WORKING PAPER

NORMALIZED CONVERGENCE IN STOCHASTIC OPTIMIZATION

Yuri M. Ermoliev
Vladimir I. Norkin

November 1989
WP-89-091

IIASA
International Institute
for Applied Systems Analysis

NORMALIZED CONVERGENCE IN STOCHASTIC OPTIMIZATION

Yuri M. Ermoliev
Vladimir I. Norkin

November 1989
WP-89-091

Glushkov Institute of Cybernetics Kiev, USSR

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

A new concept of (normalized) convergence of random variables is introduced. The normalized convergence is preserved under Lipschitz transformations. This convergence follows from the convergence in mean and itself implies the convergence in probability. If a sequence of random variables satisfies a limit theorem then it is a normalized convergent sequence. The introduced concept is applied to the convergence rate study of a statistical approach in stochastic optimization.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

CONTENTS

1. Introduction.

In the probability theory there are several types of the convergence of random variables such as the convergence in mean, in probability, a.s. and in distribution. They are preserved under continuous transformations (exept the convergence in mean which is preserved under Lipschitz transformations). In this paper we introduce a new concept of normalized convergence, which seems to be useful for the study of a convergence rate of random variables. The normalized convergence is preserved under Lipschitz trasformations and with some restrictions under locally Lipschitz transformations. The rate of normalized convergence doesn't change under such transformations. The normalized convergence follows from the convergence in mean and itself implies the convergence in probability. Moreover, if a sequence of random variables satisfy a limit theorem, then it is a normalized convergent sequence.

As an application of the introduced concept we study the convergence rate of so-called statistical method of stochastic optimization.

2. Definition, consequences

Definition 2.1. Suppose (Ω,Σ,P) is a probability space, $\omega \in \Omega$. A sequence of random variables $\xi_s(\omega)$, $s=1,2,\dots,$ from Ω into a metric space X is said to be normalized convergent to a random element $\xi(\omega)$ ($\xi_s \xrightarrow{N} \xi$), if there exists a sequence of positive numbers $\nu_s \longrightarrow +\infty$ and a distribution function $H(t)$: $R^1 \longrightarrow R^1$ such that

$$\liminf_{s\to+\infty} P\{\nu_s \rho(\xi_s,\xi)<t\} \geq H(t) \quad \forall\ t\in R^1 .$$

A similar type of convergence is known from the following well known Kolmogorov's limit theorem.

Theorem 2.2 [10]. If $F(x)$ is a continuous distribution function and $\{F_s(x)\}$ is a sequence of appropriate empirical distribution functions, then

$$\lim_{s\to+\infty} P\{\sqrt{s}\|F - F_s\|<t\}=H(t),$$

where

$$\|F - F_s\| = \sup_x |F(x) - F_s(x)|,$$

$$H(t) = 1 - 2\sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2t^2}.$$

The following straightforward generalization of Kolmogorov convergence

$$\lim_{s\to+\infty} P\{\nu_s \rho(\xi_s,\xi)<t \} = H(t)$$

have been used by Lehman [11]. Unfortunatly such generalization is not sufficient for the study of convergence rates in the stochastic optimization. A special case of the normalized convergence

$$P\{ \nu_s\rho(\xi_s,\xi)<t \} \ge H(t)=\begin{cases}1-c/t, & t\ge c>0,\\ 0, & t<0,\end{cases} \qquad s=1,2,\dots,$$

have been used by Polyak [12].

The following statements show that the normalized convergence occupies a place between the convergence in mean and the convergence in probability.

Theorem 2.3. If in a metric space X the sequence of random variables $\xi_s:\Omega\longrightarrow X$ convergences in mean to a random variable $\xi:\Omega\longrightarrow X$, i.e. $\lim_{s\to+\infty} E\rho(\xi_s,\xi)=0$, then

$$\liminf_{s\to+\infty} P\{\nu_s\rho(\xi_s,\xi)<t\} \geq H(t),$$

where

$$\nu_s=1/E\rho(\xi_s,\xi), \qquad H(t)=\begin{cases}1-1/t, & t\geq 1,\\ 0, & t<1.\end{cases}$$

Proof. For $\nu_s=1/E\rho(\xi_s,\xi)$ it follows that $\lim_{s\to+\infty}\nu_s=+\infty$ and $\nu_s E\rho(\xi_s,\xi)\leq 1$ (with the agreement $+\infty\cdot 0=0$). By Chebychev inequaliny for $\nu_s<+\infty$

$$P\{\ \nu_s\rho(\xi_s,\xi)\geq t\ \} \leq \nu_s E\rho(\xi_s,\xi)/t \leq 1/t.$$

If $\nu_s=+\infty$ then $E\rho(\xi_s,\xi)=0$ and

$$P\{\ \nu_s\rho(\xi_s,\xi)\geq t>0\ \} = P\{\ \rho(\xi_s,\xi)>0\ \} = 0.$$

Therefore in all cases

$$P\{\ \nu_s\rho(\xi_s,\xi)\geq t\ \} \leq 1/t,$$

thus

$$P\{\ \nu_s\rho(\xi_s,\xi)<t\ \} \geq 1-1/t,$$

and hence

$$P\{\ \nu_s\rho(\xi_s,\xi)<t\ \} \geq H(t)=\begin{cases}1-1/t, & t\geq 1,\\ 0, & t<1.\end{cases} \qquad \blacksquare$$

Theorem 2.4. If in a metric space $\xi_s\xrightarrow{N}\xi$, then $\xi_s\xrightarrow{P}\xi$ and hence $\xi_s\xrightarrow{D}\xi$.

Proof. Suppose that

$$\liminf_{s\to+\infty} P\{ \nu_s\rho(\xi_s,\xi)<t \} \geq H(t) \qquad \forall\, t\in R^1 .$$

For arbitrary numbers t, T and for large enough s it follows that $\nu_s t \geq T$ and hence

$$P\{\rho(\xi_s,\xi)\geq t\}=P\{\nu_s\rho(\xi_s,\xi)\geq\nu_s t\}\leq P\{\nu_s\rho(\xi_s,\xi)\geq T\}=$$

$$=1-P\{\nu_s\rho(\xi_s,\xi)<T\}.$$

Then

$$\limsup_{s\to+\infty} P\{\rho(\xi_s,\xi)\geq t\} \leq 1-\liminf_{s\to+\infty} P\{\nu_s\rho(\xi_s,\xi)<T\} \leq$$

$$\leq 1-H(t).$$

Since T may be an arbitrary large number and $\lim_{s\to+\infty} H(t)=1$ then $\lim_{s\to+\infty} P\{\rho(\xi_s,\xi)\geq t\}=0.$ ■

The following facts are essential for understanding the phenomenon of normalized convergence.

Theorem 2.5. Let X be a Banach space and random variables $\xi_s,\xi,h:\Omega\to X$ satisfy a limit theorem :

$$\nu_s(\xi_s - \xi)\xrightarrow{D}h$$

for some positiv numbers $\nu_s\longrightarrow +\infty$. Then

$$\liminf_{s\to+\infty} P\{\nu_s\|\xi_s-\xi\| <t\} \geq H(t)=P\{\|h\| <t\}.$$

Proof. Since the norm is a continuous function then the sequence $\nu_s\|\xi_s-\xi\|$ converges in distribution to the random variable $\|h\|$, i.e. measures μ_s induced in X by $\nu_s\|\xi_s-\xi\|$ weak*-converge to a measure μ induced in X by $\|h\|$. Then for an open set $G=\{x\in X|\ \|x\|<t\}$ it follows (see [2], Theorem 2.1) that

$$\liminf_{s\to+\infty} \mu_s(G)=\liminf_{s\to+\infty} P\{\nu_s\|\xi_s-\xi\|<t\} \ge \mu(G)=P\{\|h\| <t\},$$

where $H(t)=P\{\|h\| <t\}$ is some distribution function.■

Theorem 2.6. Let X, Y be metric spaces and $y:X\longrightarrow Y$ be a Lipschitz mapping in $G\subseteq X$ with the constant L, i.e.

$$\rho_Y(y(x_1),y(x_2)) \le L\rho_X(x_1,x_2) \quad \forall\ x_1,x_2\in G.$$

Suppose that a sequence of random variables $\xi_s:\Omega\longrightarrow X$, $s=1,2,\dots$, is normalized convergent to a random element $\xi:\Omega\longrightarrow X$, i.e.

$$\liminf_{s\to+\infty} P\{\nu_s\rho_X(\xi_s,\xi)<t\} \ge H(t) \quad \forall\ t\in R^1.$$

Then

$$\liminf_{s\to+\infty} P\{\nu_s\rho_Y(y(\xi_s),y(\xi))<t\} \ge H(t/L) \quad \forall\ t\in R^1.$$

Proof. From the inequality

$$P\{\nu_s\rho_Y(y(\xi_s),y(\xi))<t\} \ge P\{\nu_s L\rho_X(\xi_s,\xi)<t\} = P\{\nu_s\rho_X(\xi_s,\xi)<t/L\}$$

it follows that

$$\liminf_{s\to+\infty} P\{\nu_s\rho_Y(y(\xi_s),y(\xi))<t\} \ge$$

$$\ge \liminf_{s\to+\infty} P\{\nu_s\rho_X(\xi_s,\xi)<t/L\} \ge H(t/L).■$$

Theorem 2.7. Let X, Y be metric spaces and $y:X\longrightarrow Y$ be a local Lipschitz mapping. Suppose that a sequence of random variables $\xi_s:\Omega\longrightarrow X$ is normalized convergent to a deterministic element $\xi\in X$, i.e.

$$\liminf_{s\to+\infty} P\{\nu_s\rho_X(\xi_s,\xi)<t\} \ge H(t) \quad \forall\ t\in R^1.$$

Then

$$\liminf_{s\to+\infty} P\{\nu_s\rho_Y(y(\xi_s),y(\xi))<t\} \ge H(t/L(\xi)) \quad \forall\, t\in R^1,$$

where

$$L(\xi) = \limsup_{x\to\xi} \rho_Y(y(x),y(\xi))/\rho_X(x,\xi).$$

Proof. Let

$$L_\delta = \sup_{x:\rho(x,\xi)\le\delta} \rho_Y(y(x),y(\xi))/\rho_X(x,\xi).$$

It is clear that $L_\delta<+\infty$ for small δ, L_δ decreeses monotonously when $\delta\to 0$ and $\lim_{\delta\to 0} L_\delta = L(\xi)$. We have inequalities

$$P\{\nu_s\rho_Y(y(\xi_s),y(\xi))<t\} \ge P\{\nu_s\rho_Y(y(\xi_s),y(\xi))<t,\ \rho_X(\xi_s,\xi)<\delta\} \ge$$

$$\ge P\{\nu_s L_\delta\rho_X(\xi_s,\xi)<t,\ \rho_X(\xi_s,\xi)<\delta\} =$$

$$= P\{\nu_s L_\delta\rho_X(\xi_s,\xi)<t\} - P\{\nu_s L_\delta\rho_X(\xi_s,\xi)<t,\ \rho_X(\xi_s,\xi)\ge\delta\} \ge$$

$$\ge P\{\nu_s\rho_X(\xi_s,\xi)<t/L_\delta\} - P\{\rho_X(\xi_s,\xi)\ge\delta\}.$$

By Theorem 2.4 it follows that $\xi_s\xrightarrow{P}\xi$, i.e.

$$\lim_{s\to+\infty} P\{\rho_X(\xi_s,\xi)\ge\delta\} = 0.$$

Then for any δ

$$\liminf_{s\to+\infty} P\{\nu_s\rho_X(y(\xi_s),y(\xi))<t\} \ge$$

$$\ge \liminf_{s\to+\infty} P\{\nu_s\rho_X(\xi_s,\xi)<t/L_\delta\} \ge H(t/L_\delta).$$

The left side of this inequality does not depend on δ. The right side has a limit $H(t/L(\xi))$ when $\delta\to 0$ due to the continuity H from the left. Therefore

$$\liminf_{s\to+\infty} P\{\nu_s \rho_Y(y(\xi_s), y(\xi)) < t\} \geq \lim_{\delta\to 0} H(t/L_\delta) = H(t/L(\xi)). \blacksquare$$

Remark 2.8. In fact we have proven the statement of the Theorem 2.7 for a mapping $y(x)$ such that

$$\rho_Y(y(x), y(\xi)) \leq L\rho_X(x, \xi)$$

when x is near ξ with fixed ξ and constant L.

From this theorem is observed the following straregy of a convergence study in the stochastic optimization . Suppose we have a Lipschitz functioal $\Phi(v)$ in a metric space X, for instance a marginal functional of a STO problem. If we have a sequence of estimates $v_s \longrightarrow v$ satisfying a limit theorem $\nu_s(v_s - v) \xrightarrow{D} h$ then the convergence rate of $\Phi(v_s)$ to $\Phi(v)$ follows from Theorems 2.6 and 2.7.

3. A statistical approach

Let us now discuss an application of introduced consepts to the convergence study of so-called statistical approach in the stochastic optimization (STO).

Consider the STO problem without constraints in expectations: minimize

$$F(x) = Ef(x,\theta) = \int_\Theta f(x,\theta)P(d\theta), \qquad x\in X,$$

where X is a subset of a topological space, $\theta\in\Theta$, (Θ, Σ, P) is a probability space.

Suppose that $F(x)$ is a lower semicontinuous function with the minimal value F^* and the set of optimal solutions $X^*\subset X$ (if they exist). We are also interesting in a set of approximate solutions $X^*_\varepsilon = \{x\in X \mid F(x) \leq F^* + \varepsilon\}$.

Consider a sequence of approximate programms with empirical measures P_s instead of P:

$$\text{minimize} \quad F_s(x)=\frac{1}{s}\sum_{k=1}^{s} f(x,\theta_k), \qquad x\in X,$$

where $\theta_1, \theta_2, \ldots$ are iid observations. The optimal value F_s^* and the sets of optimal solutions

$$X_s^*=\{x\in X\,|\,F_s(x)=F_s^*\} \quad \text{and} \quad X_{\varepsilon s}^*=\{x\in X\,|\,F(x)\le F_s^*+\varepsilon\}$$

depend on a random element $\omega=(\theta_1,\theta_2,\ldots)$ of a new probability space $(\Omega,\Sigma_\omega,P_\omega)$ which is a countable direct product of coppies of the space (Θ,Σ,P). The consistency question of estimates $F_s^*(\omega)$, $X_s^*(\omega)$, $X_{\varepsilon s}^*(\omega)$ is the question of a convergence of optimal values $F_s^*(\omega)$ and optimal sets $X_s^*(\omega)$, $X_{\varepsilon s}^*(\omega)$ to the true values F^*, X^* and X_ε^* (see [3], [4], [9], [15]-[18] and references in [14]). The rate of convergence F_s^*, X_s^* as a rule is defined by means of assymptotic distributions of $\sqrt{s}(F_s^*-F^*)$ and $\sqrt{s}(X_s^*-X^*)$. Such results depend on differentiability properties of some marginal values and mappings which are difficult to be varified (see [5]-[8],[14]). The concept of normalized convergence allows us to introduce another type of a convergence rate without making use of directional derivatives.

Let us rewright the approximate problem in a parametric form:

$$\text{minimize} \quad \Phi(x,y_s(x,\omega))=F(x)+y_s(x,\omega), \qquad x\in X,$$

where

$$y_s(x,\omega)=\frac{1}{s}\sum_{k=1}^{s} f(x,\theta_k)-F(x),$$

and consider a parametric problem:

minimize $\Phi(x,y(x))=F(x)+y(x), \quad x\in X.$

Let us denote

$$\Phi^*(y)= \inf \{\Phi(x,y(x)) \mid x\in X\},$$

$$X^*(y)=\{x\in X \mid \Phi(x,y(x))=\Phi^*(y)\},$$

$$X^*_\varepsilon(y)=\{x\in X \mid \Phi(x,y(x))\le\Phi^*(y)+\varepsilon\}$$

The functional $\Phi^*(y)$ and the multivalued mappings $X^*(y)$, $X^*(y)$ are defined here on the Banach space $C(X,R^1)$ of continuous functions $y:X\longrightarrow R^1$ with the norm

$$\|y\|_C=\max\{|y(x)| \mid x\in X\}.$$

Theorem 3.9. If $F(x)$ is a lower semicontinuous function on a metric compact X then $\Phi^*(y)$ is a Lipschitz functional with a constant $L=1$.

Proof. Let $y_1,y_2\in C(X,R^1)$ and

$$\Phi^*(y_1)=\Phi(x_1,y_1(x_1)), \quad \Phi^*(y_2)=\Phi(x_2,y_2(x_2)).$$

We have

$$\Phi(x,y_1(x)) = \Phi(x,y_2(x))+y_1(x)-y_2(x) = \Phi(x,y_2(x))+\sigma(x)$$

where $\sigma(x)=y_1(x)-y_2(x)$. Thus

$$\Phi(x,y_1(x))-\|\sigma\|_C \le \Phi(x,y_2(x)) \le \Phi(x,y_1(x))+\|\sigma\|_C.$$

Then on the one hand

$$\Phi^*(y_2) = \Phi(x_2,y_2(x_2)) \ge \Phi(x_2,y_1(x_2))-\|\sigma\|_C \ge$$

$$\ge \Phi(x_1,y_1(x_1))-\|\sigma\|_C = \Phi^*(y_1)-\|\sigma\|_C$$

and on the other hand

$$\Phi^*(y_1)+\|\sigma\|_C = \Phi(x_1, y_1(x_1))+\|\sigma\|_C \geq \Phi(x_1, y_2(x_1)) \geq$$

$$\geq \Phi(x_2, y_2(x_2)) = \Phi^*(y_2).$$

Combining both inequalities we have

$$\Phi^*(y_1)-\|\sigma\|_C \leq \Phi^*(y_2) \leq \Phi^*(y_1)+\|\sigma\|_C$$

or

$$|\Phi^*(y_2)-\Phi^*(y_1)| \leq \|y_2-y_1\|. \blacksquare$$

For sets A and B from a metric space let us define the values

$$\Delta(A,B) = \sup_{a\in A} \inf_{b\in B} \rho(A,B), \qquad \rho_H(A,B) = \max\,(\Delta(A,B),\ \Delta(B,A)).$$

Theorem 3.10. If $F(x)$ is a continuous function on a metric compact X then the mappings $y\longrightarrow X^*(y)$, $y\longrightarrow X^*_\varepsilon(y)$ are upper semicontinuous (and closed) and the functionals

$$y\longrightarrow\delta(y)=\Delta(X^*(y),X^*), \qquad y\longrightarrow\rho(y)=\rho_H(X^*_\varepsilon(y),X^*_\varepsilon)$$

are upper semicontinuous with the continuity at $y\equiv0$ and $\delta(0)=\rho(0)=0$.

Proof. Firstly let us show that the mapping $y\longrightarrow X^*(y)$ is closed. Suppose $\{y_n(x)\}$ converges to an element $y'\in C(X,R^1)$, $x_n\in X^*(y_n)$ and $x_n\rightarrow x'$. We need to show that $x'\in X^*(y')$. The functions $F_n(x)=F(x)+y_n(x)$ uniformly converge to a continuous function $F'(x)=F(x)+y'(x)$. In other words for any sequence $x_n\in X$, $x_n\rightarrow x'$, it follows $F_n(x_n)\rightarrow F'(x')$. Since $x_n\in X^*(y_n)$ then $F_n(x_n)\leq F_n(z)$ for all $z\in X$. Coming to the limit we have $F'(x')\leq F'(z)$ for any $z\in X$. Hence indeed $x'\in X^*(y')$. From here by $X^*(y)\subset X$ it follows that $X^*(y)$ is upper semicontinuous and the functional $y \to \delta(y)$ is continuous at $y \equiv 0 \in C(X,R^1)$. Besides this functional $y \to \delta(y)$ is a superposition of two marginal functions and its upper semicontinuity follows from marginal function theorems. Actually the function $\delta'(x)=\inf\{\|x-x'\| \,|\, x' \in X^*\}$ is continuous for a compact X^*, and the functional

$$y \longrightarrow \delta(y)=\sup\{\delta'(x) \,|\, x \in X^*(y)\}$$

is upper semicontinuous for compact valued upper semicontinuous mapping $y \to X^*(y)$ (see [1], ch.3, par.1). The proof of the analogous statements for $X^*_\varepsilon(y)$ and $\rho(y)$ is carried out similarly.■

Theorem 3.11. Let $F(x)$ be a lower semicontinuous convex function defined on a compact subset X of some Banach space B. Then the multivalued mapping $y \to X^*_\varepsilon(y)$ is lipschitzian at $y \equiv 0 \in C(X,R^1)$ for $\varepsilon > 0$, i.e. for Housdorf metric ρ_H

$$\rho_H(X^*_\varepsilon(y), X^*_\varepsilon) \le \frac{2D_X}{\varepsilon}\|y\|_C \quad \forall\, y \in C(X,R^1),$$

where D_X is a diamiter of X.

Proof. For $x_\varepsilon \in X^*_\varepsilon(y)$ by the difinition

$$\Phi(x_\varepsilon, y(x_\varepsilon)) = F(x_\varepsilon)+y(x_\varepsilon) \le \Phi^*(y)+\varepsilon.$$

Since $\Phi^*(y)$ is lipschitzian

$$F(x_\varepsilon) \le \Phi^*(0)+\|y\|_C+|y(x_\varepsilon)|+\varepsilon \le F^*+2\|y\|_C+\varepsilon,$$

i.e. $x_\varepsilon \in X^*_{\varepsilon+2\|y\|}$ or $X^*_\varepsilon(y) \subset X^*_{\varepsilon+2\|y\|}$.

It is clear that

$$\rho_H(X^*_\varepsilon, X^*_\varepsilon(y)) \le \rho_H(X^*_\varepsilon, X^*_{\varepsilon+2\|y\|}).$$

Let us show that for $\varepsilon'\geq\varepsilon$

$$\rho_H(X^*_\varepsilon, X^*_{\varepsilon'}) \leq \frac{\varepsilon'-\varepsilon}{\varepsilon} D_X ,$$

where $D_X=\sup\{\|x-x'\| \mid x\in X, x'\in X\}$ is the diamiter of the compact X. From here it follows that

$$\rho_H(X^*_\varepsilon, X^*_\varepsilon(y)) \leq \frac{2D_X}{\varepsilon}\|y\|$$

and thus the theorem is proven. Actually we have

$$\rho_H(X^*_\varepsilon, X^*_{\varepsilon'}) = \Delta(X^*_{\varepsilon'}, X^*_\varepsilon) =$$

$$= \sup_{y\in X^*_{\varepsilon'}} \inf_{z\in X^*_\varepsilon} \| y - z \| = \| y^* - z^* \|,$$

where y^* provides the suppremum of the continuous function $\phi(y)=\inf\{\|y-z\| \mid z\in X^*_\varepsilon\}$ on $X^*_{\varepsilon'}$ and z^* provides the infinum of the function $\psi(z)=\|y^*-z\|$ on the compact X^*_ε. Let us choose a point $x^*\in X^*$ and construct

$$x'=\frac{\varepsilon'-\varepsilon}{\varepsilon'} x^*+\frac{\varepsilon}{\varepsilon'} y^*.$$

From the convexity of F(x) we have

$$F(x') \leq \frac{\varepsilon'-\varepsilon}{\varepsilon'} F(x^*) + \frac{\varepsilon}{\varepsilon'} F(y^*) \leq$$

$$\leq \frac{\varepsilon'-\varepsilon}{\varepsilon'} F^* + \frac{\varepsilon}{\varepsilon'} (F^* + \varepsilon') \leq F^* + \varepsilon ,$$

i.e. $x'\in X^*_\varepsilon$. Then

$$\rho_H(X^*_\varepsilon, X^*_{\varepsilon'}) = \|y^*-z^*\| \leq \|y^*-x'\| \leq$$

$$\le \frac{\varepsilon'-\varepsilon}{\varepsilon'}\,\|y^*-x^*\| \le \frac{\varepsilon'-\varepsilon}{\varepsilon'}\,D_{X^*_{\varepsilon'}} \le \frac{\varepsilon'-\varepsilon}{\varepsilon'}\,D_X\ .\blacksquare$$

4. Consistency results, the rate of convergence

Coming back to the consistency of estimates F^*_s, X^*_s, $X^*_{\varepsilon s}$ and the rate of convergence we can see that

$$F^*_s(\omega)=\Phi^*(y_s(\cdot,\omega)),\quad X^*_s(\omega)=X^*(y_s(\cdot,\omega)),\quad X^*_{\varepsilon s}(\omega)=X^*_\varepsilon(y_s(\cdot,\omega)),$$

where

$$y_s(x,\omega) = \frac{1}{s}\sum_{k=1}^{s} f(x,\theta_k) - F(x).$$

Since $\Phi^*(y)$ and $\rho_H(X^*_\varepsilon(y),X^*_\varepsilon)$ are lipschitzian functionals at $y\equiv 0$ then the convergence $F^*_s(\omega)$ to F^* and $X^*_{\varepsilon s}(\omega)$ to X^*_ε follows from the convergence of $y_s(\cdot,\omega)\in C(X,R^1)$ to $0\in C(X,R^1)$ in some probabilistic sense (a.s., in probability, in distribution, normalized).

For any fixed point $x\in X$ by the strong law of large numbers $y_s(x,\omega)\longrightarrow 0$ P_ω-a.s. We need additional assumptions to ensure a uniform (in x) convergence $y_s(x,\omega)$ to $0\in C(X,R^1)$ in some probabilistic sense. Let us mention some of them.

A. Suppose D is a relativly open convex set in R^n and $f\colon D\times\Theta\longrightarrow R^1$ is a convex function on D for all $\theta\in\Theta$ and $f(x,\cdot)$ is integratable for all $x\in D$.

Then for any compact $X\subset D$ from the convergence $y_s(x,\omega)\longrightarrow 0$ P_ω-a.s. for rational points x it follows (see [13], theorem 10.8) a uniform convergence of $y_s(\cdot,\omega)$ to $0\in C(X,R^1)$ on X, i.e.

$$\|y_s(\cdot,\omega)\|_C = \max_{x\in X}\left|\frac{1}{s}\sum_{k=1}^{s} f(x,\theta_k)-F(x)\right| \longrightarrow 0 \qquad P_\omega\text{-a.s.}\blacksquare$$

B. Suppose X is a compact in a separable Banach space and

$$|f(x,\theta)-f(y,\theta)| \le L(\theta)\|x-y\| \quad \forall\, x,y\in X,$$

where $f(x,\cdot)$ and $L(\cdot)$ are integratable for all $x\in X$.

Then $y_s(\cdot,\omega)$ is lipschitzian in X with the constant

$L_s(\omega)=\frac{1}{S}\sum_{k=1}^{s} L(\theta_k)$, $L_s(\omega)<+\infty$ P_ω-a.s. and $y_s(\cdot,\omega)\longrightarrow 0\in C(X,R^1)$ P_ω-a.s. ■

C. Suppose X is a compact in R^n and

1) $f: X\cdot\Theta \longrightarrow R^1$ is measurable in $\theta\in\Theta$ for all $x\in X$,

2) $|f(x,\theta)-f(y,\theta)| \le L(\theta)\|x-y\| \quad \forall\, x,y\in X$,

3) $\int_\Theta L^2(\theta)P(d\theta)<+\infty$ and $\int_\Theta f^2(x,\theta)P(d\theta)<+\infty$ for some x.

Then (see [7]) there exist a Gaussian random variable h taking values in $C(X,R^1)$ such that

$$\sqrt{S}\, y_s(\cdot,\omega) \xrightarrow{D} h. \ ■$$

Lemma 4.12. Under assumptions A, B or C functions $\omega\longrightarrow F_s^*(\omega)$, $\omega\longrightarrow\Delta(X_s^*(\omega),X^*)$ and $\omega\longrightarrow\rho_H(X_{\varepsilon s}^*(\omega),X_\varepsilon^*)$ are measurable.

Proof. By Theorem 3.9 $y\longrightarrow\Phi^*(y)$ is a Lipschitz functional and by Theorem 3.10 $\delta(y)=\Delta(X^*(y),X^*)$, $\rho(y)=\rho_H(X_\varepsilon^*(y),X_\varepsilon^*)$ are upper semicontinuous (and hence Borel) functionals. Functions $f(x,\theta_k(\omega))$ are continuous in x for all ω and measurable in ω for all x (as a superposition of two measurable mappings $f(x,\cdot):\Theta\longrightarrow R^1$ and $\theta_k:\Omega\longrightarrow\Theta$). The function $y_s(\cdot,\omega)$ clearly has the same properties therefore the mapping $\omega\longrightarrow y_s(\cdot,\omega)$ is measurable as a mapping from Ω into $C(X,R^1)$ (see [7]). Finally functions $F_s^*(\omega)=\Phi^*(y_s(\cdot,\omega))$, $\Delta(X_s^*(\omega),X^*)=\delta(y_s(\cdot,\omega))$ and $\rho_H(X_{\varepsilon s}^*(\omega),X_\varepsilon^*)=\rho(y_s(\cdot,\omega))$ are measurable as superpositions of a measurable mapping $y_s(\cdot,\cdot):\Omega\longrightarrow C(X,R^1)$ and Borel functionals $\Phi^*(y)$, $\delta(y)$ and $\rho(y)$ from $C(X,R^1)$ into R^1. ■

Theorem 4.13. Under assumptions A or B $F_s^*(\omega)\longrightarrow F^*$, $\Delta(X_s^*(\omega),X^*)\longrightarrow 0$ and $\rho_H(X_{\varepsilon s}^*(\omega),X_\varepsilon^*)\longrightarrow 0$ P_ω-a.s., $s\to+\infty$.

Proof. Under assumptions A or B $y_s(\cdot,\omega)\longrightarrow 0$ P_ω-a.s. Under A the function $F(x)=Ef(x,\theta)$ is convex in a neighbourhood of X and under B $F(x)$ is lipschitzian in X, hence in both cases $F(x)$ is continuous on X. By Theorem 3.9 the functional $\Phi^*(y)$ is lipschitzian and by Theorem 3.10 the functionals $\delta(y)=\Delta(X^*(y),X^*)$ and $\rho(y)=\rho_H(X_\varepsilon^*(y),X_\varepsilon^*)$ are continuous at $y\equiv 0$. Therefore $F_s^*(\omega)=\Phi^*(y_s(\cdot,\omega))\to\Phi^*(0)=F^*$, $\delta(y_s(\cdot,\omega))\to\delta(0)=0$ and $\rho(y_s(\cdot,\omega))\to\rho(0)=0$ P_ω-a.s. ■

Theorem 4.14. Under assumptions C

1) $\lim\limits_{s\to+\infty} P_\omega\{\sqrt{s}|F_s^*(\omega)-F^*|<t\} \ge P_\omega\{\|h\|_C<t\}$,

2) $\Delta(X_s^*(\omega),X^*)\xrightarrow{P\omega} 0.$

If in addition $f(\cdot,\theta)$ is convex for all $\theta\in\Theta$ then

3) $\liminf\limits_{s\to+\infty} P_\omega\{\sqrt{s}\rho_H(X_{\varepsilon s}^*(\omega),X_\varepsilon^*)<t\} \ge H(\varepsilon t/(2D_X))$,

where D_X is a diameter of X, i.e. the values $F_s^*(\omega)-F^*$ and $\rho_H(X_{\varepsilon s}^*(\omega),X_\varepsilon^*)$ are normalized convergent to 0 with the rate of $1/\sqrt{s}$.

Proof. Under assumptions C $\sqrt{s}y_s(\cdot,\omega)\xrightarrow{D}h$ therefore by Theorem 2.5

$$\liminf_{s\to+\infty} P_\omega\{\sqrt{s}\|y_s(\cdot,\omega)\|_C<t\} \ge H(t)=P_\omega\{\|h\|_C<t\}.$$

Since $F_s^*(\omega)=\Phi^*(y_s(\cdot,\omega))$, $\rho_H(X_{\varepsilon s}^*(\omega),X_\varepsilon^*)=\rho_H(X_\varepsilon^*(y_s(\cdot,\omega)),X_\varepsilon^*)$ and

$y \to \Phi^*(y)$, $y \to \rho_H(X^*_\varepsilon(y), X^*_\varepsilon)$ are Lipschitz functionals at $y \equiv 0$ (see Theorems 3.9, 3.11) then statements 1), 3) follow from Theorem 2.7 and Remark 2.8. As for statement 2) the value $\Delta(X^*_s(\omega), X^*) = \Delta(X^*(y_s(\cdot,\omega)), X^*)$ is measurable by Lemma 4.12 and a corresponding functional $\delta(y) = \Delta(X^*(y), X^*)$ is continuous at $y \equiv 0$. That is sufficient for preserving the convergence $y_s(\cdot,\omega) \xrightarrow{P\omega} 0$ under the transformation $y \to \delta(y)$ (see [2], Corollary 2 of Theorem 5.1).

The generalization of the discussed results to the STO problems with constraints in expectations can be done directly under common regular requiements.

References.

[1]. Aubin J.-P. and Ekeland I., *Applied nonlinear analysis*, Wiley, 1984.

[2]. Billingsly P., *Convergence of Probability Measures*, Wiley, 1968.

[3]. Dupacova J., "Experience in stochastic programming models", *Proceedings of the 9-th International Mathemetical Programming Symposium*, Budapest, Akademiai Kiado, 1979, 99-105.

[4]. Dupacova J. and Wets R.J.-B., Asymptotic behavior of statistical estimators and optimal solutions for stochastic optimization problems, IIASA WP 86-41, Laxenburg, Austria, 1986.

[5]. Dupacova J. and Wets R.J.-B., Asymptotic behavior of statistical estimators and of optimal solutions of

stochastic optimization problems, II, IIASA WP 87-9, Laxenburg, Austria, 1987.

[6]. Huber P.J. The behavior of maximum likelihood estimates under non-standard conditions, *Proceedings of the 5-th Berkeley Symposium on Mathematical Statistics*, 1967, 221-233.

[7]. King A.J. Asymptotic distributions for solutions in stochastic optimization and generalized M-estimation, IIASA WP 88-58, Laxenburg, Austria, 1988.

[8]. King A.J. and Rockafellar R.T. Non-normal asymptotic behavior of solution estimates in linear-quadratic stochastic optimization, *Manuscript*, University of Washington, Seattle, 1986.

[9]. King A.J. and Wets R.J.-B., Epi-consistency of convex stochastic programs, IIASA WP 88-57, Laxenburg, Austria, 1988.

[10]. Kolmogorov A., Sulla determinazione empirica di una legge di distribuzione, *Giorn. dell 'Inst. Ital. Attuari*, **4**, 1933, 83-91.

[11]. Lehman E.L., *Theory of point estimation*, Wiley,Interscience, New York, 1983.

[12]. Polyak B.T., Convergence and Rate of Convergence in Iterative Stochastic Processes. I, *Avtomatica and Telemehanica*, 1976, No. 12, 83-94 (In Russian).

[13]. Rockafellar R.T., *Convex Analysis*, Princeton U.Press,1970.

[14]. Shapiro A., Asymptotic properties of statistical estimators in stochastic programming, *The Annals of Statistics*, 1989, 17, No.2, 841-858.

[15]. Tamm E., Chebychev type inequality for E-models of nonlinear stochastic programming, *Izvestiya Acad. Nauk. Est. SSR., Phis.-Math.*, 1978, 27, No.4, (In Russian).

[16]. Wets R., A statistical approach to the stochastic programming with (convex) simple recourse, *Working paper*, Univ. of Kentuckey, Lexington, 1979.

[17]. Yubi E., A statistical research and a method for solving a stochastic programming problem, *Izvestiya Acad. Nauk Est. SSR, Phis.-Math.*, 1977, 26, No.4 (In Russian).

[18]. Yubi E., A strong consistent estimate for the solution of a general stochastic programming problem, *Proceedings of Tallin polytechnical institute*, No.464, 1979 (In Russian).