MATHEMATICS OF OPERATIONS RESEARCH
Vol. 32, No. 1, February 2007, pp. 118–135
ISSN 0364-765X | EISSN 1526-5471 | 07 | 3201 | 0118

DOI 10.1287/moor.1060.0222

Stochastic Integer Programming: Limit Theorems and Confidence Intervals

Andreas Eichhorn, Werner Römisch
Department of Mathematics, Humboldt-University Berlin, Unter den Linden 6, 10099 Berlin, Germany
{eichhorn@math.hu-berlin.de, http://www.math.hu-berlin.de/~eichhorn,
romisch@math.hu-berlin.de, http://www.math.hu-berlin.de/~romisch}

We consider empirical approximations (sample average 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 resampling methods (bootstrap, subsampling).

Key words: stochastic programming; mixed-integer optimization; stability; sample average approximation; empirical process; Donsker class; delta method; Hadamard directional differentiability; bootstrap; subsampling
MSC2000 subject classification: Primary: 90C15; secondary: 90C11, 60F17, 62F40
OR/MS subject classification: Primary: programming/stochastic; secondary: programming/integer/theory, statistics/estimation
History: Received January 6, 2005; revised August 10, 2005.

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 can not 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 Römisch [35] for a recent survey). Most of these results consist of (Lipschitz) continuity properties of solution and optimal values with respect to certain probability metrics $d(P, Q)$ (Römisch [35], Rachev and Römisch [33], and Rachev [32]). 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 samplesize.

Asymptotic properties of statistical estimators in stochastic programming have been studied intensively. We refer to Ruszczyński and Shapiro [38, Chapters 6, 7, and 8] for various aspects and views. For two-stage stochastic programs without integrality requirements, much is known. For the empirical estimator, the papers of Dupačová and Wets [6], King and Wets [18], and Artstein and Wets [4] contain results on (epi-) consistency, laws of large numbers, and asymptotic normality. In Shapiro [41], Rubinstein and Shapiro [37, Chapter 6], King and Rockafellar [17], Pflug [26], and Shapiro [43], 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 Ermoliev and Norkin [8], Norkin [25], Kaniovski et al. [16], Pflug [27], and Shapiro and Homem-de-Mello [44]. The situation is essentially different for mixed-integer two-stage stochastic programs. In Schultz [39], 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 Pflug et al. [28], and large deviation-type results are derived in Kleywegt et al. [20] and Ahmed and Shapiro [1] for pure integer models and in Rachev and Römisch [33] and Römisch [35] for the mixed-integer case. Much of this work is based on recent developments of empirical process theory, e.g., on Talagrand's work (Talagrand [45, 46]) (see also Giné [12] and the monographs (van der Vaart and Wellner [48] and van der Vaart [47])).

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. Shapiro [43]), 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 Römisch [36] for an introductory overview) and on a recent Hadamard directional differentiability result for infimal value mappings on the space of bounded functions (Lachout [21]). 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 Futschik and Pflug [9], 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 Morita et al. [24], 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 Apolloni and Pezzella [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 §2 we present the framework of our analysis and in §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 (van der Vaart [47] and van der Vaart and Wellner [48]); see Giné [12] for an alternative presentation. In §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 §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 §6.

2. Framework. Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a arbitrary probability space and let $\xi\colon (\Omega, \mathscr{A}) \to (\mathbb{R}^k, \mathscr{B}^k)$ 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

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

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

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

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

(i) relatively complete recourse:

$$\forall (x, \xi) \in X \times \Xi \quad \exists y \in \mathbb{Z}_+^{\widehat{m}},\ \bar{y} \in \mathbb{R}_+^{\overline{m}}\colon h(\xi) - T(\xi)x = Wy + \overline{W}\bar{y}.$$

(ii) dual feasibility:

$$\exists u \in \mathbb{R}^r\colon W'u \le q, \quad \overline{W}'u \le \bar{q}.$$

Under these assumptions, it turns out that ϕ is lower semicontinuous and piecewise polyhedral on $\operatorname{dom}\phi$ (e.g., Louveaux and Schultz [22, Proposition 2] and Römisch [35, Lemma 33]).

We define the infimal value mapping

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

that maps a probability distribution on Ξ to the optimal value of the stochastic program (Equation 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 $\xi_1, \xi_2, \ldots$ 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 polyhedral structure 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 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 $X_{\mathcal{U}} = X \cap \operatorname{cl}\mathcal{U}$ with some open and bounded set $\mathcal{U} \subset \mathbb{R}^m$ that contains the solution set of Equation (1) which is assumed to be nonempty (cf. Robinson [34], Klatte [19], and Römisch [35]). 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. Limit theorem for the objectives. 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 $f_x\colon \Xi \to \mathbb{R}$ as the integrand (objective) of Equation (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:

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

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 $\mathbb{R}$:

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

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

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

or

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

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 $\forall f \in \mathcal{F}\ \forall \xi \in \Xi$: $|f(\xi)| \le K$.*

Proof. Setting $\mathcal{T} := \{Wy + \overline{W}\bar{y}\colon y \in \mathbb{Z}_+^{\hat{m}},\ \bar{y} \in \mathbb{R}_+^{\overline{m}}\}$ we get by Blair and Jeroslow [5, 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

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

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

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

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

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:

$$\ell^\infty(Y) := \left\{\psi \in \mathbb{R}^Y\colon \sup_{y\in Y} |\psi(y)| < \infty\right\}, \qquad \|\psi\|_Y := \sup_{y\in Y} |\psi(y)|.$$

Hence, since for $Q \in \mathcal{P}(\Xi)$ the set $\{Qf\colon f \in \mathcal{F}\}$ 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., van der Vaart and Wellner [48, Chapter 1], van der Vaart [47, 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:

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

Proof. We will utilize properties of mixed-integer two-stage stochastic programs that can be found in Römisch [35] as well as empirical process theory from van der Vaart and Wellner [48]. The proof consists of five parts.

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

Setting $\mathcal{T} := \{Wy + \overline{W}\bar{y}\colon y \in \mathbb{Z}_+^{\hat{m}},\ \bar{y} \in \mathbb{R}_+^{\bar{m}}\} \subset \mathbb{R}^r$ we conclude from Römisch [35, Lemma 33] that there exist $L > 0$, $\tau \in \mathbb{N}$, and $B_j \subset \widehat{\mathcal{B}}_{ph_\tau}(\mathcal{T})$ $(j \in \mathbb{N})$ such that $\mathcal{T} = \bigcup_{j\in\mathbb{N}} B_j$ and $B_i \cap B_j = \varnothing$ for $i \neq j$ and $\phi|_{B_j}$ Lipschitz continuous with uniform Lipschitz constant L. Thereby, we use the notation

$$\widehat{\mathcal{B}}_{ph_\tau}(\mathcal{T}) := \left\{\mathcal{T} \cap \bigcap_{j=1}^{\tau} H_j \mid H_j = \{\xi\colon c_j'\xi \le d_j\} \text{ or } H_j = \{\xi\colon c_j'\xi < d_j\} \text{ with } c_j \in \mathbb{R}^r,\ d_j \in \mathbb{R}\right\}$$

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 $\mathcal{T}' := \{h(\xi) - T(\xi)x\colon \xi \in \Xi,\ x \in X\}$ is bounded and $\mathcal{T}' \subset \mathcal{T}$ due to relatively complete recourse, we know from, e.g., Römisch [35, Lemma 33], that finitely many B_j are sufficient to cover $\mathcal{T}'$, i.e., it exists $\nu \in \mathbb{N}$ and $B_1, \ldots, B_\nu \in \widehat{\mathcal{B}}_{ph_\tau}(\mathcal{T}')$ such that $\mathcal{T}' = \bigcup_{j=1}^{\nu} B_j$ and $B_i \cap B_j = \varnothing$ 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 $(i = 1, \ldots, \nu)$. 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 $\Xi_{x,j} := \{\xi \in \Xi\colon h(\xi) - T(\xi)x \in B_j\}$ $(x \in X,\ j = 1, \ldots, \nu)$. Note that there is a number $\mu \in \mathbb{N}$ such that $\Xi_{x,j} \in \widehat{\mathcal{B}}_{ph_\mu}(\Xi)$ for all $x \in X$, $j = 1, \ldots, \nu$. Furthermore, we set $f_{x,j}(\xi) := c'x + \phi_j(h(\xi) - T(\xi)x)$ for $x \in X$, $j = 1, \ldots, \nu$. Finally, for $j = 1, \ldots, \nu$ we define

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

(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 $\mathcal{G}_j$ $(j = 1, \ldots, \nu)$ are uniformly bounded by one 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, $\mathcal{F}_j$ $(j = 1, \ldots, \nu)$ are uniformly bounded by some constants $K_j \ge 0$.

For the Donsker property (van der Vaart and Wellner [48, Theorem 2.5.2]) will be used, i.e., the following three conditions have to be verified for $\mathscr{H} = \mathscr{F}_j$ and $\mathscr{H} = \mathscr{G}_j$ $(j = 1, \dots, \nu)$, respectively:

(i) Existence of an envelope function:

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

(ii) $\mathscr{H}$ is "suitable measurable":[1] There exists a countable collection $\mathscr{H}' \subset \mathscr{H}$ such that every $h \in \mathscr{H}$ is the pointwise limit of a sequence h_n in $\mathscr{H}'$.

(iii) Uniform entropy condition: The *uniform entropy* given by

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

is *finite* where $\mathscr{P}_d(\Xi)$ denotes the set of all finitely discrete probability measures on Ξ and $N(\delta, \mathscr{H}, L_p(Q))$ is the *covering number* [2] of $\mathscr{H}$ in the space $L_p(Q)$.
If $\mathscr{H}$ is uniformly bounded by a constant $K \geq 0$, then obviously $N(\delta, \mathscr{H}, L_2(Q)) = 1$ for $\delta > K$, i.e., $\log N(\delta, \mathscr{H}, L_2(Q)) = 0$. Hence, if one chooses $F_{\mathscr{H}} \equiv K$ as envelope function, it suffices to verify

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

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

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

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

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 $x_n \to x_0$. Hence, since ϕ_j is continuous,

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

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

Uniform entropy condition. In van der Vaart and Wellner [48, Chapter 2.1.1], it is demonstrated that

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

where $N_{[\,]}(\delta, \mathscr{H}, L_p(Q))$ denotes the *bracketing number*[3] of the class of functions $\mathscr{H}$ in the space $L_p(Q)$. Furthermore, 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 van der Vaart and Wellner [48, Theorem 2.7.11] that

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

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 $N(\varepsilon, X, |.|) \leq c\varepsilon^{-m}$. Hence,

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

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

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

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

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

are P-measurable, i.e., for every $n \in \mathbb{N}$ and every $e \in \{-1, 1\}^n$ the mapping $(\xi_1, \dots, \xi_n) \mapsto \sup_{h \in \mathscr{H}_\delta} |\sum_{i=1}^n e_i h(\xi_i)|$ is measurable. See van der Vaart and Wellner [48, Definition 2.3.3 and Example 2.3.4] and van der Vaart [47, remark to Theorem 19.14].

[2] The covering number of $\mathscr{H}$ 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 $\mathscr{H}$ (van der Vaart and Wellner [48, Definition 2.1.5]).

[3] A 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 $N_{[\,]}(\delta, \mathscr{H}, L_p(Q))$ of a class $\mathscr{H}$ 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 $h \in \mathscr{H}$ lies between one of these brackets, i.e., $l \leq h \leq u$ (van der Vaart and Wellner [48, Definition 2.1.6]).

Envelope function. $F \equiv 1$ does the job.

Measurability. We set $\mathscr{H}'_\mu = \{\chi_B\colon B \in \widehat{\mathscr{B}}_{ph_\mu,\mathbb{Q}}(\Xi)\}$ with

$$\widehat{\mathscr{B}}_{ph_\mu,\mathbb{Q}}(\Xi) := \left\{\Xi \cap \bigcap_{j=1}^{\mu} H_j \mid H_j = \{\xi\colon c_j'\xi \le d_j\} \text{ or } H_j = \{\xi\colon c_j'\xi < d_j\} \text{ with } c_j \in \mathbb{Q}^k,\ d_j \in \mathbb{Q}\right\}$$

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 $B \in \widehat{\mathscr{B}}_{ph_\mu}(\Xi)$ there is a sequence B_n in $\widehat{\mathscr{B}}_{ph_\mu,\mathbb{Q}}(\Xi)$ such that $\chi_{B_n} \to \chi_B$ pointwise for $n \to \infty$ (note that Ξ is a bounded polyhedron).

Uniform entropy condition. We show that $\mathscr{H}_\mu$ is a so-called VC class:[4] For the set of (subgraphs of) indicator functions of open or closed half spaces ($\mu = 1$), it holds obviously that

$$V(\mathscr{H}_1) = V(\{\operatorname{sub}\chi_B\colon B \in \widehat{\mathscr{B}}_{ph_1}(\Xi)\}) \le k+2 < \infty$$

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, $\mathscr{H}_1$ is VC. And because

$$\begin{aligned}\operatorname{sub}\mathscr{H}_\mu &= \{\operatorname{sub}\chi_B\colon B \in \widehat{\mathscr{B}}_{ph_\mu}(\Xi)\}\\ &= \{\{(\xi,t) \in \Xi \times \mathbb{R}\colon t < \chi_B(\xi)\}\colon B \in \widehat{\mathscr{B}}_{ph_\mu}(\Xi)\}\\ &= \left\{\bigcap_{i=1}^{\mu} \operatorname{sub}\chi_{B_i}\colon B_i \in \widehat{\mathscr{B}}_{ph_1}(\Xi)\right\} = \operatorname{sub}\mathscr{H}_1 \sqcap \ldots \sqcap \operatorname{sub}\mathscr{H}_1,\end{aligned}$$

it holds that $\mathscr{H}_\mu$ is also VC due to van der Vaart and Wellner [48, Lemma 2.6.17 (ii)].
Theorem 2.6.7 in van der Vaart and Wellner [48] claims that in this case the following estimate is valid for all $Q \in \mathscr{P}_d(\Xi)$ with $\|F\|_{Q,2} > 0$ and for $\varepsilon \in (0,1)$:

$$N(\varepsilon\|F\|_{Q,2}, \mathscr{H}_\mu, L_2(Q)) \le c_1\varepsilon^{-c_2}$$

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

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

Since the last term is integrable for $\varepsilon \in (0,1)$, the uniform entropy condition is verified and $\mathscr{H}_\mu$ is shown to be P-Donsker. Because $\mathscr{G}_j \subset \mathscr{H}_\mu$ for $j = 1, \ldots, \nu$, each $\mathscr{G}_j$ is P-Donsker since subsets of P-Donsker classes are again P-Donsker (van der Vaart and Wellner [48, Theorem 2.10.1]).

(e) The Donsker property for $\mathscr{F}_j$ and $\mathscr{G}_j$ implies that $\mathscr{F}$ is P-Donsker:
From van der Vaart and Wellner [48, Theorems 2.10.6 and Examples 2.10.7 and 2.10.8], it follows that the class

$$\sum_{j=1}^{\nu} \mathscr{F}_j\mathscr{G}_j = \left\{\sum_{j=1}^{\nu} f_j g_j\colon f_j \in \mathscr{F}_j,\ g_j \in \mathscr{G}_j\ (j = 1, \ldots, \nu)\right\}$$

is P-Donsker since both $\mathscr{F}_j$ and $\mathscr{G}_j$ are uniformly bounded. Furthermore, because $\mathscr{F} \subset \sum_{j=1}^{\nu}\mathscr{F}_j\mathscr{G}_j$, the proof is complete since every subset of a P-Donsker class is P-Donsker as well (van der Vaart and Wellner [48, Theorem 2.10.1]). □

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

$$V(\mathscr{C}) = \inf\{n \in \mathbb{N} \mid \forall D \subset M \text{ with } \#D = n\ \exists A_D \subset D\ \forall C \in \mathscr{C}\colon A_D \ne D \cap C\}.$$

See van der Vaart and Wellner [48, Chapter 2.6] for further details.

4. Delta method and limit theorem for the optimal values. In order to get a convergence statement for the optimal value of Equation (1) in $\mathbb{R}$, i.e., weak convergence of $\sqrt{n}(v(P_n)-v(P))$, we want to apply the delta method described in, e.g., Römisch [36, Theorem 1 and Proposition 1]. For clarity, we cite these results here adapted to our framework.

DEFINITION 4.1. Let D and F be linear metric spaces. Then, Φ: $D\to F$ is called *Hadamard directionally differentiable* at $\vartheta_0\in D$ if there exists a mapping Φ'_{ϑ_0}: $D\to F$ with

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

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

The Hadamard directional derivative Φ'_{ϑ_0} is continuous and positively homogenous. However, note that linearity of Φ'_{ϑ_0} is *not* required here. By admitting a directional version of the concept of Hadamard differentiability, we follow Shapiro [42] and Rubinstein and Shapiro [37, Chapter 6] and deviate from mainstream literature (see, e.g., Gill [10], Gill and van der Vaart [11], van der Vaart [47], and van der Vaart and Wellner [48]). 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, either.

THEOREM 4.1. *Let D and F be linear metric spaces, Φ: $D\to F$ 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 Römisch [36, Theorem 1] (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 [21].

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

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

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

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

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

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

$$\Psi'_{\vartheta_0}(h)=\min\{h(x)\colon x\in S(\vartheta_0,0)\}.\tag{7}$$

PROOF. Obviously, the minimal value mapping Ψ is concave and, hence, directional differentiable for every fixed direction $h\in D$. Furthermore, $|\Psi(\vartheta_1)-\Psi(\vartheta_2)|\le\|\vartheta_1-\vartheta_2\|_\infty$ for $\vartheta_1,\vartheta_2\in D$, i.e., Ψ is Lipschitz continuous with modulus one. Then, Hadamard directional differentiability follows from Shapiro [42, Proposition 3.5]. Further, Römisch [36, Proposition 1] (again, set $\Theta=D_\Psi=D=T_\Theta(\vartheta_0)$) records the proof of formula (6) from Lachout [21] even if X were not compact.

Thus, it remains to show Equation (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 $x_n\in S(\vartheta_0,1/n)$ such that $h(x_n)\le\inf\{h(x)\colon x\in S(\vartheta_0,1/n)\}+1/n$. Then,

$$\inf\{h(x)\colon x\in S(\vartheta_0,1/n)\}\le h(x_n)\le\inf\{h(x)\colon x\in S(\vartheta_0,1/n)\}+1/n,$$

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

$$\vartheta_0(x_0)\le\liminf_{n'\to\infty}\vartheta_0(x_{n'})\le\Psi(\vartheta_0),$$

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

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

and on the other hand

$$\begin{aligned} \Psi'_{\vartheta_0}(h) = \lim_{n' \to \infty} h(x_{n'}) &\leq \lim_{n' \to \infty} (\inf\{h(x)\colon x \in S(\vartheta_0, 1/n')\} + 1/n') \\ &\leq \lim_{n' \to \infty} (\inf\{h(x)\colon x \in S(\vartheta_0, 0)\} + 1/n') \\ &= \min\{h(x)\colon x \in S(\vartheta_0, 0)\}. \quad \square \end{aligned}$$

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 $\ell^\infty(X)$. However, 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

$$\sqrt{n}(\vartheta_{P_n} - \vartheta_P) \rightsquigarrow \vartheta_{G_P} \sim \mathcal{N}(0, (Pf_xf_y - Pf_xPf_y)_{x,\, y \in X})$$

in $\ell^\infty(X)$ with $\vartheta_{G_P} = G_P f_. \in \ell^\infty(X)^\Omega$. Hence, the delta method (Theorem 4.1) can be applied for the minimal value mapping Ψ in $\ell^\infty(X)$. Inserting $\Psi(\vartheta_P) = v(P)$ and $\Psi(\vartheta_{P_n}) = v(P_n)$ leads to:

Corollary 4.1. *For the optimal value of the stochastic program (1), it holds that*

$$\sqrt{n}(v(P_n) - v(P)) \rightsquigarrow \Psi'_{\vartheta_P}(\vartheta_{G_P}) \tag{8}$$

in $\mathbb{R}$.

If we knew the distribution of $\Psi'_{\vartheta_P}(\vartheta_{G_P})$, we could give asymptotic confidence intervals for the optimal value $v(P) = \Psi(\vartheta_P)$ based on the quantiles of $\Psi'_{\vartheta_P}(\vartheta_{G_P})$ since $v(P_n)$ can be calculated by solving a finite mixed-integer linear program. In general, however, it seems too difficult to calculate the distribution analytically since Ψ'_{ϑ_P} from Equation (6) has a rather complicated shape. Hence, the above convergence statement contains, roughly speaking, two unknowns—the true optimal value $v(P)$ and the limit distribution.

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 Römisch [35, 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 $x_n \to x_0$ in X. However, to apply Equation (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 van der Vaart and Wellner [48, Example 1.5.10]): For almost all $\omega \in \Omega$ it holds that $\vartheta_{G_P}(\omega) \in \ell^\infty(X)$ is continuous with respect to the semimetric given by

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

However, in general, $x_n \to x_0$ 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., $\widehat{m} = 0$) would be an example where $x_n \to x_0$ in X implies $\rho(x_n, x_0) \to 0$ since in this case ϕ is continuous (see Wets [49]). 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

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

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 make use of Equation (8) not only in special cases, we do not continue this discussion here. We address ourselves to more general methods for getting information about the unknown distribution $\Psi'_{\vartheta_P}(\vartheta_{G_P})$.

5. Bootstrapping. Bootstrapping is a principle to gain information about (the quantiles of) an unknown limit distribution by resampling $\xi_1^*, \xi_2^*, \ldots$ from some empirical distribution P_n. From these resamples, the *bootstrap empirical measure* $P_n^* := (1/n)\sum_{j=1}^n \delta_{\xi_j^*}$ 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 §5.2.

The bootstrap method was introduced in Efron [7]. Here, we will make use of the consistency results as well as the delta method for the bootstrap derived in Giné and Zinn [14], van der Vaart [47], and van der Vaart and Wellner [48]. For further discussion and extensions of the bootstrap method see, e.g., Mammen [23], Hall [15], and Giné [13]. Note that the extensions there are different from the extension that are developed in §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 $\xi_1, \xi_2, \ldots$ are considered as fixed. The type of convergence is *conditionally on* $\xi_1, \xi_2, \ldots$ *in distribution*, which will be defined below following Giné and Zinn [14], van der Vaart [47], and van der Vaart and Wellner [48]. 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

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

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

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

if $Z_n \in \ell^\infty(\mathcal{F})^\Omega$ and $Z_0 \in \ell^\infty(\mathcal{F})^\Omega$ are measurable and tight (see Giné and Zinn [14] and van der Vaart [47, Chapter 23]). From Giné and Zinn [14], we adopt the following definition: The sequence Z_n is said to converge to Z_0 *conditionally on* $\xi_1, \xi_2, \ldots$ *in distribution* if

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

where $E[\cdot \mid \cdot]$ and $\xrightarrow{P^*}$ denote the conditional expectation and convergence in outer probability, respectively. This definition is also used in van der Vaart [47] and van der Vaart and Wellner [48]. With this notation, we are ready to cite results from Giné and Zinn [14].

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. See Giné and Zinn [14, Theorem 5.1] (and also van der Vaart [47, Theorem 23.7]). □

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 $\Phi\colon D \to \mathbb{R}$ 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. See, e.g., van der Vaart [47, Theorem 23.9] (set $D_\Phi = D$). □

Similar results had already been obtained in Arcones and Giné [3, §4]. Applied to our problem this proposition yields:

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

$$\sqrt{n}(v(P_n^*) - v(P_n)) = \sqrt{n}(\Psi(\vartheta_{P_n^*}) - \Psi(\vartheta_{P_n})) \rightsquigarrow_* \Psi'_{\vartheta_P}(\vartheta_{G_P^*}) \tag{9}$$

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 $v(P) = \min\{Pf\colon f \in \mathcal{F}\}$, i.e., the optimal value of the stochastic program (1).

Given $\xi_1, \xi_2, \ldots, \xi_n$, 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 $\xi_1^*, \xi_2^*, \ldots, \xi_n^*$ 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 $\alpha_1 < 50\%$ and $\alpha_2 < 50\%$, the interval

$$\left[v(P_n) - \frac{1}{\sqrt{n}}\zeta^*_{1-\alpha_1, m}, v(P_n) - \frac{1}{\sqrt{n}}\zeta^*_{\alpha_2, m}\right] \tag{10}$$

is an asymptotic confidence interval[5] at level $\alpha_1 + \alpha_2$ for the optimal value $v(P)$, i.e.,

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

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 is 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.

First, we record another proposition that was stated (but not actually proven) within the proof of van der Vaart [47, Theorem 23.9].

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

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

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

A proof is given in the appendix. 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 $\ell^\infty(X) \times \ell^\infty(X)$ this means

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

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

LEMMA 5.1. *Let $\Phi\colon D \to F$ 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*

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

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

The proof is given in the appendix. Note that the latter result does not require linearity of the Hadamard derivative. Putting the previous two results together leads to:

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 can not expect that the sequence $\sqrt{n}(\Psi(\vartheta_{P_n^*})-\Psi(\vartheta_{P_n}))$ converges conditionally on $\xi_1,\xi_2,\ldots$ 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 $\Phi: D\to F$ 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 $(\sqrt{n}(\bar{\vartheta}_n-\vartheta_0),\sqrt{n}(\tilde{\vartheta}_n-\vartheta_0))\rightsquigarrow(\bar{Z},\tilde{Z})$ with $\bar{Z},\tilde{Z}\in D^\Omega$. Then, it holds that*

$$\sqrt{n}\big(2\Phi\big(\tfrac{1}{2}(\bar{\vartheta}_n+\tilde{\vartheta}_n)\big)-\Phi(\tilde{\vartheta}_n)-\Phi(\vartheta_0)\big)\rightsquigarrow\Phi'_{\vartheta_0}(\bar{Z}+\tilde{Z})-\Phi'_{\vartheta_0}(\tilde{Z}).$$

Again, the proof is given in the appendix. Now, if we sample twice from P independently, i.e., given $\tilde{\xi}_1,\bar{\xi}_1,\tilde{\xi}_2,\bar{\xi}_2,\ldots\overset{iid}{\sim}P$, then, of course, with $\tilde{P}_n:=(1/n)\sum_{j=1}^n\delta_{\tilde{\xi}_j}$ and $\bar{P}_n:=(1/n)\sum_{j=1}^n\delta_{\bar{\xi}_j}$, it holds that

$$\big(\sqrt{n}(\vartheta_{\tilde{P}_n}-\vartheta_P),\sqrt{n}(\vartheta_{\bar{P}_n}-\vartheta_P)\big)\rightsquigarrow(\vartheta_{\tilde{G}_P},\vartheta_{\bar{G}_P})$$

in $\ell^\infty(X)\times\ell^\infty(X)$ 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,

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

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

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

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 in the previous section: 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

$$\left[2\Psi\big(\tfrac{1}{2}(\vartheta_{\bar{P}_n}+\vartheta_{\tilde{P}_n})\big)-\Psi(\vartheta_{\tilde{P}_n})-\frac{1}{\sqrt{n}}\zeta^*_{1-\alpha_1},\,2\Psi\big(\tfrac{1}{2}(\vartheta_{\bar{P}_n}+\vartheta_{\tilde{P}_n})\big)-\Psi(\vartheta_{\tilde{P}_n})-\frac{1}{\sqrt{n}}\zeta^*_{\alpha_2}\right]\tag{11}$$

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 estimation of the optimal value, i.e., the center of the confidence interval, and the calculation of the empirical quantiles $\zeta^*_{1-\alpha_1}$ and $\zeta^*_{\alpha_2}$, i.e., the range of the confidence interval, are carried out independently. The samples $\xi_1,\xi_2,\ldots,\xi_n$ are at no time fixed. To get one sample point for the empirical distribution function of the approximation of the limit distribution, one has to sample both $\xi_1,\xi_2,\ldots,\xi_n\overset{iid}{\sim}P$ and $\xi_1^*,\xi_2^*,\ldots,\xi_n^*\overset{iid}{\sim}P_n(\xi_1,\xi_2,\ldots,\xi_n)$.

Note that the estimation of the center of the confidence is based on a sample of size $2n$ rather than n, whereas the range of the confidence interval is based on samples of size n. This may be appreciated from a practical point of view since the center can be considered as the more important information, but for the range a high number of instances of $v(P_n)$ and $v(P_n^*)$ has to be computed. Moreover, since center and range are calculated independently, one could even use a different (bigger) n for the center than for the quantiles $\zeta^*_{1-\alpha_1}$ and $\zeta^*_{\alpha_2}$. Note that the computational effort for calculating $v(P_n)$ is known to depend highly nonlinear on n.

In some practical applications, however, sampling from P might be much more expensive than (re)sampling from a fixed empirical distribution P_n.

5.3. Subsampling. Another approach for deriving asymptotically consistent confidence intervals for estimators with limit theorems is called subsampling (Politis and Romano [29], Politis et al. [30]). This method is also based on sampling and resampling, but resampling is performed without replacement and with a lower samplesize $b = b(n) \in \mathbb{N}$, $b \ll n$. The subsampling method is more generally applicable than the bootstrapping methods, because only a basic limit theorem like in our case $\sqrt{n}(v(P_n) - v(P)) \rightsquigarrow \Psi'_{\vartheta_P}(\vartheta_{G_P})$ is required but no extra limit theorem like Equation (9) for the resampled measure.

For some sufficiently large n, this method estimates the limit distribution which is in our case $\Psi'_{\vartheta_P}(\vartheta_{G_P})$ by the empirical approximation based on the optimal values $v(P^*(n_1, \ldots, n_b))$ with respect to subsets $\{n_1, \ldots, n_b\}$ of $\{1, \ldots, n\}$ with cardinality b. The method is justified by the limit theorem (Politis and Romano [29, Theorem 2.1]) which reads in our framework

$$\binom{n}{b}^{-1} \sum_{1\le n_1<\cdots<n_b\le n} \delta_{\{\sqrt{b}(v(P^*(n_1,\ldots,n_b))-v(P_n))\}} \rightsquigarrow \Psi'_{\vartheta_P}(\vartheta_{G_P})$$

for $b, n \to \infty$ and $b/n \to 0$; see also Politis et al. [30]. Thereby, P_n denotes the empirical measure based on samples $\xi_1, \ldots, \xi_n$ and $P^*(n_1, \ldots, n_b)$ the empirical measure based on $\xi_{n_1}, \ldots, \xi_{n_b}$.

The number of summands in the previous display becomes extremely large as n and b grow. However, the result remains valid if a number $m = m(n)$ is chosen and the sum over all possible subsets is replaced by the sum over m randomly chosen subsets of $\{1, \ldots, n\}$ of cardinality b: Let $N_j^{n,b} \subset \{1, \ldots, n\}$ randomly chosen with $\# N_j^{n,b} = b$ for $j = 1, \ldots, m$. Then,

$$\frac{1}{m}\sum_{j=1}^{m} \delta_{\{\sqrt{b}(v(P_n^*(N_j^{n,b}))-v(P_n))\}} \rightsquigarrow \Psi'_{\vartheta_P}(\vartheta_{G_P}) \tag{12}$$

for $b, n, m \to \infty$ and $b/n \to 0$ (Politis and Romano [29, Corollary 2.1]). Thereby, $P_n^*(N_j^{n,b})$ denotes the empirical measure based on $\{\xi_i\colon i \in N_j^{n,b}\}$.

Similar to what was mentioned in Remark 5.1 for the extended bootstrap, the subsampling method allows us to compute the estimate for the optimal value, i.e., the center of the confidence interval, on the basis of a higher number of samples than the samplesize used for the calculation of the quantiles. This flexibility may be very useful from a computational point of view, because for the quantile calculation a high number of problem instances has to be solved and the computational effort of one instance depends highly nonlinear on the samplesize.

On the other hand, it has been noted that in some cases the subsampling method underperforms the bootstrap with respect to the accuracy of the approximation of the distribution, cf. Politis et al. [31]. This reference also contains ideas to improve the subsampling method by symmetrization. Such considerations, however, are beyond the scope of this work; in the numerical example in the next section, the method performed comparably to the bootstrap method.

6. Examples.

To demonstrate the meaning of the results derived in the previous sections, we provide some numerical evidence.

6.1. Problem (unique solution). We consider the example in Schultz et al. [40, §7]:

$$\min\left\{x'\begin{pmatrix}-1.5\\-4\end{pmatrix} + \int_{\Xi} \phi(\xi - x)\, dP(\xi)\colon x \in \{0,1,2,3,4,5\}^2\right\} \tag{13}$$

with

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

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

$$\tilde{\Xi} := \left\{\binom{5}{5}, \binom{5}{6}, \ldots, \binom{5}{15}, \binom{6}{5}, \ldots, \binom{15}{15}\right\},$$

i.e., $P(\{\eta\}) = 1/121$ for $\eta \in \tilde{\tilde{\Xi}}$. This example has the form of Equation (1) with $k = m = r = 2$, $h(\xi) = \xi$, $T(\xi) = I_2$, $\widehat{m} = 4$, $\overline{m} = 0$, and $\Xi = \text{conv}(\tilde{\tilde{\Xi}})$. The exact solution is $x = (0, 4)$ with optimal value $v(P) = -62.29$ (see Schultz et al. [40]). This solution is unique[6] and X consists of isolated points only; thus, the theory derived in §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 samplesize* 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 (10). In our example, this leads to

α (%)	Confidence interval for $v(P)$
10	$[-64.0133, -57.9733]$
5	$[-64.5867, -57.4467]$
2	$[-65.3600, -56.7400]$

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

α (%)	Number 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 $1/\sqrt{n}$

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

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 $1/\sqrt{n}$ since length times $\sqrt{n}$ is almost constant.

6.3. Problem (nonunique solution). We changed problem (13) to

$$\min\left\{x'\begin{pmatrix}-1.5\\-3.768595041\end{pmatrix} + \int_\Xi \phi(\xi - x)\, dP(\xi)\colon x \in \{0, 1, 2, 3, 4, 5\}^2\right\} \tag{14}$$

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 §5.2 to derive confidence intervals for the optimal value of problem (14). The procedure here is slightly different than that in §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 samplesize 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
⋮	⋮	⋮
−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 samplesize 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 (11) 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 validate the level of the confidence intervals empirically. We counted the number of confidence intervals covering the true optimal value −61.363636 and got:

α (%)	Number 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%, and 2%, respectively, during this approving procedure.

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

6.5. Subsampling. We applied the subsampling method from §5.3 to problem (13). Note that this method could be applied to problem (14) in the same manner since no further assumptions (such as uniqueness of the solution) need to be satisfied here.

(i) Fix $n, b, m \in \mathbb{N}$. We used $n = 150$, $b = 75$, $m = 500$.

(ii) Sample from P with samplesize n and solve the approximate problem. We got $v(P_n) = 62.16$.

(iii) Resample from P_n *without replacement* with samplesize $b < n$ (subsampling) and solve the smaller problem. This yields $v(P_b)$. Repeat this m times.

(iv) Calculate the quantiles $\zeta^*_{\alpha/2,\,b}$ and $\zeta^*_{1-\alpha/2,\,b}$ of the $v(P_b)$ values and transform them into quantiles of $v(P)$ according to formula (12). This yielded

α (%)	Confidence interval for $v(P)$
10	[−63.7800, −60.2467]
5	[−64.0733, −59.5067]
2	[−64.1800, −58.8333]

(v) We repeated the procedure (steps (ii) to (iv)) 200 times and counted the number of confidence intervals covering the true optimal value −62.29. We got:

α (%)	Number covering intervals	Ratio (%)	Average interval length
10	176/200	88	4.2451
5	192/200	96	5.0630
2	199/200	99.5	6.0207

Note that the size of the confidence intervals is of order $1/\sqrt{n}$ rather than order $1/\sqrt{b}$. This is why the average length is lower than for the bootstrap methods while the computational effort is comparable.

6.6. Technical details. These results were produced with C++ and ILOG CPLEX 9.1. We used the GNU C++ compiler *gcc* version 3.3.5 together with the ILOG Concert Technologie Library 21 on a standard Linux machine (2 GHz, 1 GB RAM, Suse Linux 9.3). As random number generator, we took the RANLIBC/StatLib library. In CPLEX, the following accuracy parameters were used: epOpt = epGap = epRHS = 10^{-6}. This means that the solutions of the approximate problems may be considered as exact.

7. Conclusions. In this paper, mathematical results on empirical approximations (sample average approximations) of two-stage mixed-integer stochastic programs have been derived. Such approximations are gained by exchanging the original underlying probability measure P by empirical measures P_n. In particular, in §3, a limit theorem for $\sqrt{n}(P_n - P)$ in the space of bounded functionals on the set of all possible integrands of the stochastic program has been proven by means of empirical process theory and by using the special piecewise Lipschitzian structure of these integrands. In §4, this result has been carried over to a limit theorem for the optimal values of the stochastic programs by means of the functional delta method. Thereby, we relied on a concept of Hadamard directional differentiability in infinite dimensional spaces such that linearity of the derivative is not required. The limit distribution is nonnormal in general and depends on the (usually unknown) solution set of the stochastic program. In §5, the applicability of resampling methods (bootstrap, subsampling) for approximating the limit distribution and deriving confidence intervals for optimal values is analyzed. The bootstrap method can not be applied straightforward because the nonlinearity of the derivative of the optimal value mapping makes the situation more involved than in standard applications. However, after some modifications of the procedure, asymptotic consistency holds. For the subsampling method, asymptotic consistency holds without modifications since the underlying theory is based on weaker assumptions. The results of the entire paper are confirmed and illustrated by studying a numerical example in §6.

The central limit theorem (clt) type results of §§3 and 4 extend the existing theory on empirical approximations of stochastic programs which is, roughly speaking, limited either to noninteger stochastic programs or to law of large number (lln) type and large deviation type results. The clt type results here may be employed to draw conclusions about the accuracy of an empirical approximation of a mixed-integer stochastic program. Resampling methods as suggested in §§5 and 6 are one possible way for making use of these results. Other methods may be developed in future work.

Appendix: Proofs. For the sake of completeness, we give the remaining proofs.

Proof of Proposition 5.2. For abbreviation, we set $Z_n := \sqrt{n}(P_n - P)$ and $Z_n^* := \sqrt{n}(P_n^* - P_n)$. Let (G_P, G_P^*) be a pair of *independent* P-Brownian bridges. Due to, e.g., van der Vaart and Wellner [48, Theorem 1.5.4] (see also van der Vaart and Wellner [48, Exercise 3, §1.5]), it suffices to show two things: (a) that the sequence (Z_n, Z_n^*) is asymptotically tight, and (b) that every finite dimensional marginal distribution of (Z_n, Z_n^*) converges to the respective marginal distribution of (G_P, G_P^*). See below.

(a) From Theorem 3.1, we know already that $Z_n \rightsquigarrow G_P$. Furthermore, it holds that $Z_n^* \rightsquigarrow G_P^*$ (unconditionally) because due to Jensen's inequality we have

$$\begin{aligned}\sup_{h\in BL_1(\ell^\infty(\mathcal{F}))} |E[h(Z_n^*)] - E[h(G_P^*)]| &= \sup_{h\in BL_1(\ell^\infty(\mathcal{F}))} |E[E[h(Z_n^*)\mid \xi_1,\xi_2,\ldots]] - E[h(G_P^*)]|\\ &\le \sup_{h\in BL_1(\ell^\infty(\mathcal{F}))} E[|E[h(Z_n^*)\mid \xi_1,\xi_2,\ldots] - E[h(G_P^*)]|]\\ &\le E\Big[\sup_{h\in BL_1(\ell^\infty(\mathcal{F}))} |E[h(Z_n^*)\mid \xi_1,\xi_2,\ldots] - E[h(G_P^*)]|\Big] \to 0\end{aligned}$$

since $\sup_h |E[h(Z_n^*)\mid \xi_1,\xi_2,\ldots] - E[h(G_P^*)]| \xrightarrow{P^*} 0$. Hence, since G_P and G_P^* are tight, both sequences Z_n and Z_n^* are asymptotically tight due to, e.g., van der Vaart and Wellner [48, Lemma 1.3.8]. For this case, van der Vaart and Wellner [48, Lemma 1.4.3] guarantees asymptotical tightness of the joint sequence (Z_n, Z_n^*).

(b) Let $l, m \in \mathbb{N}$ and $f_1,\ldots,f_l, g_1,\ldots,g_m \in \mathcal{F}$. It needs to be shown that $(Z_n f_1,\ldots,Z_n f_l, Z_n^* g_1,\ldots,Z_n^* g_m) \rightsquigarrow (G_P f_1,\ldots,G_P f_l, G_P^* g_1,\ldots,G_P^* g_m)$ in $\mathbb{R}^{l+m}$. We make use of the characterization of weak convergence in finite dimension by characteristic functions. Since G_P and G_P^* are independent and $Z_n f_\nu$ is $\sigma(\xi_1,\ldots,\xi_n)$-measurable, we can argue as follows:

$$\begin{aligned}&|E[e^{i\sum_{\nu=1}^l s_\nu Z_n f_\nu + i\sum_{\tau=1}^m t_\tau Z_n^* g_\tau}] - E[e^{i\sum_{\nu=1}^l s_\nu G_P f_\nu + i\sum_{\tau=1}^m t_\tau G_P^* g_\tau}]|\\ &\quad = |E[e^{i\sum_{\nu=1}^l s_\nu Z_n f_\nu} e^{i\sum_{\tau=1}^m t_\tau Z_n^* g_\tau}] - E[e^{i\sum_{\nu=1}^l s_\nu G_P f_\nu}]E[e^{i\sum_{\tau=1}^m t_\tau G_P^* g_\tau}]|\\ &\quad = |E[e^{i\sum_{\nu=1}^l s_\nu Z_n f_\nu} E[e^{i\sum_{\tau=1}^m t_\tau Z_n^* g_\tau}\mid \xi_1,\xi_2,\ldots]] - E[e^{i\sum_{\nu=1}^l s_\nu G_P f_\nu}]E[e^{i\sum_{\tau=1}^m t_\tau G_P^* g_\tau}]|\\ &\quad = |E[e^{i\sum_{\nu=1}^l s_\nu Z_n f_\nu}(E[e^{i\sum_{\tau=1}^m t_\tau Z_n^* g_\tau}\mid \xi_1,\xi_2,\ldots] - E[e^{i\sum_{\tau=1}^m t_\tau G_P^* g_\tau}])]\\ &\qquad + E[(e^{i\sum_{\nu=1}^l s_\nu Z_n f_\nu} - E[e^{i\sum_{\nu=1}^l s_\nu G_P f_\nu}])E[e^{i\sum_{\tau=1}^m t_\tau G_P^* g_\tau}]]|\\ &\quad \le |E[e^{i\sum_{\nu=1}^l s_\nu Z_n f_\nu}(E[e^{i\sum_{\tau=1}^m t_\tau Z_n^* g_\tau}\mid \xi_1,\xi_2,\ldots] - E[e^{i\sum_{\tau=1}^m t_\tau G_P^* g_\tau}])]|\\ &\qquad + |(E[e^{i\sum_{\nu=1}^l s_\nu G_P f_\nu}] - E[e^{i\sum_{\nu=1}^l s_\nu Z_n f_\nu}])E[e^{i\sum_{\tau=1}^m t_\tau G_P^* g_\tau}]|\\ &\quad \le E[|E[e^{i\sum_{\tau=1}^m t_\tau Z_n^* g_\tau}\mid \xi_1,\xi_2,\ldots] - E[e^{i\sum_{\tau=1}^m t_\tau G_P^* g_\tau}]|] + |E[e^{i\sum_{\nu=1}^l s_\nu G_P f_\nu}] - E[e^{i\sum_{\nu=1}^l s_\nu Z_n f_\nu}]|.\end{aligned}$$

Now, both terms in the last line of the previous display tend to zero, the second term because $Z_n \rightsquigarrow G_P$, and the first term because it holds that $|E[e^{i\sum_{\tau=1}^m t_\tau Z_n^* g_\tau}\mid \xi_1,\xi_2,\ldots] - E[e^{i\sum_{\tau=1}^m t_\tau G_P^* g_\tau}]| \xrightarrow{P^*} 0$ since $Z_n^* \rightsquigarrow_* G_P^*$ and (the real part and the imaginary part of) the functional $h(Z) := e^{i\sum_{\tau=1}^m t_\tau Z g_\tau}$ is (are) bounded and Lipschitz. □

Proof of Lemma 5.1. We define mappings

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

Then, due to the Hadamard directional differentiability of Φ, it holds that

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

for $(h_n^*,h_n)\to(h_0^*,h_0)$. Hence, the continuous mapping theorem (van der Vaart [47, Theorem 18.11]) applies and we get

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

Proof of Lemma 5.2. Again, we define mappings

$$\begin{aligned} g_n\colon\ & D\times D\to F\\ & (\bar h,\tilde h)\mapsto \sqrt{n}\left(2\Phi\left(\vartheta_0+\frac{1}{2\sqrt{n}}(\bar h+\tilde h)\right)-\Phi\left(\vartheta_0+\frac{1}{\sqrt{n}}\tilde h\right)-\Phi(\vartheta_0)\right).\end{aligned}$$

Then, due to the Hadamard directional differentiability of Φ, it holds that

$$\begin{aligned} g_n(\bar h_n,\tilde h_n) &= 2\sqrt{n}\left(\Phi\left(\vartheta_0+\frac{1}{2\sqrt{n}}(\bar h_n+\tilde h_n)\right)-\Phi(\vartheta_0)\right)-\sqrt{n}\left(\Phi\left(\vartheta_0+\frac{1}{\sqrt{n}}\tilde h_n\right)-\Phi(\vartheta_0)\right)\\ &\to \Phi'_{\vartheta_0}(\bar h_0+\tilde h_0)-\Phi'_{\vartheta_0}(\tilde h_0)\end{aligned}$$

for $(\bar h_n,\tilde h_n)\to(\bar h_0,\tilde h_0)$. Hence, the continuous mapping theorem (van der Vaart [47, Theorem 18.11]) applies again and we obtain

$$\begin{aligned} g_n\big(\sqrt{n}(\bar\vartheta_n-\vartheta_0),\sqrt{n}(\tilde\vartheta_n-\vartheta_0)\big) &= \sqrt{n}\big(2\Phi\big(\tfrac{1}{2}(\bar\vartheta_n+\tilde\vartheta_n)\big)-\Phi(\tilde\vartheta_n)-\Phi(\vartheta_0)\big)\\ &\rightsquigarrow \Phi'_{\vartheta_0}(\bar Z+\tilde Z)-\Phi'_{\vartheta_0}(\tilde Z).\quad\square\end{aligned}$$

Acknowledgments. The authors wish to thank the referees for valuable remarks and suggestions, in particular for drawing our attention to the subsampling method as well as for the idea for the proof of Proposition 5.2. This work was supported by the DFG Research Center Matheon (Mathematics for Key Technologies) in Berlin, Germany.

References

[1] Ahmed, S., A. Shapiro. 2002. The sample average approximation method for stochastic programs with integer recourse. *Optim. Online*.
[2] Apolloni, B., F. Pezzella. 1984. Confidence intervals in the solution of stochastic integer linear programming problems. *Ann. Oper. Res.* **1** 67–78.
[3] Arcones, M. A., E. Giné. 1992. On the bootstrap of M-estimators and other statistical functionals. R. Lepage, L. Billard, eds. *Exploring the Limits of the Bootstrap, Wiley Series in Probability and Mathematical Statistics*. Wiley, New York, 13–47.
[4] Artstein, Z., R. J-B. Wets. 1995. Consistency of minimizers and the SLLN for stochastic programs. *J. Convex Anal.* **2** 1–17.
[5] Blair, C. E., R. G. Jeroslow. 1977. The value function of a mixed integer program. *Discrete Math.* **19** 121–138.
[6] Dupačová, J., R. J-B. Wets. 1988. Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *Ann. Statist.* **16** 1517–1549.
[7] Efron, B. 1979. Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.
[8] Ermoliev, Y. M., V. I. Norkin. 1991. Normalized convergence in stochastic optimization. *Ann. Oper. Res.* **30** 187–198.
[9] Futschik, A., G. C. Pflug. 1995. Confidence sets for discrete stochastic optimization. *Ann. Oper. Res.* **56** 95–108.
[10] Gill, R. D. 1989. Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part 1). *Scandinavian J. Statist. Theory Appl.* **16** 97–128.
[11] Gill, R. D., A. W. van der Vaart. 1993. Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part 2). *Scandinavian J. Statist. Theory Appl.* **20** 271–288.
[12] Giné, E. 1996. Empirical processes and applications: An overview. *Bernoulli* **2** 1–28.

[13] Giné, E. 1997. Lectures on some aspects of the bootstrap. E. Giné, G. R. Grimmet, L. Saloff-Coste, eds. *Lectures on Probability Theory and Statistics, Lecture Notes in Mathematics*, Vol. 1665. Springer, Berlin, Germany, 37–151.
[14] Giné, E., J. Zinn. 1990. Bootstrapping general empirical measures. *Ann. Probab.* **18** 851–869.
[15] Hall, P. 1992. *The Bootstrap and Edgeworth Expansion*. *Springer Series in Statistics*. Springer, New York.
[16] Kaniovski, Y. M., A. J. King, R. J-B. Wets. 1995. Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems. *Ann. Oper. Res.* **56** 189–208.
[17] King, A. J., R. T. Rockafellar. 1993. Asymptotic theory for solutions in statistical estimation and stochastic programming. *Math. Oper. Res.* **18** 148–162.
[18] King, A. J., R. J-B. Wets. 1991. Epi-consistency of convex stochastic programs. *Stochastics Stochastics Rep.* **34** 83–92.
[19] Klatte, D. 1987. A note on quantitative stability results in nonlinear optimization. K. Lommatzsch, ed. *Proc. 19. Jahrestagung Mathematische Optimierung*. Sektion Mathematik, Seminarbericht Nr. 90, Humboldt-Universität Berlin, Berlin, Germany, 77–86.
[20] Kleywegt, A. J., A. Shapiro, T. Homem-de-Mello. 2001. The sample average approximation method for stochastic discrete optimization. *SIAM J. Optim.* **12** 479–502.
[21] Lachout, P. 2004. Personal communication.
[22] Louveaux, F., R. Schultz. 2003. *Stochastic Integer Programming*. *Handbooks in Operations Research and Management Science*, Vol. 10, Chapter 4. Elsevier, Amsterdam, The Netherlands, 213–266.
[23] Mammen, E. 1992. *When Does Bootstrap Work? Asymptotic Results and Simulations. Lecture Notes in Statistics*, Vol. 77. Springer, New York.
[24] Morita, H., H. Ishii, T. Nishida. 1987. Confidence region method for a stochastic programming problem. *J. Oper. Res. Soc. Japan* **30** 218–231.
[25] Norkin, V. I. 1992. Convergence of the empirical mean method in statistics and stochastic programming. *Cybernetics Systems Anal.* **28** 253–264.
[26] Pflug, G. C. 1995. Asymptotic stochastic programs. *Math. Oper. Res.* **20** 769–789.
[27] Pflug, G. C. 1999. Stochastic programs and statistical data. *Ann. Oper. Res.* **85** 59–78.
[28] Pflug, G. C., A. Ruszczyński, R. Schultz. 1998. On the Glivenko-Cantelli problem in stochastic programming: Mixed-integer linear recourse. *Math. Methods Oper. Res.* **47** 39–49.
[29] Politis, D. N., J. P. Romano. 1994. Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22** 2031–2050.
[30] Politis, D. N., J. P. Romano, M. Wolf. 1999. *Subsampling. Springer Series in Statistics*. Springer, New York.
[31] Politis, D. N., J. P. Romano, M. Wolf. 2000. Subsampling, symmetrization, and robust interpolation. *Comm. Statist. Theory Methods* **29** 1741–1757.
[32] Rachev, S. T. 1991. *Probability Metrics and the Stability of Stochastic Models*. Wiley, Chichester, UK.
[33] Rachev, S. T., W. Römisch. 2002. Quantitative stability in stochastic programming: The method of probability metrics. *Math. Oper. Res.* **27** 792–818.
[34] Robinson, S. M. 1987. Local epi-continuity and local optimization. *Math. Programming* **37** 208–223.
[35] Römisch, W. 2003. Stability of stochastic programming problems. *Handbooks in Operations Research and Management Science*, Vol. 10, Chapter 8. Elsevier, Amsterdam, The Netherlands, 483–554.
[36] Römisch, W. 2005. Delta method, infinite dimensional. S. Kotz, C. B. Read, N. Balakrishnan, B. Vidakovic, eds. Extended entry, *Encyclopedia of Statistical Sciences*, 2nd ed., Wiley, New York.
[37] Rubinstein, R. Y., A. Shapiro. 1993. *Discrete Event Systems, Sensitivity Analysis and Stochastic Optimization by the Score Function Method*. Wiley, Chichester, UK.
[38] Ruszczyński, A., A. Shapiro, eds. 2003. *Stochastic Programming*. *Handbooks in Operations Research and Management Science*, Vol. 10. Elsevier, Amsterdam, The Netherlands.
[39] Schultz, R. 1996. Rates of convergence in stochastic programs with complete integer recourse. *SIAM J. Optim.* **6** 1138–1152.
[40] Schultz, R., L. Stougie, M. H. van der Vlerk. 1998. Solving stochastic programs with integer recourse by enumeration: A framework using Gröbner basis reductions. *Math. Programming* **83** 229–252.
[41] Shapiro, A. 1989. Asymptotic properties of statistical estimators in stochastic programming. *Ann. Statist.* **17** 841–858.
[42] Shapiro, A. 1990. On concepts of directional differentiability. *J. Optim. Theory Appl.* **66** 477–487.
[43] Shapiro, A. 2000. Statistical inference of stochastic optimization problems. S. Uryasev, ed. *Probabilistic Constrained Optimization: Methodology and Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 282–304.
[44] Shapiro, A., T. Homem-de-Mello. 2000. On rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs. *SIAM J. Optim.* **11** 70–86.
[45] Talagrand, M. 1994. Sharper bounds for Gaussian and empirical processes. *Ann. Probab.* **22** 28–76.
[46] Talagrand, M. 1996. The Glivenko-Cantelli problem, ten years later. *J. Theoret. Probab.* **9** 371–384.
[47] van der Vaart, A. W. 1998. *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, UK.
[48] van der Vaart, A. W., J. A. Wellner. 1996. *Weak Convergence and Empirical Processes. Springer Series in Statistics*. Springer, New York.
[49] Wets, R. J-B. 1974. Stochastic programs with fixed recourse: The equivalent deterministic program. *SIAM Rev.* **16** 309–339.