Bernoulli **6**(5), 2000, 917–938

Z-theorems: limits of stochastic equations

VLADIMIR V. ANISIMOV[1] and GEORG CH. PFLUG[2]

[1]*Bilkent University, Dept of Industrial Engineering, Bilkent 06533, Ankara, Turkey and Kiev University, Faculty of Cybernetics, Vladimirskaya Str., 64, 252017 Kiev 17, Ukraine*
[2]*University of Vienna, Institute of Statistics & Decision Support, Universitätsstrasse 5, A-1010 Wien, Austria. E-mail: georg.pflug@univie.ac.at*

Let $f_n(\theta, \omega)$ be a sequence of stochastic processes which converge weakly to a limit process $f_0(\theta, \omega)$. We show under some assumptions the weak inclusion of the solution sets $\boldsymbol{\theta}_n(\omega) = \{\theta : f_n(\theta, \omega) = 0\}$ in the limiting solution set $\boldsymbol{\theta}_0(\omega) = \{\theta : f_0(\theta, \omega) = 0\}$. If the limiting solutions are almost surely singletons, then weak convergence holds. Results of this type are called Z-theorems (zero-theorems). Moreover, we give various more specific convergence results, which have applications for stochastic equations, statistical estimation and stochastic optimization.

Keywords: asymptotic distribution; consistency; stochastic equations; stochastic inclusion

1. Introduction

Statistical estimators are often defined as minima of stochastic processes or roots of stochastic equations. The first group are called M-estimators and include the maximum-likelihood estimate, some classes of robust estimates and the solutions of general stochastic programs (see Shapiro 1993; Pflug 1995). The proof of asymptotic properties of such estimates requires conditions under which the convergence in distribution of some stochastic process $f_n(\cdot)$ to a limiting process $f_0(\cdot)$ entails that

$$\arg\min_u f_n(u) \text{ approaches } \arg\min_u f_0(u). \tag{1.1}$$

Conditions for (1.1) to hold have been given by Ibragimov and Has'minskii (1981), Salinetti and Wets (1986), Anisimov and Seilhamer (1994) and many others. These theorems are known under the name of M-theorems (minima-theorems).

Less attention has been paid to the asymptotic behaviour of solutions of stochastic equations and the related class of Z-theorems (zero-theorems). These are theorems which assert that under some conditions the weak convergence of some stochastic process $f_n(\cdot)$ to a limiting process $f_0(\cdot)$ entails that

$$\text{the solution set of } f_n(u) = 0 \text{ approaches weakly the solution set of } f_0(u) = 0. \tag{1.2}$$

A general Z-theorem for Banach space-valued processes has been given by Van der Vaart (1995). He considers the 'regular' case, i.e. the case where the limiting process is of the form $\eta_0(u) = Au + Z_0$, where A is an invertible linear operator and Z_0 is a Banach-valued random variable. Evidently, the solution of the limiting equation is $-A^{-1}Z_0$.

In this paper, we suggest a new approach which allows us to study more general models and more general limiting processes, but stick to the finite-dimensional case. In particular, we do not require the limiting process to be additively decomposable in a deterministic term, which depends on u, and a stochastic term, which does not. Examples of such undecomposable situations occur in non-regular statistical estimation models (where the condition of local asymptotic normality fails) as well as in non-smooth stochastic optimization.

The following set-up will be used in this paper. Let $f_n(\theta, \omega)$, $n > 0$, be a sequence of continuous (in θ) random functions defined on $\Theta \times \Omega_n$ with values in $\mathbb{R}^m$, where Θ is some open region in $\mathbb{R}^d$ and $(\Omega_n, A_n, \mathbf{P}_n)$ are probability spaces. We consider the stochastic equation

$$f_n(\theta, \omega) = 0 \tag{1.3}$$

and denote the set of possible solutions by $\boldsymbol{\theta}_n(\omega) = \{\theta : f_n(\theta, \omega) = 0\}$. Since f_n is continuous in θ, $(\boldsymbol{\theta}_n)$ is a sequence of random closed sets. We suppose further that the random functions f_n converge in distribution to a limit function f_0 defined on $(\Omega_0, \mathscr{A}_0, \mathbf{P}_0)$ and study the corresponding behaviour of the random closed sets $(\boldsymbol{\theta}_n)$. Since we allow the processes to be defined on different probability spaces, all results will be in the weak (distributional) sense. Conceptually, we rely on the notion of weak convergence of random closed sets. The reader is referred to Appendix 1 for a short review of this concept.

The paper is organized as follows. In Section 2 we study the notion of uniform convergence in distribution. Section 3 introduces the more general notion of band-convergence. Global convergence results are presented in Section 4. Applications to specific cases of limits of stochastic equations and to statistical estimates are contained in Sections 5 and 6. In Appendix A we have gathered together some facts about setwise convergence. Appendix B contains a new result on asymptotic inclusion of random sets.

2. Uniform convergence

We begin with a rather simple lemma for deterministic functions.

Lemma 2.1.

(i) If a sequence of deterministic functions $g_n(\theta)$ converges uniformly on each compact set K to a limit function $g_n(\theta)$, then we have for the solution sets

$$\limsup_n \{\theta : g_n(\theta) = 0\} \subseteq \{\theta : g_0(\theta) = 0\}.$$

Here lim sup denotes the topological upper limit as defined in Appendix A. Notice that the solution sets may be empty.

(ii) Suppose that g fulfils the following condition of separateness: there exists a $\delta > 0$ such that for any $y \in \mathbb{R}^m$, $|y| < \delta$, the equation

$$g_0(u) = y \tag{2.1}$$

has a proper unique solution. Then, for large n, $\{\theta : g_n(\theta) = 0\} \neq \varnothing$ and

$$\limsup_n \{\theta : g_n(\theta) = 0\} = \theta_0,$$

where θ_0 is the unique solution of $g_0(\theta) = 0$.

Proof. Let $g_n(\theta_n) = 0$. If θ is a cluster point of θ_n, then, by uniformity,

$$g_n(\theta_n) \to g_0(\theta),$$

which implies that θ is a root of g_0. The second statement is nearly obvious. $\square$

A generalization of this result for random functions will be proved in this section. We begin with some definitions.

For any function $g(\theta)$ and any compact set $K \subset \Theta$, denote by

$$\Delta_U(c, g(\cdot), K) = \sup\{|g(q_1) - g(q_2)| : |q_1 - q_2| \leqslant c, q_1, q_2 \in K\}$$

the modulus of continuity in uniform metric on the set K.

Definition 2.1. *The sequence of random functions $f_n(\theta)$ converges weakly uniformly (U-converges) to the function $f_0(\theta)$ on the set K if, for any $k > 0$ and for any $\theta_1 \in K, \ldots \theta_k \in K$, the multidimensional distribution of $(f_n(\theta_1), \ldots, f_n(\theta_k))$ converges weakly to the distribution of $(f_0(\theta_1), \ldots, f_0(\theta_k))$ and, for any $\varepsilon > 0$,*

$$\lim_{c\downarrow 0} \limsup_{n\to\infty} \mathbf{P}_n\{\Delta_U(c, f_n(\cdot), K) > \varepsilon\} = 0.$$

In other words, the sequence of measures generated by the sequence of functions $f_n(\cdot)$ in Skorokhod space D_K weakly converges to the measure generated by $f_0(\cdot)$.

Condition A. *We say that the random process $f(u, \omega)$ fulfils a condition of separateness if there exists a $\delta > 0$ such that, for any $y \in \mathbb{R}^m$, $|y| < \delta$, the equation*

$$f(u, \omega) = y \tag{2.2}$$

has for almost all ω a proper unique solution.

Definition 2.2. *A sequence $(\boldsymbol{\theta}_n)$ of random closed sets is called stochastically included in $\boldsymbol{\theta}_0$ in the limit if, for every collection of compact sets $K_1, \ldots, K_\ell$ and arbitrary ℓ,*

$$\limsup_n \mathbf{P}_n\{\boldsymbol{\theta}_n \cap K_1 \neq \varnothing, \ldots, \boldsymbol{\theta}_n \cap K_\ell \neq \varnothing\} \leqslant \mathbf{P}_0\{\boldsymbol{\theta}_0 \cap K_1 \neq \varnothing, \ldots, \boldsymbol{\theta}_0 \cap K_\ell \neq \varnothing\}.$$

If the limiting random set $\boldsymbol{\theta}_0$ is almost surely (a.s.) a singleton $\{\theta_0\}$ and all measurable selections $\tilde{\theta}_n \in \boldsymbol{\theta}_n$ converge in distribution to θ_0, we write

$$\theta_0 = \text{w-lim}_n \boldsymbol{\theta}_n. \tag{2.3}$$

Theorem 2.1.

(i) Suppose that the sequence of random functions $f_n(\theta)$ U-converges on any compact set $K \subseteq \Theta$ to the random function $f_0(\theta)$. Then $\boldsymbol{\theta}_n$ is stochastically included in $\boldsymbol{\theta}_0 = \{\theta : f_0(\theta) = 0\}$ in the limit.

(ii) In addition, let Condition A be fulfilled. If Θ is bounded and $\boldsymbol{\theta}_0$ is a.s. a singleton $\{\theta_0\}$, then $\lim_n \mathbf{P}_n\{\boldsymbol{\theta}_n \neq \varnothing\} = 1$ and

$$\theta_0 = \text{w-lim}_n \boldsymbol{\theta}_n. \tag{2.4}$$

Proof. The proof uses Skorokhod's (1956) method of representation on a common probability space. According to this method we can construct a new sequence of random functions $f'_n(\theta, \omega)$ and $f'_0(\theta, \omega)$ on a common probability space Ω' such that $f'_n(\theta)$ and $f_n(\theta)$ have the same finite-dimensional distributions and for almost all $\omega \in \Omega$ the sequence $f'_n(\theta, \omega)$ uniformly converges to $f'_0(\theta, \omega)$ on every compact set $K \subseteq \Theta$.

By Lemma 2.1 all cluster points of $\boldsymbol{\theta}'_n = \{\theta : f'_n(\theta, \omega) = 0\}$ are contained in $\boldsymbol{\theta}'_0 = \{\theta : f'_0(\theta, \omega) = 0\}$, i.e. $\limsup_n \boldsymbol{\theta}'_n \subseteq \boldsymbol{\theta}_0$. By Lemma B.1 in Appendix B, this proves part (i).

Further, if Condition A is satisfied, then a solution of equation (1.3) exists for large n with probability close to one because of the continuity of the function $f_n(\theta, \omega)$. If $\tilde{\theta}_n(\omega)$ is a measurable selection of $\boldsymbol{\theta}'_n$ which does not tend to θ_0, then there exists a subsequence n_k such that $\tilde{\theta}_{n_k}(\omega) \to \tilde{\theta} \neq \theta_0$. Using the uniform convergence of $f_n(\theta, \omega)$ we obtain that

$$f_n(\theta_{n_k}(\omega), \omega) \to f_0(\tilde{\theta}) = 0.$$

But θ_0 is the unique root of f_0, due to Condition A, and this contradiction proves part (ii) of the theorem. □

Theorem 2.1 applies typically to consistency proofs of estimates. In this class of applications, θ_0 is a constant. However, Z-theorems may also be used for deriving the asymptotic distribution of estimates. Here is a typical result of this kind:

Theorem 2.2. *Let the assumptions of Theorem 2.1(ii) be fulfilled, and suppose that θ_0 is deterministic. Further, let there exist a $\beta > 0$ and a non-random sequence $v_n \to \infty$ such that, for any $L > 0$, the sequence of functions*

$$\eta_n(u) := v_n^{\beta} f_n(\theta_0 + v_n^{-1} u)$$

U-converges in the region $\{|u| \leqslant L\}$ to the continuous random function $\eta_0(u)$ satisfying Condition A. Then there exists a measurable selection $\hat{\theta}_n$ from $\boldsymbol{\theta}$ such that the sequence of random variables $v_n(\hat{\theta}_n - \theta_0)$ weakly converges to the proper random variable γ_0 which is the unique solution of the equation

$$\eta_0(u) = 0. \tag{2.5}$$

Remark 2.1. In regular cases the random function $\eta_0(u)$ has the form $\xi_0 + G_0 u$, where ξ_0 and G_0 are vector- and matrix-valued (possibly dependent) random variables. In this case, if the matrix G_0 is not degenerated,

$$\gamma_0 = -G_0^{-1} \xi_0.$$

Proof. As before, we can assume without loss of generality that the sequence of functions $f_n(\theta_0 + v_n^{-1} u, \omega)$ and $\eta_0(u, \omega)$ are defined on the same probability space Ω such that

$$v_n^{\beta} f_n(\theta_0 + v_n^{-1} u, \omega) = \eta_0(u, \omega) + \beta_n(u, \omega),$$

where, for each $L > 0$,

$$\sup_{|u|<L} |\beta_n(u, \omega)| \to 0 \tag{2.6}$$

for almost all $\omega \in \Omega$.

Let us consider the equation

$$\eta_0(u, \omega) = -\beta_n(u, \omega). \tag{2.7}$$

Due to Condition A and the continuity of the left- and right-hand sides in (2.7), as

$$\sup_{|u|<L} |\beta_n(u, \omega)| \leqslant \delta,$$

then at least one solution of (2.7) exists. Denote a measurable selection by $\hat{u}_n(\omega)$. Again by Condition A, $\eta_0(u, \omega)$ has an inverse $\eta_0^{-1}(u, \omega)$ in the neighbourhood of the point $\gamma_0(\omega)$, and we can write the defining equation for $\hat{u}_n(\omega)$ in the form

$$\hat{u}_n(\omega) = \eta_0^{-1}(\beta_n(u, \omega), \omega). \tag{2.8}$$

According to (2.6), the right-hand side of (2.8) tends to $\eta_0^{-1}(0, \omega) = \gamma_0(\omega)$, which is the unique solution of the equation $\eta_0(u, \omega) = 0$. This proves Theorem 2.2 because each solution $\hat{u}_n$ of (2.7) is connected to the corresponding solution $\hat{\theta}_n$ of (1.3) by the relation $\hat{\theta}_n = \theta_0 + v_n^{-1} \hat{u}_n$, i.e. $\hat{u}_n = v_n(\hat{\theta}_n - \theta_0)$. □

3. Weakening the assumptions

Uniform convergence is a rather strong property. In connection with M-theorems uniform convergence may be replaced by epi-convergence, which is the convergence of the epigraphs. Recall that the epigraph of a function $z(\theta)$ is

$$\operatorname{epi} z = \{(\alpha, \theta) : \alpha \geqslant z(\theta)\}.$$

For the purpose of Z-theorems, we introduce here the notion of the q-band of a function, which is some nonlinear band around the graph of this function.

Definition 3.1. *Let $0 \leqslant q < 1$. The q-band of a function $f(\theta)$ is*

$$\Gamma(f(\cdot), q) = cl\{(\alpha, \theta) : |\alpha - f(\theta)| \leqslant q|f(\theta)|, \theta \in \Theta\},$$

where $cl\{B\}$ denotes the closure of the set B.

Lemma 3.1. *Let $g_n(\theta)$, $g_0(\theta)$ be continuous functions and $\boldsymbol{\theta}_n = \{\theta : g_n(\theta) = 0\}$, $\boldsymbol{\theta}_0 = \{\theta : g_0(\theta) = 0\}$. If $\limsup_n \Gamma(g_n(\cdot), 0) \subseteq \Gamma(g(\cdot), q)$ for some $0 < q < 1$, then $\limsup_n \boldsymbol{\theta}_n \subseteq \boldsymbol{\theta}_0$.*

Proof. Let $u_n \in \boldsymbol{\theta}_n$ and u be a cluster point of (u_n). We have to show that $u \in \boldsymbol{\theta}_0$. Since $(0, u_n) \in \Gamma(g_n, 0)$ and $(0, u)$ is a cluster point of $(0, u_n)$, it follows that $(0, u) \in \Gamma(g(\cdot), q)$, i.e. $|g(u)| \leqslant q|g(u)|$, whence $g(u) = 0$ and therefore $u \in \boldsymbol{\theta}_0$. □

Definition 3.2. *Let $f_n(\theta)$ and $f_0(\theta)$ be stochastic processes on $\mathbb{R}^d$. We say that the sequence $f_n(\cdot)$ band-converges to the process $f_0(\cdot)$ if, for some $0 < q < 1$, $\Gamma(f_n(\cdot), 0)$ is stochastically included in $\Gamma(f_0(\cdot), q)$ in the limit.*

Theorem 3.1. *Let the sequence $f_n(\cdot)$ band-converge to the process $f_0(\cdot)$ and let $\boldsymbol{\theta}_n$ be the set of zeros of $f_n(\theta)$ and $\boldsymbol{\theta}_0$ be the set of zeros of $f_0(\theta)$. Then $\boldsymbol{\theta}_n$ is stochastically included in $\boldsymbol{\theta}_0$ in the limit.*

Proof. Suppose that the theorem is false. Then there are compact sets $K_1, \ldots, K_\ell$ such that

$$\limsup_n \mathbf{P}_n\{\boldsymbol{\theta}_n \cap K_1 \neq \varnothing, \ldots, \boldsymbol{\theta}_n \cap K_\ell \neq \varnothing\} > \mathbf{P}_0\{\boldsymbol{\theta} \cap K_1 \neq \varnothing, \ldots, \boldsymbol{\theta} \cap K_\ell \neq \varnothing\}.$$

In particular, there is a subsequence (n_i) such that

$$\lim_{n_i} \mathbf{P}_{n_i}\{\boldsymbol{\theta}_{n_i} \cap K_1 \neq \varnothing, \ldots, \boldsymbol{\theta}_{n_i} \cap K_\ell \neq \varnothing\} > \mathbf{P}_0\{\boldsymbol{\theta} \cap K_1 \neq \varnothing, \ldots, \boldsymbol{\theta} \cap K_\ell \neq \varnothing\}. \quad (3.1)$$

$\Gamma(f_{n_i}, 0)$ is a sequence of random closed sets which contains a weakly convergent subsequence $\Gamma(f_{n_i'}, 0)$. By Skohorod's theorem, we may construct versions on a common probability space which converge pointwise, i.e. $\Gamma'(f_{n_i'}, 0) \to \Gamma_0$ a.s. Furthermore, since by assumption Γ_0 is stochastically smaller than $\Gamma(f_0, q)$, we may by Theorem B.1 (Appendix B) assume that there is a version such that $\Gamma_0' \subseteq \Gamma'(f_0, q)$ a.s. Thus $\lim_{n_i'} \Gamma'(f_{n_i'}, 0) \subseteq \Gamma'(f_0, q)$. Therefore, for this version, by Lemma 3.1, $\limsup \boldsymbol{\theta}_{n_i'} \subseteq \boldsymbol{\theta}$, which contradicts (3.1). □

Remark 3.1. The assumptions of Theorem 3.1 are fulfilled if the sequence $f_n(t)$ converges uniformly to f_0. By Skorokhod embedding, we may without loss of generality assume that $\sup_u |f_n(u) - f_0(u)| \to 0$ a.s. If (α_n, u_n) are such that $|\alpha_n - f_n(u_n)| \leqslant q|f_n(u)|$, then every cluster point (α, u) of this sequence satisfies $|\alpha - f_0(u)| \leqslant q|f_0(u)|$, which completes the argument.

Example 3.1. Theorem 3.1 is not included in Theorem 2.1. Here is an example. Let $f_n(\theta, \omega) = \bar{f}_n(\theta)(1 + \xi_n(\omega))$, where the deterministic functions $\bar{f}_n$ uniformly converge to a continuous limit function $\bar{f}$. Let $0 < q < 1$. If

$$\mathbf{P}_n\{|\xi_n| < q\} \to 1$$

as $n \to \infty$, the assumptions of Theorem 3.1 are fulfilled, but not necessarily those of Theorem 2.1.

4. Global convergence

The result of Theorem 2.2 is valid only for some solution (not any) which belongs to a close neighbourhood of the order $O(v_n^{-1})$ of the point θ_0. It is possible to show that there are examples where the conditions of Theorem 2.2 are fulfilled and there exist solutions θ_n' such that θ_n' are of order ε_n, where ε_n converges arbitrarily slow to zero. That is why it is important to find additional conditions that guarantee the convergence for the sequence $v_n(\hat{\theta}_n - \theta_0)$ for all solutions $\hat{\theta}_n$. The following theorem gives such conditions:

Theorem 4.1. *Suppose that the conditions of Theorem 2.2 hold and there exists* $c_0 > 0$ *such that, for any sequence* $\delta_n > 0$ *with the properties* $\delta_n \to 0$, $v_n\delta_n \to \infty$,

$$\lim_{L\to\infty} \liminf_{n\to\infty} \mathbf{P}_n \left\{ \inf\left\{ v_n^\beta \left| f_n\left(\theta_0 + \frac{u}{v_n}\right)\right| : L \leqslant |u| \leqslant v_n\delta_n \right\} > c_0 \right\} = 1. \tag{4.1}$$

Then, for any solution $\hat{\theta}_n$ *of* (1.3), *the sequence* $v_n(\hat{\theta}_n - \theta_0)$ *weakly converges to the unique solution* γ_0 *of* (2.5).

Proof. According to Theorem 2.1(ii), with probability close to one, the set of possible solutions of (1.3) belongs to some δ_n-neighbourhood of the point θ_0, where $\delta_n \to 0$. Then under condition (4.1), with probability close to one, the set of possible solutions of (1.3) belongs to the region $\{|\theta - \theta_0| < L/v_n\}$ for L large.

Let us now consider in a new scale of variables the sequence of functions $\eta_n(u) = v_n^\beta f_n(\theta_0 + v_n^{-1}u)$. This sequence U-converges in the region $\{|u| \leqslant L\}$ to the function $\eta_0(u)$.

Now we can construct sequences $\eta_n'(u, \omega)$ and $\eta_0'(u, \omega)$ on the same probability space Ω', having the same distributions as $\eta_n(u)$ and $\eta_0(u)$ and such that $\eta_n'(u, \omega)$ converges uniformly to $\eta_0'(u, \omega)$ for all $\omega \in \Omega_0$, where $\mathbf{P}(\Omega_0) = 1$. Introduce

$$G(L) = \{\omega : \inf\{|\eta_n(u)| : L \leqslant |u| \leqslant v_n\delta_n\} > c_0/2 \text{ for sufficiently large } n\}$$

and

$$D(L) = \{\omega : |\gamma_0(\omega)| < L\},$$

where $\gamma_0(\omega)$ is a solution of the equation

$$\eta_0(u, \omega) = 0. \tag{4.2}$$

For any $\omega \in G(L)$ and large n, the set of possible solutions of $f_n(\theta) = 0$ belongs to the region $\{|\theta - \theta_0| < L/v_n\}$. Then according to Theorem 2.1, for any $\omega \in D(L) \cap G(L) \cap \Omega_0$, $\lim_n \mathbf{u}_n(\omega) = \gamma_0(\omega)$, where $\mathbf{u}_n(\omega)$ is the set of possible solutions of the equation

$$\eta_n(u, \omega) = 0. \tag{4.3}$$

We note that the corresponding solutions of (1.3) and (4.3) are connected by the relation $\tilde{u}_n = v_n(\tilde{\theta}_n - \theta_0)$. As, according to Theorem 2.2, γ_0 is a proper unique solution of (4.2), this implies that $\mathbf{P}(D(L)) \to 1$ as $L \to \infty$, and correspondingly, according to (4.1), $\mathbf{P}(G(L)) \to 1$. This proves the statement of Theorem 4.1. $\square$

Condition (4.1) is of rather general character and we now consider a typical situation for which this condition is true. Suppose without loss of generality that we have a representation

$$f_n(\theta) = \tilde{f}_n(\theta) + \eta_n(\theta),$$

where $\tilde{f}_n(\theta)$ is some deterministic function.

Theorem 4.2. *Let Theorem 2.1(ii) and the following conditions hold:*

(i) There exists $\beta > 0$ and a non-random sequence $v_n \to \infty$ such that, for any $L > 0$, the sequence of deterministic functions $v_n^\beta \tilde{f}_n(\theta_0 + v_n^{-1}u)$ U*-converges in the region $\{|u| \leqslant L\}$ to the continuous function $\varphi_0(u)$.*

(ii) The sequence $v_n^\beta \eta_n(\theta_0)$ weakly converges to a proper random variable η_0.

(iii) The function $\varphi_0(u)$ satisfies Condition A in the following form: for any $y \in \mathbb{R}^m$ the equation

$$\varphi_0(u) = y \tag{4.4}$$

has a unique solution.

(iv) There exists $c_0 > 0$ such that, for any sequence $\delta_n > 0$ with the condition that $\delta_n \to 0$, $v_n \delta_n \to \infty$,

$$\lim_{L\to\infty} \liminf_{n\to\infty} \inf_{L \leqslant |u| \leqslant v_n\delta_n} v_n^\beta \left| \tilde{f}_n\left(\theta_0 + \frac{u}{v_n}\right) - \tilde{f}_n(\theta_0)\right| \geqslant c_0. \tag{4.5}$$

(v) For any sequence $\delta_n \to 0$ and any $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left\{ v_n^\beta \sup_{|z|\leqslant\delta_n} |\eta_n(\theta_0 + z) - \eta_n(\theta_0)| > \varepsilon \right\} = 0. \tag{4.6}$$

Then for any solution $\hat{\theta}_n$ of (1.3) the sequence $v_n(\hat{\theta}_n - \theta_0)$ weakly converges to the unique solution γ_0 of the equation

$$\varphi_0(u) + \eta_0 = 0.$$

Remark 4.1. If, for some $a > 0$, $0 < \varepsilon \leqslant \beta$ and any $u \in \mathbb{R}^r$,

$$|\tilde{f}_n(\theta_0 + u) - \tilde{f}_n(\theta_0)| \geqslant a|u|^\varepsilon + \alpha_n(u), \tag{4.7}$$

where

$$\sup_{|u|\leqslant\delta_n} v_n^\beta |\alpha_n(u)| \to 0,$$

then condition (4.5) is satisfied.

Proof. It is easy to see that under conditions (i)–(iii) of Theorem 4.2 the conditions of Theorem 2.2 are satisfied, but with $\eta_0(u)$ replaced by $\varphi_0(u) + \eta_0$. Then conditions (4.5) and (4.6) imply condition (4.1) of Theorem 4.1 and the statement of Theorem 4.2 follows from Theorems 2.2 and 4.1. □

Example 4.1. Let the function $f_0(\theta)$, $\theta \in \Theta \subset \mathbb{R}^r$, be of the form $f_0(\theta) = A\Lambda(\theta)$, where $\Lambda(\theta)$ is a diagonal matrix with elements sign $\theta_i|\theta_i|^\beta$, $i = 1, \ldots, r$ and θ_i are the components of the vector $\theta = (\theta_1, \ldots, \theta_r)$. Suppose, further, that the functions $f_n(\theta)$ are of the form

$$f_n(\theta) = f_0(\theta) + \frac{1}{n^\gamma}\zeta(\theta),$$

where $\zeta(\theta)$, $\theta \in \Theta$, is an arbitrary random function that is continuous at the point $\theta = 0$ with probability one and bounded in probability in each compact region and $\gamma > 0$. If the matrix A is invertible, then, for $n \to \infty$, the relation (2.4) holds with $\theta_0 = 0$ and also

$$\text{w-lim}_{n\to\infty} n^{\gamma/\beta}\boldsymbol{\theta}_n = \kappa,$$

where the random vector $\kappa = (\kappa_1, \ldots, \kappa_r)$ is of the form

$$\kappa_i = \text{sign}\,\tilde{\zeta}_i|\tilde{\zeta}_i|^{1/\beta},$$

and $\tilde{\zeta}_i$, $i = 1, \ldots, r$ are components of the vector $\tilde{\zeta} = A^{-1}\zeta(0)$.

Remark 4.2. If, in particular, $\beta = 1$ and the variable $\zeta(0)$ has a multidimensional Gaussian distribution with mean a and covariance matrix B^2, then the variable κ also has multidimensional Gaussian distribution with mean $A^{-1}a$ and covariance matrix $A^{-1}B^2(A^{-1})^{\mathrm{T}}$.

Proof. Under our conditions the sequence of functions $f_n(\theta)$ U-converges in each compact region $K \subset \Theta$ to the function $f_0(\theta)$. That implies the first part of the statement.

Further, as the function $\zeta(\theta)$ is continuous at the point 0, the sequence $\sup_{|u|\leq L}|\zeta(v_n^{-1}u) - \zeta(0)|$ U-converges to 0 for any $L > 0$ and any sequence $v_n \to \infty$, and it is true that, for any $L > 0$, the sequence of functions $n^\gamma f_n(n^{-\gamma/\beta}u)$ U-converges in the region $\{|u| \leq L\}$ to the continuous random function $\eta_0(u) = A\Lambda(u) + \zeta(0)$. It is obvious that the equation

$$A\Lambda(u) + \zeta(0) = 0$$

has a unique solution κ and conditions of Theorem 2.2 are satisfied. Now to prove a global convergence it is sufficient to check condition (4.7) in Remark 4.1.

We can write

$$|A\Lambda(\theta)| = |\theta|^\beta|A\Lambda(e_\theta)|, \tag{4.8}$$

where $e_\theta = |\theta|^{-1}\theta$ is a unit vector. Denote

$$a = \inf_{|e|=1}|A\Lambda(e)|.$$

As the matrix A is invertible and the function $\Lambda(\theta)$ is continuous, we obtain that $a > 0$. Then from (4.8) we obtain

$$|A\Lambda(\theta)| \geq a|\theta|^\beta,$$

which proves the second part of our statement. □

5. Solutions of stochastic equations

In this section we consider applications of our results to the study of the behaviour of approximately calculated solutions of deterministic equations under stochastic noise. Let us consider the following model. Suppose that we want to find a solution of a deterministic equation

$$f(\theta) = 0, \tag{5.1}$$

where $f(\theta)$ is some continuous function, $\theta \in \Theta$, and Θ is some bounded region in $\mathbb{R}^r$, but we can only observe the function $f(\theta)$ with random errors in the form:

$$r_k(\theta) = f(\theta) + \xi_k(\theta), \qquad 1 \leqslant k \leqslant n,$$

where $\{\xi_k(\theta), \theta \in \Theta\}$, $k \geqslant 1$, are jointly independent families of random functions that are measurable in θ, continuous with probability one and satisfy $\mathrm{E}\xi_k(\cdot) = 0$. It is natural to approximate $f(\theta)$ by

$$f_n(\theta) = \frac{1}{n}\sum_{k=1}^{n} r_k(\theta) = f(\theta) + \eta_n(\theta),$$

where

$$\eta_n(\theta) = \frac{1}{n}\sum_{k=1}^{n} \xi_k(\theta).$$

We study the asymptotic behaviour of solutions of the equation

$$f_n(\theta) = 0. \tag{5.2}$$

As before, denote by $\boldsymbol{\theta}_0$ the set of possible solutions to (5.1) and by $\boldsymbol{\theta}_n$ the set of possible solutions to (5.2).

Theorem 5.1. *Let families of random variables $\{\xi_k(\theta), \theta \in \Theta\}$ be independent (for different k) and identically distributed. Suppose also that the following conditions hold:*

(i) For any $\varepsilon > 0$ and any compact set $K \subset \Theta$,

$$\lim_{c\downarrow 0} \limsup_{n\to\infty} \mathbf{P}_n\{\Delta_U(c, \eta_n(\cdot), K) > \varepsilon\} = 0. \tag{5.3}$$

(ii) The function $f(\theta)$ satisfies the condition that there exists $\delta > 0$ such that the equation

$$f(\theta) = y,$$

at each $|y| < \delta_0$, has at least one solution, and there exists an inner point $\theta_0 \in \Theta$ such that $f(\theta_0) = 0$.

Then, as $n \to \infty$, $\mathbf{P}_n\{\boldsymbol{\theta}_n \neq \varnothing\} \to 1$ *and* $\boldsymbol{\theta}_n$ *is stochastically included in* $\boldsymbol{\theta}_0$ *in the limit.*

Proof. We represent the function $f_n(\theta)$ in the form

$$f_n(\theta) = f(\theta) + \eta_n(\theta).$$

By the law of large numbers it follows that, at each $\theta \in \Theta$,

$$\text{P-lim}_{n\to\infty}\eta_n(\theta) = 0, \tag{5.4}$$

where P-lim denotes convergence in probability, and condition (5.3) implies that the sequence of functions $\eta_n(\theta)$ U-converges to 0 on each compact set K, and correspondingly that the sequence $f_n(\cdot)$ U-converges to $f(\cdot)$. Then our statement follows directly from Theorem 2.1. □

Condition (5.3) is rather general and sometimes difficult to check. We now give some more concrete conditions sufficient for it.

Corollary 5.1. *Let*

$$\lim_{c\downarrow 0} \mathrm{E}\Delta_U(c, \xi_1(\cdot), K) = 0, \tag{5.5}$$

for any compact set $K \subset \Theta$. *Then condition (5.3) holds.*

Proof. By

$$\Delta_U(c, \eta_n(\cdot), K) \leqslant \frac{1}{n}\sum_{k=1}^{n} \Delta_U(c, \xi_k(\cdot), K) \tag{5.6}$$

and Chebyshev's inequality we obtain that

$$P\{\Delta_U(c, \eta_n(\cdot), K) > \varepsilon\} \leqslant \frac{1}{\varepsilon}\mathrm{E}\Delta_U(c, \xi_1(\cdot), K).$$

This relation, together with (5.5), implies condition (5.3) of Theorem 5.1. □

Remark 5.1. Condition (5.5) is satisfied if there exists a matrix derivative $\nabla_\theta \xi_1(\theta)$ and, for any compact set $K \subset \Theta$,

$$\sup_{\theta\in K} \mathrm{E}|\nabla_\theta \xi(\theta)| \leqslant C_K < \infty.$$

Now let us consider the asymptotic distribution of the solutions.

Theorem 5.2. *Suppose that the assumptions of Theorem 5.1 and the following conditions hold:*

(i) *For some* $\beta > 0$ *uniformly in the unit sphere* $\{e : |e| = 1\}$,

$$h^{-\beta}(f(\theta_0 + he) - f(\theta_0)) \to A(e)e \tag{5.7}$$

as $h \downarrow 0$ *(here A(e) is some matrix possibly depending on the vector e).*

(ii) For some $\gamma, \frac{1}{2} \leqslant \gamma < 1$,

$$\text{w-lim}_n \frac{1}{n^{\gamma}} \sum_{k=1}^{n} \xi_k(\theta_0) = \zeta, \tag{5.8}$$

where ζ *is a random vector with a stable distribution with parameter* $1/\gamma$.

(iii) For each $L > 0$,

$$\lim_{n\to\infty} \mathbf{P}_n\{\{\sup|q_n(u)| : |u| \leqslant Ln^{-(1-\gamma)/\beta}\} > \varepsilon\} = 0, \tag{5.9}$$

where

$$q_n(u) = \frac{1}{n^{\gamma}} \sum_{k=1}^{n} (\xi_k(\theta_0 + u) - \xi_k(\theta_0)).$$

(iv) For each $y \in \mathbb{R}^r$, *a solution of the equation*

$$A\left(\frac{u}{|u|}\right)|u|^{\beta-1}u = y$$

exists and is unique.

Then there exists a subsequence of solutions $\tilde{\theta}_n$ *of (1.3) such that*

$$\text{w-lim}_n n^{(1-\gamma)/\beta}(\tilde{\theta}_n - \theta_0) = \gamma_0, \tag{5.10}$$

where γ_0 *is the unique solution of the equation*

$$A\left(\frac{u}{|u|}\right)|u|^{\beta-1}u + \zeta = 0.$$

Proof. We have to study the behaviour of the function $v_n^{\beta} f_n(\theta_0 + v_n^{-1}u)$. Let us choose v_n in the form $v_n = n^{(1-\gamma)/\beta}$. Then

$$v_n^{\beta} f_n(\theta_0 + v_n^{-1}u) = v_n^{\beta}(f(\theta_0 + v_n^{-1}u) - f(\theta_0)) + q_n(v_n^{-1}u) + \frac{1}{n^{\gamma}} \sum_{k=1}^{n} \xi_k(\theta_0). \tag{5.11}$$

From condition (5.7) it follows that the first item on the right in (5.11) converges uniformly on u in each bounded region $\{|u| \leqslant L\}$ to the function

$$A\left(\frac{u}{|u|}\right)|u|^{\beta-1}u,$$

the second item uniformly converges to 0, and the last one weakly converges to the variable ζ. This means that the right-hand side of (5.11) converges uniformly in u in each bounded region $\{|u| \leqslant L\}$ to the function

$$A\left(\frac{u}{|u|}\right)|u|^{\beta-1}u + \zeta.$$

The statement of Theorem 5.2 now follows directly from Theorem 2.2. $\square$

Now let us consider conditions of global convergence.

Theorem 5.3. *Suppose that the assumptions of Theorem 5.2 holds, but with condition (iii) replaced by the following:*

(iii)′ For any sequence $\delta_n > 0,\ \delta_n \to 0,$

$$\lim_{n\to\infty} P\left\{ \sup_{|v|\leqslant\delta_n} |q_n(v)| > \varepsilon \right\} = 0, \tag{5.12}$$

and also

$$a = \inf_{|e|=1} |A(e)| > 0. \tag{5.13}$$

Then w-lim$v_n(\boldsymbol{\theta}_n - \theta_0) = \gamma_0$, *where* $v_n = n^{(1-\gamma)/\beta}$.

Proof. It easy to see that under our assumptions conditions (i)–(iii) and (v) of Theorem 4.2 hold. Then according to (5.7) and (5.13), at small enough v, we obtain

$$|f(\theta_0 + v) - f(\theta_0)| = |A(v/|v|)|v|^{\beta-1}v + o(|v|^\beta)| \geqslant a|v|^\beta - |o(|v|^\beta)|.$$

This relation and Remark 4.1 (see (4.7)) imply the theorem. □

We now give, for particular cases, sufficient conditions for checking condition (iii) of Theorem 5.2.

Remark 5.2. If, for any $L > 0$,

$$\lim_{n\to\infty} n^{1-\gamma} \mathrm{E} \sup\{|\xi_1(\theta_0 + n^{-(1-\gamma)/\beta}u) - \xi_1(\theta_0)| : |u| \leqslant L\} = 0, \tag{5.14}$$

then (5.9) holds. The proof is based on the same arguments as the proof of Theorem 5.1.

Example 5.1. Let the function $f(\theta)$ be continuously differentiable and let $\nabla_\theta f(\theta)$ denote its matrix derivative, i.e.

$$\lim_{h\to 0} h^{-1}(f(\theta + hz) - f(\theta)) \to \nabla_\theta f(\theta) z, \tag{5.15}$$

for any vector $z \in \mathbb{R}^r$. Suppose that condition (5.9) holds, that

$$\mathrm{E}\xi_1(\theta_0)\xi_1(\theta_0)^{\mathrm{T}} = B^2, \tag{5.16}$$

and that the matrix $G = \nabla_\theta f(\theta_0)$ is invertible. Then the statement of Theorem 5.2 holds, where $\beta = 1$, $\gamma = \frac{1}{2}$ and vector γ_0 has a Gaussian distribution with mean 0 and covariance matrix GB^2G^{T}. It is easy to check that the sequence of functions $\sqrt{n} f_n(\theta_0 + n^{-1/2}u)$ converges uniformly on u in each bounded region $\{|u| \leqslant L\}$ to the function $Gu + N(0, B^2)$, where $N(0, B^2)$ is a vector that has a Gaussian distribution with mean 0 and covariance matrix B^2. This implies our statement.

Example 5.2. Let us now consider a special case of errors of the form

$$\xi_k(\theta) = G(\theta)\xi_k, \qquad k \geqslant 1, \tag{5.17}$$

where $G(\theta)$ is some matrix function, and $\xi_k,\ k \geqslant 1$, is the sequence of independently and identically distributed random vectors in $\mathbb{R}^r$ such that $\mathrm{E}\xi_k = 0$. Suppose that condition (iii) of Theorem 5.1 holds and $G(\theta)$ is some continuous function. Then (2.4) holds. Suppose, further, that conditions (i) and (iv) of Theorem 5.2 hold and the variables ξ_k satisfy condition (5.8). Then (5.10) of Theorem 5.2 holds, where γ_0 is a unique solution of the equation

$$A\left(\frac{u}{|u|}\right)|u|^{\beta-1}u + G(\theta_0)\zeta = 0. \tag{5.18}$$

It is easy to see that

$$\Delta_U(c,\, \eta_n(\cdot),\, K) \leqslant \Delta_U(c,\, G(\cdot),\, K)\frac{1}{n}\left|\sum_{k=1}^{n}\xi_k\right|.$$

But $G(\theta)$ is uniformly continuous on each compact set K, and the variable $(1/n)|\sum_{k=1}^{n}\xi_k|$ converges to 0 in probability according to the law of large numbers. This implies the statement of the first part. In order to prove the second part, we need to check condition (iii) of Theorem 5.2. We choose v_n in the form $v_n = n^{(1-\gamma)/\beta}$. Then, due to construction (5.17), we see that

$$\sup\{|q_n(u)| : |u| \leqslant Lv_n^{-1}\} \leqslant \sup_{|u|\leqslant L}\left|G\left(\theta_0 + \frac{u}{v_n}\right) - G(\theta_0)\right|\left|n^{-\gamma}\sum_{k=1}^{n}\xi_k\right|. \tag{5.19}$$

Now the variable $|n^{-\gamma}\sum_{k=1}^{n}\xi_k|$ is bounded by probability according to condition (5.8) and, for any fixed $L > 0$ uniformly in the region $|u| \leqslant L$,

$$\sup_{|u|\leqslant L}\left|G\left(\theta_0 + \frac{u}{v_n}\right) - G(\theta_0)\right| \to 0,$$

which implies, according to Theorem 5.2, the second part of our statement.

6. Moment estimators

Now let us consider applications of the Z-theorems to problems of statistical parameter estimation by the method of moments. Let $s_{nk},\ 0 \leqslant k \leqslant n$, be a triangular (random or non-random) system with values in $\mathbb{R}^r$. Also let $\{\gamma_k(\alpha),\ \alpha \in \mathbb{R}^r\},\ k \geqslant 0$, be parametric families of random variables with values in $\mathbb{R}^m$, which are jointly independent and independent of (s_{nk}). For simplicity, suppose that the distributions of random variables $\gamma_k(\alpha)$ do not depend on k. We observe variables s_{nk} and $y_{nk} = \gamma_k(s_{nk}),\ k \leqslant n$, where n is the number of observations. Suppose now that expectations of the variables $\{\gamma_k(\alpha),\ \alpha \in \mathbb{R}^r\}$ exist and belong to the parametric family of functions $\{g(\theta,\alpha),\ \theta \in \Theta,\ \alpha \in \mathbb{R}^r\}$ and $\mathrm{E}\gamma_1(\alpha) = g(\theta_0, \alpha)$, where θ_0 is some inner point in the region Θ. The moment estimator is the solution of the equation

$$n^{-1}\sum_{k=1}^{n} g(\theta,\, s_{nk}) - n^{-1}\sum_{k=1}^{n} y_{nk} = 0. \tag{6.1}$$

Denote as before by $\boldsymbol{\theta}_n$ the set of possible solutions of (6.1). Now we study its asymptotic behaviour as $n \to \infty$.

Theorem 6.1. *Suppose the following conditions hold:*

(i) There exists a continuous deterministic function $s(t)$ on the interval $[0, 1]$ *such that the sequence s_{nk} satisfies the relation*

$$\text{P-lim}_{n\to\infty} \max_{0\leqslant k\leqslant n} |s_{nk} - s(k/n)| = 0. \tag{6.2}$$

(ii) The variables $\gamma_k(\alpha)$ satisfy the following condition: for any $L > 0$,

$$\lim_{N\to\infty} \sup_{|\alpha|\leqslant L} \mathrm{E}|\gamma_1(\alpha)|\chi\{|\gamma_1(\alpha)| > N\} = 0. \tag{6.3}$$

(iii) The function $g(\theta, \alpha)$ is continuous on both arguments (θ, α) and there exists a $\delta > 0$ such that the equation

$$\int_0^1 g(\theta, s(u))\,\mathrm{d}u - \int_0^1 g(\theta_0, s(u))\,\mathrm{d}u = v$$

has a unique solution for any $|v| < \delta$.

Then $\lim_n \mathbf{P}_n\{\boldsymbol{\theta}_n \neq \varnothing\} = 1$ *and* $\text{w-lim}_n \boldsymbol{\theta}_n = \theta_0$.

Proof. It can be easily seen that under conditions (6.2) and (6.3), the second term on the left-hand side of (6.1) converges in probability to $\int_0^1 g(\theta_0, s(u))\,\mathrm{d}u$. The first term converges for any $L > 0$ uniformly in $|\theta| \leqslant L$ to $\int_0^1 g(\theta, s(u))\,\mathrm{d}u$. Our statement now follows from Theorem 2.1. □

Let us now consider the asymptotic distribution of the estimates.

Theorem 6.2. *Suppose that the assumptions of Theorem 6.1 and the following conditions hold:*

(i) There exists a family of continuous (in both arguments) matrices $A(e, \alpha)$ such that, for some $\beta > 0$ and for any $L > 0$ uniformly in the region $\{(e, \alpha) : |e| = 1, |\alpha| \leqslant L\}$, as $h \downarrow 0$,

$$h^{-\beta}(g(\theta_0 + he, \alpha) - g(\theta_0, \alpha)) \to A(e, \alpha)e. \tag{6.4}$$

(ii) There exists a continuous function $a(\lambda, \alpha)(a(0, \alpha) = 0)$ such that for some γ, $1 < \gamma \leqslant 2$, as $h \to 0$,

$$\mathrm{E}\exp\{\mathrm{i}h\langle\lambda, \gamma_1(\alpha) - g(\alpha)\rangle\} = 1 + h^{\gamma}a(\lambda, \alpha) + o(h^{\gamma}, \alpha), \tag{6.5}$$

where, for any $L > 0$, $\lim_{h\to 0}\sup_{|\alpha|<L} h^{-\gamma}o(h^{\gamma}, \alpha) \to 0$.

(iii) For each $y \in \mathbb{R}^r$ a solution of the equation

$$\tilde{A}\left(\frac{u}{|u|}\right)|u|^{\beta-1}u = y$$

exists and is unique, where $\tilde{A}(e) = \int_0^1 A(e, s(v))\,\mathrm{d}v$.

Then there exists a solution $\hat{\theta}_n$ *of (6.1) such that*

$$\text{w-lim}\, n^{(\gamma-1)/(\gamma\beta)}(\hat{\theta}_n - \theta_0) = \gamma_0, \tag{6.6}$$

where γ_0 *is the unique solution of the equation*

$$\tilde{A}\left(\frac{u}{|u|}\right)|u|^{\beta-1}u + \zeta = 0$$

and the vector ζ *has a stable distribution with characteristic function*

$$\mathrm{E}\exp\{\mathrm{i}\langle\lambda,\, \zeta\rangle\} = \exp\left\{\int_0^1 a(\lambda,\, s(v))\,\mathrm{d}v\right\}. \tag{6.7}$$

Proof. Denote by $f_n(\theta)$ the left-hand side of (6.1). Put $v_n = n^{(\gamma-1)/\gamma\beta}$. Then we can write

$$v_n^{\beta} f_n(\theta_0 + v_n^{-1}u)$$
$$= n^{-1}\sum_{k=1}^{n} v_n^{\beta}(g(\theta_0 + v_n^{-1}u,\, s_{nk}) - g(\theta_0,\, s_{nk})) - n^{-1/\gamma}\sum_{k=1}^{n}(\gamma_k(s_{nk}) - g(\theta_0,\, s_{nk})). \tag{6.8}$$

It is not hard to prove, using conditions (6.2) and (6.5) and the continuity of the function $a(\lambda,\, \alpha)$, that the second term on the right-hand side of (6.8) weakly converges to the variable ζ (see (6.7)). The first term can be represented in the form

$$n^{-1}\sum_{k=1}^{n} A\left(\frac{u}{|u|},\, s_{nk}\right)|u|^{\beta-1}u + o(1),$$

and this term U-converges in the variable u, for any bounded region $\{|u| \leqslant L\}$, to the value $\tilde{A}(u/|u|)|u|^{\beta-1}u$. This implies our statement. □

Corollary 6.1. *Suppose that the conditions of Theorem 6.1 hold and there exist a continuous matrix of partial derivatives* $R(\theta,\, \alpha) = \nabla_\theta g(\theta,\, \alpha)$ *and a continuous matrix of second moments* $B^2(\alpha) = \mathrm{E}(\gamma_1(\alpha) - g(\alpha))(\gamma_1(\alpha) - g(\alpha))^{\mathrm{T}}$. *Suppose, further, that the matrix* $\int_0^1 R(\theta_0,\, s(u))\,\mathrm{d}u$ *is not degenerate and the variables* $\gamma_k(\alpha)$ *satisfy a Lindeberg condition in the following form: for any* $L > 0$,

$$\lim_{N\to\infty}\sup_{|\alpha|\leqslant L}\mathrm{E}|\gamma_1(\alpha)|^2\chi\{|\gamma_1(\alpha)| > N\} = 0. \tag{6.9}$$

Then there exists a solution $\hat{\theta}_n$ *of (6.1) such that the sequence* $\sqrt{n}(\hat{\theta}_n - \theta_0)$ *weakly converges to a Gaussian distribution with mean 0 and covariance matrix* $\tilde{R}^{-1}\tilde{B}^2(\tilde{R}^{-1})^{\mathrm{T}}$, *where*

$$\tilde{R} = \int_0^1 R(\theta_0,\, s(v))\,\mathrm{d}v, \qquad \tilde{B}^2 = \int_0^1 B^2(s(v))\,\mathrm{d}v.$$

Proof. We put $v_n = \sqrt{n}$, $\beta = 1$. Then it can be easily seen, using conditions (6.2) and (6.9) and the continuity of the function $B(\alpha)$, that the second term on the right of (6.8) weakly converges to the variable $\int_0^1 B(s(v))\,\mathrm{d}w(v)$, where $w(v)$ is a standard Wiener process in $\mathbb{R}^r$. The first term can be represented in the form

$$n^{-1}\sum_{k=1}^{n} R(\theta_0 + n^{-1/2}q_{nk}u,\, s_{nk})\,u,$$

where $|q_{nk}| \leqslant 1$, $k \geqslant 0$, and this term U-converges in u to the value $\int_0^1 R(\theta_0,\, s(v))\,\mathrm{d}v\,u$ in any bounded region $\{|u| \leqslant L\}$. Then, according to Theorem 2.2, there exists a solution $\hat{\theta}_n$ such that the sequence $\sqrt{n}(\hat{\theta}_n - \theta_0)$ weakly converges to the variable

$$\left[\int_0^1 R(\theta_0,\, s(t))\,\mathrm{d}t\right]^{-1}\int_0^1 B(s(v))\,\mathrm{d}w(v),$$

which has a Gaussian distribution with mean 0 and covariance matrix $\tilde{R}^{-1}\tilde{B}^2(\tilde{R}^{-1})^{\mathrm{T}}$. □

Remark 6.1. Condition (6.2) is satisfied for rather wide classes of stochastic systems that develop in a recurrent fashion (for instance, Markov systems) and it is oriented on non-stationary (transient) conditions. An average principle for general stochastic recurrent sequences is given in Anisimov (1991). Analogous results can be obtained in stationary cases under the condition that there exists a probability measure $\pi(A)$ on the Borel field of $\mathbb{R}^r$ such that, for any bounded measurable function $\varphi(\alpha)$, $\alpha \in \mathbb{R}^r$,

$$\text{P-lim}_{n\to\infty} n^{-1}\sum_{k=1}^{n}\varphi(s_{nk}) = \int_{R^r}\varphi(\alpha)\pi(\mathrm{d}\alpha) \tag{6.10}$$

(for instance s_{nk} can be a Markov ergodic sequence). Using the same technique, we can study the behaviour of maximum-likelihood and least-squares estimators. We mention that asymptotic properties of maximum-likelihood estimators constructed by observations on trajectories of recurrent processes of semi-Markov type, on the base of the same technique (analysis of maximum-likelihood equations), are studied in Anisimov and Orazklychev (1993).

Appendix A: Some properties of random closed sets

We review here some basic facts of random set theory; the reader is referred to Salinetti and Wets (1986) for more details.

Let $\mathscr{C}$ be the class of all closed sets in $\mathbb{R}^d$. For closed sets, we introduce the notions of lim inf and lim sup (in the topological sense):

$$\liminf_n C_n = \{u : \exists \text{ a sequence } (u_n) \text{ with } u_n \in C_n \text{ such that } u_n \to u\},$$

$$\limsup_n C_n = \{u : \exists \text{ a subsequence } (u_{n_i}) \text{ with } u_{n_i} \in C_{n_i} \text{ such that } u_{n_i} \to u\}.$$

We say that C_n converges in the Painlevé–Kuratowski sense to C, if

$$\limsup_n C_n = \liminf_n C_n = C.$$

In this case we write $\lim_n C_n = C$.

The topology of set convergence is metrizable, and $\mathscr{C}$ endowed with this metric is compact. A subbasis of this topology is given by the classes $\{C\colon C \cap K = \emptyset\}$ and $\{C\colon C \cap G \neq \emptyset\}$, where K runs through all compact and G runs through all open sets.

The pertaining Borel σ-algebra in $\mathscr{C}$ is called the Effros σ-algebra $\mathscr{E}_{\mathscr{C}}$.

A random closed set $A(\omega)$ is a random function defined on some probability space $(\Omega, \mathscr{A}, P)$ with values in $\mathscr{C}$, which is $\mathscr{A}$-$\mathscr{E}_{\mathscr{C}}$ measurable. The distribution of the random set $A(\omega)$ is the induced probability measure on $(\mathscr{C}, \mathscr{E}_{\mathscr{C}})$. Weak convergence of random closed sets is defined as usual for random variables with values in a metric space.

Appendix B: Stochastic inclusion

We recall first the notion of stochastic ordering for real-valued random variables. A random variable X_1 is called stochastically smaller than X_2 if, for all t,

$$G_{X_1}(t) := P\{X_1 \geqslant t\} \leqslant P\{X_2 \geqslant t\} =: G_{X_2}(t).$$

If X_1 is stochastically smaller than X_2, then we may construct versions X_1', X_2' on some new common probability space, such that X_i' coincides with X_i in distribution ($i = 1, 2$) and $X_1' \leqslant X_2'$ a.s. (Simply take $(X_1', X_2') = (F_{X_1}^{-1}(U), F_{X_2}^{-1}(U))$ for a random variable U uniformly distributed on [0,1].) Moreover, we may also define the concept of stochastic ordering in the limit: a sequence of random variables (X_n) is called stochastically smaller than X_0 in the limit if for all t

$$\limsup_n P_n\{X_n \geqslant t\} \leqslant P\{X_0 \geqslant t\}.$$

The sequence (X_n) is stochastically smaller than X_0 in the limit if and only if all weak cluster points of (X_n) are stochastically smaller than X_0.

We will now present a completely analogous set-up for random sets, where the relevant order structure is set inclusion.

Definition B.1 *(cf. Pflug 1992, Definition 1.1). Let A_1, A_2 be two random closed sets. A_1 is said to be stochastically included in A_2 if, for every collection of compact sets $K_1, \ldots, K_\ell$, ℓ arbitrary,*

$$P\{A_1 \cap K_1 \neq \emptyset, \ldots, A_1 \cap K_1 \neq \emptyset\} \leqslant P\{A_2 \cap K_\ell \neq \emptyset, \ldots, A_2 \cap K_\ell \neq \emptyset\}.$$

Remark B.1. Since all finite unions of open balls are monotone limits of compact sets, we may also equivalently define A_1 to be stochastically smaller than A_2 if, for every collection of open balls (B_{ij}),

$$P\left\{A_1 \cap \bigcup_j B_{1j} \neq \varnothing, \ldots, A_1 \cap \bigcup_j B_{\ell j} \neq \varnothing\right\} \leqslant P\left\{A_2 \cap \bigcup_j B_{1j} \neq \varnothing, \ldots, A_2 \cap \bigcup_j B_{\ell j} \neq \varnothing\right\}.$$

Remark B.2. Suppose that two random sets A_1 and A_2 are defined on the same probability space and that $A_1 \subseteq A_2$ a.s. Then trivially A_1 is stochastically included in A_2.

There is – as in the case of stochastic ordering of real variables – a construction which shows that the converse is also true:

Theorem B.1. *Let A_1 and A_2 be two random sets such that A_1 is stochastically included in A_2. Then there is a probability space $(\Omega', \mathscr{A}', P')$ and two random sets A_1' and A_2' such that A_i coincides in distribution with A_i' for $i = 1, 2$ and $A_1' \subseteq A_2'$ a.s.*

Proof. Let $\{B_i\}_{i \in \mathbf{N}}$ be the countable collection of all open balls with rational centres and rational radii in $\mathbb{R}^d$. Notice that, for all closed sets C,

$$C = \bigcap_{C \cap B_i = \varnothing} B_i^c,$$

where B^c denotes the complement of B. Let $x_C \in \{0, 1\}^{\mathbf{N}}$ be the characteristic vector of C, i.e.

$$[x_C]_i = \begin{cases} 1 & \text{if } C \cap B_i \neq \varnothing, \\ 0 & \text{if } C \cap B_i = \varnothing. \end{cases}$$

Set $x_{C_1} \leqslant x_{C_2}$ if and only if $[x_{C_1}]_i \leqslant [x_{C_2}]_i$ for all i. Obviously $C_1 \subseteq C_2$ if and only if $x_{C_1} \leqslant x_{C_2}$.

The random sets A_1 and A_2 induce probability measures P_1 and P_2 on the infinite hypercube $\{0, 1\}^{\mathbf{N}}$. We will construct a coupling P' of P_1 and P_2 on $\{0, 1\}^{\mathbf{N}} \times \{0, 1\}^{\mathbf{N}}$.

Let us first consider the case of the finite collection $B_1, \ldots, B_n$. Let μ_1 and μ_2 be the measures which are induced via the characeristic vectors on the finite hypercube $\{0, 1\}^n$. Call a subset G of the hypercube monotonic, if $x \in G$ and $x \leqslant y$ implies that $y \in G$.

We claim that the assumptions imply that $\mu_1(G) \leqslant \mu_2(G)$ for all monotonic sets G. Let $x^{(1)}, \ldots, x^{(s)}$ be the minimal elements in G. Since G is finite, the set of minimal elements is also finite. Then $G = \cup_{i=1}^s \{y \colon x \leqslant y\}$, which corresponds to the set $\cup_{i=1}^s \cap_{x_j^{(i)}=1} B_j$. By Remark B.1, μ_1 is smaller than μ_2 on exactly this class of sets.

The existence of a coupling can be seen from a graph-theoretic argument. We construct a special graph with $2 + 2^{n+1}$ nodes. Imagine two hypercubes $\{0, 1\}^n$, where node x from the first and node y from the second hypercube are connected by an oriented arc if $x \leqslant y$. Assign the capacity ∞ to these arcs. Finally, add two artificial nodes to the graph: a source which is connected to each node x of the first hypercube with capacity $\mu_1(x)$, and a sink which is reachable from each node y of the second hypercube with capacity $\mu_2(y)$. We claim that every cut in this graph has capacity at least 1. Suppose that we cut the arcs which lead from the source to the nodes $(x)_{x \in I}$ of the first hypercube. Then, in order to cut the sink from the source, we have to cut at least the arcs leading from the nodes $(y)_{y\in G}$ to the sink, where $G = \{y\colon \exists z \notin I \text{ such that } z \leqslant y\}$. (To cut arcs with infinite capacity does not work.) The capacity of this cut is

$$\sum_{x\in I}\mu_1(x) + \sum_{y\in G}\mu_2(y) = 1 - \sum_{x\notin I}\mu_1(x) + \sum_{y\in G}\mu_2(y) \geqslant 1 - \sum_{x\in G}\mu_1(x) + \sum_{y\in G}\mu_2(y) \geqslant 1,$$

since G is a monotone set.

The minimal capacity of a cut is 1. Thus by the max-flow-min-cut theorem, there is a flow of size 1 from the source to the sink. Let $\nu(x, y)$ be such a flow (it need not be unique). Notice that $\nu(x, y) \geqslant 0$, $\sum_y \nu(x, y) = \mu_1(x)$ and $\sum_x \nu(x, y) = \mu_2(y)$. We may interpret ν as a probability measure. Since a flow is only possible if $x \leqslant y$, we have that $x \leqslant y$ ν-a.s.

For a general countable class of balls, we make the above construction for each n, i.e. we construct a sequence (ν_n) of coupling measures on pairs of hypercubes $\{0, 1\}^n \times \{0, 1\}^n$. We may select a subsequence $(\nu_{n_i^{(1)}})$ such that the induced marginal distributions on the first coordinates converge, a further subsequence $(\nu_{n_i^{(2)}})$ such that the marginal distributions of the first two coordinates converge, and so on. Let $P' = \lim_k \nu_{n_k^{(k)}})$. P' is a probability measure on $\Omega' = \{0, 1\}^{\mathbf{N}} \times \{0, 1\}^{\mathbf{N}}$. It is evident that ν has marginals P_1 and P_2 and $x \leqslant y$ P'-a.s. On Ω' we construct the two new random sets by

$$A_1'(x, y) = \bigcap_{x_i=0} B_i^c, \qquad A_2'(x, y) = \bigcap_{y_i=0} B_i^c.$$

We have that $A_1' \subseteq A_2'$ a.s. and that the A_i' have the same distributions as A_i, $i = 1, 2$. $\square$

Definition B.2 *(see Definition 2.2). A sequence A_n of random sets is called stochastically included in A_0 in the limit if, for every collection of compact sets $K_1, \ldots, K_\ell$,*

$$\limsup_n P\{A_n \cap K_1 \neq \varnothing, \ldots, A_n \cap K_\ell \neq \varnothing\} \leqslant P\{A_0 \cap K_1 \neq \varnothing, \ldots, A_0 \cap K_\ell \neq \varnothing\}.$$

Remark B.3. An equivalent definition is as follows: a sequence A_n of random sets is stochastically included in A_0 in the limit if all cluster points of the sequence (A_n) are stochastically included in A_0.

Lemma B.1. *If A_n, A_0 are defined on the same probability space and* $\limsup A_n \subseteq A_0$ *a.s., then A_n is stochastically included in A_0.*

Proof. Let $K_1, \ldots, K_\ell$ be a collection of compact sets and suppose that

$$A_n \cap K_1 \neq \varnothing, \ldots, A_n \cap K_\ell \neq \varnothing$$

for infinitely many n. Then also, since $\limsup A_n \subseteq A_0$, i.e. since A_0 contains all cluster points of subsequences from A_n,

$$A_0 \cap K_1 \neq \varnothing, \ldots, A_0 \cap K_\ell \neq \varnothing.$$

Thus

$$\bigcap_N \bigcup_{n \leqslant N} \{\omega : A_n(\omega) \cap K_1 \neq \varnothing, \ldots, A_n(\omega) \cap K_\ell \neq \varnothing\} \subseteq$$
$$\{\omega : A_0(\omega) \cap K_1 \neq \varnothing, \ldots, A_0(\omega) \cap K_\ell \neq \varnothing\},$$

which implies that

$$\limsup P_n\{\omega : A_n(\omega) \cap K_1 \neq \varnothing, \ldots, A_n(\omega) \cap K_\ell \neq \varnothing\}$$
$$\leqslant P\{\omega : A_0(\omega) \cap K_1 \neq \varnothing, \ldots, A_0(\omega) \cap K_\ell \neq \varnothing\}. \qquad \square$$

Lemma B.2. *Suppose that A_0 is a.s. a singleton, i.e. $A_0 = a_0$, a random variable. If A_n is stochastically included in A_0 in the limit, then every measurable selection $\tilde{a}_n \in A_n$ converges in distribution to a_0.*

Proof. It suffices to show that, for every measurable selection,

$$\limsup_n P_n\{\tilde{a}_n \in K\} \leqslant P\{a_0 \in K\}$$

for every compact K. This is, however, clear since

$$\limsup_n P_n\{\tilde{a}_n \in K\} \leqslant \limsup_n P_n\{A_n \cap K \neq \varnothing\} \leqslant P_n\{a_0 \in K\}. \qquad \square$$

Acknowledgements

The first author was supported, in part, by the International Soros Science Education Program.

References

Anisimov, V.V. (1991) Averaging principle for switching recurrent sequences. *Theory Probab. Math. Statist.*, **45**, 3–12.

Anisimov, V.V. and Seilhamer, A.V. (1994) Asymptotic properties of extremal sets of random fields. *Theory Probab. Math. Statist.*, **51**, 1–9.

Anisimov, V.V. and Orazklychev, A. (1993) Asymptotic parameter estimation of recurrent processes of semi-Markov type. *Theory Probab. Math. Statist.*, **49**, 1–13.

Ibragimov, I.A. and Has'minskii, R.Z. (1981) *Statistical Estimation – Asymptotic Theory*. New York: Springer-Verlag.

Pflug, G. (1992) Asymptotic dominance and confidence for solutions of stochastic programs. *Czechoslovak J. Oper. Res.*, **1**(1), 21–30.

Pflug, G. (1995) Asymptotic stochastic programs. *Math. Oper. Res.*, **18**(4), 829–845.

Salinetti, G. and Wets, R.J.B. (1986) On the convergence in distribution of measurable multifunctions (random sets), normal integrands, stochastic processes and stochastic infima. *Math. Oper. Res.*, **11**(3), 385–419.

Shapiro, A. (1993) The asymptotic behavior of optimal solutions in stochastic programs. *Math. Oper. Res.*, **18**(4), 829–845.

Skorokhod, A.V. (1956) Limit theorems for random processes. *Theory Probab. Appl.*, **1**, 289–319.

van der Vaart, A.W. (1995) Efficiency of infinite dimensional M-estimators. *Statist. Neerlandica*, **49**(1), 9–30.

Received August 1998