Consistency of sample estimates of risk averse stochastic programs

Alexander Shapiro*

School of Industrial & Systems Engineering,

Georgia Institute of Technology,

765 Ferst Drive, Atlanta, GA 30332.

Abstract. In this paper we study asymptotic consistency of law invariant convex risk measures and the corresponding risk averse stochastic programming problems for independent identically distributed data. Under mild regularity conditions we prove a Law of Large Numbers and epiconvergence of the corresponding statistical estimators. This can be applied in a straightforward way to establishing convergence w.p.1 of sample based estimators of risk averse stochastic programming problems.

Key Words: law invariant convex and coherent risk measures, stochastic programming, Law of Large Numbers, consistency of statistical estimators, epiconvergence, sample average approximation.

Mathematics Subject Classification: Primary: 62F12, 90C15, 60F25,
Secondary: 91B30, 62G30

*e-mail: ashapiro@isye.gatech.edu. This research was partly supported by the NSF award DMS-0914785 and ONR award N000140811104.

1 Introduction

In many practical situations one has to make decisions facing uncertainty of the future. This raises the question of how to make such decisions in some optimal way, i.e., how to formulate an appropriate optimization problem. In the classical approach of stochastic programming one specifies a probability distribution of the uncertain parameters and optimize (say minimize) a relevant objective function on *average*. Since this does not take into account deviations of possible realization of the objective from its expected value, this approach is referred to as *risk neutral*. Recently considerable attention was attracted to *risk averse* formulations of stochastic programs.

To be specific let us consider the following optimization problem:

$$\min_{x\in\mathcal{X}} G(x,\xi), \tag{1.1}$$

depending on parameter vector $\xi \in \Xi \subset \mathbb{R}^d$. Here $\mathcal{X} \subset \mathbb{R}^n$ is the feasible set of decision variables and $G : \mathcal{X} \times \mathbb{R}^d \to \mathbb{R}$ is the objective function. An optimal solution of this problem depends on a particular realization of ξ, which is not known at the time when decision x should be made. If ξ is modeled as a random vector with a specified probability distribution, then we can average the objective by taking the expectation $g(x) := \mathbb{E}[G(x,\xi)]$, and consequently to minimize $g(x)$ subject to the feasibility constraints $x \in \mathcal{X}$. If this procedure is repeated many times, under more or less similar conditions, such a decision will be optimal on average and can be justified by the Law of Large Numbers. However, for a particular realization of the random vector ξ and given x, value of random variable $G_x(\xi) = G(x,\xi)$ can be quite different from the expected value $g(x)$. This gives a motivation for considering an appropriate functional ρ other than the expectation, defined on a space of random variables, and hence to consider the corresponding risk averse problem (e.g., [8, Chapter 6]):

$$\min_{x\in\mathcal{X}} \{g(x) := \rho(G_x)\}. \tag{1.2}$$

In order to formalize the concept of risk functionals we proceed as follows. Suppose that random vector $\xi = \xi(\omega)$ is defined on a probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{Z} := L_p(\Omega, \mathcal{F}, P)$, $p \in [1,\infty)$, be the space of random variables $Z : \Omega \to \mathbb{R}$ having finite p-th order moments, and $\rho : \mathcal{Z} \to \mathbb{R}$ be a (real valued) functional referred to as a *risk measure*. Assume that the random variable $G_x(\xi(\omega))$ belongs to the space $\mathcal{Z}$ for all $x \in \mathcal{X}$, and hence $\rho(G_x)$ is well defined.

In applications the functional ρ is often constructed in such a way as to provide a compromise between optimization on average and controlling the involved risk. Therefore something like "mean-risk" could be a better term for ρ. Anyway the terminology of risk measures became somewhat standard, so we will follow it here. In the influential paper of Artzner et al [2] it was suggested that a "good" risk measure should satisfy the following conditions (axioms).

(A1) *Monotonicity:* If $Z, Z' \in \mathcal{Z}$ and $Z \succeq Z'$, then $\rho(Z) \geq \rho(Z')$.

(A2) *Convexity:*

$$\rho(tZ + (1-t)Z') \leq t\rho(Z) + (1-t)\rho(Z')$$

for all $Z, Z' \in \mathcal{Z}$ and all $t \in [0,1]$.

(A3) *Translation Equivariance:* If $a \in \mathbb{R}$ and $Z \in \mathcal{Z}$, then $\rho(Z + a) = \rho(Z) + a$.

(A4) *Positive Homogeneity:* If $t \geq 0$ and $Z \in \mathcal{Z}$, then $\rho(tZ) = t\rho(Z)$.

The notation $Z \succeq Z'$ means that $Z(\omega) \geq Z'(\omega)$ for a.e. $\omega \in \Omega$. Risk measures ρ satisfying the above axioms (A1)-(A4) were called *coherent* in [2]. If a risk measure satisfies axioms (A1)-(A3), but not necessarily (A4), it is called *convex* (cf., [5]). It is said that ρ is *law invariant* if $\rho(Z)$ depends only on the distribution of Z, i.e., if $Z, Z' \in \mathcal{Z}$ have the same cumulative distribution function then $\rho(Z) = \rho(Z')$. Throughout the paper we denote by $F(z) := P(Z \leq z)$ the cumulative distribution function (cdf) of a considered random variable $Z \in \mathcal{Z}$.

An important example of risk measures is the Average Value-at-Risk (also called Conditional Value-at-Risk) measure

$$\mathsf{AVaR}_\alpha(Z) := \inf_{t \in \mathbb{R}} \left\{ t + (1-\alpha)^{-1} \mathbb{E}[Z - t]_+ \right\}, \tag{1.3}$$

where $\alpha \in [0,1)$. Defined on $\mathcal{Z} := L_1(\Omega, \mathcal{F}, P)$, this is a (real valued) law invariant coherent risk measure. Given a sample $Z_1, ..., Z_N$ of random variable Z, the corresponding empirical cdf is $\hat{F}_N(z) = N^{-1} \sum_{i=1}^N \mathbf{1}_{\{Z_i \leq z\}}$. Consequently $\mathsf{AVaR}_\alpha(Z)$ can be estimated by replacing the cdf F of Z with its empirical estimate $\hat{F}_N$, that is by replacing the expectation $\mathbb{E}[Z - t]_+$ in (1.3) with its sample average estimate $N^{-1} \sum_{i=1}^N [Z_i - t]_+$. We assume throughout this paper that the sample $Z_1, ..., Z_N$ is iid *(independent identically distributed).* Since a law invariant risk measure ρ can be considered as a function of the cdf $F(z) = P(Z \leq z)$, we also write $\rho(F)$ to denote the corresponding value $\rho(Z)$. By replacing F with its empirical estimate $\hat{F}_N$, we obtain the estimate $\rho(\hat{F}_N)$ to which we refer as the *sample* or *empirical* estimate of $\rho(F)$.

In a similar way the "true" risk averse optimization problem (1.2) can be approximated by its empirical estimate. That is, let $\xi_i = \xi_i(\omega)$, $i = 1, ..., N$, be an iid sample of the random vector $\xi = \xi(\omega)$, defined on the same probability space. For $x \in \mathbb{R}^n$ consider the sample estimate $\hat{g}_N(x)$

given by $\rho(\hat{F}_N)$, with $\hat{F}_N$ being the empirical cdf associated with the sample $G(x,\xi_1),...,G(x,\xi_N)$. Consequently we obtain the following approximation of the problem (1.2)

$$\operatorname*{Min}_{x\in\mathcal{X}} \hat{g}_N(x). \tag{1.4}$$

Note that $\hat{g}_N(x) = \hat{g}_N(x,\omega)$ is a random function, sometimes we suppress dependence on ω in the notation.

The goal of this paper is to investigate convergence properties of the sample estimates of law invariant convex risk measures and the corresponding optimization problems of the form (1.4). A need for such statistical inference appears in two somewhat different situations. It could be that the random sample represents given (collected) data while the "true" distribution is not known. On the other hand, in some cases the random sample is generated by Monte Carlo techniques for the purpose of numerical integration. In the context of solving the optimization problem (1.4) this is the approach of the so-called Sample Average Approximation (SAA) method (see, e.g., [8, Chapter 5]). Although conceptually different these two applications involve the same statistical inference.

Let us recall the following basic duality result associated with convex risk measures. Recall that the space $\mathcal{Z} = L_p(\Omega,\mathcal{F},P)$, $p\in[1,\infty)$, equipped with the norm $\|Z\|_p = \left(\int_\Omega |Z(\omega)|^p dP(\omega)\right)^{1/p}$, becomes a Banach space, and its dual space is $\mathcal{Z}^* = L_q(\Omega,\mathcal{F},P)$, where $q\in(1,\infty]$ is such that $1/p+1/q=1$. A (real valued) convex risk measure $\rho:\mathcal{Z}\to\mathbb{R}$ is continuous (in the norm topology of $\mathcal{Z}$) and there exists a convex set $\mathfrak{A}\subset\mathcal{Z}^*$ of probability density functions such that

$$\rho(Z) = \sup_{\zeta\in\mathfrak{A}} \int_\Omega Z(\omega)\zeta(\omega)dP(\omega) - \rho^*(\zeta), \;\; \forall Z\in\mathcal{Z}, \tag{1.5}$$

where $\rho^*:\mathcal{Z}^*\to\overline{\mathbb{R}}$ is the conjugate of the functional ρ and $\mathfrak{A} = \operatorname{dom}(\rho^*)$. The dual representation (1.5) follows from the classical Fenchel-Moreau theorem. Originally it was derived in [2], and the following up literature (cf., [5]), for space $\mathcal{Z} = L_\infty(\Omega,\mathcal{F},P)$. For spaces $\mathcal{Z} = L_p(\Omega,\mathcal{F},P)$, $p\in[1,\infty)$, this representation was derived in [7] and it was shown there that monotonicity (axiom (A1)) and convexity (axiom (A2)) imply continuity of the (real valued) risk measure ρ in the norm topology of the space $L_p(\Omega,\mathcal{F},P)$.

We assume throughout this paper that the probability space $(\Omega,\mathcal{F},P)$ is *atomless* and *complete*. Then we can set $\Omega := [0,1]$ equipped with its Lebesgue sigma algebra $\mathcal{F}$ (cf., [3, p. 25]), and uniform probability distribution (measure) P. For a cdf F we denote by $F^{-1}(\omega) = \inf\{z : F(z)\ge\omega\}$ its left-side quantile. Note that $Z(\cdot) := F^{-1}(\cdot)$ is a real valued monotonically nondecreasing left-side continuous function on the interval (0,1), and (in case of $\Omega = [0,1]$) can be considered as an element of the space $L_p(\Omega, \mathcal{F}, P)$, provided $\int_0^1 |Z(\omega)|^p d\omega$ is finite.

2 Law of Large Numbers for sample estimates of convex risk measures

In this section we investigate convergence with probability one (w.p.1) of sample estimates of convex risk measures.

Theorem 2.1 *Let $\mathcal{Z} := L_p(\Omega, \mathcal{F}, P)$, $p \in [1, \infty)$, and $\rho : \mathcal{Z} \to \mathbb{R}$ be a (real valued) law invariant convex risk measure. Then $\rho(\hat{F}_N)$ converges to $\rho(F)$ w.p.1 as $N \to \infty$.*

Proof. Recall that since the space $(\Omega, \mathcal{F}, P)$ is assumed to be atomless and complete, we can work with $\Omega = [0,1]$ equipped with its Lebesgue sigma algebra $\mathcal{F}$ and uniform probability distribution P. Consider $Z(\cdot) := F^{-1}(\cdot)$. We can view Z as an element of the space $\mathcal{Z} = L_p(\Omega, \mathcal{F}, P)$. Consider $\hat{Z}_N := \hat{F}_N^{-1}$ the left-side quantile of the empirical cdf $\hat{F}_N$ (based on iid sample $Z_1, ..., Z_N$). Recall that $\rho(Z) = \rho(F)$ and $\rho(\hat{Z}_N) = \rho(\hat{F}_N)$. Note that the function $\hat{Z}_N(\cdot)$ is piecewise constant and hence $\hat{Z}_N \in \mathcal{Z}$ as well. Consider the set $\mathfrak{C} \subset [0,1]$ of points where function $Z(\cdot)$ is discontinuous. Since $Z(\cdot)$ is monotonically nondecreasing, the set $\mathfrak{C}$ is countable and hence has Lebesgue measure zero. By Glivenko-Cantelli theorem we have that w.p.1 $\hat{F}_N$ converges to F uniformly on $\mathbb{R}$. It follows that w.p.1 $\hat{Z}_N$ converges pointwise to Z on the set $[0,1] \setminus \mathfrak{C}$, and hence w.p.1

$$\lim_{N\to\infty} \int_0^1 |Z(\omega) - \hat{Z}_N(\omega)|^p d\omega = \int_0^1 \lim_{N\to\infty} |Z(\omega) - \hat{Z}_N(\omega)|^p d\omega = 0, \tag{2.1}$$

where the interchangeability of the limit and integral operators is justified provided that w.p.1 the sequence $|Z(\cdot) - \hat{Z}_N(\cdot)|^p$ is uniformly integrable (e.g., [3, p. 217]).

Let us show that the uniform integrability indeed holds. We have (triangle inequality) that $\|Z - \hat{Z}_N\|_p \le \|Z\|_p + \|\hat{Z}_N\|_p$, i.e.,

$$\left(\int_0^1 |Z(\omega) - \hat{Z}_N(\omega)|^p d\omega\right)^{1/p} \le \left(\int_0^1 |Z(\omega)|^p d\omega\right)^{1/p} + \left(\int_0^1 |\hat{Z}_N(\omega)|^p d\omega\right)^{1/p}. \tag{2.2}$$

The first term in the right hand side of (2.2) is constant, therefore it is sufficient to verify uniform integrability of $|\hat{Z}_N(\cdot)|^p$. We have that

$$\int_0^1 |\hat{Z}_N(\omega)|^p d\omega = \int_0^\infty |z|^p d\hat{F}_N(z) = \frac{1}{N}\sum_{i=1}^N |Z_i|^p,$$

and by the Law of Large Numbers $N^{-1}\sum_{i=1}^N |Z_i|^p$ converges w.p.1 to $\mathbb{E}|Z|^p$. Since $Z \in L_p(\Omega, \mathcal{F}, P)$ we have that $\mathbb{E}|Z|^p$ is finite. It follows that $\int_0^1 |\hat{Z}_N(\omega)|^p d\omega$ converges w.p.1 to a finite limit, which implies that w.p.1 $|\hat{Z}_N(\cdot)|^p$ is uniformly integrable.

By (2.1) this shows that $\hat{Z}_N$ converges to Z w.p.1 in the norm topology of the space $L_p(\Omega, \mathcal{F}, P)$. It remains to recall that the axioms (A1) and (A2) imply that the risk measure $\rho : \mathcal{Z} \to \mathbb{R}$ is continuous in the norm topology of $\mathcal{Z} = L_p(\Omega, \mathcal{F}, P)$ (cf., [7]), and hence $\rho(\hat{Z}_N)$ converges to $\rho(Z)$ w.p.1. ∎

Convergence w.p.1 of empirical estimates of law invariant coherent risk measures was studied in Wozabal and Wozabal [9] by employing techniques based on Kusuoka representation of law invariant coherent risk measures. The above proof is more direct and does not involve growth conditions used in [9, Theorem 3.4].

3 Convergence of statistical estimates of risk averse stochastic programs

We proceed now to investigation of uniform type convergence of empirical estimates of risk measures. Recall that a sequence of functions $\phi_k : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to epiconverge to a function ϕ, denoted $\phi_k \xrightarrow{e} \phi$, if for any point $\bar{x} \in \mathbb{R}^n$ the following two conditions are satisfied (e.g., [6, p.241]):

(i) for any sequence x_k converging to $\bar{x}$ it holds that

$$\liminf_{k\to\infty} \phi_k(x_k) \geq \phi(\bar{x}), \tag{3.1}$$

(ii) there exists a sequence x_k converging to $\bar{x}$ such that

$$\limsup_{k\to\infty} \phi_k(x_k) \leq \phi(\bar{x}). \tag{3.2}$$

Consider now the objective function $g(x) = \rho(G_x)$ of the "true" problem (1.2) and its sample estimate $\hat{g}_N(x) = \hat{g}_N(x, \omega)$ based on iid sample $\xi_i = \xi_i(\omega)$, $i = 1, ..., N$. With some abuse of the notation we write $G(x, \omega)$ for the function $G(x, \xi(\omega))$. We investigate now epiconvergence w.p.1 of $\hat{g}_N$ to g. Consider the following conditions.

(C1) For every $x \in \mathbb{R}^n$, random variable $G_x(\omega) = G(x, \omega)$ belongs to the space $\mathcal{Z} = L_p(\Omega, \mathcal{F}, P)$, $p \in [1, \infty)$.

(C2) The function $G(x, \omega)$ is random lower semicontinuous, i.e., the epigraphical multifunction $\omega \mapsto \operatorname{epi} G(\cdot, \omega)$ is closed valued and measurable.

(C3) For every $\bar{x} \in \mathbb{R}^n$ there is a neighborhood $\mathcal{V}_{\bar{x}}$ of $\bar{x}$ and a function $h \in \mathcal{Z}$ such that $G(x, \cdot) \geq h(\cdot)$ for all $x \in \mathcal{V}_{\bar{x}}$.

Some remarks about the above regularity conditions are now in order. Condition (C1) means that for every $x \in \mathbb{R}^n$ the corresponding random variable G_x has finite p-th order moment. For a thorough discussion of random lower semicontinuous functions and related measurability questions we can refer to [6, Chapter 14]. (Random lower semicontinuous functions are called *normal integrands* in [6].) In particular, a set-valued mapping $A : \Omega \rightrightarrows \mathbb{R}^m$ is measurable if for every open set $O \subset \mathbb{R}^m$ the set $A^{-1}(O)$ is $\mathcal{F}$-measurable, [6, Definition 14.1]. It could be noted that closeness of the epigraph $\operatorname{epi} G(\cdot, \omega)$ means that the function $G(\cdot, \omega)$ is lower semicontinuous. A sufficient condition for $G(x, \omega)$ to be random lower semicontinuous is that $G(x, \omega)$ is measurable in ω for every x and continuous in x for a.e. ω. Such functions are called *Carathéodory integrands* in [6, Example 14.29].

Theorem 3.1 *Let $\rho : \mathcal{Z} \to \mathbb{R}$ be a law invariant convex risk measure and $g(x) := \rho(G_x)$. Suppose that conditions* (C1)-(C3) *hold. Then the function $g(\cdot)$ is lower semicontinuous and $\hat{g}_N \xrightarrow{e} g$ w.p.1.*

Proof. Note that since the risk measure $\rho(\cdot)$ is real valued, it follows from (C1) that the function $g(\cdot)$ is real valued.

Since the risk measure ρ is convex, it has the dual representation (1.5). Consider a point $\bar{x} \in \mathbb{R}^n$. By condition (C3) we have for any $\zeta \in \mathfrak{A}$ and $x \in \mathcal{V}_{\bar{x}}$ that $\zeta(\cdot)G(x, \cdot) \geq \zeta(\cdot)h(\cdot)$. Moreover, since $h \in \mathcal{Z}$ and $\zeta \in \mathcal{Z}^*$ we have that $\int_\Omega \zeta(w)h(\omega)dP(\omega)$ is finite. Hence by Fatou's lemma it follows that for a sequence $x_k \to \bar{x}$,

$$\liminf_{k\to\infty} \int_\Omega \zeta(\omega)G(x_k, \omega)dP(\omega) \geq \int_\Omega \liminf_{k\to\infty} \zeta(\omega)G(x_k, \omega)dP(\omega) \geq \int_\Omega \zeta(\omega)G(\bar{x}, \omega)dP(\omega),$$

where the last inequality follows by lower semicontinuity of $G(\cdot, \omega)$ (condition (C2)). That is, function $x \mapsto \int_\Omega \zeta(\omega)G(x, \omega)dP(\omega) - \rho^*(\zeta)$ is lower semicontinuous. Since by (1.5), $g(x) = \rho(G_x)$ is given by a maximum of such functions, it follows that $g(x)$ it is also lower semicontinuous.

In order to show that $\hat{g}_N \xrightarrow{e} g$ w.p.1 we need to verify the conditions (i) and (ii). That is, to verify the condition (i) we have to show that there exists a set $\Delta \subset \Omega$ of measure zero such that for any point $\bar{x} \in \mathbb{R}^n$ and any sequence x_N converging to $\bar{x}$ it holds that

$$\liminf_{N\to\infty} \hat{g}_N(x_N, \omega) \geq g(\bar{x}), \quad \forall \omega \in \Omega \setminus \Delta. \tag{3.3}$$

Note that the set Δ should not depend on $\bar{x}$. Similarly for the condition (ii).

To verify condition (i) we proceed as follows. For some sequence $\gamma_k \downarrow 0$ of positive numbers consider $V_k := \{x \in \mathbb{R}^n : \|x - \bar{x}\| < \gamma_k\}$, and let

$$\mathcal{G}_k(\omega) := \inf_{x \in V_k} G(x,\omega), \;\; k \in \mathbb{N}.$$

Since $G(x,\omega)$ is random lower semicontinuous (condition (C2)), we have that $\mathcal{G}_k(\omega)$ is measurable (cf., [6]). By (C3) we have $\mathcal{G}_k(\cdot) \geq h(\cdot)$ for all k large enough (such that $V_k \subset \mathcal{V}_{\bar{x}}$). Of course, we also have that $G(x,\cdot) \geq \mathcal{G}_k(\cdot)$ for any $x \in V_k$. It follows that $\mathcal{G}_k \in \mathcal{Z}$. Consider the dual representation (1.5) of ρ and let

$$\bar{\zeta} \in \arg\max_{\zeta \in \mathfrak{A}} \{\mathbb{E}[\zeta(\omega)G(\bar{x},\omega)] - \rho^*(\zeta)\}$$

(note that such maximizer exists, [7]). That is, $\bar{\zeta} \in \mathfrak{A}$ and

$$g(\bar{x}) = \mathbb{E}[\bar{\zeta}(\omega)G(\bar{x},\omega)] - \rho^*(\bar{\zeta}).$$

Since $\mathcal{G}_k \in \mathcal{Z}$ we have by (1.5) that

$$\rho(\mathcal{G}_k) \geq \mathbb{E}[\bar{\zeta}(\omega)\mathcal{G}_k(\omega)] - \rho^*(\bar{\zeta}).$$

By condition (C3), $\bar{\zeta}(\cdot)\mathcal{G}_k(\cdot)$ is bounded from below, on a neighborhood of $\bar{x}$, by integrable function $\bar{\zeta}(\cdot)h(\cdot)$, and hence applying Fatou's lemma we have

$$\liminf_{k\to\infty} \mathbb{E}\left[\bar{\zeta}(\omega)\mathcal{G}_k(\omega)\right] \geq \mathbb{E}\big[\liminf_{k\to\infty} \bar{\zeta}(\omega)\mathcal{G}_k(\omega)\big], \tag{3.4}$$

and by lower semicontinuity of $G(\cdot,\omega)$

$$\mathbb{E}\big[\liminf_{k\to\infty} \bar{\zeta}(\omega)\mathcal{G}_k(\omega)\big] - \rho^*(\bar{\zeta}) \geq \mathbb{E}[\bar{\zeta}(\omega)G(\bar{x},\omega)] - \rho^*(\bar{\zeta}) = g(\bar{x}). \tag{3.5}$$

We obtain that

$$\liminf_{k\to\infty} \rho(\mathcal{G}_k) \geq g(\bar{x}). \tag{3.6}$$

Now let us choose $\varepsilon > 0$. By (3.6) there exists $\bar{k} = \bar{k}(\varepsilon)$ such that

$$\rho(\mathcal{G}_{\bar{k}}) \geq g(\bar{x}) - \varepsilon. \tag{3.7}$$

Let $\hat{\rho}_N(\mathcal{G}_{\bar{k}})$ be the empirical estimate of $\rho(\mathcal{G}_{\bar{k}})$ based on the same sample as the sample used for the estimate $\hat{g}_N(\cdot)$, i.e., $\hat{\rho}_N(\mathcal{G}_{\bar{k}}) = \rho(\hat{H}_N)$, where $\hat{H}_N$ is the empirical cdf of the sample $Y_i = \inf_{x\in V_{\bar{k}}} \bar{G}(x,\xi_i)$, $i = 1,...,N$. By Theorem 2.1 we have that $\hat{\rho}_N(\mathcal{G}_{\bar{k}}) \to \rho(\mathcal{G}_{\bar{k}})$ w.p.1 as $N \to \infty$. Hence for a.e. $\omega \in \Omega$ there is $\bar{N}_{\bar{x}}(\omega)$ such that $\hat{\rho}_N(\mathcal{G}_{\bar{k}}) \geq \rho(\mathcal{G}_{\bar{k}}) - \varepsilon$, for all $N \geq \bar{N}_{\bar{x}}(\omega)$. Together with (3.7) this implies that

$$\hat{\rho}_N(\mathcal{G}_{\bar{k}}) \geq g(\bar{x}) - 2\varepsilon \tag{3.8}$$

for a.e. $\omega \in \Omega$ and $N \geq \bar{N}_{\bar{x}}(\omega)$.

For any $x \in V_{\bar{k}}$ we have (for the same random sample) that the empirical cdf of $G(x, \cdot)$ dominates the empirical cdf of $\mathcal{G}_{\bar{k}}(\cdot)$, and hence (cf., [8, Theorem 6.28])

$$\hat{g}_N(x) = \hat{\rho}_N(G_x) \geq \hat{\rho}_N(\mathcal{G}_{\bar{k}}). \tag{3.9}$$

By (3.8) it follows that

$$\inf_{x \in V_{\bar{k}}} \hat{g}_N(x) \geq g(\bar{x}) - 2\varepsilon \tag{3.10}$$

for a.e. $\omega \in \Omega$ and $N \geq \bar{N}_{\bar{x}}(\omega)$. That is, there exist $\bar{N}(\omega)$, a set $\Upsilon \subset \Omega$ of measure zero and a neighborhood V (both depending on $\bar{x}$ and ε) such that

$$\hat{g}_N(x) \geq g(\bar{x}) - 2\varepsilon, \tag{3.11}$$

for all $N \geq \bar{N}(\omega)$, $x \in V$ and $\omega \in \Omega \setminus \Upsilon$. It follows that there exists a countable number of points $x_1, ...,$ in $\mathbb{R}^n$, with the corresponding neighborhoods $V_1, ...,$ covering $\mathbb{R}^n$. Let $\Upsilon_1, ...,$ be the corresponding sets of measure zero and $\bar{\Upsilon} := \cup_{i=1}^{\infty} \Upsilon_i$. Note that the set $\bar{\Upsilon}$ has measure zero, and that $\bar{\Upsilon}$ depends on ε, but not on a particular point of $\mathbb{R}^n$. It follows that for any $\bar{x} \in \mathbb{R}^n$ there is a neighborhood W and $N^*(\omega)$ such that (3.11) holds for all $x \in W$, $N \geq N^*(\omega)$ and $\omega \in \Omega \setminus \bar{\Upsilon}$.

Consequently for any point $\bar{x}$ and a sequence x_N converging to $\bar{x}$ we have that

$$\liminf_{N \to \infty} \hat{g}_N(x_N, \omega) \geq g(\bar{x}) - 2\varepsilon \tag{3.12}$$

for all $\omega \in \Omega \setminus \bar{\Upsilon}$. Choose now a sequence of positive numbers $\varepsilon_i \downarrow 0$ and let $\bar{\Upsilon}_i$ be the corresponding sets of measure zero. Set $\Delta := \cup_{i=1}^{\infty} \bar{\Upsilon}_i$. Then (3.12) implies that (3.3) holds for any $\bar{x} \in \mathbb{R}^n$ and any sequence x_N converging to $\bar{x}$. This proves that condition (i) holds for a.e. $\omega \in \Omega$.

In order to verify condition (ii) we need the following result.

Lemma 3.1 *There exists a countable set $\mathcal{D} \subset \mathbb{R}^n$ such that for any point $\bar{x} \in \mathbb{R}^n$ there exists a sequence $x_k \in \mathcal{D}$ converging to $\bar{x}$ and*

$$\limsup_{k \to \infty} g(x_k) \leq g(\bar{x}). \tag{3.13}$$

Proof. Such set $\mathcal{D}$ can be constructed as follows. Let $A \subset \mathbb{R}$ be a countable and dense subset of $\mathbb{R}$. Consider level sets $\mathcal{L}_a := \{x \in \mathbb{R}^n : g(x) \leq a\}$, and let D_a, $a \in A$, be a countable and dense subset of $\mathcal{L}_a$. (Of course, some of the sets $\mathcal{L}_a$ and D_a can be empty.) Define $\mathcal{D} := \cup_{a \in A} D_a$. Clearly the set $\mathcal{D}$ is countable. The condition (3.13) also holds. Indeed, let $a_k \in A$ be a monotonically decreasing sequence converging to $g(\bar{x})$. Note that $\bar{x} \in \mathcal{L}_{a_k}$ for all k. Therefore there exists a point

$x_k \in D_{a_k}$ such that $\|x_k - \bar{x}\| \le 1/k$. We have then that $x_k \to \bar{x}$ and $g(x_k) \le a_k$, and hence (3.13) follows. ∎

Proof of Theorem 3.1 continued. Let $\mathcal{D}$ be a set specified in Lemma 3.1. Consider a point $\bar{x} \in \mathbb{R}^n$ and let $x_k \in \mathcal{D}$ be a sequence of points converging to $\bar{x}$ such that (3.13) holds. For a given $x \in \mathbb{R}^n$ we have by Theorem 2.1 that $\hat{g}_N(x)$ converges to $g(x)$ w.p.1. That is, there exists a set Υ_x of measure zero such that $\hat{g}_N(x, \omega)$ converges to $g(x)$ for every $\omega \in \Omega \setminus \Upsilon_x$. Consider the set $\tilde{\Upsilon} := \cup_{x \in \mathcal{D}} \Upsilon_x$. Since the set $\mathcal{D}$ is countable, the set $\tilde{\Upsilon}$ has measure zero. We have that $\hat{g}_N(x, \omega)$ converges to $g(x)$ for every $x \in \mathcal{D}$ and $\omega \in \Omega \setminus \tilde{\Upsilon}$. Hence there is a sequence $N_k = N_k(\omega)$ of positive integers such that for all k,

$$|\hat{g}_{N_k}(x_k, \omega) - g(x_k)| < 1/k, \quad \forall \omega \in \Omega \setminus \tilde{\Upsilon}. \tag{3.14}$$

Now let x'_N be a sequence of points such that $x'_{N_k} = x_k$ for all k. We have then that $x'_{N_k} \to \bar{x}$ and by (3.13) and (3.14) that

$$\limsup_{k \to \infty} \hat{g}_{N_k}(x'_{N_k}, \omega) \le g(\bar{x}), \quad \forall \omega \in \Omega \setminus \tilde{\Upsilon}. \tag{3.15}$$

This shows that condition (ii) holds w.p.1. ∎

For the expectation operator, i.e., when $\rho(\cdot) = \mathbb{E}(\cdot)$, the epiconvergence result of Theorem 3.1 was proved in Artstein and Wets [1].

Suppose now that for a.e. ω the function $G(\cdot, \omega)$ is convex and let $\rho : \mathcal{Z} \to \mathbb{R}$ be a law invariant convex risk measure. Then the functions $g(x) = \rho(G_x)$ and $\hat{g}_N(x)$ are also convex. It is known that if $\phi_k : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a sequence of convex functions and $\phi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a convex lower semicontinuous function such that its domain has a nonempty interior, then the following conditions are equivalent: (a) $\phi_k \xrightarrow{\text{e}} \phi$, (b) there exists a dense subset $\mathcal{D}$ of $\mathbb{R}^n$ such that $\phi_k(x) \to \phi(x)$ for all $x \in \mathcal{D}$, (c) $\phi_k(\cdot)$ converges uniformly to $\phi(\cdot)$ on every compact set $C \subset \mathbb{R}^n$ that does not contain boundary points of $\text{dom}(\phi)$ (cf., [6, Theorem 7.17]). This together with Theorem 2.1 imply the following result in a straightforward way. Recall that by condition (C1) the function $g(x)$ is real valued and hence $\text{dom}(g) = \mathbb{R}^n$.

Theorem 3.2 *Let $\rho : \mathcal{Z} \to \mathbb{R}$ be a law invariant convex risk measure and $g(x) := \rho(G_x)$. Suppose that condition* (C1) *holds, and for a.e. ω the function $G(\cdot, \omega)$ is convex. Then the functions $g(\cdot)$ and $\hat{g}_N(\cdot)$ are convex and $\hat{g}_N(\cdot)$ converges w.p.1 uniformly to $g(\cdot)$ on every compact set.*

The results of Theorem 3.1 can be applied in a straightforward way to investigation of consistency of the sample approximation (1.4) of the "true" problem (1.2).

Theorem 3.3 *Let $\rho : \mathcal{Z} \to \mathbb{R}$ be a law invariant convex risk measure. Suppose that conditions* (C1)-(C3) *hold and the set $\mathcal{X}$ is nonempty and compact. Then the optimal value of the SAA problem* (1.4) *converges w.p.1 to the optimal value of the "true" problem* (1.2)*, and the distance from an optimal solution of* (1.4) *to the set of optimal solutions of* (1.2) *converges w.p.1 to zero as* $N \to \infty$

Proof. Since the function $g(x)$ is lower semicontinuous and the set $\mathcal{X}$ is compact, problem (1.2) has a nonempty set $\mathfrak{S}$ of optimal solutions and its optimal value ϑ^* is finite. Since $\hat{g}_N \xrightarrow{\text{e}} g$ w.p.1, it follows that the optimal value $\hat{\vartheta}_N$ of problem (1.4) converges w.p.1 to ϑ^*, and the distance from an optimal solution of (1.4) to the set $\mathfrak{S}$ converges w.p.1 to zero as $N \to \infty$ (cf., [6, Chapter 7, E]). ∎

The assumption of compactness of the set $\mathcal{X}$ in Theorem 3.3 is needed in order to ensure existence of optimal solutions of the true and SAA optimization problems, and that optimal solutions of the SAA problems stay w.p.1 in a bounded set. In some problems this can be verified by ad hoc methods.

For risk neutral stochastic programming problems the epiconvergence approach was used in Dupačová and Wets [4] to study asymptotic consistency of the corresponding statistical estimators.

References

[1] Z. Artstein and R.J.B.Wets. Consistency of minimizers and the SLLN for stochastic programs, *Journal of Convex Analysis*, 2 (1996), 1–17.

[2] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, Coherent measures of risk, *Mathematical Finance*, 9 (1999), 203–228.

[3] P. Billingsley, *Probability and Measure*, Wiley, Third Edition, New York, 1995.

[4] J. Dupačová and R.J.B. Wets, Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems, *The Annals of Statistics*, 16 (1988), 1517–1549.

[5] H. Föllmer and A. Schied, *Stochastic Finance: An Introduction in Discrete Time*, Walter de Gruyter, Berlin, 2002 (Second Edition 2004).

[6] R.T. Rockafellar and R.J.-B. Wets, *Variational Analysis*, Springer, New York, 1998.

[7] A. Ruszczyński and A. Shapiro, Optimization of convex risk functions, *Mathematics of Operations Research*, 31 (2006), 433–452.

[8] A. Shapiro, D. Dentcheva and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, SIAM, Philadelphia, 2009.

[9] D. Wozabal and N. Wozabal, Asymptotic consistency of risk functionals, *Journal of Nonparametric Statistics*, 21 (2009), 977–990.