Notes on Asymptotic Properties of Approximated Stochastic Programs *

JITKA DUPAČOVÁ
Dept. of Probability and Mathematical Statistics, Charles University, Prague, Czech Republic
e-mail: dupacova@karlin.mff.cuni.cz

Abstract. For various reasons, the underlying probability measure in stochastic programming models must be frequently substituted by a suitable approximation. This in turn requires to investigate stability of solutions of these models with respect to the probability measure. This paper is devoted to a discussion about asymptotic properties of empirical stochastic programs where the true probability measure is replaced by its empirical counterpart.

Keywords: Stochastic programs, approximated probability distributions, asymptotic analysis, Sample Average Approximation

JEL classification: C44
AMS classification: 90C15

1. Introduction

The basic model of stochastic programming

$$\min_{x \in \mathcal{X}(P)} E_P f(x, \omega) \tag{1}$$

is identified by

- a known probability distribution P of random parameter ω whose support Ω is a closed subset of $\mathbb{R}^s$;
- a given, nonempty, closed set $\mathcal{X}(P) \subset \mathbb{R}^n$ of decisions x;
  In this paper we shall assume that $\mathcal{X}$ does not depend on P and at the same time, P does not depend on x; hence, we can replace $\mathcal{X}(P)$ by $\mathcal{X}$.
- a preselected random objective $f : \mathcal{X}(P) \times \Omega \to \mathbb{R}$ (or $\bar{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$) interpreted as a loss caused by decision x when scenario ω occurs. *We assume that as a function of ω, f is measurable and its expectation $F(x, P) = E_P f(x, \omega)$ exists $\forall x \in \mathcal{X}$.* The form of f may be quite complicated (e.g. for multistage problems). For convex $\mathcal{X}$, a frequent assumption is that f is lower semicontinuous and convex with respect to x, i.e., f is a *convex normal integrand.*

*This research was supported by the Grant Agency of the Czech Republic under Grants P201/12/0083 and P402/11/0150.

We denote $\mathcal{X}^*(P) \subset \mathcal{X}$, with $x^*(P)$ a generic element, the set of optimal solutions and $\varphi(P)$ the optimal value of the objective function in (1).

To apply a stochastic programming model it is necessary first to to get $f, P, \mathcal{X}(P)$, etc., having in mind that the goal is to support decisions, and to analyze the model properties (e.g. existence of expectation, convexity). Contrary to statistical problems where f is a tool for estimating the true parameter value, in stochastic programming the random objective function $f(x, \omega)$ reflects goals of the decision problem and all minimizers of (1) are equally acceptable for decision making. Their uniqueness is not required.

For the sake of numerical tractability, various approximation schemes are used to solve the stochastic programs. For instance, one may approximate P by a discrete probability distribution, say P^ν, based on a sample from P or on historical data, by a probability distribution belonging to a given parametric family with an estimated parameter, approximation may reflect an additional information, etc. The choice depends on the type of the stochastic program to be solved, on the available information and data. Without additional analysis, however, it will be dangerous to use the obtained output (the optimal value and the optimal solutions of the approximate stochastic program) to replace the sought solution of the "true" problem. We refer to [8, 11, 25, 28, 33] for an overview of suitable output analysis methods that treat partly also robustness of the optimal value and optimal solutions in case that the true probability distribution P is not fully known. In such situations, it is not enough to rely on asymptotic results.

In this paper we shall mostly focus on properties of optimal solutions of expectation type stochastic programs when an empirical probability distribution, say P^ν, based on a sample $\omega^1, \dots, \omega^\nu$ is used to approximate the true probability distribution P. Then the expectation of the objective function $f(x, \omega)$ in (1) is replaced by the sample average, the Sample Average Approximation (SAA) problem

$$\min_{x \in \mathcal{X}} F(x, P^\nu) = \min_{x \in \mathcal{X}} \frac{1}{\nu} \sum_{j=1}^{\nu} f(x, \omega^j) \tag{2}$$

is solved, and the corresponding optimal solutions $x^*(P^\nu)$ are used at the place of the true optimal solution $x^*(P)$.

The first task is to study the behavior of the optimal values $\varphi(P^\nu)$ and optimal solutions $x^*(P^\nu)$ of (2) when $\{P^\nu, \nu = 1, \dots\}$ is a sequence of probability distributions converging to P. There is a vast statistical literature dealing with similar problems which arise in the context of maximum likelihood estimation, cf. [38], and were extended to M-estimation, e.g. [14]. Consistency results were obtained for unconstrained problems with some generalizations to problems with smooth equality constraints, cf. [1]. In stochastic programming, however, the set $\mathcal{X}$ is defined by inequalities. The first step to consistency results for stochastic programming can be found e.g. in [16] and an exponential rate of convergence was proved in [17], see section 2.1. In section 2.2 we shall present consistency results based on epi-convergence as developed in [12] in their relation to other consistency results.

The next step, asymptotic analysis of the empirical optimal value and optimal solutions $x^*(P^\nu)$, will be discussed in section 3. As we shall see, asymptotic normality of empirical optimal solutions cannot be expected in general; see e.g. [10, 12, 32, 34].

In section 4 we shall briefly indicate parallel results for the parametric case, i.e. when the approximate solutions are obtained by solving the stochastic program for probability distribution with estimated parameters; cf. [7, 36]. We conclude by brief remarks on non-asymptotic results for empirical stochastic programs.

2. Consistency of empirical solutions

Consider a sample space $(Z, \mathcal{F}, \mu)$ with an increasing sequence of σ-fields $(\mathcal{F}^\nu)_{\nu=1}^\infty$ contained in $\mathcal{F}$. For an increasing sample size the sample path ζ leads to a sequence of $\mathcal{F}^\nu$-measurable probability distributions $\{P^\nu(\bullet, \zeta), \nu = 1, 2, \dots\}$ on $(\Omega, \mathcal{B})$ based on information collected up to ν.
The optimal value $\varphi(P^\nu)$ and optimal solutions $x^*(P^\nu)$ of the approximate stochastic program

$$\min_{x \in \mathcal{X}} F(x, P^\nu) := E_{P^\nu} f(x, \omega) \tag{3}$$

based on $P^\nu(\bullet, \zeta)$ depend on the used sample path ζ and *all presented results hold true for almost all sample paths ζ, i.e., μ-a.s.*

The empirical probability distributions are a special case with the sample path $\zeta = \{\omega_1, \omega_2, \dots\}$ obtained by the simple random sampling from $(\Omega, \mathcal{B}, P)$, $\mu = P^\infty$, and the empirical stochastic program is the Sample Average Approximation problem (2).

2.1. Classical consistency results

(i) If $P^\nu \to P$ weakly and $f(x, \bullet)$ is a continuous bounded function of ω for all $x \in \mathcal{X}$, the pointwise convergence of expected value objectives $F(x, P^\nu) \to F(x, P)\, \forall x \in \mathcal{X}$ follows directly from the definition of weak convergence.

(ii) If $\mathcal{X}$ is compact and the convergence of expectations in (i) is *uniform* on $\mathcal{X}$ then μ-a.s. convergence of optimal values $\varphi(P^\nu) \to \varphi(P)$ follows. It implies convergence in probability, i.e. for all $t > 0$

$$\mu\{\zeta : |\varphi(P^\nu) - \varphi(P)| > t\} \to 0 \text{ for } \nu \to \infty.$$

(iii) A stronger result

$$\mu\{\zeta : \nu^\beta |\varphi(P^\nu) - \varphi(P)| > t\} \to 0 \text{ for } \nu \to \infty$$

suggests a rate of convergence of the empirical optimal value. The exponent β depends, *inter alia*, on existence of finite moments of the random objective function; consult [19] and references therein. Existence of a finite moment generating function is sufficient to guarantee an exponential convergence rate.

(iv) If in addition to (i), (ii), $\mathcal{X}$ is *convex* and $f(\bullet, \omega)$ is strictly convex on $\mathcal{X}$ then the (unique) optimal solutions $x^*(P^\nu)$ of (3) converge μ-a.s. to the unique optimal solution $x^*(P)$ of the original problem

$$\min_{x \in \mathcal{X}} F(x, P) := E_P f(x, \omega). \tag{4}$$

This approach was used for instance in early papers of Kaňková [16, 17].

Notice that merely the pointwise convergence of the empirical expectations does not imply consistency of optimal values. Convexity of $f(x, \bullet)$ helps; (e.g. [13, 24, 33]); in this case, consistency of empirical optimal values follows from the pointwise Law of Large Numbers (LLN) and boundedness assumptions (e.g. [13, 24, 33]).

2.2. General consistency results

A general consistency result based on the notion of epi-convergence of lower semi-continuous (lsc.) functions was proved in [12]. The main step is to prove that the approximate objective functions $F(x, P^\nu)$ epi-converge to the true objective function in (4), which in turn implies convergence results for optimal values and for *sets of optimal solutions* [27].

Definition 1 (Epi-convergence). *A sequence of functions $\{u^\nu : \mathbb{R}^n \to \bar{\mathbb{R}}, \nu = 1, \dots\}$ is said to epi-converge to $u : \mathbb{R}^n \to \bar{\mathbb{R}}$ if for all $x \in \mathbb{R}^n$ the two following properties hold true:*

$$\liminf_{\nu\to\infty} u^\nu(x^\nu) \geq u(x) \text{ for all sequences } x^\nu \to x \tag{5}$$

and for some sequence of $\tilde{x}^\nu$ converging to x

$$\limsup_{\nu\to\infty} u^\nu(\tilde{x}^\nu) \leq u(x). \tag{6}$$

Pointwise convergence implies condition (6), additional assumptions are needed to get validity of condition (5). For example, pointwise convergence of lsc. *convex* functions u, u^ν with $\operatorname{int} \operatorname{dom}(u) \neq \emptyset$ implies epi-convergence. See e.g. Corollary 4 of [41].

Epi-convergence implies that any cluster point $\hat{x}$ of any sequence $\{x^\nu, \nu = 1, \dots\}$ with $x^\nu \in \arg\min u^\nu$ belongs to $\arg\min u$ and convergence of optimal values follows; consult [27].

To formulate the consistency result, we shall assume

a. $\mathcal{X} \subset R^n$ is a nonempty closed set independent of P,

b. $f(x, \omega)$ is a random lower semicontinuous function, i.e. f is jointly measurable and $f(\bullet, \omega)$ is lower semicontinuous for all $\omega \in \Omega$,

c. $P^\nu \to P$ weakly.

To get *epi-convergence* of expectations $F(x, P^\nu) \to F(x, P)$, additional assumptions on convergence of $P^\nu \to P$ and on properties of f are needed. These assumptions mimic to a certain extent those of the classical consistency result:

d. continuity of $f(x, \bullet)$ on Ω for all $x \in \mathcal{X}$;

e. uniform integrability (asymptotic neglibility, tightness) of probability distributions P, P^ν with respect to functions $f(x, \bullet)\, \forall x \in \mathcal{X}$; this replaces assumption of bounded integrands $f(x, \bullet)\, \forall x$.

f. local (lower) Lipschitz property of $f(\bullet, \omega)$ for all $\omega \in \Omega$;
in case of $f(\bullet, \omega)$ convex for all $\omega \in \Omega$, this assumption is not needed.

Proposition 2 (cf. Theorems 3.7, 3.8 of [12]). *Under assumptions* **a**–**f**, $F(x, P^\nu)$ *are* μ*-a.s. proper random lsc. functions and* $F(x, P)$ *is* μ*-a.s. both epi-limit and pointwise limit of* $F(x, P^\nu)$ *for* $\nu \to \infty$.

Epi-convergence of objective functions, cf. [27], implies the consistency result:

Proposition 3 (cf. Theorem 3.9 of [12]). *Under assumptions* **a**–**f** *we have that* μ*-a.s.*

$$\limsup_{\nu \to \infty} \varphi(P^\nu) \le \varphi(P),$$

$\mathcal{X}^*(P^\nu)$ *is a closed-valued* $\mathcal{F}^\nu$*-measurable multifunction and any cluster point* $\hat{x}$ *of any sequence* $\{x^*(P^\nu), \nu = 1, 2, \dots\}$ *of optimal solutions* $x^*(P^\nu) \in \mathcal{X}^*(P^\nu)$ *belongs to* $\mathcal{X}^*(P)$.

In particular, if there is a compact set $\mathcal{D} \subset \mathbb{R}^n$ *such that* μ*-a.s.,* $\mathcal{X}^*(P^\nu) \cap \mathcal{D} \neq \emptyset$ *for* $\nu = 1, 2, \dots,$ *and* $\{x^*(P)\} = \mathcal{X}^*(P) \cap \mathcal{D}$, *then there exists a measurable selection* $x^*(P^\nu)$ *of* $\mathcal{X}^*(P^\nu)$ *such that* $x^*(P) = \lim_{\nu \to \infty} x^*(P^\nu)$ *for* μ*-almost all* ζ *and also* $\varphi(P) = \lim_{\nu \to \infty} \varphi(P^\nu)$ μ*-a.s.*

The problem is much simpler if the underlying probability distribution P is discrete.

Application 1. Consistency result for a convex polyhedral stochastic program with a discrete true distribution.
Let $\omega^1, \dots, \omega^N$ be the atoms of P and $\pi_j, j = 1, \dots, N$, their probabilities, let $\mathcal{X}$ be a nonempty bounded convex polyhedron, $f(x, \omega)$ a continuous function of ω on $\text{conv}\{\omega^1, \dots, \omega^N\}$ and a piece-wise linear convex function of x on $\mathcal{X}$, i.e. the type of the random objective function which is common for two-stage stochastic linear programs.

This implies that $F(x, P) := \sum_{j=1}^N \pi_j f(x, \omega^j)$ is also a piece-wise linear convex function, hence, there exists a finite number of bounded nonoverlapping convex polyhedra $\mathcal{X}^k, k = 1, \dots, K$, such that $\mathcal{X} = \cup_{k=1}^K \mathcal{X}^k$ and $F(x, P)$ is linear on each of $\mathcal{X}^k$. Then the set of optimal solutions $\mathcal{X}^*(P)$ evidently intersects the set $\overline{\mathcal{X}}(P)$ of all extremal points of $\mathcal{X}^k, k = 1, \dots, K$.

Assume that the true probability distribution P is estimated by empirical distributions P^ν based on finite samples of sizes ν from P, carried by subsets of $\{\omega^1, \dots, \omega^N\}$. The empirical objective functions $F(x, P^\nu)$ are also convex, piece-wise linear and the sets of the related extremal points $\overline{\mathcal{X}}(P^\nu)$ intersect $\overline{\mathcal{X}}(P)$. This means that the assumptions of Proposition 2 are fulfilled with the compact set $\mathcal{D} = \overline{\mathcal{X}}(P)$. Consequently, with probability one, any cluster point of any sequence of points $x^\nu \in \mathcal{X}^*(P^\nu) \cap \overline{\mathcal{X}}(P)$ is an optimal solution of the true problem.

Assume in addition that there is a *unique optimal solution* $x^*(P)$ of the true problem

$$\min_{x \in \mathcal{X}} \sum_{j=1}^N \pi_j f(x, \omega^j).$$

In this case there is a measurable selection $x^*(P^\nu)$ from $\mathcal{X}^*(P^\nu) \cap \overline{\mathcal{X}}(P)$ such that with probability 1, $\lim_{\nu \to \infty} x^*(P^\nu) = x^*(P)$. Due to the special form of the objective functions and of the sets $\mathcal{X}^*(P^\nu) \cap \overline{\mathcal{X}}(P)$, this is equivalent to

$$x^*(P^\nu) \equiv x^*(P)\, \mu\text{-a.s. for } \nu \text{ large enough.} \tag{7}$$

It means that *for ν large enough the empirical problem provides μ-a.s. the exact optimal solution of the true problem.* The sample size ν needed to get the above result depends on the data path ζ and on the structure of the problem and it can be estimated, see [33, 35]. For example, in our Application 1 with a unique true optimal solution, there is an exponential rate of convergence for (7).

Comments 1.

- For convex functions $f(\bullet, \omega)$, convex $\mathcal{X}$ and for empirical probability distributions P^ν, epi-convergence of $F(x, P^\nu)$ to $F(x, P)$ follows from the strong law of large numbers for sums of random closed sets and the consistency result can be extended from $\mathbb{R}^s$ to reflexive Banach spaces, cf. [22].
- The general consistency result holds true not only for the empirical probability distributions based on i.i.d. sequences of observations. Hence, it can be a starting point for proving consistency under various weaker assumptions about approximate probability distributions P^ν. For the sake of robustness it is important, *inter alia*, to relax the independence assumptions and to get results for slightly dependent ω^j's. An idea can be to replace the independence assumption by mixing conditions. This was done in [18] along with the rate of convergence.
- For extensions to problems with expectation type constraints see e.g. [25, 33], for consistency of complete local minimizing sets consult [26].
- An important generalization is to discontinuous integrands $f(x, \bullet)$. In such cases, uniform integrability is not sufficient for semicontinuity of integrals $F(x, P^\nu)$. A suitable additional condition is that the probability of the set of discontinuity points of $f(x, \bullet)$ for the true problem is zero; cf. [3]. See [30] for an application to approximated integer stochastic programs or probabilistic programs.
- An alternative approach to consistency proofs is via uniform convergence of expectations, cf. [34]. Its disadvantage is that it leads to difficulties when extending consistency results to problems whose constraints depend on probability distributions.

3. Asymptotic distribution

Under assumption that consistency holds true one may try to derive an asymptotic distribution. See [40] for the first attempts in this direction.

3.1. Asymptotic normality of empirical optimal values

For empirical stochastic program (2), asymptotic normality of the objective function $\frac{1}{\nu}\sum_{j=1}^{\nu} f(x, \omega^j)$ at each point $x \in \mathcal{X}$ is a consequence of the Central Limit Theorem (CLT) if the variance of $f(x, \omega)$ is finite. Also asymptotic normality of the optimal value $\varphi(P^\nu)$ can be proved under relatively weak assumptions, such as compact $\mathcal{X} \neq \emptyset$, unique true optimal solution $x^*(P)$ and $f(\bullet, \omega)$ Lipschitz continuous $\forall\, \omega$, with finite expectation $E\{f(\hat{x}, \omega)\}^2$ at a point $\hat{x} \in \mathcal{X}$; see e.g. [32, 33, 34]. This result allows us to construct approximate confidence intervals for the true optimal value.

For inference based on these approximate confidence intervals one should realize that the empirical optimal value $\varphi(P^\nu)$ has a one-directional bias in the sense that

$$E_\mu \varphi(P^\nu) \le \varphi(P).$$

(Empirical point estimate of $E_\mu \varphi(P^\nu)$ follows from the LLN.) Asymptotic confidence interval for this lower bound for the true optimal value $\varphi(P)$ can be obtained from CLT. An upper bound for the true optimal value $\varphi(P)$ is the expected value of the objective function evaluated at an arbitrary point $\hat{x} \in \mathcal{X}$. Again, its point estimate follows from LLN and an asymptotic confidence interval from CLT. This idea was elaborated in [23] and applied e.g. to a bond portfolio management problem [4]. The resulting bounds are important for designing stopping rules and for testing quality of a "candidate" solution.

3.2. Asymptotic distribution of empirical optimal solutions

Asymptotic normality of constrained maximum likelihood estimators was proved in [1] using the classical Lagrangian approach for problems with explicit equality constraints. It turns out, however, that *in the presence of general constraints asymptotic normality of empirical optimal solutions $x^*(P^\nu)$ cannot be expected* even when all solution sets $\mathcal{X}^*(P)$ and $\mathcal{X}^*(P^\nu)\,\forall\nu$ are singletons. This was observed already in [5] and detailed also in [13].

It is possible to prove that under reasonable assumptions, the asymptotic distribution of unique consistent empirical optimal solutions $x^*(P^\nu)$ is conically normal being projection of normal distribution on a convex cone. Additional assumptions are needed to get asymptotic normality. It may hold true when the true optimal solution $x^*(P)$ is an interior point of $\mathcal{X}$ or when the problem reduces on a neighborhood of the true optimal solution $x^*(P)$ to one with affine constraints. This occurs e.g. when $\mathcal{X}$ is a convex polyhedron with nondegenerated vertices and the strict complementarity conditions are valid at the true solution. A less evident sufficient condition for asymptotic normality was derived in [12], Theorem 4.1. It turns out, however, that it is again connected with the strict complementarity conditions [10].

The general tool for derivation of asymptotic distribution is the generalized δ-theorem, cf. [32], Theorem 2.1, or [20, 25]. It requires certain differentiability property of optimal solution map x^* at P, and a suitable version of CLT for (generalized) gradients of the empirical objective function. Under various assumptions, the empirical optimal solutions converge in distribution to the optimal solution of a randomly perturbed convex quadratic program. The most general results concerning these *conically normal* asymptotic distributions are contained e.g. in papers [20, 21].

Comments 2.

- CLT for $\nabla_x F(x, P^\nu)$ is obtained e.g. for $f(\bullet, \omega)$ convex $C^{1,1}$-functions for all ω, with square integrable Lipschitz constants, with a finite nonsingular variance matrix $V = \mathrm{var}[\nabla_x f(x^*(P), \omega)]$, a finite expectation $E\|\nabla_x f(x^*(P), \omega)\|^2$ and for empirical probability distributions P^ν.

• Differentiability assumptions concerning $f(x,\omega)$ or $F(x,P)$ restrict applicability of the above results. To an extent, they can be relaxed; see for instance [21] for stochastic program with a linear-quadratic convex recourse.
• Using Lagrangians the asymptotic results can be extended to (1) with the set $\mathcal{X}(P)$ of feasible solution described by smooth or convex inequalities, see e.g. [12, 25, 32, 33].
• Conically normal asymptotic distribution of optimal solutions of approximate problems can be also obtained for certain types of empirical distributions based on slightly dependent observations, cf. [39].
• For nonnormal asymptotic distributions and for rates of convergence different from $(\sqrt{\nu})^{-1}$ see e.g. [25]. The key property is again validity of a version of CLT for (generalized) gradients of approximate objective functions.
• If there are multiple optimal solutions asymptotic results concerning convergence of sets of optimal solutions can be treated via asymptotic results for suitably chosen distances of these sets; see e.g. [29].

4. Asymptotic results for a parametric family

Assume now that the true probability distribution P is known to belong to a parametric family $\mathcal{P} = \{\mathcal{P}_\theta\}$ of probability distributions indexed by a parameter vector θ belonging to an open set $\Theta \subset \mathbb{R}^q$. The objective function now depends on θ, $F(x, P_\theta) := F(x,\theta)$, and (1) is a standard parametric program $\min_{x\in\mathcal{X}} F(x,\theta)$. Assume that the optimal value $\varphi(\theta)$ exists for all $\theta \in \Theta$ and is a *continuous* function of θ on a neighborhood of the true parameter value, say, θ_0. Having a statistical estimate θ^ν of θ_0 and knowing its asymptotic properties we can obtain parallel asymptotic properties of the optimal value $\varphi(\theta^\nu)$, cf. [31]:

Proposition 4. *Whenever $\theta^\nu \to \theta_0$ with probability 1 or in probability, then $\varphi(\theta^\nu) \to \varphi(\theta_0)$ with probability 1 or in probability, respectively.*

This assertion can be complemented by the *rates of convergence* and an asymptotic distribution can be based on the δ-theorem:

Proposition 5. *Let θ^ν be an asymptotically normal estimate of θ_0, i.e., $\sqrt{\nu}(\theta^\nu - \theta_0) \sim N(0,\Sigma)$, and φ be continuously differentiable at θ_0 with $\nabla\varphi(\theta_0) \neq 0$. Then $\varphi(\theta^\nu)$ is asymptotically normal,*

$$\sqrt{\nu}(\varphi(\theta^\nu) - \varphi(\theta_0)) \sim N(0, \nabla\varphi(\theta_0)^\top \Sigma \nabla\varphi(\theta_0)). \tag{8}$$

Similar assertions can be proved for optimal solutions provided that these solutions are unique, continuous, differentiable on a neighborhood of θ_0. However, even for unique optimal solutions special assumptions related to the “true” stochastic program are needed. They are not always realistic and their verification is not straightforward. See [7, 9] for asymptotic results based on parametric programming stability, such as:

Proposition 6. *Let $\mathcal{X}$ be a closed convex polyhedral set with* $\operatorname{int}\mathcal{X} \neq \emptyset$ *and $x^*(\theta_0)$ be the unique optimal solution of* $\min_{x\in\mathcal{X}} F(x,\theta_0)$. *Let* $\nabla_x f(x,\theta)$, $\nabla_{xx} f(x,\theta)$, $\nabla_{x\theta} f(x,\theta)$

exist and be jointly continuous on a neighborhood of $[x^*(\theta_0), \theta_0]$ *with* $\nabla_{xx} f(x^*(\theta_0), \theta_0)$ *positive definite. Let* θ^ν *be an asymptotically normal estimate of* θ_0.
Then $\sqrt{\nu}(x^*(\theta^\nu) - x^*(\theta_0))$ *converges in distribution to a mixture of normal distributions conditioned by convex polyhedral sets (the stability sets of the related quadratic program).*

Comments 3.
• The accuracy of the normal approximation can be estimated via the Berry-Esseén theorem provided that there exist higher moments of ω. See [7] for an example which was applied in [2].
• An asymptotic expansion of the density of $\sqrt{\nu}(x^*(\theta^\nu) - x^*(\theta_0))$ was derived in [36].
• The quality of the parametric approach depends on the right choice of the parametric model. If the model is accepted and P_0 is known to belong to the specified parametric family of probability distributions, this information can be exploited when choosing a suitable sample based approximation of P_0.

5. Non-asymptotic results

The asymptotic results help to get reliable solutions via a numerical procedure assuming that the probability distribution is known and an arbitrarily large sample can be used. In situations where the probability distribution is not known, but there is at disposal a finite sample from this distribution, non-asymptotic confidence bounds valid for any sample size are of interest. They can be based on various probability inequalities. For example, using the Chernoff bound (see e.g. [31]), exponential rate of convergence of consistent empirical objective functions $F(x, P^\nu)$ to $F(x, P)$ holds true: For arbitrary real numbers a, t,

$$P(F(x, P^\nu) - F(x, P) \geq a) \leq e^{-ta}[M(t/\nu)]^\nu$$

provided that the moment generating function $M(t)$ of deviations $f(x, \omega) - F(x, P)$ is bounded. Under additional assumptions, exponential convergence

$$P(F(x^*(P^\nu), P) - \varphi(P) \geq \epsilon) \leq \alpha e^{-\beta\nu} \tag{9}$$

can be proved and extended also to exponential convergence for deviations of the empirical optimal values from the true optimum and deviations of unique minimizers, cf. [6]. This idea appears already in the paper [17] who used the Hoeffding inequality to prove an exponential rate for empirical optimal values under boundedness assumption. A general possibility is to apply the Large Deviations Theory, cf. [15, 25, 33].

The above results provide information about the quality of the already obtained solution and may support construction of stopping rules or indicate the need for larger sample sizes. To apply them, it is necessary to estimate parameters, such as α, β in (9); an approach is delineated in [15].

Comments 4.
• The results were achieved mostly for i.i.d. observations. For robustness reasons

one is again interested in "slightly" dependent variables. Generalizations under various mixing assumptions are indicated e.g. in [18], see also [6].

• The rates of convergence can be obtained also for sets of empirical optimal solutions, e.g. [15, 29].

• For small sample sizes it pays to exploit all available information to provide an approximation which is a good substitute of the true problem. Several possibilities of exploitation of such "soft" information were delineated already in [42]. For example, information about the growth condition for the true objective function and about the tail behavior of the true probability distribution is essential for construction of universal confidence sets for optimal solutions which are applicable for any sample size; cf. [37].

References

[1] Aitchinson, J., and Silvey, S. V.: Maximum likelihood estimation of parameters subject to restraints, *Ann. Math. Statist.* **29** (1958), 813–828.

[2] Anandalingam, G.: A stochastic programming process model for investment planning, *Comput. Oper. Res.* **14** (1987), 521–536.

[3] Artstein, Z., and Wets, R. J.-B.: Stability results for stochastic programs and sensors, allowing for discontinuous objective functions, *SIAM J. Optimization* **4** (1994), 537–550.

[4] Bertocchi, M., Dupačová, J., and Moriggia, V.: Sensitivity of bond portfolio's behavior with respect to random movements in yield curve: A simulation study, *Annals of Oper. Res.* **99** (2000), 267–286.

[5] Chernoff, H.: On the distribution of the likelihood ratio, *Ann. Math. Statist.* **25** (1954), 573–578.

[6] Dai, L., Chen, C. H., and Birge, J.L.: Convergence properties of two-stage stochastic programming, *J. Optim. Theory. Appl.* **106** (2000), 489–509.

[7] Dupačová, J.: Stability in stochastic programming with recourse – estimated parameters, *Math. Progr.* **28** (1984), 72–83.

[8] Dupačová, J.: Stability and sensitivity analysis for stochastic programming, *Annals of Oper. Res.* **27** (1990), 115–142.

[9] Dupačová, J.: On statistical sensitivity analysis in stochastic programming, *Annals of Oper. Res.* **30** (1991), 199–214.

[10] Dupačová, J.: On non-normal asymptotic behavior of optimal solutions for stochastic programming problems and on related problems of mathematical statictics, *Kybernetika* **27** (1991), 38–52.

[11] Dupačová, J.: Output analysis for approximated stochastic programs. In: *Stochastic Optimization: Algorithms and Applications* (S. Uryasev, and P. M. Pardalos, eds.), Kluwer, 2001, 1–29.

[12] Dupačová, J., and Wets, R. J.-B.: Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems, *Ann. Stat.* **16** (1988), 1517–1549.

[13] Haberman, S. J.: Concavity and estimation, *Ann. Stat.* **17** (1988), 1631–1661.

[14] Huber, P.: The behavior of maximum likelihood estimates under nonstandard conditions. In: *Proc. Fifth Berkley Symp. Math. Statist. Prob.* **1**, Univ. California Press, 1967, 799–821.

[15] Kaniovski, Y. M., King, A. J., and Wets, R. J.-B.: Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems, *Annals of Oper. Res.* **56** (1995), 189-208.

[16] Kaňková, V.: Optimum solution of a stochastic optimization problem with unknown parameters. In:*Transactions of the 7th Prague Conference 1974*, Academia, Prague, 1977, 239–244.

[17] Kaňková, V.: An approximative solution of a stochastic optimization problem. In: *Transactions of the 8th Prague Conference 1978*, Academia, Prague, 1978, 349–353.

[18] Kaňková, V.: A note on estimates in stochastic programming, *JCAM* **56** (1994), 97–112.

[19] Kaňková, V.: Empirical estimates in economic and financial problems via heavy tails. In: *Proceedings of the* 30^{th} *international conference Mathematical Methods in Economics* (J. Ramík, and D. Stavárek, eds.), Karviná: Silesian University, School of Business Administration, 2012, 396–401.

[20] King, A. J.: Generalized delta theorems for multivalued mappings and measurable selections, *Math. of Oper. Res.* **14** (1989), 720–736.

[21] King, A. J., and Rockafellar, R. T.: Asymptotic theory for solutions in statistical estimation and stochastic programming, *Math. of Oper. Res.* **18** (1993), 148–162.

[22] King, A. J., and Wets, R. J.-B.: Epi-consistency of convex stochastic programs, *Stochastics and Stochastics Reports* **34** (1990), 83–92.

[23] Mak Wai-Kei, Morton, D. P., and Wood, R. K.: Monte Carlo bounding techniques for determining solution quality in stochastic programs, *OR Letters* **24** (1999), 47–56.

[24] Niemiro, W.: Asymptotics for M-estimators defined by convex minimization, *Ann. Stat.* **20** (1992), 1514–1533.

[25] Pflug, G. Ch.: Stochastic optimization and statistical inference. Chapter 7 in: *Stochastic Programming* (A. Ruszczynski, and A. Shapiro, eds.), Handbooks in OR & MS **10**. Elsevier, 1995, 427–482.

[26] Robinson, S. M.: Analysis of sample-path optimization, *Math. of Oper. Res.* **21** (1996), 513–528.

[27] Rockafellar, R. T., and Wets, R. J.-B.: *Variational Analysis*, Springer, New York, 1997.

[28] Römisch, W.: Stability of stochastic programming. Chapter 8 in: *Stochastic Programming* (A. Ruszczynski, and A. Shapiro, eds.), Handbooks in OR & MS **10**. Elsevier, 2003, 483–554.

[29] Römisch, W., and Schultz, R.: Lipschitz stability for stochastic programs with complete recourse, *SIAM J. Optimization* **6** (1996), 531–547.

[30] Schultz, R.: Rates of convergence in stochastic programs with complete integer recourse, *SIAM J. Optimization*, **6** (1996), 1138–1152.

[31] Serfling, R. J.: *Approximation Theorems in Mathematical Statistics*, Wiley, New York, 1980.

[32] Shapiro, A.: Asymptotic analysis of stochastic programs, *Annals of Oper. Res.* **30** (1991), 169–186.

[33] Shapiro, A.: Monte Carlo sampling methods. Chapter 6 in: *Stochastic Programming* (A. Ruszczynski, and A. Shapiro, eds.), Handbooks in OR & MS **10**. Elsevier 2003, 353–425.

[34] Shapiro, A.: Statistical inference, Chapter 5 in: A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming.* MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2009, 155–252.

[35] Shapiro, A., and Homem-de-Mello, T.: On rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs, *SIAM J. Optimization* **11** (2000), 70–86.

[36] Vogel, S.: Stability results for stochastic programming problems, *Optimization* **19** (1988), 269–288.

[37] Vogel, S.: Universal confidence sets for solutions of optimization problems. *SIAM J. Optimization* **19** (2008), 1467–1488.

[38] Wald, A.: Note on consistency of the maximum likelihood estimates, *Ann. Math. Statist.* **30** (1940), 389–407.

[39] Wang, L., and Wang, J.: Limit distribution of statistical estimators for stochastic programs with dependent samples, *ZAMM* **19** (1999), 257–266.

[40] Wets, R. J.-B.: A statistical approach to the solution of stochastic programs with (convex) simple recourse. Working Paper, Univ. of Kentucky, 1979.

[41] Wets, R. J.-B.: Convergence of convex functions, variational inequalities and convex optimization problems. In: *Variational Inequalities and Complementarity Problems. Theory and Applications* (R. W. Cottle et al., eds.) Wiley, New York, 1980, 375–403.

[42] Wets, R. J.-B.: Statistical estimation from an optimization viewpoint, *Annals of Oper. Res.* **85** (1999), 79–101.