

Stochastic integer programming: Limit theorems and confidence intervals*

Andreas Eichhorn

Department of Mathematics, Humboldt-University Berlin, Unter den Linden 6, 10099 Berlin, Germany
email: eichhorn@math.hu-berlin.de <http://www.math.hu-berlin.de/~eichhorn>

Werner Römisch

Department of Mathematics, Humboldt-University Berlin, Unter den Linden 6, 10099 Berlin, Germany
email: romisch@math.hu-berlin.de <http://www.math.hu-berlin.de/~romisch>

We consider empirical approximations of two-stage stochastic mixed-integer linear programs and derive central limit theorems for the objectives and optimal values. The limit theorems are based on empirical process theory and the functional delta method. We also show how these limit theorems can be used to derive confidence intervals for optimal values via a certain modification of the bootstrapping method.

Key words: Stochastic programming; mixed-integer optimization; stability; Hadamard directional differentiability; Donsker class; delta method; bootstrap

MSC2000 Subject Classification: Primary: 90C15; Secondary: 90C11, 60F17, 62F40

OR/MS subject classification: Primary: Programming/Stochastic; Secondary: Programming/Integer/Theory, Statistics/Estimation

1. Introduction Stochastic optimization problems take into account random influence. In this paper it is assumed that this can be described by means of a probability distribution P on \mathbb{R}^k with some $k \in \mathbb{N}$. We consider two-stage linear mixed-integer stochastic programs where the sum of the first stage cost and the expectation with respect to P of the second stage cost has to be minimized. However, in most applications P is not known exactly. Moreover, even if P is given, it might happen that the stochastic program cannot be solved due to technical limitations and one has to use a simpler approximating distribution that makes the problem solvable. Hence, one often has to deal with statistical models and approximations Q of P . Of course, since solutions and optimal values of the original problem containing the distribution P are of interest, it is necessary to have statements at hand about stability of stochastic programs with respect to perturbations of P .

There are a number of such stability results in literature, see [28] for a recent survey. Most of these results consist of (Lipschitz) continuity properties of solution and optimal values with respect to certain probability metrics $d(P, Q)$. Especially in the case that P is unknown, this may in the end not be completely satisfactory, because in this case the distance $d(P, Q)$ is, of course, also unknown. Hence, the question arises, whether it is possible to prove statistical statements about the accuracy of solution and optimal values. In particular, confidence sets may be of interest. Of course, such statistical statements require the availability of some statistical estimates associated with P , e.g., independent identically distributed (iid) samples of P . The latter are often called empirical estimates and they can be understood as the so-called empirical measure $Q = P_n$ with $n \in \mathbb{N}$ denoting the sample size.

Asymptotic properties of statistical estimators in stochastic programming have been studied intensively. We refer to Chapters 6,7 and 8 in [31] for various aspects and views. For two-stage stochastic programs without integrality requirements much is known. For the empirical estimator the papers [5], [14] and [3] contain results on (epi-) consistency, laws of large numbers and on asymptotic normality. In [34], [30, Chapter 6], [13], [22] and [36] limit theorems for optimal values and solutions are derived by imposing uniqueness of solutions and certain differentiability properties of objectives and/or integrands. Convergence rates and large deviation type results are derived, e.g., in [7], [21], [12], [23] and [37]. The situation is essentially different for mixed-integer two-stage stochastic programs. In [32] conditions are given implying consistency, convergence rates and a law of the iterated logarithm for optimal values. Glivenko-Cantelli results for the objective are established in [24] and large deviation type results are derived in [16] and [1] for pure integer models and in [26, 28] for the mixed-integer case. Much of this work is based on recent developments of empirical process theory, e.g., on Talagrand's work [38, 39] (see also the monographs [41, 40]).

*This work was supported by the DFG Research Center MATHEON *Mathematics for key technologies* in Berlin.

In this paper, we extend the earlier work by deriving a uniform limit theorem for the objective of mixed-integer two-stage stochastic programs. Its proof is again based on recent results of empirical process theory. While Banach spaces of continuous functions play an important role for such limit theorems in case of two-stage stochastic programs without integrality constraints (cf. [36]), the Banach space of bounded functions has to be used in the mixed-integer situation. More precisely, it is shown that the family of integrands forms a so-called Donsker class in the Banach space of bounded functions. As a consequence, a limit theorem for optimal values is derived by relying on the infinite-dimensional delta method (see [29] for an introductory overview) and on a recent Hadamard directional differentiability result for infimal value mappings on the space of bounded functions [17]. Furthermore, since the Hadamard directional derivative is not linear in general, special bootstrap techniques are developed that allow to compute approximate confidence intervals for optimal values.

So far there is some special work about confidence sets for solutions and optimal values of stochastic programs. In [8], a stochastic program with finite decision space is considered. Confidence sets for the solution set are derived by estimating the objective for each possible decision and selecting the presumably best decisions according to some statistical selection procedure. In [20], a certain simple two-stage stochastic program is analyzed for the case that $P = P_\theta$ is contained in the parametric family of normal distributions and that a confidence set of the unknown parameter vector θ is given. It is suggested to calculate the worst case solution with θ varying in the given confidence region. In [2], a stochastic integer program without first stage decision is considered. For such problems, optimization can be carried out scenariowise. To approximate the distribution of the optimal value a method based on order statistics is suggested where only a finite number of deterministic programs has to be solved.

In this paper, we analyze statistical behavior of the objective of general linear two-stage stochastic programs (possibly with integer requirements). We assume that the underlying probability distribution P is unknown and *that we are able to sample from it independently*. In Section 2, we present the framework of our analysis and in Section 3 our main result, a limit theorem for the objective of the stochastic program, is proven by means of empirical process theory. Thereby, we are geared to the monographs [40] and [41]. In Section 4, this limit theorem is carried forward to the optimal value of the stochastic program by means of the functional delta method. These results are used in Section 5 to derive a general method for calculating confidence intervals for the optimal value by means of resampling techniques (bootstrap-like methods). Finally, some numerical examples are presented in Section 6.

2. Framework Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an arbitrary probability space and let $\xi : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^k, \mathcal{B}^k)$ a measurable random vector with support $\Xi \subset \mathbb{R}^k$ which is assumed to be polyhedral and bounded and let $P = \mathbb{P}^\xi$ be the probability distribution of ξ . We consider the stochastic mixed-integer program

$$\min \left\{ c'x + \int_{\Xi} \phi(h(\xi) - T(\xi)x) dP(\xi) : x \in X \right\} \quad (1)$$

with $X \subset \mathbb{R}^m$ compact, $c \in \mathbb{R}^m$, $T : \Xi \rightarrow \mathbb{R}^{r \times m}$ and $h : \Xi \rightarrow \mathbb{R}^r$ affinely linear. The function $\phi : \mathbb{R}^r \rightarrow \mathbb{R}$ contains the second stage problem given by

$$\phi(t) := \min \{ q'y + \bar{q}'\bar{y} : Wy + \bar{W}\bar{y} = t, y \in \mathbb{Z}_+^{\hat{m}}, \bar{y} \in \mathbb{R}_+^{\bar{m}} \} \quad (2)$$

with $q \in \mathbb{R}^{\hat{m}}$, $\bar{q} \in \mathbb{R}^{\bar{m}}$, $W \in \mathbb{Q}^{r \times \hat{m}}$ and $\bar{W} \in \mathbb{Q}^{r \times \bar{m}}$. It is assumed that (1) satisfies

- (i) relatively complete recourse:
 $\forall (x, \xi) \in X \times \Xi \exists y \in \mathbb{Z}_+^{\hat{m}}, \bar{y} \in \mathbb{R}_+^{\bar{m}} : h(\xi) - T(\xi)x = Wy + \bar{W}\bar{y}.$
- (ii) dual feasibility:
 $\exists u \in \mathbb{R}^r : W'u \leq q, \bar{W}'u \leq \bar{q}.$

Under these assumptions it turns out that ϕ is lower semicontinuous and piecewise polyhedral on $\text{dom}\phi$ (e.g., [18, Proposition 2], [28, Lemma 33]).

We define the infimal value mapping

$$\begin{aligned} v : \mathcal{P}(\Xi) &\rightarrow \mathbb{R} \\ Q &\mapsto v(Q) := \min \left\{ c'x + \int_{\Xi} \phi(h(\xi) - T(\xi)x) dQ(\xi) : x \in X \right\} \end{aligned}$$

that maps a probability distribution on Ξ to the optimal value of the stochastic program (1). We are interested in the asymptotic behavior of $v(P) - v(P_n)$ where P_n is the empirical distribution according

to independent samples ξ_1, ξ_2, \dots of the original distribution P , i.e.,

$$P_n = \frac{1}{n} \sum_{j=1}^n \delta_{\xi_j}.$$

REMARK 2.1 *The lower semicontinuity and the piecewise polyhedrality of ϕ is also valid if X is not bounded but closed. However, the results that are derived below need compactness of X , so we impose it throughout in this paper. If X is not bounded, the standard technique in perturbation analysis of optimization problems consists in localizing the problem, i.e., to replace the unbounded X by $X_{\mathcal{U}} = X \cap \text{cl}\mathcal{U}$ with some open and bounded set $\mathcal{U} \subset \mathbb{R}^m$ that contains the solution set of (1) which is assumed to be nonempty (cf. [27, 15]). Then, however, the localized infimal value at a perturbed probability distribution Q (e.g., P_n) does not coincide with $v(Q)$ in general, but represents the (local) infimal value attained at some locally optimal solution.*

3. Central Limit Theorem for the Objective In this section we are going to prove a central limit theorem for the objective function by means of empirical process theory and asymptotic statistics. In order to make the notation of the previous section fit to the notation that is used in asymptotic statistics we have to reformulate the stochastic program (1). For $x \in X$ we define the function $f_x : \Xi \rightarrow \mathbb{R}$ as the integrand of (1):

$$f_x(\xi) := c'x + \phi(h(\xi) - T(\xi)x)$$

Further, we define the class \mathcal{F} as the set of all possible integrands of the stochastic program:

$$\mathcal{F} := \{f_x : x \in X\}$$

Thus we can understand the distributions $P, Q \in \mathcal{P}(\Xi)$ and $P_n \in \mathcal{P}(\Xi)^\Omega$ (with Ω denoting the randomness of the sampling procedure) as mappings from \mathcal{F} to \mathbb{R} :

$$Qf := \int_{\Xi} f(\xi) dQ(\xi), \quad P_n f := \frac{1}{n} \sum_{j=1}^n f(\xi_j)$$

for $f \in \mathcal{F}$. With these notations (1) reads

$$v(Q) = \min \{Qf_x : x \in X\} \tag{3}$$

or

$$v(Q) = \min \{Qf : f \in \mathcal{F}\} \tag{4}$$

Due to our assumptions about X and Ξ it turns out that the class \mathcal{F} is uniformly bounded.

LEMMA 3.1 *There exists a constant K such that $\forall f \in \mathcal{F} \forall \xi \in \Xi : |f(\xi)| \leq K$.*

PROOF. Setting $\mathcal{T} := \{Wy + \bar{W}\bar{y} : y \in \mathbb{Z}_+^{\hat{m}}, \bar{y} \in \mathbb{R}_+^{\hat{m}}\}$ we get by [4, Theorem 2.1] that there exist real numbers $a, b \in \mathbb{R}$ such that for all $t, \tilde{t} \in \mathcal{T}$ the following estimate holds

$$|\phi(t) - \phi(\tilde{t})| \leq a|t - \tilde{t}| + b. \tag{5}$$

Since X and Ξ are bounded and $h(\cdot)$ and $T(\cdot)$ are affinely linear, also the set $\mathcal{T}' := \{h(\xi) - T(\xi)x : \xi \in \Xi, x \in X\}$ is bounded. Furthermore, it holds that $\mathcal{T}' \subset \mathcal{T}$ because relatively complete recourse was assumed. Thus, (5) implies that ϕ is bounded on \mathcal{T}' . Thus

$$\begin{aligned} |f_x(\xi)| &\leq \|c\| \|x\| + \|\phi(h(\xi) - T(\xi)x)\| \\ &\leq \|c\| \max_{\bar{x} \in X} \|\bar{x}\| + \sup_{t \in \mathcal{T}'} \|\phi(t)\| =: K \end{aligned}$$

for every $f \in \mathcal{F}, \xi \in \Xi$. □

For an arbitrary set Y we introduce the linear normed space $\ell^\infty(Y)$ of all real-valued bounded functions on Y and the supremum-norm, respectively:

$$\ell^\infty(Y) := \{\psi \in \mathbb{R}^Y : \sup_{y \in Y} |\psi(y)| < \infty\}, \quad \|\psi\|_Y := \sup_{y \in Y} |\psi(y)|$$

Hence, since for $Q \in \mathcal{P}(\Xi)$ the set $\{Qf : f \in \mathcal{F}\}$ is bounded in \mathbb{R} , we can write: $Q \in \ell^\infty(\mathcal{F})$. Analogously, we have $P_n \in \ell^\infty(\mathcal{F})^\Omega$ with Ω denoting the randomness of the sampling procedure. Our main result now is a statement about weak convergence of $\sqrt{n}(P_n - P)$ in this space $\ell^\infty(\mathcal{F})$. Since, however, the mapping $P_n(\cdot)$ from Ω to $\ell^\infty(\mathcal{F})$ is not measurable in general, we have to rely on the generalized weak convergence concept abbreviated by \rightsquigarrow for sequences of arbitrary maps (e.g., [41, Chapter 1], [40, Chapter 18]).

THEOREM 3.1 *The class \mathcal{F} is P -Donsker, i.e., in $\ell^\infty(\mathcal{F})$ we have the weak convergence*

$$\sqrt{n}(P_n - P) \rightsquigarrow G_P,$$

where $G_P \in \ell^\infty(\mathcal{F})^\Omega$ is a P -Brownian Bridge, i.e., G_P is measurable, tight and Gaussian:

$$G_P \sim \mathcal{N}\left(0, (Pfg - PfPg)_{f,g \in \mathcal{F}}\right).$$

PROOF. We will utilize properties of mixed-integer two-stage stochastic programs that can be found in [28] as well as empirical process theory from [41]. The proof consists of 5 parts.

a) First we show that the function ϕ from (2) and, as a consequence, also the functions $f \in \mathcal{F}$ have a piecewise Lipschitzian structure:

Setting $\mathcal{T} := \{Wy + \bar{W}\bar{y} : y \in \mathbb{Z}_+^{\bar{m}}, \bar{y} \in \mathbb{R}_+^{\bar{m}}\} \subset \mathbb{R}^r$ we conclude from [28, Lemma 33] that there exist $L > 0$, $\tau \in \mathbb{N}$, and $B_j \subset \hat{\mathcal{B}}_{ph_\tau}(\mathcal{T})$ ($j \in \mathbb{N}$) such that $\mathcal{T} = \cup_{j \in \mathbb{N}} B_j$ and $B_i \cap B_j = \emptyset$ for $i \neq j$ and $\phi|_{B_j}$ Lipschitz continuous with uniform Lipschitz constant L . Thereby, we use the notation

$$\hat{\mathcal{B}}_{ph_\tau}(\mathcal{T}) := \left\{ \mathcal{T} \cap \bigcap_{j=1}^{\tau} H_j \mid \begin{array}{l} H_j = \{\xi : c'_j \xi \leq d_j\} \text{ or} \\ H_j = \{\xi : c'_j \xi < d_j\} \text{ with } c_j \in \mathbb{R}^r, d_j \in \mathbb{R} \end{array} \right\}$$

for intersections of \mathcal{T} and at most τ open or closed half-spaces, i.e., polyhedra with at most τ faces where each face may be included or excluded. Moreover, since $\mathcal{T}' := \{h(\xi) - T(\xi)x : \xi \in \Xi, x \in X\}$ is bounded and $\mathcal{T}' \subset \mathcal{T}$ due to relatively complete recourse, we know from, e.g., [28, Lemma 33], that finitely many B_j are sufficient to cover \mathcal{T}' , i.e., it exists $\nu \in \mathbb{N}$ and $B_1, \dots, B_\nu \in \hat{\mathcal{B}}_{ph_\tau}(\mathcal{T}')$ such that $\mathcal{T}' = \cup_{j=1}^{\nu} B_j$ and $B_i \cap B_j = \emptyset$ for $i \neq j$ and $\phi|_{B_j}$ Lipschitz continuous with Lipschitz constant L . Let ϕ_j be a Lipschitz extension of $\phi|_{B_j}$ from B_j to \mathbb{R} preserving the Lipschitz constant L ($i = 1, \dots, \nu$). Then ϕ can be written as

$$\phi(t) = \sum_{j=1}^{\nu} \phi_j(t) \chi_{B_j}(t)$$

with $\chi_{B_j}(t)$ denoting the indicator function taking value 1 if $t \in B_j$ and 0 otherwise. Thus, every $f_x \in \mathcal{F}$ can be written as

$$f_x(\xi) = c'x + \sum_{j=1}^{\nu} \phi_j(h(\xi) - T(\xi)x) \chi_{B_j}(h(\xi) - T(\xi)x)$$

Now, we set $\Xi_{x,j} := \{\xi \in \Xi : h(\xi) - T(\xi)x \in B_j\}$ ($x \in X, j = 1, \dots, \nu$). Note that there is a number $\mu \in \mathbb{N}$ such that $\Xi_{x,j} \in \hat{\mathcal{B}}_{ph_\mu}(\Xi)$ for all $x \in X, j = 1, \dots, \nu$. Furthermore, we set $f_{x,j}(\xi) := c'x + \phi_j(h(\xi) - T(\xi)x)$ for $x \in X, j = 1, \dots, \nu$. Finally, for $j = 1, \dots, \nu$ we define

$$\begin{aligned} \mathcal{F}_j &:= \{f_{x,j} : x \in X\} &= \{c'x + \phi_j(h(\xi) - T(\xi)x) : x \in X\} \\ \mathcal{G}_j &:= \{\chi_{\Xi_{x,j}} : x \in X\} &= \{\chi_{\{\xi \in \Xi : h(\xi) - T(\xi)x \in B_j\}} : x \in X\} \end{aligned}$$

b) Next, it will be shown that each of these 2ν classes is uniformly bounded and the criterion that will be used below to prove the Donsker property for these classes will be formatted:

Clearly, the classes \mathcal{G}_j ($j = 1, \dots, \nu$) are uniformly bounded by 1 since they contain indicator functions only.

Since \mathcal{T}' is bounded and ϕ_j is Lipschitz continuous with modulus L we have that ϕ_j is bounded on \mathcal{T}' . Hence, \mathcal{F}_j ($j = 1, \dots, \nu$) are uniformly bounded by some constants $K_j \geq 0$.

For the Donsker property [41, Theorem 2.5.2] will be used, i.e., the following three conditions have to be verified for $\mathcal{H} = \mathcal{F}_j$ and $\mathcal{H} = \mathcal{G}_j$ ($j = 1, \dots, \nu$), respectively:

(i) Existence of an envelope function:

$$\exists F_{\mathcal{H}} \in \mathbb{R}^{\Xi} : P^*(F_{\mathcal{H}}^2) < \infty, F_{\mathcal{H}}(\xi) \geq |h(\xi)| \quad \forall h \in \mathcal{H}, \xi \in \Xi.$$

(ii) \mathcal{H} is ‘‘suitable measurable’’¹:

There exists a countable collection $\mathcal{H}' \subset \mathcal{H}$ such that every $h \in \mathcal{H}$ is the pointwise limes of a sequence h_n in \mathcal{H}' .

¹The measurability condition here is stronger than necessary but easy to verify. In the original version it is required that the classes

$$\mathcal{H}_\infty^2 := \{(h - g)^2 : h, g \in \mathcal{H}\}, \quad \mathcal{H}_\delta := \{h - g : h, g \in \mathcal{H}, \|h - g\|_{P,2} \leq \delta\} \quad (\delta > 0)$$

are P -measurable, i.e., for every $n \in \mathbb{N}$ and every $e \in \{-1, 1\}^n$ the mapping $(\xi_1, \dots, \xi_n) \mapsto \sup_{h \in \mathcal{H}_\delta} |\sum_{i=1}^n e_i h(\xi_i)|$ is measurable. See Definition 2.3.3 and Example 2.3.4 in [41] and remark to Theorem 19.14 in [40].

(iii) Uniform entropy condition:

The *uniform entropy* given by

$$\int_0^\infty \sup \left\{ \sqrt{\log N \left(\varepsilon \|F_{\mathcal{H}}\|_{Q,2}, \mathcal{H}, L_2(Q) \right)} : Q \in \mathcal{P}_d(\Xi), 0 < QF_{\mathcal{H}}^2 < \infty \right\} d\varepsilon$$

is *finite* where $\mathcal{P}_d(\Xi)$ denotes the set of all finitely discrete probability measures on Ξ and $N(\delta, \mathcal{H}, L_p(Q))$ is the *covering number*² of \mathcal{H} in the space $L_p(Q)$.

If \mathcal{H} is uniformly bounded by a constant $K \geq 0$ then, obviously, $N(\delta, \mathcal{H}, L_2(Q)) = 1$ for $\delta > K$, i.e., $\log N(\delta, \mathcal{H}, L_2(Q)) = 0$. Hence, if one chooses $F_{\mathcal{H}} \equiv K$ as envelope function, it suffices to verify

$$\int_0^1 \sup \left\{ \sqrt{\log N(\varepsilon K, \mathcal{H}, L_2(Q))} : Q \in \mathcal{P}_d(\Xi) \right\} d\varepsilon < \infty$$

(note that in this case $\|F_{\mathcal{H}}\|_{Q,2} = K$ and $QF_{\mathcal{H}}^2 = K^2$ for all $Q \in \mathcal{P}_d(\Xi)$). Hence, for $\mathcal{H} = \mathcal{F}_j$ and $\mathcal{H} = \mathcal{G}_j$ ($j = 1, \dots, \nu$) it is sufficient to verify finiteness of the latter integral.

c) We start with verifying these three conditions for the classes $\mathcal{H} = \mathcal{F}_j$ for arbitrary $j \in \{1, \dots, \nu\}$:

1. Envelope function:

As stated above, \mathcal{F}_j is uniformly bounded by a constant $K_j \geq 0$, i.e., $F_{\mathcal{F}_j} \equiv K_j$ is an envelope function for \mathcal{F}_j with $P^*(F_{\mathcal{F}_j}) = K_j < \infty$

2. Measurability:

Of course, since $X \subset \mathbb{R}^m$, there exists a countable dense subset $X' \subset X$. Thus, for arbitrary $x_0 \in X$ there is a sequence x_n in X' such that $x_n \rightarrow x_0$. Hence, since ϕ_j is continuous,

$$c'x_n + \phi_j(h(\xi) - T(\xi)x_n) \rightarrow c'x_0 + \phi_j(h(\xi) - T(\xi)x_0)$$

for every $\xi \in \Xi$, i.e., $f_{x_n,j} \rightarrow f_{x_0,j}$ pointwise. Thus, $\mathcal{H}' := \{f_{x,j} : x \in X'\}$ is a suitable countable subset of \mathcal{H} .

3. Uniform entropy condition:

In Chapter 2.1.1 in [41] it is demonstrated that

$$N(\varepsilon K_j, \mathcal{F}_j, L_2(Q)) \leq N_{[]} (2\varepsilon K_j, \mathcal{F}_j, L_2(Q))$$

where $N_{[]}(\delta, \mathcal{H}, L_p(Q))$ denotes the *bracketing number*³ of the class of functions \mathcal{H} in the space $L_p(Q)$. Further, for $x, \bar{x} \in X$ it holds that

$$\begin{aligned} |f_{x,j}(\xi) - f_{\bar{x},j}(\xi)| &= |c'(x - \bar{x}) + \phi_j(h(\xi) - T(\xi)x) - \phi_j(h(\xi) - T(\xi)\bar{x})| \\ &\leq (\|c\| + L \|T(\xi)\|) \|x - \bar{x}\|, \end{aligned}$$

i.e., the functions $f_{x,j}$ are Lipschitz in the parameter x . Thus, we get by means of [41, Theorem 2.7.11] that

$$N_{[]} (2\varepsilon K_j, \mathcal{F}_j, L_2(Q)) \leq N(\varepsilon, X, |\cdot|),$$

where the right-hand side is the covering number of the set X in \mathbb{R}^m which does not depend on the measure Q . Because X is compact there exists a constant $c \geq 0$ such that $N(\varepsilon, X, |\cdot|) \leq c\varepsilon^{-m}$. Hence,

$$\int_0^1 \sup \left\{ \sqrt{\log N(\varepsilon K_j, \mathcal{F}_j, L_2(Q))} : Q \in \mathcal{P}_d(\Xi) \right\} d\varepsilon \leq \int_0^1 \sqrt{\log c\varepsilon^{-m}} d\varepsilon < \infty$$

thus the third condition holds and \mathcal{F}_j is shown to be P -Donsker.

d) Now we will prove the Donsker property for \mathcal{G}_j . Therefore, we verify the three conditions for the set $\mathcal{H}_\mu = \left\{ \chi_B : B \in \hat{\mathcal{B}}_{ph_\mu}(\Xi) \right\}$ and note that $\mathcal{G}_j \subset \mathcal{H}_\mu$ for $j = 1, \dots, \nu$.

1. Envelope function:

$F \equiv 1$ does the job.

²The covering number of \mathcal{H} in the space $L_p(Q)$ is defined as the minimum number of open balls in $L_p(Q)$ with radius δ that are needed to cover \mathcal{H} (Definition 2.1.5 in [41]).

³A δ -bracket is a pair of functions $l, u \in L_p(Q)$ such that $l(\xi) \leq u(\xi) \forall \xi \in \Xi$ and $\|u - l\|_{Q,p} < \delta$. The bracketing number $N_{[]}(\delta, \mathcal{H}, L_p(Q))$ of a class \mathcal{H} in the space $L_p(Q)$ is defined as the minimum number of δ -brackets $[l, u]$ in $L_p(Q)$ that is needed such that every $h \in \mathcal{H}$ lies between one of these brackets, i.e., $l \leq h \leq u$ ([41, Definition 2.1.6]).

2. Measurability

We set $\mathcal{H}'_\mu = \left\{ \chi_B : B \in \hat{\mathcal{B}}_{ph_\mu, \mathbb{Q}}(\Xi) \right\}$ with

$$\hat{\mathcal{B}}_{ph_\mu, \mathbb{Q}}(\Xi) := \left\{ \Xi \cap \bigcap_{j=1}^\mu H_j \mid \begin{array}{l} H_j = \{ \xi : c_j' \xi \leq d_j \} \text{ or} \\ H_j = \{ \xi : c_j' \xi < d_j \} \text{ with } c_j \in \mathbb{Q}^k, d_j \in \mathbb{Q} \end{array} \right\}$$

the set of intersection of Ξ and polyhedra being described by rational coefficients and having at most μ faces where each face may be included or excluded. It is easy to see that for each $B \in \hat{\mathcal{B}}_{ph_\mu}(\Xi)$ there is a sequence B_n in $\hat{\mathcal{B}}_{ph_\mu, \mathbb{Q}}(\Xi)$ such that $\chi_{B_n} \rightarrow \chi_B$ pointwise for $n \rightarrow \infty$ (note that Ξ is a bounded polyhedron).

3. Uniform entropy condition:

We show that \mathcal{H}_μ is a so-called VC class⁴: For the set of (subgraphs of) indicator functions of open or closed half-spaces ($\mu = 1$) it holds obviously that

$$V(\mathcal{H}_1) = V\left(\left\{ \text{sub}\chi_B : B \in \hat{\mathcal{B}}_{ph_1}(\Xi) \right\}\right) \leq k + 2 < \infty$$

because given $k + 2$ different points in \mathbb{R}^k it is never possible to separate linearly each subset of these points from the rest. Thus, \mathcal{H}_1 is VC. And because

$$\begin{aligned} \text{sub}\mathcal{H}_\mu &= \left\{ \text{sub}\chi_B : B \in \hat{\mathcal{B}}_{ph_\mu}(\Xi) \right\} \\ &= \left\{ \{(\xi, t) \in \Xi \times \mathbb{R} : t < \chi_B(\xi)\} : B \in \hat{\mathcal{B}}_{ph_\mu}(\Xi) \right\} \\ &= \left\{ \bigcap_{i=1}^\mu \text{sub}\chi_{B_i} : B_i \in \hat{\mathcal{B}}_{ph_1}(\Xi) \right\} = \text{sub}\mathcal{H}_1 \cap \dots \cap \text{sub}\mathcal{H}_1 \end{aligned}$$

it holds that \mathcal{H}_μ is also VC due to [41, Lemma 2.6.17 (ii)].

Theorem 2.6.7 in [41] claims that in this case the following estimate is valid for all $Q \in \mathcal{P}_d(\Xi)$ with $\|F\|_{Q,2} > 0$ and for $\varepsilon \in (0, 1)$:

$$N\left(\varepsilon \|F\|_{Q,2}, \mathcal{H}_\mu, L_2(Q)\right) \leq c_1 \varepsilon^{-c_2}$$

with some constants $c_1, c_2 \geq 0$ depending on $V(\mathcal{H}_\mu)$ only. Note that the right-hand side does not depend on Q . Thus

$$\begin{aligned} &\sup \left\{ \sqrt{\log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{H}_\mu, L_2(Q)\right)} : Q \in \mathcal{P}_d(\Xi), 0 < QF^2 < \infty \right\} \\ &\leq \sqrt{\log c_1 \varepsilon^{-c_2}} = \sqrt{\log c_1 + c_2 \log \varepsilon^{-1}} \leq \sqrt{\log c_1} + \sqrt{c_2 \log \varepsilon^{-1}} \leq \sqrt{\log c_1} + \sqrt{c_2 \varepsilon^{-1}} \end{aligned}$$

Since the last term is integrable for $\varepsilon \in (0, 1)$ the uniform entropy condition is verified and \mathcal{H}_μ is shown to be P -Donsker. Because $\mathcal{G}_j \subset \mathcal{H}_\mu$ for $j = 1, \dots, \nu$ each \mathcal{G}_j is P -Donsker since subsets of P -Donsker classes are again P -Donsker ([41, Theorem 2.10.1]).

e) The Donsker property for \mathcal{F}_j and \mathcal{G}_j implies that \mathcal{F} is P -Donsker:

From Theorems 2.10.6 and Examples 2.10.7 and 2.10.8 in [41] it follows that the class

$$\sum_{j=1}^\nu \mathcal{F}_j \mathcal{G}_j = \left\{ \sum_{j=1}^\nu f_j g_j : f_j \in \mathcal{F}_j, g_j \in \mathcal{G}_j (j = 1, \dots, \nu) \right\}$$

is P -Donsker since both, \mathcal{F}_j and \mathcal{G}_j , are uniformly bounded. Furthermore, because $\mathcal{F} \subset \sum_{j=1}^\nu \mathcal{F}_j \mathcal{G}_j$, the proof is complete since every subset of a P -Donsker-class is P -Donsker as well ([41, Theorem 2.10.1]). \square

4. Delta Method In order to get a convergence statement for the optimal value of (1) in \mathbb{R} , i.e., weak convergence of $\sqrt{n}(v(P_n) - v(P))$, we want to apply the delta method described, e.g., in Theorem 1 and Proposition 1 of [29]. For clarity, we cite these results here adapted to our framework.

⁴A set of functions \mathcal{F} is called a VC class (*Vapnik-Cervonenkis-class*) if the corresponding set of subgraphs $\text{sub}\mathcal{F} := \{\text{sub}f : f \in \mathcal{F}\}$ is a VC class of subsets of $\Xi \times \mathbb{R}$ with $\text{sub}f = \{(x, t) \in \Xi \times \mathbb{R} : t < f(x)\}$. A set \mathcal{C} of subsets of some set M is called VC class if its VC-index $V(\mathcal{C})$ is finite, i.e., $V(\mathcal{C}) < \infty$ with

$$V(\mathcal{C}) = \inf \{n \in \mathbb{N} \mid \forall D \subset M \text{ with } \#D = n \exists A_D \subset D \forall C \in \mathcal{C} : A_D \neq D \cap C\}$$

See [41] Chapter 2.6 for further details.

DEFINITION 4.1 *Let D and F be linear metric spaces. Then $\Phi : D \rightarrow F$ is called Hadamard directionally differentiable at $\vartheta_0 \in D$ if there exists a mapping $\Phi'_{\vartheta_0} : D \rightarrow F$ with*

$$\lim_{n \rightarrow \infty} \frac{\Phi(\vartheta_0 + t_n h_n) - \Phi(\vartheta_0)}{t_n} = \Phi'_{\vartheta_0}(h)$$

for all $h \in D$ and all suitable sequences $(h_n) \in D^{\mathbb{N}}$ and $(t_n) \in \mathbb{R}^{\mathbb{N}}$ such that $t_n \downarrow 0$ and $h_n \rightarrow h$.

The Hadamard directional derivative Φ'_{ϑ_0} is continuous and positively homogenous. But, note that linearity of Φ'_{ϑ_0} is not required here. By admitting a directional version of the concept of Hadamard differentiability we follow [35] and [30, Chapter 6] and deviate from mainstream literature (see, e.g., [40, 41]). We do so because for Φ we have the infimal value mapping in mind. It will be shown below that it is Hadamard directionally differentiable in our sense with nonlinear derivative. Moreover, linearity is not required for the delta method, too.

THEOREM 4.1 *Let D and F be linear metric spaces, $\Phi : D \rightarrow F$ Hadamard directionally differentiable at $\vartheta_0 \in D$. Let further $Z, \vartheta_n \in D^{\Omega}$ and $\sqrt{n}(\vartheta_n - \vartheta_0) \rightsquigarrow Z$. Then we have*

$$\sqrt{n}(\Phi(\vartheta_n) - \Phi(\vartheta_0)) \rightsquigarrow \Phi'_{\vartheta_0}(Z).$$

PROOF. We refer to Theorem 1 in [29], (set $r_n = \sqrt{n}$ and $\Theta = D_{\Phi} = D$, thus $T_{\Theta}(\vartheta_0) = D$). \square

The second result provides the Hadamard directional differentiability of the infimal value mapping. Its first part is due to Lachout [17].

PROPOSITION 4.1 *Set $D := \ell^{\infty}(X)$ and $F := \mathbb{R}$ and define the infimal value mapping*

$$\begin{aligned} \Psi : \ell^{\infty}(X) &\rightarrow \mathbb{R} \\ \vartheta &\mapsto \inf \{\vartheta(x) : x \in X\} \end{aligned}$$

and the ε -solution set $S(\vartheta, \varepsilon) := \{x \in X \mid \vartheta(x) \leq \Psi(\vartheta) + \varepsilon\}$ for $\varepsilon \geq 0$.

Then Ψ is Hadamard directionally differentiable in every $\vartheta_0 \in D$ with

$$\begin{aligned} \Psi'_{\vartheta_0} : D &\rightarrow \mathbb{R} \\ h &\mapsto \lim_{\varepsilon \downarrow 0} \inf \{h(x) : x \in S(\vartheta_0, \varepsilon)\} \end{aligned} \quad (6)$$

Moreover, if $\vartheta_0 \in \ell^{\infty}(X)$ is lower semicontinuous and $h \in \ell^{\infty}(X)$ is continuous, then it holds that

$$\Psi'_{\vartheta_0}(h) = \min \{h(x) : x \in S(\vartheta_0, 0)\} \quad (7)$$

PROOF. Proposition 1 in [29] (again, set $\Theta = D_{\Psi} = D = T_{\Theta}(\vartheta_0)$) records the proof of Hadamard directional differentiability and formula (6) from [17] even if X is an arbitrary set. (Remember that in this paper here X was assumed to be a compact subset of \mathbb{R}^m .) Thus, it remains to show (7):

Let $\vartheta_0 \in \ell^{\infty}(X)$ be lower semicontinuous and $h \in \ell^{\infty}(X)$ continuous. Of course, representation (6) holds. For $n \in \mathbb{N}$ choose $x_n \in S(\vartheta_0, \frac{1}{n})$ such that $h(x_n) \leq \inf \{h(x) : x \in S(\vartheta_0, \frac{1}{n})\} + \frac{1}{n}$. Then

$$\inf \{h(x) : x \in S(\vartheta_0, \frac{1}{n})\} \leq h(x_n) \leq \inf \{h(x) : x \in S(\vartheta_0, \frac{1}{n})\} + \frac{1}{n},$$

thus $\Psi'_{\vartheta_0}(h) = \lim_{n \rightarrow \infty} h(x_n)$ since $\inf \{h(x) : x \in S(\vartheta_0, \frac{1}{n})\} \rightarrow \Psi'_{\vartheta_0}(h)$. Because $S(\vartheta_0, \frac{1}{n+1}) \subset S(\vartheta_0, \frac{1}{n}) \subset X$ and X is compact there exists a subsequence $x_{n'}$ converging to some $x_0 \in X$ in \mathbb{R}^m . And because $\vartheta_0(x_{n'}) \leq \Psi(\vartheta_0) + \frac{1}{n'}$ and ϑ_0 is lower semicontinuous it holds that

$$\vartheta_0(x_0) \leq \liminf_{n' \rightarrow \infty} \vartheta_0(x_{n'}) \leq \Psi(\vartheta_0),$$

hence $x_0 \in S(\vartheta_0, 0)$, thus on the one hand

$$\Psi'_{\vartheta_0}(h) = \lim_{n' \rightarrow \infty} h(x_{n'}) = h(x_0) \geq \min \{h(x) : x \in S(\vartheta_0, 0)\}$$

and

$$\begin{aligned} \Psi'_{\vartheta_0}(h) = \lim_{n' \rightarrow \infty} h(x_{n'}) &\leq \lim_{n' \rightarrow \infty} \left(\inf \{h(x) : x \in S(\vartheta_0, \frac{1}{n'})\} + \frac{1}{n'} \right) \\ &\leq \lim_{n' \rightarrow \infty} \left(\inf \{h(x) : x \in S(\vartheta_0, 0)\} + \frac{1}{n'} \right) \\ &= \min \{h(x) : x \in S(\vartheta_0, 0)\} \end{aligned}$$

on the other hand. \square

At first glance this framework seems not to fit for our purpose, since we have mappings Q on $\ell^\infty(\mathcal{F})$ rather than on $\ell^\infty(X)$. But if we define for $Q \in \mathcal{P}(\Xi)$

$$\vartheta_Q(x) := Qf_x = c'x + \int_{\Xi} \phi(h(\xi) - T(\xi)x) dQ(\xi)$$

we have $\vartheta_Q \in \ell^\infty(X)$ and $v(Q) = \Psi(\vartheta_Q)$ for all $Q \in \mathcal{P}(\Xi)$. The convergence $\sqrt{n}(P_n - P) \rightsquigarrow G_P$ in $\ell^\infty(\mathcal{F})$ means

$$\sqrt{n}(\vartheta_{P_n} - \vartheta_P) \rightsquigarrow \vartheta_{G_P} \sim \mathcal{N}\left(0, (Pf_x f_y - Pf_x P f_y)_{x,y \in X}\right)$$

in $\ell^\infty(X)$ with $\vartheta_{G_P} = G_P f. \in \ell^\infty(X)^\Omega$. This leads to

COROLLARY 4.1 *For the optimal value of the stochastic program (1) it holds that*

$$\sqrt{n}(v(P_n) - v(P)) = \sqrt{n}(\Psi(\vartheta_{P_n}) - \Psi(\vartheta_P)) \rightsquigarrow \Psi'_{\vartheta_P}(\vartheta_{G_P}).$$

If we knew quantiles of the distribution of $\Psi'_{\vartheta_P}(\vartheta_{G_P})$ we could give asymptotic confidence intervals for the optimal value $v(P) = \Psi(\vartheta_P)$ since $v(P_n)$ can be calculated by solving a finite mixed-integer linear program. In general, it seems too difficult to calculate the distribution analytically since Ψ'_{ϑ_P} from (6) has a rather complicated shape. Thus empirical methods are needed.

REMARK 4.1 *Only in special cases the simpler formula (7) can be applied. The condition that for $Q \in \mathcal{P}(\Xi)$ the elements ϑ_Q are lower semicontinuous on X is always satisfied due to the lower semicontinuity of ϕ (see [28] Lemma 33) together with Fatou's Lemma:*

$$\begin{aligned} \liminf Qf_{x_n} &= c'x_0 + \liminf \int_{\Xi} \phi(h(\xi) - T(\xi)x_n) dQ(\xi) \\ &\geq c'x_0 + \int_{\Xi} \liminf \phi(h(\xi) - T(\xi)x_n) dQ(\xi) \\ &\geq c'x_0 + \int_{\Xi} \phi(h(\xi) - T(\xi)x_0) dQ(\xi) = Qf_{x_0} \end{aligned}$$

for $x_n \rightarrow x_0$ in X . However, to apply (7) it would have to be shown in addition, that the P -Brownian Bridge G_P (and accordingly ϑ_{G_P}) has continuous sample paths. Indeed, there is a continuity property for ϑ_{G_P} because G_P is tight (see Example 1.5.10 in [41]): For almost all $\omega \in \Omega$ it holds that $\vartheta_{G_P}(\omega) \in \ell^\infty(X)$ is continuous with respect to the semi-metric given by

$$\rho(x_0, x_1) := \left(P(f_{x_0} - f_{x_1})^2 - (P(f_{x_0} - f_{x_1}))^2 \right)^{1/2}$$

But, in general, $x_n \rightarrow x_0$ in $X \subset \mathbb{R}^m$ does not imply continuity with respect to ρ , hence $\vartheta_{G_P}(\omega)$ is not necessarily continuous.

The special case that the second stage problem contains no integrality (i.e., $\hat{m} = 0$) would be an example where $x_n \rightarrow x_0$ in X implies $\rho(x_n, x_0) \rightarrow 0$ since in this case ϕ is continuous (see [42]). Another example would be the case where X consists of isolated points only. For such examples it holds indeed

$$\Psi'_{\vartheta_P}(\vartheta_{G_P}) = \inf \{G_P f \mid f \in \mathcal{F}, Pf = v(P)\}.$$

If it is known in addition that the solution set $S(P) := \{x \in X : Pf_x = v(P)\}$ of the stochastic program (1) is a singleton, i.e., $\#S(P) = 1$, $S(P) = \{x^*\}$, then we get $\Psi'_{\vartheta_P}(\vartheta) = \vartheta(x^*)$, i.e., Ψ'_{ϑ_P} is a linear mapping in this case. Moreover, due to the definition of the P -Brownian Bridge G_P , it holds that

$$\Psi'_{\vartheta_P}(\vartheta_{G_P}) = G_P f_{x^*} \sim \mathcal{N}\left(0, Pf_{x^*}^2 - (Pf_{x^*})^2\right),$$

i.e., we know that the limit is normally distributed with zero mean and unknown variance (since both, x^* and P , are unknown).

Since our goal is to calculate confidence intervals not only in special cases, we do not continue this discussion here and address ourselves to more general methods.

5. Bootstrapping Bootstrapping is a principle to gain information about the quantiles of an unknown limit distribution by resampling ξ_1^*, ξ_2^*, \dots from some empirical distribution P_n . From these resamples the *bootstrap empirical measure* $P_n^* := \frac{1}{n} \sum_{j=1}^n \delta_{\xi_j^*}$ is constructed. For our problem, the unknown distribution is the limit distribution of $\sqrt{n}(v(P_n) - v(P))$.

It will be shown below that under certain conditions it holds that $\sqrt{n}(P_n^* - P_n)$ converges in some sense to the same limit as $\sqrt{n}(P_n - P)$. The mathematical backbone of this method is the independence of the sampling and the resampling procedure. The convergence of $\sqrt{n}(P_n^* - P_n)$ can be carried over to convergence statements about $\sqrt{n}(v(P_n^*) - v(P_n))$ in several ways. However, a delta method statement like Theorem 4.1 can only be given for the case that Φ'_{ϑ} is linear. For the general case, an alternative method is suggested in Section 5.2.

The bootstrap method was introduced in [6]. Here, we will make use of the consistency results as well as the delta method for the bootstrap derived in [40] and [41]. For further discussion and extensions of the bootstrap method see, e.g., [19, 11, 9]. Note that the extensions there are different from the extension that are developed in Section 5.2.

5.1 Classical Bootstrap The classical bootstrap method rests upon a statement about convergence of the bootstrap empirical measure in $\ell^\infty(\mathcal{F})$ where the samples ξ_1, ξ_2, \dots are considered as fixed. The type of convergence is “conditionally to ξ_1, ξ_2, \dots in distribution”, which will be defined below following [40] and [41]. To motivate this definition we first define for a normed space D , e.g., $D = \ell^\infty(\mathcal{F})$, the set of bounded Lipschitz functions

$$BL_\gamma(D) := \{h \in [-1, 1]^D : |h(z_1) - h(z_2)| \leq \gamma \|z_1 - z_2\| \ \forall z_1, z_2 \in D\}$$

and we note that for $D = \ell^\infty(\mathcal{F})$ weak convergence can be characterized by

$$Z_n \rightsquigarrow Z_0 \Leftrightarrow \sup_{h \in BL_1(\ell^\infty(\mathcal{F}))} |E[h(Z_n) - h(Z_0)]| \rightarrow 0,$$

if $Z_n \in \ell^\infty(\mathcal{F})^\Omega$ and $Z_0 \in \ell^\infty(\mathcal{F})^\Omega$ are measurable and tight (see [10], [40, Chapter 23]). In accordance with [40] and [41] we define that Z_n converges to Z_0 *conditionally to ξ_1, ξ_2, \dots in distribution* if

$$Z_n \rightsquigarrow_* Z_0 \Leftrightarrow \sup_{h \in BL_1(\ell^\infty(\mathcal{F}))} |E[h(Z_n) - h(Z_0) \mid \xi_1, \xi_2, \dots]| \xrightarrow{P} 0,$$

where $E[\cdot \mid \cdot]$ and \xrightarrow{P} denote the conditional expectation and convergence in probability, respectively. With these notations we are ready to cite two results from [40].

THEOREM 5.1 *If \mathcal{F} is P -Donsker then $\sqrt{n}(P_n^* - P_n) \rightsquigarrow_* G_P^*$ in $\ell^\infty(\mathcal{F})$. The limit G_P^* is a P -Brownian Bridge, thus, it has the same distribution as the limit G_P in Theorem 3.1.*

PROOF. Theorem 23.7 in [40]. □

At this point one would expect a delta method theorem similar to Theorem 4.1 but for the bootstrap case. However, for such a statement we need additionally that Φ'_{ϑ_0} is linear.

PROPOSITION 5.1 *Let D be a normed space, $\vartheta_0 \in D$ and let $\Phi : D \rightarrow \mathbb{R}$ be Hadamard directionally differentiable at ϑ_0 with derivative Φ'_{ϑ_0} being linear. Let further $\vartheta_n \in D^\Omega$ and $\vartheta_n^* \in D^\Omega$ and $Z \in D^\Omega$ and $\sqrt{n}(\vartheta_n^* - \vartheta_n) \rightsquigarrow_* Z$ and $\sqrt{n}(\vartheta_n - \vartheta_0) \rightsquigarrow Z$. Then:*

$$\sqrt{n}(\Phi(\vartheta_n^*) - \Phi(\vartheta_n)) \rightsquigarrow_* \Phi'_{\vartheta_0}(Z).$$

PROOF. Theorem 23.9 in [40] (set $D_\Phi = D$). □

Applied to our problem this means

COROLLARY 5.1 *If Ψ'_{ϑ_P} is linear, then it holds that*

$$\sqrt{n}(v(P_n^*) - v(P_n)) = \sqrt{n}(\Psi(\vartheta_{P_n^*}) - \Psi(\vartheta_{P_n})) \rightsquigarrow_* \Psi'_{\vartheta_P}(\vartheta_{G_P^*})$$

in \mathbb{R} with G_P^ being a P -Brownian Bridge in $\ell^\infty(\mathcal{F})$.*

The limit $\Psi'_{\vartheta_P}(\vartheta_{G_P^*})$ is the same as in Corollary 4.1. This fact can be used to approximate the distribution $\Psi'_{\vartheta_P}(\vartheta_{G_P})$ and to derive confidence intervals for the (unknown) value $v(P) = \min\{Pf : f \in \mathcal{F}\}$, i.e., the optimal value of the stochastic program (1).

Given $\xi_1, \xi_2, \dots, \xi_n$, i.e. given P_n , with n fixed sufficiently large, the distribution $\Psi'_{\vartheta_P}(\vartheta_{G_P})$ can be approximated by some empirical distribution of $\sqrt{n}(v(P_n^*) - v(P_n))$ gained from sufficiently many resampled n -tuples $\xi_1^*, \xi_2^*, \dots, \xi_n^*$ from P_n . This means: if $\zeta_{\alpha, m}^*$ is a lower α -quantile of an empirical distribution of $\sqrt{n}(v(P_n^*) - v(P_n))$ gained from m (sufficiently large) resamples then for $\alpha_1 < 50\% > \alpha_2$ the interval

$$\left[v(P_n) - \frac{1}{\sqrt{n}} \zeta_{1-\alpha_1, m}^*, v(P_n) - \frac{1}{\sqrt{n}} \zeta_{\alpha_2, m}^* \right] \quad (8)$$

is an asymptotic confidence interval⁵ at level $\alpha_1 + \alpha_2$ for the optimal value $v(P)$, i.e.,

$$\liminf_{n, m \rightarrow \infty} \mathbb{P} \left(v(P) \in \left[v(P_n) - \frac{1}{\sqrt{n}} \zeta_{1-\alpha_1, m}^*, v(P_n) - \frac{1}{\sqrt{n}} \zeta_{\alpha_2, m}^* \right] \right) \geq 1 - \alpha_1 - \alpha_2.$$

5.2 Extended Bootstrap As seen in the previous sections, the classical empirical delta method for bootstrapping works only if the Hadamard directional derivative of Φ at ϑ_0 is linear. As discussed in Remark 4.1, for the infimal value mapping Ψ this is only the case under strong additional assumptions. The question arises, whether there's another method to derive confidence intervals that works without this assumption of linearity. The answer is yes, but, of course, this is more involved and more expensive in terms of computation, too.

First of all, we cite another result from [40] that will be needed below.

PROPOSITION 5.2 *In $\ell^\infty(\mathcal{F})$ it holds that*

$$(\sqrt{n}(P_n - P), \sqrt{n}(P_n^* - P_n)) \rightsquigarrow (G_P, G_P^*)$$

with G_P and G_P^* being independent P -Brownian Bridges.

PROOF. See Proof of Theorem 23.9 in [40]. □

Note that $G_P^* \sim G_P$. Note further, that this is a convergence statement about ordinary weak convergence, i.e., unconditional. Of course, in $\ell^\infty(X)$ this means

$$(\sqrt{n}(\vartheta_{P_n} - \vartheta_P), \sqrt{n}(\vartheta_{P_n^*} - \vartheta_{P_n})) \rightsquigarrow (\vartheta_{G_P}, \vartheta_{G_P^*})$$

Next, we establish a kind of alternative delta method suitable for this framework.

LEMMA 5.1 *Let $\Phi : D \rightarrow F$ be Hadamard directionally differentiable in $\vartheta_0 \in D$ and let $\vartheta_n^*, \vartheta_n \in D^\Omega$ ($n \in \mathbb{N}$) be given satisfying $(\sqrt{n}(\vartheta_n - \vartheta_0), \sqrt{n}(\vartheta_n^* - \vartheta_n)) \rightsquigarrow (Z, Z^*)$ with $Z, Z^* \in D^\Omega$. Then it holds that*

$$\sqrt{n}(\Phi(\vartheta_n^*) - \Phi(\vartheta_n)) \rightsquigarrow \Phi'_{\vartheta_0}(Z^* + Z) - \Phi'_{\vartheta_0}(Z)$$

PROOF. We define mappings

$$\begin{aligned} g_n : D \times D &\rightarrow F \\ (h^*, h) &\mapsto \sqrt{n} \left(\Phi \left(\vartheta_0 + \frac{1}{\sqrt{n}}(h^* + h) \right) - \Phi \left(\vartheta_0 + \frac{1}{\sqrt{n}}h \right) \right) \end{aligned}$$

then due to the Hadamard directional differentiability of Φ it holds that

$$\begin{aligned} &g_n(h_n^*, h_n) \\ &= \sqrt{n} \left(\Phi \left(\vartheta_0 + \frac{1}{\sqrt{n}}(h_n^* + h_n) \right) - \Phi(\vartheta_0) \right) - \sqrt{n} \left(\Phi \left(\vartheta_0 + \frac{1}{\sqrt{n}}h_n \right) - \Phi(\vartheta_0) \right) \\ &\rightarrow \Phi'_{\vartheta_0}(h_0^* + h_0) - \Phi'_{\vartheta_0}(h_0) \end{aligned}$$

for $(h_n^*, h_n) \rightarrow (h_0^*, h_0)$. Hence, the continuous mapping theorem [40, Theorem 18.11] applies and we get

$$g_n(\sqrt{n}(\vartheta_n^* - \vartheta_n), \sqrt{n}(\vartheta_n - \vartheta_0)) = \sqrt{n}(\Phi(\vartheta_n^*) - \Phi(\vartheta_n)) \rightsquigarrow \Phi'_{\vartheta_0}(Z^* + Z) - \Phi'_{\vartheta_0}(Z) \quad \square$$

Note that the latter result does not require linearity of the Hadamard derivative. Putting the previous two results together leads to

⁵Note that $\zeta_{1-\alpha_1, m}^* \geq \zeta_{\alpha_2, m}^*$ since $1 - \alpha_1 > 50\% > \alpha_2$, so we have indeed that $\left[v(P_n) - \frac{1}{\sqrt{n}} \zeta_{1-\alpha_1, m}^*, v(P_n) - \frac{1}{\sqrt{n}} \zeta_{\alpha_2, m}^* \right]$ is an interval.

COROLLARY 5.2 *In \mathbb{R} it holds that*

$$\sqrt{n}(v(P_n^*) - v(P_n)) = \sqrt{n}(\Psi(\vartheta_{P_n^*}) - \Psi(\vartheta_{P_n})) \rightsquigarrow \Psi'_{\vartheta_P}(\vartheta_{G_P^*} + \vartheta_{G_P}) - \Psi'_{\vartheta_P}(\vartheta_{G_P})$$

with G_P and G_P^* being independent P -Brownian Bridges in $\ell^\infty(\mathcal{F})$.

This result also shows that, if Ψ'_{ϑ_P} is not linear, one cannot expect that the sequence $\sqrt{n}(\Psi(\vartheta_{P_n^*}) - \Psi(\vartheta_{P_n}))$ converges conditionally to ξ_1, ξ_2, \dots in distribution to $\Psi'_{\vartheta_P}(\vartheta_{G_P})$ or $\Psi'_{\vartheta_P}(\vartheta_{G_P^*})$. However, it is possible to define another sequence containing the unknown value $\Psi(\vartheta_P)$ that converges to the same limit $\Psi'_{\vartheta_P}(\vartheta_{G_P^*} + \vartheta_{G_P}) - \Psi'_{\vartheta_P}(\vartheta_{G_P})$. This idea is developed in the following modified version of Theorem 4.1.

LEMMA 5.2 *Let $\Phi : D \rightarrow F$ be Hadamard directionally differentiable at $\vartheta_0 \in D$ and let $\tilde{\vartheta}_n, \bar{\vartheta}_n \in D^\Omega$ ($n \in \mathbb{N}$) be given satisfying $(\sqrt{n}(\bar{\vartheta}_n - \vartheta_0), \sqrt{n}(\tilde{\vartheta}_n - \vartheta_0)) \rightsquigarrow (\tilde{Z}, \bar{Z})$ with $\tilde{Z}, \bar{Z} \in D^\Omega$. Then it holds that*

$$\sqrt{n}\left(2\Phi\left(\frac{1}{2}(\bar{\vartheta}_n + \tilde{\vartheta}_n)\right) - \Phi(\tilde{\vartheta}_n) - \Phi(\vartheta_0)\right) \rightsquigarrow \Phi'_{\vartheta_0}(\tilde{Z} + \bar{Z}) - \Phi'_{\vartheta_0}(\tilde{Z})$$

PROOF. Again, we define mappings

$$g_n : D \times D \rightarrow F \\ (\bar{h}, \tilde{h}) \mapsto \sqrt{n}\left(2\Phi\left(\vartheta_0 + \frac{1}{2\sqrt{n}}(\bar{h} + \tilde{h})\right) - \Phi\left(\vartheta_0 + \frac{1}{\sqrt{n}}\tilde{h}\right) - \Phi(\vartheta_0)\right)$$

then due to the Hadamard directional differentiability of Φ it holds that

$$g_n(\bar{h}_n, \tilde{h}_n) \\ = 2\sqrt{n}\left(\Phi\left(\vartheta_0 + \frac{1}{2\sqrt{n}}(\bar{h}_n + \tilde{h}_n)\right) - \Phi(\vartheta_0)\right) - \sqrt{n}\left(\Phi\left(\vartheta_0 + \frac{1}{\sqrt{n}}\tilde{h}_n\right) - \Phi(\vartheta_0)\right) \\ \rightarrow \Phi'_{\vartheta_0}(\bar{h}_0 + \tilde{h}_0) - \Phi'_{\vartheta_0}(\tilde{h}_0)$$

for $(\bar{h}_n, \tilde{h}_n) \rightarrow (\bar{h}_0, \tilde{h}_0)$. Hence, the continuous mapping theorem [40, Theorem 18.11] applies again and we obtain

$$g_n\left(\sqrt{n}(\bar{\vartheta}_n - \vartheta_0), \sqrt{n}(\tilde{\vartheta}_n - \vartheta_0)\right) = \sqrt{n}\left(2\Phi\left(\frac{1}{2}(\bar{\vartheta}_n + \tilde{\vartheta}_n)\right) - \Phi(\tilde{\vartheta}_n) - \Phi(\vartheta_0)\right) \\ \rightsquigarrow \Phi'_{\vartheta_0}(\tilde{Z} + \bar{Z}) - \Phi'_{\vartheta_0}(\tilde{Z}).$$

□

If we sample twice from P independently, i.e., given $\tilde{\xi}_1, \bar{\xi}_1, \tilde{\xi}_2, \bar{\xi}_2, \dots \stackrel{iid}{\sim} P$, then, of course, with $\tilde{P}_n := \frac{1}{n} \sum_{j=1}^n \delta_{\tilde{\xi}_j}$ and $\bar{P}_n := \frac{1}{n} \sum_{j=1}^n \delta_{\bar{\xi}_j}$, it holds that

$$\sqrt{n}(\vartheta_{\tilde{P}_n} - \vartheta_P) \rightsquigarrow \vartheta_{\tilde{G}_P} \quad \text{and} \quad \sqrt{n}(\vartheta_{\bar{P}_n} - \vartheta_P) \rightsquigarrow \vartheta_{\bar{G}_P}$$

in $\ell^\infty(X)$ with two independent P -Brownian Bridges \tilde{G}_P and \bar{G}_P and $\vartheta_{\tilde{P}_n} \in \ell^\infty(X)^\Omega$ defined by $\vartheta_{\tilde{P}_n}(x) = \sum_{j=1}^n f_x(\tilde{\xi}_j)$ and $\vartheta_{\bar{P}_n}$ analogously. Thus,

$$\sqrt{n}\left(2\Psi\left(\frac{1}{2}(\vartheta_{\tilde{P}_n} + \vartheta_{\bar{P}_n})\right) - \Psi(\vartheta_{\tilde{P}_n}) - \Psi(\vartheta_P)\right) \rightsquigarrow \Psi'_{\vartheta_P}(\vartheta_{\tilde{G}_P} + \vartheta_{\bar{G}_P}) - \Psi'_{\vartheta_P}(\vartheta_{\tilde{G}_P})$$

Because both pairs, \tilde{G}_P and \bar{G}_P as well as G_P^* and G_P , are independent, it holds that $(\tilde{G}_P, \bar{G}_P) \sim (G_P^*, G_P)$, hence

$$\Psi'_{\vartheta_P}(\vartheta_{\tilde{G}_P} + \vartheta_{\bar{G}_P}) - \Psi'_{\vartheta_P}(\vartheta_{\tilde{G}_P}) \sim \Psi'_{\vartheta_P}(\vartheta_{G_P^*} + \vartheta_{G_P}) - \Psi'_{\vartheta_P}(\vartheta_{G_P}).$$

Since we can approximate the distribution of $\Psi'_{\vartheta_P}(G_P^* + G_P) - \Psi'_{\vartheta_P}(G_P)$ by sampling and resampling without knowing P or $\Psi(\vartheta_P)$, we can construct confidence intervals in a similar way as above: Let ζ_α^* be a lower α -quantile of (an approximation of) $\Psi'_{\vartheta_P}(\vartheta_{G_P^*} + \vartheta_{G_P}) - \Psi'_{\vartheta_P}(\vartheta_{G_P})$, then for $\alpha_1 < 50\% > \alpha_2$ it holds that

$$\left[2\Psi\left(\frac{1}{2}(\vartheta_{\tilde{P}_n} + \vartheta_{\bar{P}_n})\right) - \Psi(\vartheta_{\tilde{P}_n}) - \frac{1}{\sqrt{n}}\zeta_{1-\alpha_1}^*, 2\Psi\left(\frac{1}{2}(\vartheta_{\tilde{P}_n} + \vartheta_{\bar{P}_n})\right) - \Psi(\vartheta_{\tilde{P}_n}) - \frac{1}{\sqrt{n}}\zeta_{\alpha_2}^*\right] \quad (9)$$

is an approximate confidence interval at level $\alpha_1 + \alpha_2$ for the optimal value $\Psi(\vartheta_P) = v(P)$.

REMARK 5.1 *Here, in contrast to classical bootstrapping, the samples $\xi_1, \xi_2, \dots, \xi_n$ are at no time fixed. To get one sample point for the empirical distribution function of the approximation of $\Psi'_{\vartheta_P}(\vartheta_{G_P^*} + \vartheta_{G_P}) - \Psi'_{\vartheta_P}(\vartheta_{G_P})$ one has to sample both, $\xi_1, \xi_2, \dots, \xi_n \stackrel{iid}{\sim} P$ and $\xi_1^*, \xi_2^*, \dots, \xi_n^* \stackrel{iid}{\sim} P_n(\xi_1, \xi_2, \dots, \xi_n)$. Hence, the computational effort is twice as high as for the classical bootstrap method. Moreover, in practice, sampling from P might be much more expensive than (re)sampling from the empirical distribution P_n .*

6. Examples To demonstrate the significance of the results derived above, we provide some numerical evidence.

6.1 Problem (unique solution) We consider the example in [33, Section 7]:

$$\min \left\{ x' \begin{pmatrix} -1.5 \\ -4 \end{pmatrix} + \int_{\Xi} \phi(\xi - x) dP(\xi) : x \in \{0, 1, 2, 3, 4, 5\}^2 \right\} \quad (10)$$

with

$$\phi(t) := \min \left\{ y' \begin{pmatrix} -16 \\ -19 \\ -23 \\ -28 \end{pmatrix} : y \in \{0, 1\}^4, \begin{pmatrix} 2 & 3 & 4 & 5 \\ 6 & 1 & 3 & 2 \end{pmatrix} y \leq t \right\}$$

and ξ being uniformly distributed on the two dimensional integer grid between 5 and 15

$$\Xi' := \left\{ \begin{pmatrix} 5 \\ 5 \end{pmatrix}, \begin{pmatrix} 5 \\ 6 \end{pmatrix}, \dots, \begin{pmatrix} 5 \\ 15 \end{pmatrix}, \begin{pmatrix} 6 \\ 5 \end{pmatrix}, \dots, \begin{pmatrix} 15 \\ 15 \end{pmatrix} \right\},$$

i.e., $P(\{\eta\}) = 1/121$ for $\eta \in \Xi'$. This example has the form of (1) with $k = m = r = 2$, $h(\xi) = \xi$, $T(\xi) = I_2$, $\hat{m} = 4$, $\bar{m} = 0$ and $\Xi = \text{conv}(\Xi')$. The exact solution is $x = (0, 4)$ with optimal value $v(P) = -62.29$ (see [33]). This solution is unique⁶ and X consists of isolated points only, thus the theory derived in Section 5.1 holds here (see Remark 4.1).

6.2 Classical Bootstrapping Suppose that we don't know the distribution P but we are able to sample from it. Further, suppose we know that the solution is unique. The classical bootstrap procedure for deriving confidence intervals for the optimal value works as follows:

- (i) Fix $n \in \mathbb{N}$, sample from P and solve the approximated problem. We used $n = 75$ and got $v(P_n) = -61.2667$
- (ii) Resample from P_n using the *same sample-size* n and solve the new problem. Repeat this m times to get an empirical distribution function of $\sqrt{n}(v(P_n^*) - v(P_n))$ conditional to P_n . We worked with $m = 500$ and obtained

$v(P_n^*)$	$\sqrt{n}(v(P_n^*) - v(P_n))$
-58.64	22.7476
-61.8533	-5.08068
-58.56	23.4404
\vdots	\vdots
-63.9867	-23.5559

- (iii) Calculate the quantiles at level $1 - \alpha_1$ and α_2 (α_j small) of the empirical distribution function of the $\sqrt{n}(v(P_n^*) - v(P_n))$ values. We used $\alpha_1 = \alpha_2 = \alpha/2$ and got

α	$[\zeta_{\alpha/2, m}^*, \zeta_{1-\alpha/2, m}^*]$
10%	[-28.5211, 23.7868]
5%	[-33.0822, 28.7520]
2%	[-39.2021, 35.4493]

- (iv) Convert these quantiles to quantiles for the optimal value $v(P)$ according to formula (8). In our example this leads to

α	confidence interval for $v(P)$
10%	[-64.0133, -57.9733]

⁶Uniqueness is only required for the first stage solution x , so we don't claim that $y(\xi)$ is unique, too. We "proved" the uniqueness of x with CPLEX by calculating the 36 solutions of the problem with x fixed at $(0, 0)$, $(0, 1)$, ... and, indeed, it turned out that -62.29 is only reached for $x = (0, 4)$.

5%	[-64.5867, -57.4467]
2%	[-65.3600, -56.7400]

- (v) This procedure was repeated 200 times in order to approve the level of the confidence intervals empirically. Counting the number of confidence intervals covering the true optimal value -62.29 leads to

α	# covering intervals	ratio	average interval length
10%	180/200	90%	6.01218
5%	188/200	94%	7.1325
2%	197/200	98.5%	8.4506

Of course, enlarging n leads to smaller confidence intervals. Because $\sqrt{n}(v(P_n^*) - v(P_n))$ has approximately the same probability distribution as the fixed random element $\Psi'_{\vartheta_P}(\vartheta_{G_P})$ we can expect a decrease of order $\frac{1}{\sqrt{n}}$ for the size of the confidence intervals of $v(P)$. For $\alpha = 5\%$ we got

sample-size n	average interval length	length* \sqrt{n}
50	8.5852	60.70
75	7.1325	61.77
150	4.8139	58.95
200	4.3117	60.98
300	3.4112	59.08

so indeed the decrease is approximately of order $\frac{1}{\sqrt{n}}$ since length times \sqrt{n} is almost constant.

6.3 Problem (non-unique solution) We changed problem (10) to

$$\min \left\{ x' \begin{pmatrix} -1.5 \\ -3.768595041 \end{pmatrix} + \int_{\Xi} \phi(\xi - x) dP(\xi) : x \in \{0, 1, 2, 3, 4, 5\}^2 \right\} \quad (11)$$

with ϕ , P and Ξ as above. Here, the solution is no longer unique. The optimal value -61.363636 is attained at $x = (0, 3)$ and at $x = (0, 4)$. Thus, classical bootstrapping is not theoretically justified here.

6.4 Extended Bootstrapping We applied the extended bootstrap method developed in section 5.2 to derive confidence intervals for the optimal value of problem (11). The procedure here is slightly different than that in Section 6.2. The main difference is that the approximation of the limit distribution is carried out independently from the estimation of the center of the confidence interval. The procedure works as follows:

- (i) Fix $n \in \mathbb{N}$. We used $n = 75$.
- (ii) Sample from P and solve the approximate problem. We got $v(P_n) = -60.5725$. Resample from P_n using the same sample-size n and solve the resulting problem. We got $v(P_n^*) = -59.2439$.
- (iii) Repeat the previous step (sampling *and* resampling) m times to obtain an empirical distribution function of $\sqrt{n}(v(P_n^*) - v(P_n))$. We chose $m = 500$ and obtained

$v(P_n)$	$v(P_n^*)$	$\sqrt{n}(v(P_n^*) - v(P_n))$
-60.5725	-59.2439	11.506
-62.9277	-63.2391	-2.69685
-57.1905	-57.6172	-3.69504
\vdots	\vdots	\vdots
-65.3144	-65.403	-0.767256

- (iv) Calculate the quantiles at level $1 - \alpha_1$ and α_2 (α_j small) of the empirical distribution function

of the $\sqrt{n}(v(P_n^*) - v(P_n))$ values. We used $\alpha_1 = \alpha_2 = \alpha/2$ and got

α	$[\zeta_{\alpha/2,m}^*, \zeta_{1-\alpha/2,m}^*]$
10%	$[-29.4241, 23.6338]$
5%	$[-35.0455, 29.3868]$
2%	$[-42.5669, 32.6578]$

- (v) Sample independently from P with sample-size n twice to get \tilde{P}_n and \bar{P}_n . Calculate $v(\tilde{P}_n)$ and $v(\frac{1}{2}(\tilde{P}_n + \bar{P}_n))$. We got $2v(\frac{1}{2}(\tilde{P}_n + \bar{P}_n)) - v(\tilde{P}_n) = 63.1277$. Using the quantiles from the previous step formula (9) leads to

α	confidence interval for $v(P)$
10%	$[-65.8567, -59.7301]$
5%	$[-66.5210, -59.0810]$
2%	$[-66.8987, -58.2125]$

- (vi) We repeated the previous step 200 times in order to approve the level of the confidence intervals empirically. We counted the number of confidence intervals covering the true optimal value -61.363636 . We got:

α	# covering intervals	ratio	average interval length
10%	182/200	91%	6.1277
5%	191/200	95,5%	7.4400
2%	195/200	97,5%	8.6862

Note that the quantiles $\zeta_{\alpha/2,m}^*$ and $\zeta_{1-\alpha/2,m}^*$ can remain fixed for $\alpha = 10\%, 5\%, 2\%$, respectively, during this approving procedure.

Of course, enlarging n leads again to a decrease of order $\frac{1}{\sqrt{n}}$ for the size of the confidence intervals for $v(P)$.

6.5 Technical Details These results were produced with ILOG CPLEX 8.0 and the ILOG Concert Technology Interface for C++. We used the GNU C++ compiler gcc version 3.0.4 on a Suse Linux system. As random number generator we took the RANLIBC/StatLib library. In CPLEX, the following accuracy parameters were used: $\text{epOpt} = \text{epGap} = \text{epRHS} = 10^{-6}$. This means that the solutions of the approximate problems may be considered as exact.

References

- [1] S. Ahmed and A. Shapiro, *The sample average approximation method for stochastic programs with integer recourse*, Optimization Online (2002).
- [2] B. Apolloni and F. Pezzella, *Confidence intervals in the solution of stochastic integer linear programming problems*, Ann. Oper. Res. **1** (1984), 67–78.
- [3] Z. Artstein and R. J.-B. Wets, *Consistency of minimizers and the SLLN for stochastic programs*, J. Convex Anal. **2** (1995), 1–17.
- [4] C. E. Blair and R. G. Jeroslow, *The value function of a mixed integer program*, Discrete Math. **19** (1977), 121–138.
- [5] J. Dupačová and R. J.-B. Wets, *Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems*, Ann. Statist. **16** (1988), 1517–1549.
- [6] B. Efron, *Bootstrap methods: Another look at the jackknife*, Ann. Statist. **7** (1979), 1–26.
- [7] Y. M. Ermoliev and V. I. Norikin, *Normalized convergence in stochastic optimization*, Ann. Oper. Res. **30** (1991), 187–198.
- [8] A. Futschik and G. C. Pflug, *Confidence sets for discrete stochastic optimization*, Ann. Oper. Res. **56** (1995), 95–108.
- [9] E. Giné, *Lectures on some aspects of the bootstrap*, Lectures on Probability Theory and Statistics (E. Giné, G.R. Grimmet, L. Saloff-Coste, eds.), Lecture Notes in Mathematics Vol. 1665, Springer, Berlin, 1997, pp. 37–151.

- [10] E. Giné and J. Zinn, *Bootstrapping general empirical measures*, Ann. Probab. **18** (1990), 851–869.
- [11] P. Hall, *The Bootstrap and Edgeworth Expansion*, Springer Series in Statistics, Springer, New York, 1992.
- [12] Y. M. Kaniovski, A. J. King, and R. J.-B. Wets, *Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems*, Ann. Oper. Res. **56** (1995), 189–208.
- [13] A. J. King and R. T. Rockafellar, *Asymptotic theory for solutions in statistical estimation and stochastic programming*, Math. Oper. Res. **18** (1993), 148–162.
- [14] A. J. King and R. J.-B. Wets, *Epi-consistency of convex stochastic programs*, Stoch. Stoch. Rep. **34** (1991), 83–92.
- [15] D. Klatte, *A note on quantitative stability results in nonlinear optimization*, Proceedings 19. Jahrestagung Mathematische Optimierung (K. Lommatzsch, ed.), Humboldt-Universität Berlin, Sektion Mathematik, Seminarbericht Nr. 90, 1987, pp. 77–86.
- [16] A. J. Kleywegt, A. Shapiro, and T. Homem-de-Mello, *The sample average approximation method for stochastic discrete optimization*, SIAM J. Optim. **12** (2001), 479–502.
- [17] P. Lachout, Personal communication, 2004.
- [18] F. Louveaux and R. Schultz, *Stochastic Integer Programming*, Chapter 4 in [31], 2003, pp. 213–266.
- [19] E. Mammen, *When Does Bootstrap Work? Asymptotic Results and Simulations*, Lecture Notes in Statistics Vol. 77, Springer, New York, 1992.
- [20] H. Morita, H. Ishii, and T. Nishida, *Confidence region method for a stochastic programming problem*, J. Oper. Res. Soc. Japan **30**, 218–231 (1987).
- [21] V. I. Norikin, *Convergence of the empirical mean method in statistics and stochastic programming*, Cybernet. Systems Anal. **28** (1992), 253–264.
- [22] G. C. Pflug, *Asymptotic stochastic programs*, Math. Oper. Res. **20** (1995), 769–789.
- [23] G. C. Pflug, *Stochastic programs and statistical data*, Ann. Oper. Res. **85** (1999), 59–78.
- [24] G. C. Pflug, A. Ruszczyński, and R. Schultz, *On the Glivenko-Cantelli problem in stochastic programming: Mixed-integer linear recourse*, Math. Methods Oper. Res. **47** (1998), 39–49.
- [25] S. T. Rachev, *Probability Metrics and the Stability of Stochastic Models*, Wiley, Chichester, 1991.
- [26] S. T. Rachev and W. Römisch, *Quantitative stability in stochastic programming: The method of probability metrics*, Math. Oper. Res. **27** (2002), 792–818.
- [27] S. M. Robinson, *Local epi-continuity and local optimization*, Math. Program. **37** (1987), 208–223.
- [28] W. Römisch, *Stability of Stochastic Programming Problems*, Chapter 8 in [31], 2003, pp. 483–554.
- [29] W. Römisch, *Delta method, infinite dimensional*, extended entry, Encyclopedia of Statistical Sciences, Second Edition, Wiley, 2004 (to appear).
- [30] R. Y. Rubinstein and A. Shapiro, *Discrete Event Systems, Sensitivity Analysis and Stochastic Optimization by the Score Function Method*, Wiley, Chichester, 1993.
- [31] A. Ruszczyński and A. Shapiro, eds., *Stochastic Programming*, Handbooks in Operations Research and Management Science Vol. 10, Elsevier, Amsterdam, 2003.
- [32] R. Schultz, *Rates of convergence in stochastic programs with complete integer recourse*, SIAM J. Optim. **6** (1996), 1138–1152.
- [33] R. Schultz, L. Stougie, and M. H. van der Vlerk, *Solving stochastic programs with integer recourse by enumeration: A framework using Gröbner basis reductions*, Math. Program. **83** (1998), 229–252.
- [34] A. Shapiro, *Asymptotic properties of statistical estimators in stochastic programming*, Ann. Statist. **17** (1989), 841–858.
- [35] A. Shapiro, *On concepts of directional differentiability*, J. Optim. Theory Appl. **66** (1990), 477–487.
- [36] A. Shapiro, *Statistical inference of stochastic optimization problems*, Probabilistic Constrained Optimization: Methodology and Applications (S. Uryasev, ed.), Kluwer Academic Publishers, Dordrecht, 2000, pp. 282–304.
- [37] A. Shapiro and T. Homem-de-Mello, *On rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs*, SIAM J. Optim. **11** (2000), 70–86.
- [38] M. Talagrand, *Sharper bounds for Gaussian and empirical processes*, Ann. Probab. **22** (1994), 28–76.
- [39] M. Talagrand, *The Glivenko-Cantelli problem, ten years later*, J. Theoret. Probab. **9** (1996), 371–384.
- [40] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 1998.
- [41] A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes*, Springer Series in Statistics, Springer, New York, 1996.
- [42] R. J.-W. Wets, *Stochastic programs with fixed recourse: The equivalent deterministic program*, SIAM Rev. **16** (1974), 309–339.