delta^{n,\psi^1,\psi^2}(\cdot \mid \omega, t) = \int I_{\{\psi_\delta^{1,n}, \psi_\delta^{2,n} \in \cdot\}} q_{\omega,t}^n(dx)$$

(the measure $q_{\omega,t}^n(dx)$ is defined by (3.117)).

Fix the sequence $\psi_\delta^0 \in \Psi_\delta^0$ and introduce the set

$$\Phi^{\psi_\delta^0} = \Big\{ \psi \in \Psi_\delta^0 : Q_\delta^{n,\psi,\psi^0} \Rightarrow Q_\delta^{\psi,\psi^0} \Big\}. \tag{3.156}$$

Let $\Psi_\delta \subset \Psi_\delta^0$ be the set of sequences $\psi_\delta = \{\psi_\delta^n\}_{n \geq 1}$ with properties (1), (2), (3), and (4) with $\alpha = \delta$ and, in addition,

(5)

$$\widehat{\psi}_\delta^n = 0 \quad \text{for each} \;\; n \geq 1, \tag{3.157}$$

(6)

$$\begin{aligned} (Q_\delta^{n,\psi} \Rightarrow Q_\delta^{\psi}) \Rightarrow \mathcal{L}^{Q_\delta^{\psi}} \Big(&\{\nu : \nu \;\text{ does not have a unique median}\} \\ &\cup \{\nu : \nu \;\text{ is degenerated in }\; 0\} \cup \Big\{ \nu : \int y \, \nu(dy) \neq 0 \Big\} \Big) = 0. \end{aligned} \tag{3.158}$$

Remark 3.21. Note that from (2) it follows that $\psi_\alpha \in \Phi^{\psi_\alpha}$ (take $\psi_\alpha^1 = \psi_\alpha^2 = \psi_\alpha$) and, therefore,

$$\overline{Q}^{n,\psi_\alpha} \xrightarrow{W} \overline{Q}^{\psi_\alpha}, \quad \alpha = c, \pi, b,$$
$$Q^{n,\psi_\delta} \Rightarrow Q^{\psi_\delta}.$$

Now $\Psi = (\Psi_\alpha, \, \alpha = c, \pi, \delta, b)$.

(ii) Definition of the class $\mathcal{H}_\Psi$. Let $A_\alpha \subset \Psi^0_\alpha$, $\alpha = c, \pi, \delta, b$.

Define for each $\alpha = c, \pi, \delta, b$

$$\mathcal{H}(A_\alpha) = \left\{ \{H^n_\alpha\}_{n\geq 1} : \begin{array}{ll} (1) & \sup\limits_n \int (H^n_\alpha)^2 d\mu^n_\alpha < \infty \\ (2) & \int \psi^n_\alpha H^n_\alpha d\mu^n_\alpha \to, \ \ \forall \psi_\alpha \in A_\alpha \end{array} \right\}. \tag{3.159}$$

The symbol "$a_n \to$" means that the sequence $\{a_n\}_{n\geq 1}$ has a finite limit.

Let $\mathcal{H}(A) = (\mathcal{H}(A_\alpha),\ \alpha = c, \pi, \delta, b)$, where $A = (A_\alpha,\ \alpha = c, \pi, \delta, b)$.

We say that the sequence

$$H = \{H^n\}_{n\geq 1} = \{H^n_\alpha,\ \alpha = c, \pi, \delta, b\}_{n\geq 1} \in B$$

if for each sequence of martingales $\{N^n_\theta\}_{n\geq 1}$ defined by relation (3.122) with $H^n_\alpha \in L_2(\mu^n_\alpha)$, $n \geq 1$, $\alpha = c, \pi, \delta, b$, condition (3.87) of Proposition 3.6 is satisfied.

Denote for each $n \geq 1$

$$\mathcal{K}^n = \{H^n : N^n_\theta \in \mathcal{M}^2(P^n_\theta);\ \langle N^n_\theta \rangle_T \leq K_n\ (P^n_\theta\text{-a.s.}), c_n(\theta)\Delta N^n_\theta \geq R_n > -1\}, \tag{3.160}$$

where the martingale N^n_θ $(= N^n_\theta(H^n))$ is defined by relation (3.122) and K_n and R_n are real numbers. Recall that in (3.122) the index n is omitted.

Define

$$\mathcal{H}_\Psi = \{H = \{H^n\}_{n\geq 1} : H^n \in \mathcal{H}^n_R \cap \mathcal{K}^n,\ n \geq 1;\ H \in \mathcal{H}(\Psi) \cap B\}, \tag{3.161}$$

where $\mathcal{H}^n_R$ is defined by (3.127).

Finally, suppose that

$$\mathcal{L}^{Q^l_\delta}(\{\nu : \nu \ \text{ is degenerate}\} \cup \{\nu : \nu(\operatorname{med}\nu) > 0\}) = 0. \tag{3.162}$$

Remark 3.22. Note that in the case of the discrete time, the property $H^n \in \mathcal{K}^n$ is reduced to the property $\sup\limits_{n,\omega,t} \lambda^n(\omega, t) < \infty$. See also Remark 3.4.

Remark 3.23. Consider, at the qualitative level, the assumptions used in the definition of classes Ψ and $\mathcal{H}_\Psi$. We begin with the class Ψ. Assumption (3.150) reflects the fact that in this work we do not come out of the framework of the L_2-theory. Assumption (3.151) strengthens the admissibility property (see Lemma 3.6), which statistically means that the "unbounded" score functions cannot be optimal, and they a priori are excluded from our consideration. Next, we see that for $\alpha = c, \pi, b$ and for $\alpha = \delta$ the conditions differ from each other and, therefore, they are given separately.

An additional assumption in the case of $\alpha = \delta$ is

$$\widehat{\psi}_\delta = 0,$$

which corresponds to the conditionally centering property of the score function $\psi^n_i(z, \theta \mid x)$ in discrete time models (see Subsec. 3.2.1, the definition of CLAN estimators, (3)). This results in a necessity of considering the parameter $\beta^n - \beta(Q^{n,l}_\delta)$, which is a nonlinear functional defined on the space of probability distributions. Note that technical condition (3.158) and the similar condition (3.162) ensure the continuity of the functional β^n with respect to the topology of a weak convergence of distributions (see Lemma 3.4).

The presence of the parameter β^n in relation (3.134) for the optimal score function $\psi^{*,n}$ complicates the verifying of the ergodic properties and results in a necessity of considering a generalized weak convergence of distributions. This convergence reduces to a weak convergence of distributions on the metric space of probability measures or, roughly speaking, to a weak convergence of distributions of distributions, see (3.156) and (3.158) (cf. (3.152) and (3.154)).

Also, we note that conditions (3.141), (3.142), (3.146), (3.147), and (3.148) are based on condition (3.150) and on property (4) of the definition of the class Ψ_α and, therefore, they appear to be natural.

The necessity of condition (3.153) is obvious. Condition (3.154), which is based on relation (3.152) strengthens, on one hand, the ergodicity properties (3.79) and (3.80) and, on the other hand, ensures asymptotic "homogeneity" of score functions in the class Ψ.

This implies the necessity of the requirement of convergence of joint distributions both in (3.152) and (3.156).

Condition (3.155) ("completeness" of the class Ψ) makes the class of score functions richer.

Proceed to the class $\mathcal{H}_\Psi$, namely, to relation (3.161). Begin with the end of the formula. Belonging of the sequence H to the set $\mathcal{H}(\Psi) \cap B$ guarantees, on one hand, ergodic coordination of the classes Ψ and $\mathcal{H}_\Psi$ (see (3.159), (2)) and, on the other hand, imposes boundedness type conditions on this sequence (see conditions (3.159), (1), and (3.87)). In particular, belonging to the set B implies the contiguity of the sequence of alternative measures $\{\widetilde{P}^n_\theta\}$ with respect to the sequence of basic measures $\{P^n_\theta\}$. Further, belonging of each term H^n of the sequence H to the set $\mathcal{K}^n$ implies a uniform integrability of the exponential martingale $\mathcal{E}(c_n(\theta)N^n_\theta)$ and also the fact that $\inf_t \mathcal{E}_t(c_n(\theta)N^n_\theta) > 0$. These facts imply the property $\widetilde{P}^n_\theta \sim P^n_\theta$.

The nontriviality of the optimization problem (3.101) follows from the relation $H^n \in \mathcal{H}^n_R$.

Finally, we note that the above definitions of the classes Ψ and $\mathcal{H}_\Psi$ give statistical sense to the risk functional $D(\psi, H, \theta)$, to the optimization problem, and, as a result, to the whole problem of robust estimation considered in this chapter.

Let $\psi^* = \{\psi^{*,n}\}_{n\geq 1}$ be the sequence of score functions constructed in Theorem 3.5 by relation (3.134).

Theorem 3.6. *The sequence ψ^* is $(\Psi, \mathcal{H}_\Psi)$-optimal.*

Proof. The method of proving Theorem 3.4 developed in Subsection 3.2.3, is general and can be used with small changes in the considered case.

We illustrate this fact, taking as an example the proof of Lemma 3.9.

Introduce the functions

$$H^{n,\varepsilon}_\alpha = \frac{R I_{\{\psi^n_\alpha > \eta_\alpha(1-\varepsilon)\}}}{\displaystyle\int I_{\{\psi^n_\alpha > \eta_\alpha(1-\varepsilon)\}} d\mu^n_\alpha}, \quad \alpha = c, \pi, b;$$

$$H^{n,\varepsilon}_\delta = \frac{R\left(I_{\{\psi^n_\delta > \eta_\delta(1-\varepsilon)\}} - \displaystyle\int I_{\{\psi^n_\delta > \eta_\delta(1-\varepsilon)\}} q^n_{\omega,t}(dx)\right)}{\displaystyle\int I_{\{\psi^n_\delta > \eta_\delta(1-\varepsilon)\}} d\mu^n_\delta},$$

where $\eta_\alpha := \operatorname*{ess\,sup}_{\overline{Q}^\psi_\alpha} |x|$.

Obviously, $\widehat{H}^{n,\varepsilon}_\delta = 0$ and for each $n \geq 1$

$$c^2_n(\theta)\langle N^{n,\varepsilon}_\theta\rangle_T = \sum_\alpha \int (H^{n,\varepsilon}_\alpha)^2 d\widetilde{\mu}^n_\alpha \leq \sum_\alpha \frac{R^2 \displaystyle\int I_{\{u > \eta_\alpha(1-\varepsilon)\}} \widetilde{Q}^{n,\psi}_\alpha(\omega, du)}{\left(\displaystyle\int I_{\{\psi^n_\alpha > \eta_\alpha(1-\varepsilon)\}} d\mu^n_\alpha\right)^2} \leq K_n \quad P^n_\theta\text{-a.s.}, \tag{3.163}$$

for some constant K_n, $0 < K_n < \infty$, thanks to (3.151).

Thus, the process $\mathcal{E}(c_n(\theta)N^{n,\varepsilon}_\theta)$ is square integrable and hence a uniformly integrable martingale. In particular, $E^n_\theta \mathcal{E}(c_n(\theta)N^{n,\varepsilon}_\theta) = 1$.

Moreover, obviously, $c_n(\theta)\Delta N^{n,\varepsilon}_\theta \geq R_n > -1$ for some constant R_n.

Indeed, $H_\alpha^{n,\varepsilon} \ge 0$, $\alpha = c, \pi, b$, and

$$H_\delta^{n,\varepsilon} \ge \frac{-R\displaystyle\int I_{\{\psi_\delta^n > \eta_\delta(1-\varepsilon)\}} q_{\omega,t}^n(dx)}{\displaystyle\int I_{\{\psi_\delta^n > \eta_\delta(1-\varepsilon)\}} d\mu_\delta^n}.$$

Hence if we denote by $\lambda^n(\omega, t)$ the last expression without the sign "$-$," we obtain that $c_n(\theta)\lambda^n(\omega, t) < 1$ for some $n \ge 1$, by virtue of condition (3.151) and the fact that

$$\inf_n \int I_{\{\psi_\delta^n > \eta_\delta(1-\varepsilon)\}} d\mu_\delta^n > 0.$$

The latter follows from the definition of η_δ.

Without loss of generality, we assume that $c_n(\theta)\lambda^n(\omega, t) < 1$ for any $n \ge 1$ (see also Remark 3.4).

Therefore, $\widetilde{P}_\theta^n := \mathcal{E}(c_n(\theta) N_\theta^{n,\varepsilon}) \cdot P_\theta^n$ is a probability measure, equivalent for each $n \ge 1$ to the measure P_θ^n, $\widetilde{P}_\theta^n \sim P_\theta^n$.

We prove now that the sequence $\{c_n^2(\theta)\langle N_\theta^{n,\varepsilon}\rangle_T\}_{n\ge 1}$ is stochastically bounded with respect to the sequence of measures $\{\widetilde{P}_\theta^n\}_{n\ge 1}$.

Denote

$$b_\alpha^n = \int I_{\{\psi_\alpha^n > \eta_\alpha(1-\varepsilon)\}} d\mu_\alpha^n, \quad \alpha = c, \pi, \delta, b.$$

From (3.163) we obtain that for all $n \ge 1$,

$$P_\theta^n\{\omega : c_n^2(\theta)\langle N_\theta^{n,\varepsilon}\rangle_T \le K_n\} = 1.$$

But, as we have proved above, $\widetilde{P}_\theta^n \sim P_\theta^n$. Hence

$$\widetilde{P}_\theta^n\{\omega : c_n^2(\theta)\langle N_\theta^{n,\varepsilon}\rangle_T \le K_n\} = 1.$$

Thus,

$$\overline{\lim_{n\to\infty}}\, \widetilde{P}_\theta^n\{c_n^2(\theta)\langle N_\theta^{n,\varepsilon}\rangle_T > d\} \le \overline{\lim_{n\to\infty}}\, \widetilde{P}_\theta^n\{K_n > d\} \to 0 \quad \text{as } d \to \infty,$$

since the sequence of numbers $\{K_n\}_{n\ge 1}$ is bounded. Indeed, $\inf\limits_n b_\alpha^n > 0$ by the definition of η_α and the sequence $\{r_\alpha^n\}_{n\ge 1}$ from (3.151) is bounded for each $\alpha = c, \pi, \delta, b$.

Now, according to Remark 3.19, if $\eta_j = \max\limits_\alpha\{\eta_\alpha\}$, we take as $\overline{H}^{n,\varepsilon}$ the function

$$\overline{H}^{n,\varepsilon} = (0, \overline{H}_j^{n,\varepsilon}, 0, 0),$$

where

$$\overline{H}_j^{n,\varepsilon} = \frac{R I_{\{\psi_j^n > \eta_j(1-\varepsilon)\}}}{\int I_{\{\psi_j^n > \eta_j(1-\varepsilon)\}} d\mu_j^n}.$$

Note that condition (3.127) is trivially satisfied.

Hence, we obtain

$$\overline{H}^{n,\varepsilon} \in \overline{\mathcal{H}}_\Psi.$$

The assertion follows. □

Remark 3.24. The assertions of Corollaries 3.3 and 3.4 with obvious notational changes are also true.

CHAPTER 4

THE ROBBINS–MONRO-TYPE STOCHASTIC DIFFERENTIAL EQUATIONS. CONVERGENCE OF SOLUTIONS AND STRONG CONSISTENCY

The Robbins–Monro-type (RM-type) SDEs are introduced. This type of SDEs naturally includes both generalized RM stochastic approximation algorithms with martingale noises and recursive estimation procedures for general statistical models. The approach of the investigation of the a.s. convergence as $t \to \infty$ of the strong solution $Z = (Z_t)_{t\geq 0}$ of such type of equations is proposed. This approach is based on the new description of the convergence sets of semimartingales and nonstandard representation of the process of bounded variation (in the decomposition of the special semimartingale $(Z_t^2)_{t\geq 0}$) in the form of the difference of two increasing predictable processes.

4.1. Specification of the Model. Standard and Nonstandard Representations

1. Let the following objects be given on the stochastic basis $(\Omega, \mathcal{F}, F = (\mathcal{F}_t)_{t\geq 0}, P)$:

(a) a random field $H = \{H_t(u),\ t \geq 0,\ u \in \mathbb{R}_1\} = \{H_t(\omega, u),\ t \geq 0,\ \omega \in \Omega,\ u \in \mathbb{R}_1\}$ with properties

(i) for each $u \in \mathbb{R}_1$, the process $H(u) = (H_t(u))_{t\geq 0} \in \mathcal{P}$ (i.e., is predictable);

(ii) for each $t \geq 0$,

$$(A) \qquad \begin{aligned} H_t(u) = 0 \quad &\text{if} \quad u = 0, \\ H_t(u)u < 0 \quad &\text{for all} \quad u \neq 0 \end{aligned}$$

P-a.s.;

(b) a random field $M = \{M(t,u),\ t \geq 0,\ u \in \mathbb{R}_1\} = \{M(\omega, t, u),\ \omega \in \Omega,\ t \geq 0,\ u \in \mathbb{R}_1\}$ such that for each $u \in \mathbb{R}_1$ the process $M(u) = (M(t,u))_{t\geq 0} \in \mathcal{M}^2_{\text{loc}}$;

(c) a predictable increasing process $K = (K_t)_{t\geq 0}$ (i.e., $K \in \mathcal{V}^+ \cap \mathcal{P}$).

We restrict the consideration to the following particular cases: for each $u \in \mathbb{R}_1$,

1°. $M(u) \equiv m \in \mathcal{M}^2_{\text{loc}}$.

2°. $M(u) = f(u) \cdot m + g(u) \cdot n$, where $m \in \mathcal{M}^c_{\text{loc}}$, $n \in \mathcal{M}^{d,2}_{\text{loc}}$, the predictable processes $f(u) = f(t,u))_{t\geq 0}$ and $g(u) = (g(t,u))_{t\geq 0}$ are such that the corresponding stochastic integrals are well defined, and $M(u) \in \mathcal{M}^2_{\text{loc}}$.

3°. $M(u) = \varphi(u) \cdot m + W(u) * (\mu - \nu)$, where $m \in \mathcal{M}^c_{\text{loc}}$, the process $\varphi(u) = (\varphi(t,u))_{t\geq 0}$ is predictable, μ is an integer-valued random measure on $(\Omega \times \mathbb{R}_+ \times E, \mathcal{F} \otimes \mathcal{B}(\mathbb{R}_+) \times \mathcal{E})$, ν is its P-compensator, $(E, \mathcal{E})$ is the Blackwell space, and $W(u) = (W(t,x,u),\ t \geq 0,\ x \in E) \in \mathcal{P} \otimes \mathcal{E}$ (here all stochastic integrals are assumed to be well defined, and $M(u) \in \mathcal{M}^2_{\text{loc}}$).

Later on, by the symbol $\int_0^t M(ds, u_s)$, where $u = (u_t)_{t\geq 0}$ is a predictable process, we denote the following stochastic integrals:

$$\int_0^t f(s, u_s)\, dm_s + \int_0^t g(s, u_s)\, dn_s \quad (\text{in case } 2^\circ)$$

or

$$\int_0^t \varphi(s, u_s)\, dm_s + \int_0^t \int_E W(s, x, u_s)(\mu - \nu)(ds, dx) \quad (\text{in case } 3^\circ),$$

provided the latter are well defined.

Consider the stochastic equation (RM procedure)

$$Z_t = Z_0 + \int_0^t H_s(Z_{s-})\, dK_s + \int_0^t M(ds, Z_{s-}), \quad t \geq 0, \quad Z_0 \in \mathcal{F}_0, \tag{4.1}$$

or the differential form

$$dZ_t = H_t(Z_{t-})dK_t + M(dt, Z_{t-}), \quad Z_0 \in \mathcal{F}_0. \tag{4.2}$$

We say that SDE (4.1) is of the Robbins–Monro type, since the drift coefficient satisfies the specific condition (A).

Assume that there exists a unique strong solution $Z = (Z_t)_{t\geq 0}$ of Eq. (4.1) on the whole time interval $[0, \infty)$, $\widetilde{M} \in \mathcal{M}^2_{\mathrm{loc}}$, where

$$\widetilde{M}_t := \int_0^t M(ds, Z_{s-}).$$

Certain sufficient conditions for this can be found in [23, 25, 44].

We study the problem of P-a.s. convergence $Z_t \to 0$ as $t \to \infty$.

2. We need some facts concerning semimartingale convergence sets [118].

For completeness we formulate them.

Let $X_\infty = \lim\limits_{t\to\infty} X_t$, and let $\{X \to\}$ denote a set where X_∞ exists and is a finite random variable.

Denote by $\mathcal{V}^+(\mathcal{V})$ a set of processes $A = (A_t)_{t\geq 0}$, $A_0 = 0$, $A \in F \cap D$ (i.e., the process A is F-adapted with cadlag trajectories) with nondecreasing (bounded variation on each interval $[0, t]$) trajectories. We write $X \in \mathcal{P}$ if X is a predictable process. Denote by S_P a class of special semimartingales, i.e., $X \in S_P$ if $X \in F \cap D$ and

$$X = X_0 + A + M,$$

where $A \in \mathcal{V} \cap \mathcal{P}$, $M \in \mathcal{M}_{\mathrm{loc}}$.

If $\Gamma_1, \Gamma_2 \in \mathcal{F}$, then $\Gamma_1 = \Gamma_2$ (P-a.s.) or $\Gamma_1 \subseteq \Gamma_2$ (P-a.s.) which means that $P(\Gamma_1 \Delta \Gamma_2) = 0$ or $P(\Gamma_1 \cap (\Omega \setminus \Gamma_2)) = 0$, respectively, where Δ denotes the symmetric difference of sets.

Let $X \in S_P$. We set $A = A^1 - A^2$, where $A^1, A^2 \in \mathcal{V}^+ \cap \mathcal{P}$. Denote

$$\widehat{A} = (1 + X_- + A^2_-)^{-1} \circ A^1 \quad \left(:= \int_0^{\cdot} (1 + X_{s-} + A^2_{s-})^{-1} dA^1_s\right).$$

Theorem 4.1. *Let $X \in S_P$, $X \geq 0$. Then*

$$\{\widehat{A}_\infty < \infty\} \subseteq \{X \to\} \cap \{A^2_\infty < \infty\} \quad (P\text{-}a.s.).$$

Corollary 4.1.

$$\{A^1_\infty < \infty\} = \{(1 + X_-)^{-1} \circ A^1_\infty < \infty\} = \{\widehat{A}_\infty < \infty\} \quad (P\text{-}a.s.).$$

Remark 4.1. The theorems below have been proved in [67].

Introduce the following assumptions:

(1) $EX_0 < \infty$;

(2) one of the conditions (α) and (β) below is satisfied:

(α) there exists $\varepsilon > 0$ such that $A^1_{t+\varepsilon} \in \mathcal{F}_t$ for all $t > 0$,

(β) for any predictable Markov moment σ,

$$E\Delta A^1_\sigma I_{\{\sigma < \infty\}} < \infty.$$

Theorem (A). *Let $X \in S_p$, $X \geq 0$, $X = X_0 + A^1 - A^2 + M$, $A^1, A^2 \in \mathcal{V}^+ \cap \mathcal{P}$, $M \in \mathcal{M}_{\mathrm{loc}}$, and let assumptions* (1) *and* (2) *be satisfied. Then*

$$\{A^1_\infty < \infty\} \subseteq \{X \to\} \cap \{A^2_\infty < \infty\} \quad (P\text{-}a.s.).$$

Theorem (B). *Let $X \in S_p$, $X \geq 0$, $X = X_0 + X_- \circ B + A^1 - A^2 + M$, $B, A^1, A^2 \in \mathcal{V}^+ \cap \mathcal{P}$, $M \in \mathcal{M}_{\mathrm{loc}}$, and let assumptions* (1) *and* (2) *(only for A^1) be satisfied. Then*

$$\{A^1_\infty < \infty\} \cap \{B_\infty < \infty\} \subseteq \{X \to\} \cap \{A^2_\infty < \infty\} \quad (P\text{-}a.s.).$$

Theorem (B) follows from Theorem (A) if we apply the latter to the process $X \cdot \mathcal{E}^{-1}(B)$, where $\mathcal{E}(B)$ is the Dolean exponential, i.e., the solution of the equation $Y = 1 + Y_- \circ B$.

Now we reject assumptions (1) and (2) in both theorems based only on these theorems. Indeed, consider the process $Y = 1 + X$. Obviously, $\{X \to\} = \{Y \to\}$ (P-a.s.). Introduce the process

$$\widetilde{A} = \frac{1}{Y_-} \circ A^1 \in \mathcal{V}^+ \cap \mathcal{P}.$$

We have

$$X = X_0 + A^1 - A^2 + M \Rightarrow Y = Y_0 + A^1 - A^2 + M \equiv Y_0 + Y_- \circ \widetilde{A} - A^2 + M,$$

i.e., we obtain the decomposition of the process Y with $A^1 = 0$ and $B = \widetilde{A}$ (see Theorem (B)).

Now Theorem (B) yields that only under assumption (1),

$$\{\widetilde{A}_\infty < \infty\} \subseteq \{Y \to\} \cap \{A^2_\infty < \infty\} = \{X \to\} \cap \{A^2_\infty < \infty\} \quad (P\text{-a.s.}).$$

But $\{\widetilde{A}_\infty < \infty\} = \{A^1_\infty < \infty\}$ (P-a.s.), since $\{A^1_\infty < \infty\} \subseteq \{\widetilde{A}_\infty < \infty\}$ (P-a.s.) and

$$\{A^1_\infty < \infty\} \cap \{X \to\} \cap \{A^2_\infty < \infty\} = \{\widetilde{A}_\infty < \infty\} \cap \{X \to\} \cap \{A^2_\infty < \infty\} \quad (P\text{-a.s.}).$$

Thus, we obtain that the assertion of Theorem (A) is true without assumption (2). Applying the just-proved fact to $X \cdot \mathcal{E}^{-1}(B)$, we obtain that the assertion of Theorem (B) is also true without assumption (2).

Further, we denote

$$X^l = X_0 \wedge l + A^1 - A^2 + M,$$

for each constant $l > 0$. Obviously, $EX^l_0 = F(X_0 \wedge l) \le l$.

Then, by Theorem (A),

$$\{A^1_\infty < \infty\} \subseteq \{X^l \to\} \cap \{A^2_\infty < \infty\} \quad (P\text{-a.s.}).$$

Hence $\{A^1_\infty < \infty\} \subseteq \{A^2_\infty < \infty\}$ and

$$\begin{aligned}\{A^1_\infty < \infty\} \subseteq \{X^l \to\} &= \{X^l \to\} \cap (\{X = X^l\} + \{X \ne X^l\}) \\ &= \{X \to\} \cap \{X = X^l\} + \{X^l \to\} \cap \{X \ne X^l\} \\ &\subseteq \{X \to\} \cap \{X_0 \le l\} + \{X_0 > l\} \quad (P\text{-a.s.})\end{aligned}$$

for each $l > 0$. Now we note that $\{X_0 \le l\} \uparrow \Omega$ (P-a.s.) and $\{X_0 > l\} \downarrow \varnothing$ (P-a.s.) as $l \to \infty$.

Thus,

$$\begin{aligned}\{A^1_\infty < \infty\} &\subseteq \{X \to\} \cap \lim_{l\to\infty} \{X_0 \le l\} + \lim_{l\to\infty} \{X_0 > l\} \\ &= \{X \to\} \cap \Omega + \varnothing = \{X \to\} \quad (P\text{-a.s.}).\end{aligned}$$

Hence, we conclude that the assertion of Theorem (A) is true without condition (1) as well.

Corollary. *The assertions of Theorems* (A) *and* (B) *are true without assumptions* (1) *and* (2).

Remark 4.2. In [91], Theorem (A) was proved without assumptions (1) and (2). But the proof is not correct.

Apply Theorem 4.1 to the semimartingale $X_t = Z_t^2$, $t \ge 0$. Using the Itô formula, we obtain the following for the process $(Z_t^2)_{t\ge0}$:

$$dZ_t^2 = dA_t + dN_t, \tag{4.3}$$

where

$$\begin{aligned}dA_t &= \alpha_t(Z_{t-})dK_t + \beta_t(Z_{t-})dK_t^d + d\langle\widetilde{M}\rangle_t, \\ dN_t &= 2Z_{t-}d\widetilde{M}_t + H_t(Z_{t-})\Delta K_t d\widetilde{M}_t^d + d([\widetilde{M}]_t - \langle\widetilde{M}\rangle_T),\end{aligned}$$

with

$$\alpha_t(u) := 2H_t(u)u,$$
$$\beta_t(u) := H_t^2(u)\Delta K_t.$$

Note that $A = (A_t)_{t\geq 0} \in \mathcal{V} \cap \mathcal{P}$, $N \in \mathcal{M}_{\text{loc}}$.

Represent the process A in the form

$$A_t = A_t^1 - A_t^2 \tag{4.4}$$

with

(1) $$\begin{cases} dA_t^1 = \beta_t(Z_{t-})dK_t^d + d\langle \widetilde{M} \rangle_t, \\ -dA_t^2 = \alpha_t(Z_{t-})dK_t, \end{cases}$$

or

(2) $$\begin{cases} dA_t^1 = [\alpha_t(Z_{t-})I_{\{\Delta K_t \neq 0\}} + \beta_t(Z_{t-})]^+ dK_t^d + d\langle \widetilde{M} \rangle_t, \\ -dA_t^2 = \{\alpha_t(Z_{t-})I_{\{\Delta K_t = 0\}} - [\alpha_t(Z_{t-})I_{\{\Delta K_t \neq 0\}} + \beta_t(Z_{t-})]^- \}dK_t, \end{cases}$$

where $[a]^+ = \max(0, a)$ and $[a]^- = -\min(0, a)$.

As follows from (A), $\alpha_t(Z_{t-}) \leq 0$ for all $t \geq 0$ and, therefore, representation (4.3) (1) directly corresponds to the usual standard form of the process A (in (4.3), $A = A^1 - A^2$, with A^1 and A^2 from (4.4) (1)). Therefore, representation (4.4) (1) of the process A is said to be standard, while representation (4.4) (2) is said to be nonstandard.

4.2. The Convergence Theorem

Introduce the following group of conditions: for all $u \in \mathbb{R}_1$ and $t \in [0, \infty)$,

(B) (i) $\langle M(u) \rangle \ll K$,

(ii) $h_t(u) \leq B_t(1 + u^2)$, $B_t \geq 0$, $B = (B_t)_{t\geq 0} \in \mathcal{P}$, $B \circ K_\infty < \infty$, where $h_t(u) = \dfrac{d\langle M(u) \rangle_t}{dK_t}$;

(I) (i) (i$_1$) $I_{\{\Delta K_t \neq 0\}} |H_t(u)| \leq C_t(1 + |u|)$, $C_t \geq 0$, $C = (C_t)_{t\geq 0} \in \mathcal{P}$, $C \circ K_t < \infty$,

(i$_2$) $C^2 \Delta K \circ K_\infty^d < \infty$,

(ii) for each $\varepsilon > 0$,

$$\inf_{\varepsilon \leq |u| \leq 1/\varepsilon} |\alpha(u)| \circ K_\infty = \infty;$$

(II) (i) $[\alpha_t(u)I_{\{\Delta K_t \neq 0\}} + \beta_t(u)]^+ \leq D_t(1 + u^2)$, $D_t \geq 0$, $D = (D_t)_{t\geq 0} \in \mathcal{P}$, $D \circ K_\infty^d < \infty$,

(ii) for each $\varepsilon > 0$,

$$\inf_{\varepsilon \leq |u| \leq 1/\varepsilon} \{|\alpha(u)| I_{\{\Delta K = 0\}} + [\alpha(u) I_{\{\Delta K \neq 0\}} + \beta(u)]^- \} \circ K_\infty = \infty.$$

Remark 4.3. In the above-mentioned case 1° for $M(u) \equiv m \in \mathcal{M}_{\text{loc}}^2$, we do not require the condition $\langle m \rangle \ll K$ and replace condition (B) by

$$(B') \qquad \langle m \rangle_\infty < \infty.$$

Remark 4.4. Everywhere we assume that all conditions are satisfied P-a.s.

Remark 4.5. It is obvious that (I) (ii) $\Rightarrow C \circ K_\infty = \infty$.

Theorem 4.2. *Let conditions* (A), (B), (I) *or* (A), (B), (II) *be satisfied. Then*

$$Z_t \to 0 \quad P\text{-}a.s. \quad \text{as } t \to \infty.$$

Proof. Assume, for example, that conditions (A), (B), and (I) are satisfied. Then by virtue of Corollary 4.1 and (4.3) with standard representation (4.4) (1) of the process A, we obtain

$$\{(1+Z_-^2)^{-1}\circ A_\infty^1<\infty\}\subseteq\{Z^2\to\}\cap\{A_\infty^2<\infty\}. \tag{4.5}$$

But from conditions (B) and (I) (i) we find that

$$\{(1+Z_-^2)^{-1}\circ A_\infty^1<\infty\}=\Omega\quad (P\text{-a.s.})$$

and, therefore,

$$\{Z^2\to\}\cap\{A_\infty^2<\infty\}=\Omega\quad (P\text{-a.s.}). \tag{4.6}$$

Denote $Z_\infty^2=\lim\limits_{t\to\infty}Z_t^2$, $N=\{Z_\infty^2>0\}$, and assume that $P(N)>0$. In this case, from (I) (ii), by simple arguments, we obtain

$$P(|\alpha(Z_-)|\circ K_\infty=\infty)>0,$$

which contradicts Eq. (4.6). Hence $P(N)=0$.

The proof of the second case is similar. □

In the following propositions the relationship between conditions (I) and (II) is given.

Proposition 4.1. (I) $\Rightarrow$ (II).

Proof. From (I) (i_1) we find

$$[\alpha_t(u)I_{\{\Delta K_t\neq 0\}}+\beta_t(u)]^+\le\beta_t(u)\le C_t^2\Delta K_t(1+u^2),$$

and if we take $D_t=C_t^2\Delta K_t$, then (II) (i) follows from (I) (i_2).

Further, from (I) (i_1) we obtain

$$|\alpha_t(u)|I_{\{\Delta K_t=0\}}+[\alpha_t(u)+\beta_t(u)]^-I_{\{\Delta K_t\neq 0\}}\ge|\alpha_t(u)|-\beta_t(u)\ge|\alpha_t(u)|-C_t^2\Delta K_t\left(1+\frac{1}{\varepsilon^2}\right)$$

for each $\varepsilon>0$ and u with $\varepsilon\le|u|\le\dfrac{1}{\varepsilon}$.

Now (II) (ii) follows from (I) (i_2) and (I) (ii). □

Proposition 4.2. *Under* (I) (i) *we have* (I) (ii) $\Leftrightarrow$ (II) (ii).

The proof immediately follows from the previous proposition and the trivial implication (II) (ii) $\Rightarrow$ (I) (ii).

4.3. Simple Sufficient Conditions for (I) and (II)

Introduce the following group of conditions: for each $u\in\mathbb{R}_1$ and $t\in[0,\infty)$,

(S.1)

(i) (i_1) $G_t|u|\le|H_t(u)|\le\widetilde{G}_t|u|,\ \ G_t\ge 0,\ \ G=(G_t)_{t\ge 0},\ \ \widetilde{G}=(\widetilde{G}_t)_{t\ge 0}\in\mathcal{P},\ \ \widetilde{G}\circ K_t<\infty,$

(i_2) $\widetilde{G}^2\Delta K\circ K_\infty^d<\infty;$

(ii) $G\circ K_\infty=\infty;$ (4.7)

(S.2)

$$\text{(i)}\quad \widetilde{G}[-2+\widetilde{G}\Delta K]^+\circ K_\infty^d<\infty; \tag{4.8}$$

$$\text{(ii)}\quad G\{2I_{\{\Delta K=0\}}+[-2+\widetilde{G}\Delta K]^-I_{\{\Delta K\neq 0\}}\}\circ K_\infty=\infty. \tag{4.9}$$

Proposition 4.3.

$$\text{(S.1)}\Rightarrow\text{(I)},$$
$$\text{(S.1)}\,(\mathrm{i}_1),\ \text{(S.2)}\Rightarrow\text{(II)}.$$

Proof. The first implication is obvious. As for the second one, we note that

$$\begin{aligned}\alpha_t(u)I_{\{\Delta K_t\neq 0\}} + \beta_t(u) &= -2|H_t(u)||u|I_{\{\Delta K_t\neq 0\}} + H_t^2(u)\Delta K_t \\ &\leq |H_t(u)||u|[-2I_{\{\Delta K_t\neq 0\}} + \widetilde{G}_t\Delta K_t].\end{aligned} \tag{4.10}$$

Thus,

$$\begin{aligned}[\alpha_t(u)I_{\{\Delta K_t\neq 0\}} + \beta_t(u)]^+ &\leq |H_t(u)||u|[-2I_{\{\Delta K_t\neq 0\}} + \widetilde{G}_t\Delta K_t]^+ \\ &\leq \widetilde{G}_t[-2I_{\{\Delta K_t\neq 0\}} + \widetilde{G}_t\Delta K_t]^+|u^2|\end{aligned}$$

and (II) (i) follows from inequality (4.8) if we take

$$D_t = \widetilde{G}_t[-2 + \widetilde{G}_t\Delta K_t]^+ I_{\{\Delta K_t\neq 0\}}.$$

Further, from inequality (4.10) we find that

$$|\alpha_t(u)|I_{\{\Delta K_t=0\}} + [\alpha_t(u)I_{\{\Delta K_t\neq 0\}} + \beta_t(u)]^- \geq u^2 G_t\{2I_{\{\Delta K_t=0\}} + [-2I_{\{\Delta K_t\neq 0\}} + \widetilde{G}_t\Delta K_t]^-\}$$

and (II) (ii) follows from (4.9). □

Remark 4.6. (a) (S.1) ⇒ (S.2);

(b) under (S.1) (i) we have (S.1) (ii) ⇔ (S.2) (ii);

(c) (S.2) (ii) ⇒ (S.1) (ii).

Summarizing the above results, we come to the following conclusions: (a) if condition (S.1) (ii) is not satisfied, neither is condition (S.2) (ii); (b) if conditions (S.1) (i_1) and (S.1) (ii) are satisfied but condition (S.1) (i_2) is violated, then conditions (S.2) (i) and (S.2) (ii) can nevertheless be satisfied.

In this case, the nonstandard representation (4.4) (2) is useful.

Remark 4.7. Denote

$$\widetilde{G}_t\Delta K_t = 2 + \delta_t, \quad \delta_t \geq -2 \quad \text{for all } \ t \in [0,\infty).$$

It is obvious that if $\delta_t \leq 0$ for all $t \in [0,\infty)$, then $[-2 + \widetilde{G}_t\Delta K_t]^+ = 0$. Therefore, condition (S.2) (i) is trivially satisfied and (S.2) (ii) takes the form

$$G\{2I_{\{\Delta K=0\}} + |\delta|I_{\{\Delta K\neq 0\}}\} \circ K_\infty = \infty. \tag{4.11}$$

Note that if $G\min(2,|\delta|)\circ K_\infty = \infty$, then (4.11) holds, and the simplest sufficient condition for (4.11) is

$$G \circ K_\infty = \infty, \quad |\delta_t| \geq \text{const} > 0$$

for all $t \geq 0$.

Remark 4.8. Let conditions (A), (B), and (I) be satisfied. Since we apply Theorem 4.1 and its corollary on the semimartingale convergence sets, we get rid of many "usual" restrictions such as "moment" restrictions, boundedness of the regression function, etc.

Remark 4.9. As an example of a nonstandard representation we tried to show to what extent one of numerous possible representations of the process A from (3.2) can be useful. Obviously, starting from the purposes of statistical problems, some other useful representations are possible.

4.4. Special Models

1. Discrete time.

(a) Recursive MLE in parametric statistical models. Let $X_0, X_1, ..., X_n, ...$ be observations taking values in a measure space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that there exist regular conditional densities of distributions (w.r.t. a measure μ) $f_i(x_i, \theta \mid x_{i-1}, ..., x_0)$, $i \le n$, $n \ge 1$, and $\theta \in \mathbb{R}_1$ is the parameter to be estimated. Denote by P_θ the corresponding distribution on $(\Omega, \mathcal{F}) := (\mathcal{X}^\infty, \mathcal{B}(\mathcal{X}^\infty))$. Identify the process $X = (X_i)_{i \ge 0}$ with the coordinate process and denote $\mathcal{F}_n = \sigma(X_i,\, i \le n)$. If $\psi = \psi(X_i, X_{i-1}, ..., X_0)$ is a r.v., then by $E_\theta(\psi \mid \mathcal{F}_{i-1})$ we mean the following version of the conditional expectation:

$$E_\theta(\psi \mid \mathcal{F}_{i-1}) := \int \psi(z, X_{i-1}, ..., X_0) f_i(z, \theta \mid X_{i-1}, ..., X_0)\, \mu(dz)$$

if the last integral exists.

Assume that the usual regularity conditions are satisfied and denote

$$\dot{f}_i(x_i, \theta \mid x_{i-1}, ..., x_0) := \frac{\partial}{\partial \theta} f_i(x_i, \theta \mid x_{i-1}, ..., x_0),$$

the maximum likelihood scores by

$$l_i(\theta) = \frac{\dot{f}_i}{f_i}(X_i, \theta \mid X_{i-1}, ..., X_0),$$

and the empirical Fisher information by

$$I_n(\theta) := \sum_{i=1}^{n} E_\theta(l_i^2(\theta) \mid \mathcal{F}_{i-1}).$$

Also, we denote

$$b_n(\theta, u) := E_\theta(l_n(\theta + u) \mid \mathcal{F}_{n-1})$$

and show that for each $\theta \in \mathbb{R}_1$, $n \ge 1$,

$$b_n(\theta, 0) = 0 \quad (P\text{-a.s.}). \tag{4.12}$$

Consider the following recursive procedure:

$$\theta_n = \theta_{n-1} + I_n^{-1}(\theta_{n-1}) l_n(\theta_{n-1}), \quad \theta_0 \in \mathcal{F}_0.$$

Fixing θ and denoting $Z_n = \theta_n - \theta$, we rewrite the last equation in the form

$$Z_n = Z_{n-1} + I_n^{-1}(\theta + Z_{n-1}) b_n(\theta, Z_{n-1}) + I_n^{-1}(\theta + Z_{n-1}) \Delta m_n, \quad Z_0 = \theta_0 - \theta, \tag{4.13}$$

where $\Delta m_n = \Delta m(n, Z_{n-1})$ with $\Delta m(n, u) = l_n(\theta + u) - E_\theta(l_n(\theta + u) \mid \mathcal{F}_{n-1})$.

Note that algorithm (4.13) is embedded into the stochastic approximation scheme (4.1) with

$$H_n(u) = I_n^{-1}(\theta + u) b_n(\theta, u) \in \mathcal{F}_{n-1},$$
$$\Delta K_n = 1,$$
$$\Delta M(n, u) = I_n^{-1}(\theta + u) \Delta m(n, u).$$

This example clearly shows the necessity of considering random fields $H_n(u)$ and $M(n, u)$.

Remark 4.10. Let $\theta \in \Theta \subset \mathbb{R}_1$, where Θ is an open proper subset of $\mathbb{R}_1$. It can be possible to define the objects $l_n(\theta)$ and $I_n(\theta)$ only on the set Θ, but for each fixed $\theta \in \Theta$ the objects $H_n(u)$ and $M(n, u)$ are well-defined functions of the variable u on the whole $\mathbb{R}_1$. Then under the conditions of Theorem 4.2, $\theta_n \to \theta$ P_θ-a.s. as $n \to \infty$, starting from an arbitrary θ_0. The example given below illustrates this situation. The same example also shows the efficiency of representation (4.3) (2).

(b) The Calton–Watson branching process with immigration. Let the process

$$X_i = \sum_{j=1}^{X_{i-1}} Y_{i,j} + 1, \quad i = 1, 2, ..., n, \quad X_0 = 1,$$

be observed, where $Y_{i,j}$ are i.i.d. random variables having the Poisson distribution with the parameter θ, $\theta > 0$, to be estimated. If $\mathcal{F}_i = \sigma(X_j,\, j \le i)$, then

$$P_\theta(X_i = m \mid \mathcal{F}_{i-1}) = \frac{(\theta X_{i-1})^{m-1}}{(m-1)!} e^{-\theta X_{i-1}}, \quad i = 1, 2, ...; \; m \ge 1,$$

whence we have

$$l_i(\theta) = \frac{X_i - 1 - \theta X_{i-1}}{\theta}, \quad I_n(\theta) = \theta^{-1} \sum_{i=1}^{n} X_{i-1}.$$

The recursive procedure has the form

$$\theta_n = \theta_{n-1} + \frac{X_n - 1 - \theta_{n-1} X_{n-1}}{\sum\limits_{i=1}^{n} X_{i-1}}, \quad \theta_0 \in \mathcal{F}_0, \tag{4.14}$$

and if, as usual, $Z_n = \theta_n - \theta$, then

$$Z_n = Z_{n-1} - \frac{Z_{n-1} X_{n-1}}{\sum\limits_{i=1}^{n} X_{i-1}} + \frac{\varepsilon_n}{\sum\limits_{i=1}^{n} X_{i-1}}, \tag{4.15}$$

where $\varepsilon_n = X_n - 1 - \theta X_{n-1}$ is a P_θ-square integrable martingale-difference. In fact, $E_\theta(\varepsilon_n \mid \mathcal{F}_{n-1}) = 0$, $E_\theta(\varepsilon_n^2 \mid \mathcal{F}_{n-1}) = \theta X_{n-1}$. In this case, $H_n(u) = -uX_{n-1} \Big/ \sum\limits_{i=1}^{n} X_{i-1}$, $\Delta M(n, u) = \varepsilon_n \Big/ \sum\limits_{i=1}^{n} X_{i-1}$, $\Delta K = 1$, and, therefore, they are well defined on the whole $\mathbb{R}_1$.

Now we show that the solution of Eq. (4.14) coincides with the MLE

$$\widehat{\theta}_n = \frac{\sum\limits_{i=1}^{n} (X_i - 1)}{\sum\limits_{i=1}^{n} X_{i-1}}.$$

It is easy to see that $(\widehat{\theta}_n)_{n \ge 1}$ is strongly consistent for all $\theta > 0$. Indeed,

$$\widehat{\theta}_n - \theta = \frac{\sum\limits_{i=1}^{n} \varepsilon_i}{\sum\limits_{i=1}^{n} X_{i-1}}$$

and the desirable follows from the strong law of large numbers for martingales and from the well-known fact (see, e.g., [31]) that for all $\theta > 0$,

$$\sum_{i=1}^{\infty} X_{i-1} = \infty \quad (P_\theta\text{-a.s.}). \tag{4.16}$$

Derive this result as a corollary of Theorem 4.2.

First, we note that for each $\theta > 0$, conditions (A) and (B′) below are satisfied. Indeed,

(A) $$H_n(u)u = \frac{-u^2 X_{n-1}}{\sum\limits_{i=1}^{n} X_{i-1}} < 0$$

for all $u \neq 0$ $(X_i > 0,\ i \geq 0)$;

$$\text{(B}'\text{)} \qquad \langle m \rangle_\infty = \theta \sum_{n=1}^{\infty} \frac{X_{n-1}}{\left(\sum\limits_{i=1}^{n} X_{i-1}\right)^2} < \infty,$$

thanks to (4.16).

Now, to illustrate the efficiency of condition (II), let us consider two cases:

(1) $0 < \theta \leq 1$ and (2) θ is arbitrary, i.e., $\theta > 0$.

In case (1), condition (I) is satisfied. In fact,

$$|H_n(u)| = \left(X_{n-1} \Big/ \sum_{i=1}^{n} X_{i-1}\right) |u|$$

and

$$\sum_{n=1}^{\infty} X_{n-1}^2 \Big/ \left(\sum_{i=1}^{n} X_{i-1}\right)^2 < \infty$$

P_θ-a.s. But if $\theta > 1$, then the last series diverges and condition (I) (i) is not satisfied.

On the other hand, the proof of the desirable convergence by verifying condition (II) is almost trivial. Indeed, using Remark 4.7 and taking

$$\widetilde{G}_n = G_n = X_{n-1} \Big/ \sum_{i=1}^{n} X_{i-1},$$

we obtain $\sum\limits_{n=1}^{\infty} G_n = \infty$ P_θ-a.s., for all $\theta > 0$. Moreover, $\delta_n = -2 + \widetilde{G}_n < 0$, $|\delta_n| \geq 1$.

2. RM algorithm with a deterministic regression function. Consider a particular case of algorithm (4.1) with $H_t(\omega, u) = \gamma_t(\omega) R(u)$, where the process $\gamma = (\gamma_t)_{t \geq 0} \in \mathcal{P}$, $\gamma_t \geq 0$ for all $t \geq 0$, $dM(t,u) = \gamma_t dm_t$, $m \in \mathcal{M}^2_{\text{loc}}$, i.e.,

$$dZ_t = \gamma_t R(Z_{t-}) dK_t + \gamma_t dm_t, \quad Z_0 \in \mathcal{F}_0.$$

(a). Let the following conditions be satisfied:

$$\begin{aligned}
&\text{(A)} && R(0) = 0, \quad R(u)u < 0 \ \text{ for all } \ u \neq 0, \\
&\text{(B}'\text{)} && \gamma^2 \circ \langle m \rangle_\infty < \infty, \\
&\text{(1)} && |R(u)| \leq C(1 + |u|), \quad C > 0 \ \text{ is constant}, \\
&\text{(2)} && \text{for each } \ \varepsilon > 0, \quad \inf_{\varepsilon \leq u \leq 1/\varepsilon} |R(u)| > 0, \\
&\text{(3)} && \gamma \circ K_t < \infty, \quad \forall t \geq 0, \quad \gamma \circ K_\infty = \infty, \\
&\text{(4)} && \gamma^2 \Delta K \circ K^d_\infty < \infty.
\end{aligned}$$

Then $Z_t \to 0$ P-a.s. as $t \to \infty$.

Indeed, it is easy to see that (A), (B$'$), (1)–(4) $\Rightarrow$ (A), (B), and (I) of Theorem 4.2.

In [72], this result has been proved on the basis of the theorem on the semimartingale convergence sets mentioned in Remark 4.1. In the case where $K^d \neq 0$, this automatically leads to the "moment" restrictions and also the additional assumption $|R(u)| \leq \text{const}$.

(b). Let, as in case (a), conditions (A) and (B′) be satisfied. Moreover, assume that for each $u \in \mathbb{R}_1$ and $t \in [0, \infty)$:

$$\begin{aligned}
&(1) \quad \alpha_t(u) + \beta_t(u) \le 0, \\
&(2) \quad \text{for all } \varepsilon > 0, \ \ I_\varepsilon := \inf_{\varepsilon \le u \le 1/\varepsilon} \{-(\alpha(u) + \beta(u))\} \circ K_\infty = \infty.
\end{aligned}$$

Then $Z_t \to 0$ P-a.s. as $t \to \infty$.

Indeed, it is not hard to verify that (1), (2) $\Rightarrow$ (II).

The question arises whether is it possible for (1) and (2) to be satisfied.

Suppose, in addition, that

$$C_1|u| \le |R(u)| \le C_2|u|, \quad C_1, C_2 \quad \text{are constants,} \tag{4.17}$$

$$\begin{aligned}
&(3) \qquad 2 - C_2\gamma_t\Delta K_t \ge 0, \\
&(4) \qquad \gamma(2 - C_2\gamma\Delta K) \circ K_\infty = \infty.
\end{aligned}$$

Then (3) $\Rightarrow$ (1) and (4) $\Rightarrow$ (2).

Indeed,

$$\begin{aligned}
\alpha_t(u) + \beta_t(u) &\le C_1\gamma_t|u|^2[-2 + C_2\gamma_t\Delta K_t] \le 0, \\
I_\varepsilon &\ge C_1\varepsilon^2\{\gamma(2 - C_2\gamma\Delta K) \circ K_\infty\} = \infty.
\end{aligned}$$

Remark 4.11. (4) $\Rightarrow \gamma \circ K_\infty = \infty$.

In [74], the convergence $Z_t \to 0$ P-a.s. as $t \to \infty$ was proved under the following conditions:

(A) $R(0) = 0$, $R(u)u < 0$ for all $u \neq 0$;

(M) there exists a nonnegative predictable process $r = (r_t)_{t \ge 0}$ integrable with respect to the process $K = (K_t)_{t \ge 0}$ on any finite interval $[0, t]$ with the following properties:

$$\begin{aligned}
&(a) \quad r \circ K_\infty = \infty; \\
&(b) \quad A^1_\infty = \gamma^2\mathcal{E}^{-1}(-r \circ K) \circ \langle m \rangle_\infty < \infty; \\
&(c) \quad \text{all jumps of the process } \ A^1 \ \text{ are bounded;} \\
&(d) \quad r_t u^2 + \gamma_t^2 \Delta K_t R^2(u) \le -2\gamma_t R(u)u, \ \ \text{for all } \ u \in \mathbb{R}^1 \ \text{ and } \ t \in [0, \infty).
\end{aligned}$$

Show that (M) $\Rightarrow$ (B′), (1), and (2).

It is obvious that (b) $\Rightarrow$ (B′). Further, (d) $\Rightarrow$ (1). Finally, (2) follows from (a) and (d) thanks to the relation

$$I_\varepsilon = \inf_{\varepsilon \le |u| \le 1/\varepsilon} -(\alpha(u) + \beta(u)) \circ K_\infty \ge \varepsilon^2 r \circ K_\infty = \infty.$$

The implication is proved.

In the particular case where (4.7) holds and for all $t \ge 0$, $\gamma_t\Delta K_t \le q$, $q > 0$ is a constant, and C_1 and C_2 in (4.17) are chosen so that $2C_1 - qC_2^2 > 0$ if we take $r_t = b\gamma_t$, $b > 0$ with $b < 2C_1 - qC_2^2$, then (a) and (d) are satisfied if $\gamma \circ K_\infty = \infty$.

But these conditions imply (3) and (4). In fact, on one hand, $0 < 2C_1 - qC_2^2 \le C_1(2 - qC_2)$ and, therefore, item (3) follows, since $2 - C_2\gamma_t\Delta K_t \ge 2 - qC_2 > 0$. On the other hand, item (4) follows from $\gamma(2 - C_2\gamma\Delta K) \circ K_\infty \ge (2 - qC_2)\gamma \circ K_\infty = \infty$.

From what was said above, we can conclude that if conditions (A), (B′), (4.17), $\gamma_t\Delta K_t \le q$, $q > 0$, $2 - qC_2 > 0$, and $\gamma \circ K_\infty = \infty$ are satisfied, then the desirable convergence $Z_t \to 0$ P-a.s. takes place and, therefore, there is no need for choosing the process $r = (r_t)_{t \ge 0}$ with properties (M) (cf. [74], Remark 3.3 and Sec. 4).

(c) Linear model (see, e.g., [72]). Consider the linear RM procedure

$$dZ_t = b\gamma_t Z_{t-} dK_t + \gamma_t dm_t, \quad Z_0 \in \mathcal{F},$$

where $b \in B \subseteq (-\infty, 0)$ and $m \in \mathcal{M}^2_{\text{loc}}$.

Assume that

$$\gamma^2 \circ \langle m \rangle_\infty < \infty, \tag{4.18}$$

$$\gamma \circ K_\infty = \infty, \tag{4.19}$$

$$\gamma^2 \Delta K \circ K^d < \infty.$$

Then for each $b \in B$, conditions (A), (B′), and (I) are satisfied. Hence,

$$Z_t \to 0 \quad P\text{-a.s.} \quad \text{as } t \to \infty. \tag{4.20}$$

Now let (4.18) and (4.19) be satisfied, but $P(\gamma^2 \Delta K \circ K^d = \infty) > 0$.

At the same time, assume that $B = [b_1, b_2]$, $-\infty < b_1 \le b_2 < 0$ and for all $t \ge 0$, $\gamma_t \Delta K_t < |b_1|^{-1}$.

Then for each $b \in B$, (4.20) holds.

Indeed,

$$\begin{aligned}[\alpha_t(u) I_{\{\Delta K \neq 0\}} + \beta_t(u)]^+ &= |b|\gamma_t u^2 [-2 + |b|\gamma_t \Delta K_t I_{\{\Delta K \neq 0\}}]^+ \\ &\le I_{\{\Delta K \neq 0\}} |b| \gamma_t u^2 [-2 + |b_1|\gamma_t \Delta K_t]^+ = 0\end{aligned}$$

and, therefore, (II) (i) is satisfied.

On the other hand,

$$\begin{aligned}&\inf_{\varepsilon \le |u| \le 1/\varepsilon} u^2 \{2\gamma |b| I_{\{\Delta K \neq 0\}} + b\gamma[2 - |b|\gamma \Delta K] I_{\{\Delta K \neq 0\}}\} \circ K_\infty \\ &\ge \varepsilon^2 |b| \gamma [2 - |b|\gamma \Delta K] \circ K_\infty \ge \varepsilon^2 |b| \gamma \circ K_\infty = \infty.\end{aligned}$$

Thus, (II) (ii) is also satisfied.

REFERENCES

1. A. Albert and L. Gardner, *Stochastic Approximation and Nonlinear Regression,* M. I. T. Press, Cambridge, Massachusetts (1967).
2. D. Allinger and S. Mitter, "New results on the innovation problem for nonlinear filtering," *Stochastics,* **4**, 339–348 (1981).
3. C. Anulova, A. Veretennikov, N. Krylov, R. Liptser, and A. Shiryaev, "Stochastic calculus," In: *Progress in Science and Technology, Contemporary Problems of Mathematics* [in Russian], Vol. 45, All-Union Institute for Scientific and Technical Information (VINITI), Akad. Nauk SSSR, Moscow (1989), pp. 5–256.
4. R. Beran, "An efficient and robust adaptive estimator of location," *Ann. Statist.*, **6**, 292–313 (1978).
5. P. Bickel, "Quelques aspects de la statistique robuste," In: *Lecture Notes in Math.*, **876**, Springer, New York (1981), pp. 2–68.
6. P. Bickel, "On adaptive estimation," *Ann. Statist.*, **10**, 647–671 (1982).
7. P. Bickel, C. Klaasen, Y. Ritov, and J. Wellner, *Efficient and Adaptive Estimation of Semiparametric Models,* The John Hopkins Univ. Press (1993).
8. P. Billingsley, *Convergence of Probability Measures,* Wiley, New York (1968).
9. M. Bockshtein, "Theorems on existence and uniqueness of solutions of systems of ordinary differential equations," *Uch. Zap. NGU, Matematika*, **15**, 3–72 (1939).
10. N. Chan and C. Wei, "Asymptotic inference for weakly nonstationary *AR*(1) processes," *Ann. Statist.*, **15**, 1050–1063 (1987).
11. R. Chitashvili, "On the existence of an innovation process for a class of Itô processes," In: *Teor. Sluch. Process.*, Kiev, **10**, 105–116 (1982).

12. R. Chitashvili, “On the existence of an innovation process for a component of a diffusion-type process,” *Teor. Veroyatnost. Primenen.*, **XXX**, No. 1, 44–57 (1985).
13. B. Cirelson, “A stochastic differential equation without strong solution,” *Teor. Veroyatnost. Primenen.*, **XX**, No. 2, 427–430 (1975).
14. J. M. C. Clark, “Conditions for the one-to-one correspondence between an observation process and its innovations,” *Center of Comput. and Automat., Imperial College,* London, Technical Report 1 (1918).
15. E. Coddington and N. Levinson, *Theory of Differential Equations,* Moscow (1952).
16. D. Cox, “Partial likelihood,” *Biometrika*, **64**, 269–276 (1975).
17. D. Cox, “Gaussian likelihood estimation for nearly nonstationary $AR(1)$ process,” *Ann. Statist.*, **19**, 1129–1164 (1991).
18. D. Cox and I. Llatas, “Maximum likelihood type estimation for nearly nonstationary autoregression time series,” *Ann. Statist.*, **19**, 1109–1128 (1991).
19. H. Cramer, *Mathematical Methods of Statistics* [Russian translation], Mir, Moscow (1975).
20. K. Dzaparidze and D. Spreij, “On optimality of regular projective estimators in semimartingale models,” *Stochastics*, **43**, Nos. 3–4, 161–178 (1993).
21. E. Eliott, *Stochastic Calculus and Application,* Springer, New York (1982).
22. P. Frost and T. Kailath, “An innovation approach to least squares estimation. Part III. Nonlinear estimation in white Gaussian noise,” In: *IEEE Trans. Automat. Control*, **AC-16**, No. 3, 217–226 (1968).
23. L. Gal’čuk, “On the uniqueness and existence of solution of stochastic equations with respect to semimartingales,” *Theory Probab. Appl.* **23**, 782–795 (1978).
24. I. Gikhman and A. Skorokhod, *Stochastic Differential Equations,* Springer-Verlag, New York (1972).
25. I. Gikhman and A. Skorokhod, *Stochastic Differential Equations and Their Applications* [in Russian], Naukova Dumka, Kiev (1982).
26. P. Greenwood and A. Shiryaev, “Asymptotic minimaxity of a sequential estimator for a first order autoregressive model,” *Stochastics and Stochast. Reports*, **38**, No. 1, 49–65 (1992).
27. P. Greenwood and W. Weffelmeyer, “Efficiency of estimators for partially specified filtered models,” *Preprint in Statistics, Univ. of Cologne,* No. 119 (1989).
28. P. Greenwood and W. Weffelmeyer, “Partially and fully specified semimartingale models and efficiency,” *Preprint in Statistics, Univ. of Cologne,* No. 124 (1989).
29. P. Greenwood and W. Weffelmeyer, “Efficient estimating equation for nonparametric filtered models,” In: *Statist. Inference in Stochast. Process.*, Basel, New York (1990), pp. 107–141.
30. P. Greenwood and W. Weffelmeyer, “Asymptotic minimax results for statistic processes familiar with critical points,” *Stochastic Process. Appl.*, **44**, No. 1, 107–116 (1993).
31. A. Gushchin, “On asymptotic optimality of parameter estimators under the LAQ condition,” *Probab. Theory Appl.*, **40**, No. 2, 286–300 (1995).
32. F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics. The Approach Based on Influence Functions,* Wiley, New York (1986).
33. O. Hossjer, “Robust linear regression by means of M- and R-statistics,” *Doctoral Thesis., Uppsala Univ. Publ. Depart. Math.* (1991).
34. P. Huber, “The behavior of maximum likelihood estimates under nonstandard conditions,” In: *Proc. 5th Berkeley Symp. on Math. Statist. and Probab.*, **1**, Univ. of California Press, Berkley (1967), pp. 221–233.
35. D. Huber, *Robust Statistics,* Wiley, New York (1981).
36. I. Ibragimov and R. Khasminskii, *Statistical Estimation. Asymptotic Theory,* Springer, New York (1981).
37. N. Ikeda and Sh. Watanabe, *Stochastic Differential Equations and Diffusion Processes,* North-Holland, Amsterdam (1981).

38. K. Ito and H. McKean, *Diffusion Processes and Their Sample Paths,* Springer-Verlag, Berlin (1965).
39. J. Jacod, "Calcul stochastique et problémes de martingales," In: *Lecture Notes in Math.*, **714**, Springer, Berlin–Heidelberg–New York (1979).
40. J. Jacod, "Partial likelihood process and asymptotic normality," *Laboratoire de Probabilites Universite P. et M. Curie,* Paris (1987).
41. J. Jacod, "Regularity, partial regularity, partial information process for a filtered statistical model," *Laboratoire de Probabilites Universite P. et M. Curie,* Paris (1989).
42. J. Jacod, "Filtered statistical models and Hellinger process," *Stochastic Process. Appl.*, **32**, 3–45 (1989).
43. J. Jacod, "Convergence of filtered statistical models and Hellinger process," *Stochastic Process. Appl.*, **32**, 47–68 (1989).
44. J. Jacod and J. Memin, "Weak and strong solutions of stochastic differential equations: existence and stability. Stochastic integrals," In: *Lecture Notes in Math.*, **851**, Springer-Verlag, Berlin (1981) pp. 169–212.
45. J. Jacod and A. Shiryaev, *Limit Theorems for Stochastic Processes,* Springer, New York (1987).
46. P. Jegenathan, "On the asymptotic behavior of least square estimators in AR series with roots near the unit circle. I, II," *Dept. Statistics, Univ. Michigan*, Technical Report 150 (1990).
47. T. Jeulin and M. Yor, "Filtration des ponts browniens et equations differentielles lineaires," In: *Semin. Probab.* XXIV, *Lecture Notes in Math.*, **1426**, Springer, Berlin (1990), pp. 227–265.
48. Yu. Kabanov, R. Liptser, and A. Shiryaev, "Absolute continuity of locally absolute continuous probability distributions. I, II," *Mat. Sb.*, **35**, 631–680 (1978); **36**, 31–58 (1979).
49. T. Kailath, "An innovation approach to least squares estimation," In: *IEEE Trans. Automat. Control*, **13**, 654–656 (1968).
50. G. Kallianpur, "A linear stochastic system with discontinuous control," In: *Proc. Internat. Symp. on Stochast. Differential Equations,* Kyoto (1978), pp. 127–140.
51. I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus,* Springer-Verlag (1991).
52. A. Kirillov and A. Gvishiani, *Theorems and Problems of Functional Analysis* [in Russian], Nauka, Moscow (1979).
53. H. Kone and S. Levental, "Weak convergence of the residual empirical process in explosive autoregression," *Ann. Statist.*, **17**, No. 4, 1784–1794 (1989).
54. N. Krylov, *Controlled Diffusion-Type Processes* [in Russian], Nauka, Moscow (1977).
55. N. Krylov, "On the coincidence of σ-algebra in the filtration problem for diffusion processes," *Teor. Veroyatn. Primenen.*, **XXIV**, No. 4, 771–780 (1979).
56. H. Kunita, "On the decomposition of solutions of stochastic differential equations," In: *Lecture Notes in Math.*, **851**, pp. 213–255, Springer, Berlin (1981).
57. H. Künsch, "Infinitesimal robustness for autoregression processes," *Ann. Statist.*, **12**, 843–863 (1984).
58. Yu. Kutoyants, *Parameter Estimation for Stochastic Processes,* Hedermann Verlag, Berlin (1984).
59. L. Le Cam and G. Yang, *Asymptotics in Statistics: Some Basic Concepts,* Springer-Verlag, New York (1990).
60. A. Le Breton, *Dissertation,* Grenoble (1976).
61. A. Le Breton, "Parameter estimation in a linear stochastic differential equation," In: *Trans. 7th Prague Conf. Inform. Theor. A*, Riedel, Dordrecht (1977), pp. 353–366.
62. J. F. Le Gall, "Applications du temps local aux equations differentielles stochastiques unidimentionnelles," In: *Semin. Probab.* XVII, *Lecture Notes in Math.*, **986**, Springer (1983), pp. 15–31.
63. E. Lenglarth, "Sur la convergence presque sure des martingales locales," *C. R. Acad. Sci.*, Paris, **284**, No. 17, A1085–A1088 (1977).
64. L. Ljung, C. Pflug, and H. Walk, *Stochastic Approximation and Optimization of Random Systems,* DMV-Seminar, Bd. 17, Birkhäuser Verlag, Basel–Boston–Berlin (1992).

65. R. Liptser, "A strong law of large numbers for local martingales," *Stochastics,* **3**, No. 3, 217–228 (1980).
66. R. Liptser and A. Shiryaev, *Statistics of Random Processes. General Theory,* Springer-Verlag, New York (1974).
67. R. Liptser and A. Shiryaev, *Theory of Martingales,* Kluwer Academic Publ., Dordrecht (1989).
68. C. Lotockii, "Stochastic approximation type robust algorithms (continuous time)," *Teor. Veroyatn. Primenen.*, **40**, No. 2, 324–346 (1995).
69. H. Luschgy, "Asymptotic inference for semimartingale models with singular parameter points," *J. Statist. Plann. Inference*, **39**, No. 2, 155–186 (1994).
70. D. Martin and V. J. Yohai, "Influence functional for time series," *Ann. Statist.*, **14**, 781–819 (1986).
71. H. McKean, *Stochastic Integrals,* Academic Press, New York (1969).
72. A. Melnikov, "Procedures of stochastic approximation for semimartingales," In: *Statistics and Control of Random Processes, Trudy Mat. Inst. Steklov.*, Nauka, Moscow (1989), pp. 147–156.
73. A. Melnikov and A. Rodkina, "Martingale approach to the procedures of stochastic approximation," In: H. Niemi et al. (Eds.), *Frontier in Pure and Appl. Probab.*, **1**, TVP/VSP (1993), pp. 165–182.
74. A. Melnikov, A. Rodkina, and E. Valkeila, "On a general class of stochastic approximation algorithms," In: H. Niemi et al. (Eds.), *Frontier in Pure and Appl. Probab.*, **1**, TVP/VSP (1993), pp. 183–196.
75. A. Melnikov, "Stochastic differential equations: nonsmooth coefficients, regression models, and stochastic approximation," *Usp. Mat. Nauk*, **51**, No. 5 (311), 43–136 (1996).
76. A. Melnikov and E. Valkeila, *On the Exact Behaviour of Stochastic Approximation Procedures,* Univ. of Helsinki, Finland, Depart. Math. (1997), pp. 1–10.
77. P. Meyer, "Sur une probléme de filtration," In: *Lecture Notes in Math.*, **321**, 223–247. Springer, Berlin (1973).
78. M. Nevelson and R. Khas'minskii, *Stochastic Approximation and Recursive Estimation* [in Russian], Nauka, Moscow (1972).
79. M. Perlman, "The limiting behaviour of multiple roots of the likelihood equation," *Recent Advances in Statistics*, Academic Press, New York (1983), pp. 339–370.
80. E. Perkins, "Local time and pathwise uniqueness for stochastic differential equations," In: *Semin. Probab.* XVI, *Lecture Notes in Math.*, **920**, Springer, Berlin (1982), pp. 201–208.
81. J. Pfanzage and W. Wefelmeyer, "Contribution to a general asymptotic statistical theory," In: *Lecture Notes in Statist.*, **13**, Springer, New York (1982).
82. B. L. S. Prakasa Rao, *Asymptotic Theory of Statistical Inference,* Wiley (1987).
83. P. Protter, "Semimartingales and stochastic differential equations," In: *Lecture Notes 3rd Chilean Winter School in Probab. and Statist.*, Technical Report 85-25, Dep. Statist., Purdue Univ. (1984).
84. D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion,* Springer-Verlag, Berlin (1994).
85. H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, **22**, No. 1, 400–407 (1951).
86. H. Robbins and A. Siegmund, "A convergence theorem for nonnegative almost semimartingales," In: J. S. Rustagi (Ed.), *Optimizing Methods in Statistics,* Academic Press, New York (1971), pp. 233–257.
87. A. Shiryaev, "A stochastic equation of linear filtration of a stepwise Markov process," *Probl. Peredachi Inf.*, No. 3, 3–22 (1966).
88. A. Shiryaev, *Probability,* Springer, Berlin (1996).
89. A. Shiryaev and V. Spokoiny, *Statistical Experiments and Decisions: Asymptotic Theory,* VINITI, Moscow (1995).
90. A. Shiryaev and V. Spokoiny, "On sequential estimation of an autoregressive parameter," *Stochastics,* **60**, 219–240 (1997).

91. P. Spreij, "Approximate maximum likelihood estimation for a class of counting process models," *Multivariate Analysis,* **39**, No. 2, 236–245 (1991).
92. M. Staab, *Robust Parameter Estimation for ARMA Models,* Ph.D Thesis, Univ. Bayreuth, Bayreuth, Germany (1985).
93. H. Strasser, *Mathematical Theory of Statistics,* Gruyter, Berlin (1985).
94. A. Taraskin, "On the behavior of the likelihood ratio of semimartingales," *Teor. Veroyatn. Primenen.*, **29**, 452–464 (1986).
95. T. Toronjadze, "On absolute continuity of measures corresponding to a class of diffusion type processes with respect to a Wiener measure," *Bull. Acad. Sci. Georgian SSR*, **76**, No. 9, 285–288 (1974).
96. T. Toronjadze, "Construction of an innovation process for a class of Ito processes," *Bull. Acad. Sci. Georgian SSR*, **77**, No. 2, 553–558 (1975).
97. T. Toronjadze and R. Chitashvili, "On equivalence of strongly and weakly regular stochastic differential equations," *Bull. Acad. Sci. Georgian SSR*, **98**, No. 1, 37–40 (1980).
98. T. Toronjadze and R. Chitashvili, "Stochastic differential equations with unit diffusion coefficient. Theorems of existence and uniqueness of strong solutions," *Bull. Acad. Sci. Georgian SSR*, **98**, No. 3, 537–539 (1980).
99. T. Toronjadze and R. Chitashvili, "The structure of an integral funnel of all solutions of stochastic differential equations with unit diffusion coefficient," *Bull. Acad. Sci. Georgian SSR*, **99**, No. 1, 21–24 (1980).
100. T. Toronjadze and R. Chitashvili, "On local theorems of existence and uniqueness and extension theorems for one-dimensional stochastic differential equations with unit diffusions," *Bull. Acad. Sci. Georgian SSR*, **99**, No. 2, 285–288 (1980).
101. T. Toronjadze and R. Chitashvili, "On one-dimensional stochastic differential equations with unit diffusion coefficient. Structure of solutions," *Stochastics,* **4**, No. 4, 281–315 (1981).
102. T. Toronjadze, "Innovation process existence for a class of partially observable diffusion processes," In: *4th USSR–Japan Sympos. on Probab. Theor. Math. Statist.* **II** (1982), pp. 238–239.
103. T. Toronjadze, "Construction of the innovation process in a filtering problem of partially observable diffusion type," *Stochastics,* **16**, 197–216 (1986).
104. T. Toronjadze and N. Lazrieva, "Asymptotic properties of MLE in general statistical schemes," *Bull. Acad. Sci. Georgian SSR*, **123**, No. 1, 25–28 (1986).
105. T. Toronjadze and N. Lazrieva, "The Itô–Ventzel formula for semimartingales and applications to recursive estimation," *Bull. Acad. Sci. Georgian SSR*, **123**, No. 2, 253–256 (1986).
106. T. Toronjadze, "Asymptotic properties of the MLE, Itô–Ventzel formula for semimartingales and its application to recursive estimation in general statistical models," In: *Proc. First World Congress of the Bernoulli Society,* Tashkent, VNU Sci. Press (1986), pp. 63–67.
107. T. Toronjadze, "Coincidence of σ-algebras in the scheme of partially observable diffusion-type processes," *Teor. Veroyatn. Primenen.*, **33**, No. 3, 580–585 (1987).
108. T. Toronjadze, "Innovation problem for a class of Ito processes: filtration problem for multidimensional diffusion type process," *Probab. Theor. and Math. Statist. (Proc. 4th Internat. Vilnius Conf.)*, **2**, 659–671, VNU Science Press, Utrecht, Netherlands (1987).
109. T. Toronjadze and N. Lazrieva, "Joint asymptotic distribution of the MLE and M-estimators," In: *Lecture Notes in Math.*, **1299**, Springer-Verlag, Berlin (1987), pp. 259–266.
110. T. Toronjadze and N. Lazrieva, "Ito–Ventzel's formula for semimartingales, asymptotic properties of MLE and recursive estimation," In: *Lecture Notes in Control and Inform. Sci.*, **96**, pp. 346–355, Springer-Verlag, Berlin (1987).
111. T. Toronjadze and R. Chitashvili, "On innovation processes of semimartingales," In: (A. Shiryaev, Ed.), *Statistics and Control of Stochastic Processes. Steklov Math. Inst.,* Nauka, Moscow (1989), pp. 202–207.

112. T. Toronjadze and N. Lazrieva, "Asymptotic properties of M-estimators in general statistical models," In: (A. Shiryaev, Ed.), *Statistics and Control of Stochastic Processes. Steklov Math. Inst.*, Nauka, Moscow (1989), pp. 102–112.
113. T. Toronjadze, R. Chitashvili, and N. Lazrieva, "Asymptotic theory of estimation in general statistical experiment under model disturbance," *Trudy Tbilis. Mat. Inst. Razmadze,* **92**, 106–145 (1989).
114. T. Toronjadze, R. Chitashvili, and N. Lazrieva, "Asymptotic theory of M-estimators in general statistical models," *Centre for Mathematics and Computer Sciences,* Amsterdam, Netherlands, Report BS-R9019 (1990), pp. 1–31.
115. T. Toronjadze, R. Chitashvili, and N. Lazrieva, "Asymptotic theory of M-estimators in general statistical models. On asymptotic behaviour of estimators in the presence of nuisance parameters," *Centre for Mathematics and Computer Sciences,* Amsterdam, Netherlands, Report BS-R9020 (1990), pp. 1–45.
116. T. Toronjadze and N. Lazrieva, "On stable M-estimators in the partial likelihood scheme," In: *New Trends in Probability and Statistics,* **1**, VSP, Utrecht and Tokyo, Mokslas, Vilnius (1991), pp. 567–597.
117. T. Toronjadze and N. Lazrieva, "Robust estimators in statistical models with filtration. Shrinking neighborhoods," *Seminarberichte, Fachbereich Mathematik,* Fernuniversität, Hagen, Germany, **48**, 50–68 (1994).
118. T. Toronjadze, N. Lazrieva, and T. Sharia, "The Robbins–Monro type stochastic differential equations. Convergence of solutions," *Stochastics,* **61**, Nos. 1–2, 67–87 (1997).
119. T. Toronjadze and N. Lazrieva, "Robust estimators in discrete time statistical models. Contiguous alternatives," *Proc. A. Razmadze Math. Inst.,* **115**, 58–96 (1997).
120. T. Toronjadze and N. Lazrieva, "Influence functionals for discrete time statistical models. Weakly contiguous alternatives," *Proc. A. Razmadze Math. Inst.,* **115**, 97–120 (1997).
121. A. Veretennikov, "On the strong solutions of stochastic differential equations," *Teor. Veroyatn. Primenen.*, **24**, 354–366.
122. L. Vostrikova, "On a weak convergence of likelihood ratio processes of general statistical models," In: *Internat. Vilnius Conf. on Probab. Theor. and Math. Statist., Abstr. of Commun.*, **3**, Vilnius (1986), pp. 331–332.
123. C. Wei and J. Winnicki, "Estimation of the means in the branching process with immigration," *Ann. Statist.*, **18**, No. 4, 1757–1773 (1990).
124. J. Wellner, *Semimartingale Models: Progress and Problems,* Centre for Math. and Comput. Sci., Amsterdam, Report MS-R8614 (1986).
125. W. H. Wong, "Theory of partial likelihood," *Ann. Statist.*, **14**, No. 1, 88–124 (1976).
126. T. Yamada and Sh. Watanabe, "On the uniqueness of solutions of stochastic differential equations. I, II," *J. Math., Kyoto Univ.,* **11**, 155–167 (1971).
127. M. Yershov, "On representation of Itô processes," *Teor. Veroyatn. Primenen.*, **XVII**, 167–172 (1972).
128. M. Yershov, "Structure of solutions of stochastic equations," In: *Materials of Seminar on Theor. Random Processes,* **I**, Druskininkai (1974), pp. 28–37.
129. M. Yershov, "Causal and non-anticipating solutions of stochastic equations," In: *Lecture Notes in Control Inform. Sci.,* **16**, pp. 606–615, Springer-Verlag, Berlin (1978).
130. K. Yoshida, *Functional Analysis,* Springer, Berlin (1966).
131. A. Zvonkin, "A transformation of the state space of a diffusion process that removes the drift," *Mat. Sb.*, **22**, 129–149 (1974).
132. A. Zvonkin and N. Krylov, "On strong solutions of stochastic differential equations," In: *Proc. of Druskinenkay Seminar,* **II**, Vilnius (1975), pp. 9–88.