Noname manuscript No.
(will be inserted by the editor)

On rates of convergence for sample average approximations in the almost sure sense and in mean

Dirk Banholzer · Jörg Fliege · Ralf Werner

Received: date / Accepted: date

Abstract We study the rates at which optimal estimators in the sample average approximation approach converge to their deterministic counterparts in the almost sure sense and in mean. To be able to quantify these rates, we consider the law of the iterated logarithm in a Banach space setting and first establish under relatively mild assumptions almost sure convergence rates for the approximating objective functions, which can then be transferred to the estimators for optimal values and solutions of the approximated problem. By exploiting a characterisation of the law of the iterated logarithm in Banach spaces, we are further able to derive under the same assumptions that the estimators also converge in mean, at a rate which essentially coincides with the one in the almost sure sense. This, in turn, allows to quantify the asymptotic bias of optimal estimators as well as to draw conclusive insights on their mean squared error and on the estimators for the optimality gap. Finally, we address the notion of convergence in probability to derive rates in probability for the deviation of optimal estimators and (weak) rates of error probabilities without imposing strong conditions on exponential moments. We discuss the possibility to construct confidence sets for the optimal values and solutions from our obtained results and provide a numerical illustration of the most relevant findings.

Keywords Stochastic programming · Sample average approximation · Almost sure rates of convergence · Rates of convergence in mean · Law of the iterated logarithm

Dirk Banholzer
Department of Mathematical Sciences, University of Southampton
Southampton, SO17 1BJ, UK
E-mail: dirk.banholzer@soton.ac.uk

Jörg Fliege
Department of Mathematical Sciences, University of Southampton
Southampton, SO17 1BJ, UK
E-mail: J.Fliege@soton.ac.uk

Ralf Werner
Institut für Mathematik, Universität Augsburg
86159 Augsburg, Germany
E-mail: ralf.werner@math.uni-augsburg.de

1 Introduction

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space on which we consider the stochastic programming problem

$$\min_{x \in \mathcal{X}} \Big\{ f(x) := \mathbb{E}_{\mathbb{P}}[h(x, \xi)] \Big\}, \tag{1}$$

where $\mathcal{X} \subset \mathbb{R}^n$ denotes a nonempty finite-dimensional compact set with the usual (Euclidean) metric, ξ a random vector whose distribution $\mathbb{P}^{\xi}$ is supported on a set $\Xi \subset \mathbb{R}^m$, and $h : \mathcal{X} \times \Xi \to \mathbb{R}$ a function depending on some parameter $x \in \mathcal{X}$ and the random vector ξ. For f to be well-defined, we assume for every $x \in \mathcal{X}$ that $h(x, \cdot)$ is measurable with respect to the Borel σ-algebras $\mathcal{B}(\Xi)$ and $\mathcal{B}(\mathbb{R})$, and that it is $\mathbb{P}^{\xi}$-integrable.

The stochastic problem (1) may arise in various applications from a broad range of areas, such as finance and engineering, where deterministic approaches turn out to be unsuitable for formulating the actual problem. Quite frequently, the problem is encountered as a first-stage problem of a two-stage stochastic program where $h(x, \xi)$ describes the optimal value of a subordinate second-stage problem, see, e.g., Shapiro et al (2014). Naturally, problem (1) may also be viewed as a self-contained problem, in which h directly results from modelling a stochastic quantity.

Unfortunately, in many situations, the distribution of the random function $h(\cdot, \xi)$ is not known exactly, such that the expected value in (1) cannot be evaluated readily and therefore needs to be approximated in some way. Using Monte Carlo simulation, a common approach (see, e.g., Homem-de-Mello and Bayraksan (2014) for a recent survey) consists of drawing a sample of i.i.d. random vectors $\xi_1, \ldots, \xi_N$, $N \in \mathbb{N}$, from the same distribution as ξ, and considering the *sample average approximation* (SAA) problem

$$\min_{x \in \mathcal{X}} \left\{ \hat{f}_N(x) := \frac{1}{N} \sum_{i=1}^{N} h(x, \xi_i) \right\} \tag{2}$$

as an approximation to the original stochastic programming problem (1). Since the SAA problem (2) depends on the set of random vectors $\xi_1, \ldots, \xi_N$, its optimal value $\hat{f}_N^*$ is an estimator of the optimal value f^* of the original problem (1), and a solution $\hat{x}_N^*$ from the set of optimal solutions $\widehat{\mathcal{X}}_N^* := \arg\min_{x \in \mathcal{X}} \hat{f}_N(x)$ is an estimator of a solution x^* from the set of optimal solutions $\mathcal{X}^*$ of the original problem (1). For a particular realisation of the random sample, the approximating problem (2) represents a deterministic problem instance, which can then be solved by adequate optimisation algorithms. For this purpose, one usually assumes that the set $\mathcal{X}$ is described by (deterministic) equality and inequality constraints.

An appealing feature of the SAA approach is its sound convergence properties, which have been discussed in a variety of publications. Considering the consistency of SAA estimators, which is typically deemed to be a minimal requirement for any good estimator, Dupačová and Wets (1988) show in a rather general way that the sequence of approximating objective function $\{\hat{f}_N\}$ epi-converges to the true objective f, which allows to infer the strong consistency of optimal values and of sets of optimal solutions (Rockafellar and Wets (1998)). A similar approach to consistency based on the concept of epi-convergence has been pursued by Robinson (1996),

whereas Bates and White (1985) (cf. also Shapiro et al (2014), Chapter 5) take an alternative approach and derive the strong consistency of the optimal estimators by first establishing the almost sure uniform convergence of $\{\hat{f}_N\}$ to f. Clearly, the strong consistency of optimal estimators implies their weak consistency.

Given consistency, it is reasonable to further investigate the rates of convergence at which the SAA estimators approach their original counterparts as N tends to infinity. In this regard, Shapiro (1989, 1990, 1991) and King and Rockafellar (1993), among others, provide necessary and sufficient conditions for a characterisation of the asymptotic distribution of the estimators (inter alia, uniqueness of x^* is assumed in the case of optimal solutions), from which it immediately follows that $\{\hat{f}_N^*\}$ and $\{\hat{x}_N^*\}$ converge in distribution to their deterministic counterparts at a rate of $1/\sqrt{N}$. In particular, the findings of the former author are essentially based on the central limit theorem in Banach spaces, to which the delta method with a first and second order expansion of the minimum value function is then applied, while the latter use a generalised implicit function theorem to achieve these results.

Rates of convergence have also been studied for the convergence in probability with respect to different purposes. Especially, once having obtained rates of convergence in distribution, it is easy to see that the normalising sequences $\{\sqrt{N}(\hat{f}_N^* - f^*)\}$ and $\{\sqrt{N}(\hat{x}_N^* - x^*)\}$ stay bounded in probability as $N \to \infty$, thus providing insights on the inner deviation rate for optimal estimators, cf. Pflug (2003). Moreover, the rates of error probabilities, i.e. the deviation probabilities between the optimal estimators and their corresponding unknown true values, have been quantified, due to their practical relevance. This has been addressed, for instance, by Vogel (1988, 1992) who uses a large deviation approach to estimate the probability that the solution set of an approximating problem is not contained in an ϵ-neighbourhood of the original solution set in a standard stochastic programme and to estimate the probability of particular events of both solution sets in a multiobjective programming framework, respectively. Further results concerning rates of error probabilities have also been provided by Kaniovski et al (1995) and Dai et al (2000), where exponential bounds for the error probabilities of optimal values and solutions are derived by means of the theory of large deviations. To obtain these results the authors have to make the rather strong but unavoidable assumption of an existing moment generating function with a finite value in a neighbourhood of zero. However, this assumption then allows to derive conservative estimates for the sample size required to solve the original problem to a given accuracy with overwhelming probability, see, e.g., Shapiro (2003) or Shapiro et al (2014), Sections 5.3 and 7.2.10, for further details. Further results on exponential rates of convergence are obtained by Shapiro and Homem-de-Mello (2000) in the setting of a convex, piecewise smooth function h and a discrete distribution $\mathbb{P}^\xi$, and by Homem-de-Mello (2008) in case the underlying sample of random vectors is non-i.i.d.. Eventually, Vogel (2017) considers approximations of solution sets in probability with (inner) rate of convergence and (outer) tail behaviour function within a general multiobjective framework. These results then serve as a prerequisite to construct universal confidence sets for the optimal value and optimal solutions, see, e.g., Pflug (2003) and Vogel (2008b). However, universal confidence sets usually rely on some explicit knowledge of the random variables involved, see, e.g., Vogel (2008a). Therefore, in situations with less information available, approximate confidence sets are often considered by invoking some central limit

theorem. Especially estimators for the optimality gap (cf. Mak et al (1999)) have gained practical interest, see also Homem-de-Mello and Bayraksan (2014) for a detailed discussion.

Accordingly, all rates of convergence which have been established so far in the SAA context consider convergence in distribution or convergence in probability, cf. Table 1 for a brief overview. To the best of our knowledge, rates of convergence

	convergence	rate of convergence
almost surely	✓	✗
in mean (L_1)	✓	✗
in probability	✓	✓
in distribution	✓	✓

Table 1: Convergence results for the SAA framework for the objective functions, optimal values, and solutions, under different assumptions on h.

that hold almost surely and thus complement the strong consistency of optimal estimators with its corresponding rate have not yet been considered in the SAA framework, with very few exceptions using particular assumptions. Convergence in mean can be derived in a straightforward manner from convergence in probability and some uniform integrability condition (e.g. a finite second moment and almost sure Lipschitz continuity of h in x). However, no rates for this type of convergence seem to have been established, as far as we are aware of. This is an important issue, as convergence in mean is the main basis to derive meaningful statements on the size of the bias of estimators. The most notable related work on almost sure rates of convergence is Homem-de-Mello (2003), which used the slightly different setting of a variable SAA (VSAA), where in each iteration k the objective function is approximated by an estimator $\hat{f}_{N_k}$ with a newly drawn random sample of (potentially) different size N_k. In particular, the author derives for any $x \in \mathcal{X}$ pointwise sample path bounds on the error $|\hat{f}_{N_k}(x) - f(x)|$, which in turn allows to infer almost sure rates of convergence for objective functions. Yet, as the obtained rates hold pointwise, they only apply to finite feasible sets $\mathcal{X}$ and cannot be generalised to universal compact sets that we consider here. Further related results outside the SAA framework can be found, for instance, in He and Wang (1995) in the context of M-estimators. Their approach, however, differs considerable from ours in that the obtained results are based on necessary and sufficient first order optimality conditions assuming a sufficiently smooth convex objective and no constraints. Also, their main result is that an optimal estimator satisfies the law of the iterated logarithm – a statement which actually excludes faster rates of convergence in the almost sure sense.

In this paper, we aim at closing the gaps described above, providing rates of convergence in the almost sure sense and in mean, where possible. As it has to be expected, rates of convergence that hold almost surely may be derived by means of the law of the iterated logarithm (LIL), which characterises the extreme fluctuations occurring in a sequence of averages and thus complements the strong law of large numbers and the central limit theorem (CLT). In particular, by applying the LIL in a Banach space setting, we are able to obtain rates for objective function values, optimal values, and solutions, similar to the technique that has already

been applied in the form of the functional CLT to obtain asymptotic distributions of the respective quantities, see, e.g., Shapiro (1991). Moreover, we also obtain convergence in mean of the approximating objective functions and of the optimal estimators, including their associated rates of convergence. This appears to be one advantage of using the LIL in Banach spaces. The rates essentially coincide with the almost sure rates of convergence and may be used to quantify the asymptotic bias of the optimal estimators. Further, it is possible to show that the mean squared errors of optimal estimators converge to zero with known rates of convergence, which again may be used in the particular case of optimal values to show that the size of the confidence set for the optimality gap of f^* converges to zero at a known rate. As the LIL in Banach spaces also provides an interesting implication on convergence in probability, we discuss rates of convergence of this kind as well. We derive rates of convergence in probability (i.e. 'inside the probability') for the deviation of optimal values and solutions, and weak rates of error probabilities (i.e. 'outside the probability') by which we are able to decrease the gap between rates obtained from first or second moments and rates obtained via exponential moments. At last, we exploit the inferred rates of convergence in probability to define confidence sets for the optimal values and solutions, albeit without known coverage probability.

The remainder of this paper is organised as follows. In Section 2, we set the stage for later results and briefly review basic concepts of random variables with values in a Banach space, as well as the CLT and the LIL in Banach spaces. To better compare our findings, Section 3 first outlines known results on the convergence in distribution of the SAA estimators and its corresponding rates. In analogy to these results, we then derive within the same setting by virtue of the LIL rates of convergence for the SAA estimators that hold almost surely and in mean. In Section 4, we establish immediate consequences of the obtained rates of convergence in the almost sure sense and in mean, providing an improved analysis of the estimator for the optimality gap and the construction of confidence sets. In Section 5, we illustrate some selected results by a numerical simulation, while Section 6 contains our conclusions.

2 Probability in Banach Spaces

We first introduce some basic concepts of Banach space valued random variables and corresponding results of limit theorems in Banach spaces to be used throughout this paper. For a more detailed discussion on these subjects and further references, let us refer to the excellent monograph of Ledoux and Talagrand (1991).

2.1 Banach Space Valued Random Variables

Let B denote a separable Banach space, i.e. a vector space over the field of real numbers equipped with a norm $\|\cdot\|$ with which the space is complete and which contains a countable dense subset. Its topological dual is denoted by B' and duality is given by $g(y) = \langle g, y \rangle$ for $g \in B'$, $y \in B$. The dual norm of $g \in B'$ is also denoted by $\|g\|$ for convenience.

A random variable X on B, or B-valued random variable in short, is a measurable mapping from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ into B equipped with its Borel σ-algebra $\mathcal{B}(B)$ generated by the open sets of B. Thus, for every Borel set $U \in B$, we have $X^{-1}(U) \in \mathcal{F}$. A B-valued random variable X is said to be strongly (or Bochner) integrable if the real-valued random variable $\|X\|$ is integrable, i.e. $\mathbb{E}_{\mathbb{P}}[\|X\|] < \infty$. The variable is said to be weakly (or Pettis) integrable if for any $g \in B'$ the real-valued random variable $g(X)$ is integrable and there exists a unique element $y \in B$ such that $g(y) = \mathbb{E}_{\mathbb{P}}[g(X)] = \int g(X) \mathrm{d}\mathbb{P}$. If this is the case, then the element y is denoted by $\mathbb{E}_{\mathbb{P}}[X]$ and called the expected value of X. A sufficient condition for its existence is that $\mathbb{E}_{\mathbb{P}}[\|X\|] < \infty$. Given that $\mathbb{E}_{\mathbb{P}}[g(X)] = 0$ and $\mathbb{E}_{\mathbb{P}}[g^2(X)] < \infty$ for all $g \in B'$, the covariance function of X is defined by $(\mathbb{C}\mathrm{ov}\, X)(g_1, g_2) := \mathbb{E}_{\mathbb{P}}[g_1(X) g_2(X)]$, $g_1, g_2 \in B'$, which is a nonnegative symmetric bilinear form on B'.

The familiar notions of convergence of random variables on the real line extend in a straightforward manner to Banach spaces. As such, a sequence $\{X_N\}$ of random variables with values in B converges in distribution (or weakly) to a random variable X, denoted by $X_N \xrightarrow{d} X$, if for any bounded and continuous function $\psi : B \to \mathbb{R}$, $\mathbb{E}_{\mathbb{P}}[\psi(X_N)] \to \mathbb{E}_{\mathbb{P}}[\psi(X)]$ as $N \to \infty$. Moreover, $\{X_N\}$ converges in probability to X, in brief $X_N \xrightarrow{p} X$, if for each $\epsilon > 0$, $\lim_{N\to\infty} \mathbb{P}(\|X_N - X\| > \epsilon) = 0$. The sequence is said to be bounded in probability if, for each $\epsilon > 0$, there exists $M_\epsilon > 0$ such that $\sup_N \mathbb{P}(\|X_N\| > M_\epsilon) < \epsilon$. Similarly, $\{X_N\}$ is said to converge $\mathbb{P}$-almost surely to a B-valued random variable X if $\mathbb{P}(\lim_{N\to\infty} X_N = X) = 1$, and it is $\mathbb{P}$-almost surely bounded if $\mathbb{P}(\sup_N \|X_N\| < \infty) = 1$. Finally, denoting by $L_1(B) = L_1(\Omega, \mathcal{F}, \mathbb{P}; B)$ the space of all B-valued random variables X on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{E}_{\mathbb{P}}[\|X\|] < \infty$, we say that the sequence $\{X_N\}$ converges to X in $L_1(B)$ if X_N, X are in $L_1(B)$ and $\mathbb{E}_{\mathbb{P}}[\|X_N - X\|] \to 0$ as $N \to \infty$.

For a sequence $\{X_N\}$ of i.i.d. B-valued random variables with the same distribution as X, we define $S_N := \sum_{i=1}^N X_i$ for $N \in \mathbb{N}$. We write $\mathrm{Log}(x)$ to denote the function $\max\{1, \log x\}$, $x \geq 0$, and let $\mathrm{LLog}(x)$ stand for $\mathrm{Log}(\mathrm{Log}(x))$. Further, we set for $N \in \mathbb{N}$,

$$a_N := \sqrt{2N \,\mathrm{LLog}(N)} \quad \text{and} \quad b_N := \frac{a_N}{N} = \frac{\sqrt{2 \,\mathrm{LLog}(N)}}{\sqrt{N}}.$$

2.2 Basic Limit Theorems

Based on the notions of convergence of random variables, the CLT and the LIL on the real line can be extended subject to minor modifications to random variables taking values in a separable Banach space. However, the necessary and sufficient conditions for these limit theorems to hold in the Banach case are fundamentally different from those for the real line.

For the sake of generality, the following discussion is phrased in terms of a generic separable Banach space B. However, to establish rates of convergence for the SAA setup, we will from Section 3 onwards only work in the separable Banach space $C(\mathcal{X})$ of continuous functions $\psi : \mathcal{X} \to \mathbb{R}$, endowed with the supremum norm $\|\psi\|_\infty = \sup_{x\in\mathcal{X}} |\psi(x)|$, and in the separable Banach space $C^1(\mathcal{X})$ of continuously

differentiable functions ψ, defined on an open neighbourhood of the compact set $\mathcal{X}$ and equipped with the norm

$$\|\psi\|_{1,\infty} = \sup_{x\in\mathcal{X}} |\psi(x)| + \sup_{x\in\mathcal{X}} \|\nabla\psi(x)\|,$$

where $\nabla\psi(x)$ denotes the gradient of the function $\psi \in C^1(\mathcal{X})$ at the point x. Instead of the generic B-valued random variables X and the i.i.d. copies X_i, $i = 1,\ldots,N$, we will then consider the random variables $\widetilde{X} := h(\cdot,\xi) - \mathbb{E}_{\mathbb{P}}[h(\cdot,\xi)]$ and $\widetilde{X}_i := h(\cdot,\xi_i) - \mathbb{E}_{\mathbb{P}}[h(\cdot,\xi_i)]$, respectively, to which the limit theorems in the particular Banach spaces are applied.

2.2.1 The Central Limit Theorem

A random variable X with values in B is said to satisfy the *CLT* if for i.i.d. B-valued random variables $\{X_N\}$ with the same distribution as X, there exists a mean zero Gaussian random variable Z with values in B such that

$$\frac{S_N}{\sqrt{N}} \xrightarrow{d} Z, \text{ as } N \to \infty.$$

Here, by definition, a B-valued random variable Z is Gaussian if for any $g \in B'$, $g(Z)$ is a real-valued Gaussian random variable. In particular, note that all weak moments of Z thus exist for any $g \in B'$, and it follows from Fernique's theorem (see Fernique (1970)) that Z also has finite strong moments of all orders, i.e. $\mathbb{E}_{\mathbb{P}}[\|Z\|^p] < \infty$ for $p > 0$. If X satisfies the CLT in B, then for any $g \in B'$ the real-valued random variable $g(X)$ satisfies the CLT with limiting Gaussian distribution of variance $\mathbb{E}_{\mathbb{P}}[g^2(X)] < \infty$. Hence, the sequence $\{S_N/\sqrt{N}\}$ converges in distribution to a Gaussian random variable Z with the same covariance function as X, i.e. for $g_1, g_2 \in B'$, we have $(\mathbb{C}\mathrm{ov}\, X)(g_1,g_2) = (\mathbb{C}\mathrm{ov}\, Z)(g_1,g_2)$.

For general Banach spaces, no necessary and sufficient conditions such that a random variable X satisfies the CLT seem to be known. In particular, as mentioned e.g. by Kuelbs (1976a), the moment conditions $\mathbb{E}_{\mathbb{P}}[X] = 0$ and $\mathbb{E}_{\mathbb{P}}[\|X\|^2] < \infty$ are neither necessary nor sufficient for the CLT, as opposed to real-valued random variables. (See Strassen (1966) for the equivalence.) Nevertheless, sufficient conditions can be given for certain classes of random variables, such as for mean zero Lipschitz random variables X with square-integrable (random) Lipschitz constant on the spaces $C(\mathcal{X})$ and $C^1(\mathcal{X})$, see Araujo and Giné (1980), Chapter 7.

2.2.2 The Law of the Iterated Logarithm

For the LIL in Banach spaces, essentially two definitions may be distinguished. The first definition naturally arises from Hartman and Wintner's LIL for real-valued random variables, see Hartman and Wintner (1941), and says that a random variable X satisfies the *bounded LIL* if the sequence $\{S_N/a_N\}$ is $\mathbb{P}$-almost surely bounded in B, or equivalently, if the nonrandom limit (due to Kolmogorov's zero-one law)

$$\Lambda(X) := \limsup_{N\to\infty} \frac{\|S_N\|}{a_N}$$

is finite, $\mathbb{P}$-almost surely (cf. Ledoux and Talagrand (1991), Section 8.2).

Strassen's sharpened form of the LIL for random variables on the real line, see Strassen (1964), however, suggests a second natural definition of the LIL in Banach spaces, which is known as the *compact LIL*. Accordingly, X satisfies the compact LIL if the sequence $\{S_N/a_N\}$ is not only $\mathbb{P}$-almost surely bounded in B, but $\mathbb{P}$-almost surely relatively compact in B. While coinciding in finite dimensions, both definitions clearly differ from each other in the case of infinite-dimensional Banach spaces. Kuelbs (1976a) further showed that when the sequence $\{S_N/a_N\}$ is $\mathbb{P}$-almost surely relatively compact in B, then there is a convex symmetric and necessarily compact set K in B such that

$$\lim_{N\to\infty} \operatorname{dist}\left(\frac{S_N}{a_N}, K\right) = 0, \quad \text{and} \quad \operatorname{CP}\left(\left\{\frac{S_N}{a_N}\right\}\right) = K, \tag{3}$$

each $\mathbb{P}$-almost surely, where $\operatorname{dist}(y,K) = \inf_{\bar{y}\in K}\|y-\bar{y}\|$ for any point $y \in B$ and $\operatorname{CP}(\{y_N\})$ denotes the set of all limit points of the sequence $\{y_N\}$ in B. This characterisation may be seen as an equivalent definition of the compact LIL (e.g., Ledoux and Talagrand (1991), Theorem 8.5). In particular, we then have $\Lambda(X) = \sup_{y\in K}\|y\|$.

The limit set $K = K_X$ in (3) is known to be the unit ball of the reproducing kernel Hilbert space $H = H_X \subset B$ associated to the covariance of X, and can briefly be described as follows, see Kuelbs (1976a) and Goodman et al (1981) for further details. Assuming that for all $g \in B'$, $\mathbb{E}_{\mathbb{P}}[g(X)] = 0$ and $\mathbb{E}_{\mathbb{P}}[g^2(X)] < \infty$, and considering the operator $A = A_X$ defined as $A : B' \to L_2 = L_2(\Omega, \mathcal{F}, \mathbb{P})$, $Ag = g(X)$, we have

$$\|A\| = \sup_{\|g\|\leq 1} \left(\mathbb{E}_{\mathbb{P}}[g^2(X)]\right)^{1/2} =: \sigma(X), \tag{4}$$

and by a closed graph argument that A is bounded. Moreover, the adjoint $A' = A'_X$ of the operator A with $A'\zeta = \mathbb{E}_{\mathbb{P}}[\zeta X]$ for $\zeta \in L_2$ maps L_2 into $B \subset B''$. The space $A'(L_2) \subset B$ equipped with the scalar product $\langle\cdot,\cdot\rangle_X$ transferred from L_2 and given by $\langle A'\zeta_1, A'\zeta_2\rangle_X = \langle\zeta_1,\zeta_2\rangle_{L_2} = \mathbb{E}_{\mathbb{P}}[\zeta_1\zeta_2]$, with $\zeta_1, \zeta_2 \in L_2$, then determines a separable Hilbert space H. Latter space reproduces the covariance structure of X in that for $g_1, g_2 \in B'$ and any element $y = A'(g_2(X)) \in H$, we have $g_1(y) = \mathbb{E}_{\mathbb{P}}[g_1(X)g_2(X)]$. In particular, if X_1 and X_2 are two random variables with the same covariance function, it follows from the reproducing property that $H_{X_1} = H_{X_2}$. Eventually, the closed unit ball K of H, i.e. $K = \{y \in B : y = \mathbb{E}_{\mathbb{P}}[\zeta X], (\mathbb{E}_{\mathbb{P}}[\|\zeta\|^2])^{1/2} \leq 1\}$, is a bounded and convex symmetric subset of B, and it can be shown that

$$\sup_{y\in K}\|y\| = \sigma(X).$$

As the image of the (weakly compact) unit ball of L_2 under A', the set K is weakly compact. It is compact when $\mathbb{E}_{\mathbb{P}}[\|X\|^2] < \infty$, as shown by Kuelbs (1976a), Lemma 2.1, and if and only if the family of random variables $\{g^2(X) : g \in B', \|g\| \leq 1\}$ is uniformly integrable, see, e.g., Ledoux and Talagrand (1991), Lemma 8.4.

While for a real-valued or, more generally, finite-dimensional random variable X the LIL is satisfied if and only if $\mathbb{E}_{\mathbb{P}}[X] = 0$ and $\mathbb{E}_{\mathbb{P}}[\|X\|^2] < \infty$ (see Strassen (1966) and Pisier and Zinn (1978)), the moment conditions are neither

necessary nor sufficient for a B-valued random variable to satisfy the LIL in an infinite-dimensional setting, see Kuelbs (1976a). Yet, conditions for the bounded LIL to hold were initially given by Kuelbs (1977), asserting that under the hypothesis $\mathbb{E}_{\mathbb{P}}[X] = 0$ and $\mathbb{E}_{\mathbb{P}}[\|X\|^2] < \infty$, the sequence $\{S_N/a_N\}$ is $\mathbb{P}$-almost surely bounded if and only if $\{S_N/a_N\}$ is bounded in probability. Similarly, Kuelbs also showed under the same assumptions that $\{S_N/a_N\}$ is $\mathbb{P}$-almost surely relatively compact in B (and thus (3) holds for the unit ball K of the reproducing kernel Hilbert space associated to the covariance of X) if and only if

$$S_N/a_N \xrightarrow{p} 0, \text{ as } N \to \infty, \tag{5}$$

which holds if and only if

$$\mathbb{E}_{\mathbb{P}}\big[\|S_N\|\big] = \mathbf{o}(a_N). \tag{6}$$

An immediate consequence of this result is that, given the moment conditions, X satisfying the CLT implies that X also satisfies the compact LIL (Pisier, 1975), but not vice versa (Kuelbs, 1976b). Specifically, the former statement holds since convergence in distribution of $\{S_N/\sqrt{N}\}$ to a mean zero Gaussian random variable in B entails that the sequence is bounded in probability, from which then (5) follows directly.

Considering the necessary conditions for the random variable X to satisfy the LIL in Banach spaces, however, it turns out that the moment condition $\mathbb{E}_{\mathbb{P}}[\|X\|^2] < \infty$ is unnecessarily restrictive in infinite dimensions and can hence be further relaxed. This leads to the following characterisation of the LIL in Banach spaces, providing optimal necessary and sufficient conditions, cf. Ledoux and Talagrand (1988), Theorems 1.1 and 1.2. In this regard, note that since the boundedness in probability of $\{S_N/a_N\}$ comprises $\mathbb{E}_{\mathbb{P}}[X] = 0$, cf. Ledoux and Talagrand (1988), Proposition 2.3, the latter property is already omitted in condition (*ii*) of both respective statements.

Theorem 1 *(Ledoux and Talagrand, 1988). Let X be a random variable with values in a separable Banach space.*

a) The sequence $\{S_N/a_N\}$ is $\mathbb{P}$-almost surely bounded if and only if (i) $\mathbb{E}_{\mathbb{P}}[\|X\|^2/\operatorname{LLog}(\|X\|)] < \infty$, (ii) for each $g \in B'$, $\mathbb{E}_{\mathbb{P}}[g^2(X)] < \infty$, and (iii) $\{S_N/a_N\}$ is bounded in probability.

b) The sequence $\{S_N/a_N\}$ is $\mathbb{P}$-almost surely relatively compact if and only if (i) $\mathbb{E}_{\mathbb{P}}[\|X\|^2/\operatorname{LLog}(\|X\|)] < \infty$, (ii) $\{g^2(X) : g \in B', \|g\| \leq 1\}$ is uniformly integrable, and (iii) $S_N/a_N \xrightarrow{p} 0$ as $N \to \infty$.

To highlight the relation between the CLT and the compact LIL in Banach spaces by means of Theorem 1, note that if the CLT holds, then condition (iii) of assertion b) is fulfilled, as described above. Also, condition (ii) follows from the CLT, as the limiting Gaussian random variable Z with the same covariance as X has a strong second moment, due to the integrability properties of Gaussian random variables. This implies that K, the unit ball of the reproducing kernel Hilbert space associated to X, is compact and that the family $\{g^2(X) : g \in B', \|g\| \leq 1\}$ is uniformly integrable, as remarked previously. Hence, necessary and sufficient conditions for the compact LIL in the presence of the CLT reduce to condition (i) of Theorem 1b), cf. Ledoux and Talagrand (1988), Corollary 1.3.

In the subsequent analysis, we will use the compact LIL to derive almost sure convergence rates, even though the bounded LIL, guaranteeing the $\mathbb{P}$-almost sure finiteness of $\Lambda(X)$, would be sufficient to establish most of our results. However, by working with the compact LIL, we find ourselves in the same setup in which the CLT and thus convergence rates in distribution have already been established. Another advantage of using the compact LIL in our setup is the ability to describe the set of limit points K by the $\mathbb{P}$-almost sure relation $\Lambda(X) = \sup_{y \in K} \|y\| = \sigma(X)$, allowing for a better interpretation. Finally, the compact LIL also leads to slightly better convergence rates in probability.

3 Rates of Convergence

In this section, we establish rates of convergence in the almost sure sense and in mean for the SAA setting introduced in Section 1. Since our results are closely related to rates of convergence in distribution, which have mainly been investigated within the asymptotic analysis of optimal values and solutions by Shapiro (1989, 1990, 1991), we first review the main results of these studies in Section 3.1. By use of the compact LIL in the Banach spaces $C(\mathcal{X})$ and $C^1(\mathcal{X})$, we then provide in Section 3.2 our main findings on almost sure rates of convergence for estimators of optimal values and solutions. Eventually, in Section 3.3, we infer from a characterisation of the compact LIL that these quantities also convergence in mean and derive the corresponding rates of convergence. In particular, these rates can be used to quantify the asymptotic bias of optimal estimators, and to obtain quantitative estimates of the bias without the additional (strong) assumption of uniform integrability.

3.1 Rates of Convergence in Distribution

On the space $C(\mathcal{X})$, we initially make the following assumptions with respect to the random function h:

(A1) For some $x_0 \in \mathcal{X}$ we have $\mathbb{E}_{\mathbb{P}}\big[h^2(x_0, \xi)\big] < \infty$.

(A2) There exists a measurable function $G : \Xi \to \mathbb{R}_+$ such that $\mathbb{E}_{\mathbb{P}}[G^2(\xi)] < \infty$ and

$$|h(x_1, \xi) - h(x_2, \xi)| \leq G(\xi)\|x_1 - x_2\|, \quad \forall x_1, x_2 \in \mathcal{X},$$

$\mathbb{P}$-almost surely.

Assumptions (A1) and (A2) imply that $\mathbb{E}_{\mathbb{P}}[h(x, \xi)]$ and $\mathbb{E}_{\mathbb{P}}[h^2(x, \xi)]$ are finite-valued for all $x \in \mathcal{X}$. Moreover, assumption (A2) provides that f is Lipschitz continuous on $\mathcal{X}$ and, as $\mathcal{X}$ is assumed to be compact, thus guarantees that the set of minimisers $\mathcal{X}^*$ of the original problem (1) is nonempty. Further, it follows from the compactness of $\mathcal{X}$ and assumption (A2) that $\hat{f}_N^*$ and $\widehat{\mathcal{X}}_N^*$ are measurable and that the latter set is nonempty, $\mathbb{P}$-almost surely, cf. Aliprantis and Border (2006), Theorem 18.19. Above all, a particular solution $\hat{x}_N^*$ of the SAA problem (2)

may thus be viewed as a measurable selection $\hat{x}_N^* \in \widehat{\mathcal{X}}_N^*$. Eventually, both assumptions (A1) and (A2) also imply that the variance of $h(x,\xi)$ compared to that of $h(x_0,\xi)$ can only grow as fast as the quadratic distance between x and x_0.

Note that assumptions (A1) and (A2) cover the following important special cases: (i) non-smooth convex optimisation over a convex compact set, (ii) smooth convex optimisation over a convex compact set, and finally (iii) smooth global optimisation over an arbitrary compact set. However, the treatment of unbounded domains is beyond our framework. In such a setting it would be more beneficial to directly analyse the necessary first order conditions. Further, methods like stochastic gradient methods are also not covered by our setting and, as we require Lipschitz continuity of h in x, indicator functions cannot be used as h either. Finally, note that in the specific case of a two-stage stochastic program with subordinate linear second-stage, (A1) and (A2) are typically satisfied if the second stage problem has a feasible set which is $\mathbb{P}$-almost surely contained in a sufficiently large compact set, see, e.g., Shapiro et al (2014), Chapter 2.

Most notably, assumptions (A1) and (A2) are sufficient to ensure that the $C(\mathcal{X})$-valued random variable $\widetilde{X} = h(\cdot,\xi) - \mathbb{E}_{\mathbb{P}}[h(\cdot,\xi)]$ satisfies the CLT in this Banach space, see Araujo and Giné (1980), Corollary 7.17. It thus holds

$$\sqrt{N}(\hat{f}_N - f) \xrightarrow{d} \widetilde{Z}, \text{ as } N \to \infty, \tag{7}$$

where $\widetilde{Z}$ denotes a $C(\mathcal{X})$-valued mean zero Gaussian random variable which is completely defined by the covariance of $\widetilde{X}$, that is by $(\mathbb{C}\mathrm{ov}\,\widetilde{X})(g_1,g_2) = \mathbb{E}_{\mathbb{P}}[g_1(\widetilde{X})g_2(\widetilde{X})]$ for $g_1, g_2 \in C(\mathcal{X})'$. Note that assertion (7) implies that $\{\hat{f}_N\}$ converges in distribution to f, at a rate of $1/\sqrt{N}$. In particular, for any fixed $x \in \mathcal{X}$, we have that $\{\sqrt{N}(\hat{f}_N(x) - f(x))\}$ converges in distribution to a real-valued normal distributed random variable $\widetilde{Z}(x)$ with mean zero and variance $\mathbb{E}_{\mathbb{P}}[h^2(x,\xi)] - \mathbb{E}_{\mathbb{P}}[h(x,\xi)]^2$.

3.1.1 Rate of Convergence of Optimal Values

Provided that $\{\sqrt{N}(\hat{f}_N - f)\}$ converges in distribution to a random variable $\widetilde{Z}$ with values in $C(\mathcal{X})$, the convergence in distribution of $\{\sqrt{N}(\hat{f}_N^* - f^*)\}$ can be assessed using a first order expansion of the optimal value function, see Shapiro (1991). To this end, let the minimum value function $\vartheta : C(\mathcal{X}) \to \mathbb{R}$ be defined by $\vartheta(\psi) := \inf_{x\in\mathcal{X}} \psi(x)$, i.e. $\hat{f}_N^* = \vartheta(\hat{f}_N)$ and $f^* = \vartheta(f)$. Since $\mathcal{X}$ is compact, the mapping ϑ is continuous and hence measurable with respect to the Borel σ-algebras $\mathcal{B}(C(\mathcal{X}))$ and $\mathcal{B}(\mathbb{R})$. Moreover, ϑ is Lipschitz continuous with constant one, i.e. $|\vartheta(\psi_1) - \vartheta(\psi_2)| \leq \|\psi_1 - \psi_2\|_\infty$ for any $\psi_1, \psi_2 \in C(\mathcal{X})$, and it can be shown that ϑ is directionally differentiable at f with

$$\vartheta'_f(\psi) = \inf_{x\in\mathcal{X}^*(f)} \psi(x), \quad \psi \in C(\mathcal{X}), \tag{8}$$

where $\mathcal{X}^*(f) = \mathcal{X}^* = \arg\min_{x\in\mathcal{X}} f(x)$, see Danskin's theorem (e.g., Danskin (1966)). For a general definition of directional differentiability and related notions as used hereinafter, we refer to Shapiro et al (2014), Section 7.2.8. By the Lipschitz continuity and directional differentiability, it then follows that ϑ is also directionally differentiable at f in the Hadamard sense, see, e.g., Shapiro et al (2014), Proposition 7.65. Hence, an application of the first order delta method for Banach spaces with ϑ to (7) yields the following result, cf. Shapiro (1991), Theorem 3.2.

Theorem 2 *(Shapiro, 1991). Suppose that assumptions (A1)–(A2) hold. Then,*

$$\sqrt{N}(\hat{f}_N^* - f^*) \xrightarrow{d} \vartheta'_f(\widetilde{Z}), \quad \text{as } N \to \infty, \tag{9}$$

where $\widetilde{Z}$ denotes the $C(\mathcal{X})$-valued mean zero Gaussian random variable as obtained by (7) in $C(\mathcal{X})$, and ϑ'_f is given by (8). In particular, if $\mathcal{X}^*(f) = \{x^*\}$ is a singleton, then

$$\sqrt{N}(\hat{f}_N^* - f^*) \xrightarrow{d} \widetilde{Z}(x^*), \quad \text{as } N \to \infty. \tag{10}$$

Formulas (9) and (10) specify the asymptotic distribution of $\{\sqrt{N}(\hat{f}_N^* - f^*)\}$, which is asymptotically normal if uniqueness of a minimiser x^* is assumed. Moreover, both formulas allow to deduce that the speed of convergence in distribution of $\{\hat{f}_N^*\}$ to f^* can be quantified by the rate $1/\sqrt{N}$.

3.1.2 Rate of Convergence of Optimal Solutions

Under more restrictive assumptions, it is possible to specify the rate of convergence of optimal solutions as well. The derivation of this result is essentially based on the CLT in the Banach space $C^1(\mathcal{X})$, to which the delta method with a second order expansion of the optimal value function ϑ is applied. This then provides a first order expansion for optimal solutions of the SAA problem.

For keeping our exposition on convergence of optimal solutions in this and the related Subsection 3.2.2 as comprehensive as possible, we follow the general approach of Shapiro (2000). In particular, we make the following additional assumptions on the underlying random function h and its gradient $\nabla_x h$, facilitating convergence in distribution in $C^1(\mathcal{X})$:

(A3) The function $h(\cdot, \xi)$ is continuously differentiable on $\mathcal{X}$, $\mathbb{P}$-almost surely.

and

(A1') For some $x_0 \in \mathcal{X}$ we have $\mathbb{E}_{\mathbb{P}}\left[\|\nabla_x h(x_0, \xi)\|^2\right] < \infty$.

(A2') The gradient $\nabla_x h(\cdot, \xi)$ is Lipschitz continuous with constant $G_\nabla(\xi)$ on $\mathcal{X}$, $\mathbb{P}$-almost surely, and $\mathbb{E}_{\mathbb{P}}[G_\nabla^2(\xi)] < \infty$.

Assumption (A3) implies that $\hat{f}_N$ is a random variable with values in $C^1(\mathcal{X})$, and assumptions (A1)–(A3) together imply that f is continuously differentiable on $\mathcal{X}$ and that $\nabla f(x) = \mathbb{E}_{\mathbb{P}}[\nabla_x h(x, \xi)]$ for $x \in \mathcal{X}$ (e.g., Shapiro et al (2014), Theorems 7.49 and 7.53). Moreover, all assumptions (A1)–(A3) and (A1')–(A2') entail that $\widetilde{X} = h(\cdot, \xi) - \mathbb{E}_{\mathbb{P}}[h(\cdot, \xi)]$ also satisfies the CLT in the Banach space $C^1(\mathcal{X})$, such that (7) holds for a $C^1(\mathcal{X})$-valued mean zero Gaussian random variable $\widetilde{Z}$.

Note that by considering the class $C^1(\mathcal{X})$ of continuously differentiable functions and assumptions (A1')–(A2'), we implicitly assume that the objective functions f and $\hat{f}_N$ and their gradients are sufficiently well-behaved. This presents a reasonable regularity condition in order to derive general rates of convergence. If an objective function does not meet these criteria, a similar deduction becomes considerably more difficult.

Aside from conditions on h and $\nabla_x h$, let us further consider the following regularity assumptions for the original problem (1):

(B1) The problem (1) has a unique optimal solution $x^* \in \mathcal{X}$.

(B2) The function f satisfies the second-order growth condition at x^*, i.e. there exists $\alpha > 0$ and a neighbourhood V of x^* such that

$$f(x) \geq f(x^*) + \alpha \|x - x^*\|^2, \quad \forall x \in \mathcal{X} \cap V.$$

(B3) The set $\mathcal{X}$ is second order regular at x^*.

(B4) The function f is twice continuously differentiable in a neighbourhood of the point x^*.

Assumptions (B1)–(B4) represent standard second order optimality conditions to be found in common literature on perturbation analysis of optimisation problems, see, e.g., Bonnans and Shapiro (2000). While assumptions (B1) and (B4) are self-explanatory, the growth condition in assumption (B2) involves that x^* is locally optimal and that f increases at least quadratically near x^*. This condition can be ensured to hold in several ways by assuming second order sufficient conditions, as given, for instance, in Section 3.3 of Bonnans and Shapiro (2000). Finally, the second order regularity of $\mathcal{X}$ in (B3) concerns the tangent set $T^2_{\mathcal{X}}(x^*, d)$ to $\mathcal{X}$ at x^* in direction d and guarantees that it is a sufficient good second order approximation to $\mathcal{X}$ in direction d. In the context of two-stage stochastic problems as mentioned in the introduction, note that sufficient conditions for assumptions (B1) and (B2) are rather problem-specific, while (B3) and (B4) are not often satisfied.

By imposing (B1)–(B4), a second order expansion of the minimal value function ϑ, now mapping $C^1(\mathcal{X})$ into $\mathbb{R}$, can be calculated, along with a first order expansion of the associated optimal solution function $\kappa : C^1(\mathcal{X}) \to \mathbb{R}^n$, where $\kappa(\psi) \in \arg\min_{x \in \mathcal{X}} \psi(x)$, $\psi \in C^1(\mathcal{X})$. More precisely, under (B1)–(B4), ϑ is shown to be first and second order Hadamard directionally differentiable at f, with $\vartheta'_f(\psi) = \psi(x^*)$ and

$$\vartheta''_f(\psi) = \inf_{d \in C_{x^*}} \Big\{ 2d^\top \nabla \psi(x^*) + d^\top \nabla^2 f(x^*) d + \inf_{w \in T^2_{\mathcal{X}}(x^*, d)} w^\top \nabla f(x^*) \Big\}, \qquad (11)$$

for $\psi \in C^1(\mathcal{X})$, and where C_{x^*} is the critical cone of problem (1), $\nabla^2 f(x^*)$ the Hessian matrix of f at x^*, and $T^2_{\mathcal{X}}(x^*, d)$ denotes the second order tangent set to $\mathcal{X}$ at x^* in direction d (see, e.g., Shapiro (2000), Theorem 4.1). Moreover, if the problem on the right-hand side of (11) admits a unique solution $d^*(\psi)$, then the mapping κ is also Hadamard directionally differentiable at f, and $\kappa'_f(\psi) = d^*(\psi)$ holds. Hence, using a second order delta method for ϑ on the convergence (7) in $C^1(\mathcal{X})$ provides the following asymptotic results for $\{\hat{f}^*_N\}$ and $\{\hat{x}^*_N\}$, cf. Shapiro (2000), Theorems 4.2 and 4.3. Note that $\{\hat{x}^*_N\}$ denotes any sequence of measurable selections $\hat{x}^*_N$ from the sets of optimal solutions $\widehat{\mathcal{X}}^*_N$, respectively.

Theorem 3 *(Shapiro, 2000). Suppose that assumptions (A1)–(A3), (A1')–(A2') and (B1)–(B4) hold. Then,*

$$N\big(\hat{f}^*_N - \hat{f}_N(x^*)\big) \xrightarrow{d} \tfrac{1}{2}\vartheta''_f(\widetilde{Z}), \quad \text{as } N \to \infty,$$

where $\widetilde{Z}$ denotes the $C^1(\mathcal{X})$-valued mean zero Gaussian random variable as obtained by (7) *in $C^1(\mathcal{X})$, and ϑ''_f is given by* (11)*. Further, suppose that for any*

$\psi \in C^1(\mathcal{X})$, the problem on the right-hand side of (11) *has a unique solution* $d^*(\psi)$*. Then,*

$$\sqrt{N}(\hat{x}_N^* - x^*) \xrightarrow{d} d^*(\widetilde{Z}), \quad \text{as } N \to \infty. \tag{12}$$

Remark 1 It has to be noted that assertion (12) yields the usual convergence rate for an optimal solution in distribution. This, however, does not directly imply any result on convergence in mean, nor on the bias. Although the expectation of the right-hand side is finite, this is not necessarily the case for the limit of the expectations of the upscaled difference of the optimal solutions on the left. The limit of the expectations of the left-hand side only exists and equals the expectation of the right-hand side if and only if the upscaled sequence is uniformly integrable, see, e.g., Serfling (1980), Theorem 1.4A. However, making such an assumption for $\{\sqrt{N}(\hat{x}_N^* - x^*)\}$ is actually already equivalent to imposing a convergence order of $\mathbf{O}(1/\sqrt{N})$ for $\{\hat{x}_N^* - x^*\}$ to zero in the L_1-sense.

3.2 Almost Sure Rates of Convergence

We now turn to almost sure convergence and first observe that in the specific case of $C(\mathcal{X})$-valued random variables, the compact LIL is satisfied under exactly the same assumptions as the CLT in the Banach space setting, see Kuelbs (1976a), Theorem 4.4. In this context, note that the compactness of the feasible set $\mathcal{X}$ is crucial. Given assumptions (A1) and (A2), we thus have for the $C(\mathcal{X})$-valued random variable $\widetilde{X} = h(\cdot, \xi) - \mathbb{E}_{\mathbb{P}}[h(\cdot, \xi)]$ and the related sequence of i.i.d. copies $\{\widetilde{X}_i\}$ that

$$\lim_{N\to\infty} \operatorname{dist}\left(\frac{\hat{f}_N - f}{b_N}, K_{\widetilde{X}}\right) = 0, \quad \text{and} \quad \operatorname{CP}\left(\left\{\frac{\hat{f}_N - f}{b_N}\right\}\right) = K_{\widetilde{X}},$$

each $\mathbb{P}$-almost surely, where $K_{\widetilde{X}}$ denotes the unit ball of the reproducing kernel Hilbert space $H_{\widetilde{X}}$ associated to the covariance of $\widetilde{X}$ and $K_{\widetilde{X}}$ is compact. In line with Section 2, it follows from this result that

$$\Lambda(\widetilde{X}) = \limsup_{N\to\infty} \frac{\|\hat{f}_N - f\|_\infty}{b_N} = \sigma(\widetilde{X}), \tag{13}$$

$\mathbb{P}$-almost surely, where $\sigma(\widetilde{X}) = \sup_{\|g\|\leq 1}(\mathbb{E}_{\mathbb{P}}[g^2(\widetilde{X})])^{1/2}$, $g \in C(\mathcal{X})'$. Now, by virtue of Riesz's representation theorem (e.g., Albiac and Kalton (2006), Theorem 4.1.1), the dual space $C(\mathcal{X})'$ can be identified with the space $M(\mathcal{X})$ of all finite Borel measures on the compact space $\mathcal{X}$, with total variation norm $\|\mu\| = |\mu|(\mathcal{X})$, $\mu \in M(\mathcal{X})$. Moreover, for $\mu \in M(\mathcal{X})$, the extreme points of the subset defined by $\|\mu\| \leq 1$ are the Dirac measures $\mu = \pm\delta_x$, where $\delta_x(\widetilde{X}) = \widetilde{X}(x)$ for a $C(\mathcal{X})$-valued random variable $\widetilde{X}$ (e.g., Albiac and Kalton (2006), Remark 8.2.6) and $x \in \mathcal{X}$. Hence,

$$\sigma(\widetilde{X}) = \sup_{x\in\mathcal{X}} \left(\mathbb{E}_{\mathbb{P}}\left[\widetilde{X}^2(x)\right]\right)^{1/2}, \tag{14}$$

which is finite-valued by assumption.

By definition of the limit superior, equation (13) implies the following observation, specifying the speed of convergence of the approximating objective function in the almost sure sense.

Lemma 1 *Suppose that assumptions (A1)–(A2) hold. Then, for any $\epsilon > 0$, there exists a finite random variable $N^* = N^*(\epsilon) \in \mathbb{N}$ such that*

$$\forall N \geq N^* : \quad \|\hat{f}_N - f\|_\infty \leq (1+\epsilon) b_N \sigma(\widetilde{X}), \tag{15}$$

$\mathbb{P}$-almost surely. Here, $\sigma(\widetilde{X})$ is given by (14) *for the $C(\mathcal{X})$-valued random variable $\widetilde{X}$.*

In particular, inequality (15) reveals that the almost sure convergence of $\{\hat{f}_N\}$ to f occurs at a rate of $\mathbf{O}(b_N)$, $b_N = \sqrt{2\,\mathrm{LLog}(N)}/\sqrt{N}$, which is only marginally slower than the rate $1/\sqrt{N}$ obtained from convergence in distribution. To get an idea for the scale involved, note that $\sqrt{\log(\log(10^{99}))} \approx 2.33$. Yet, unlike the rate $1/\sqrt{N}$, the rate b_N holds $\mathbb{P}$-almost surely, which is a different notion of convergence than convergence in distribution. Although not explicitly stated in Lemma 1, let us emphasise that as the compact LIL holds, we also know that the almost sure rate of convergence of $\{\hat{f}_N\}$ to f is exactly b_N and cannot be faster.

Remark 2 Note that it is not possible to exactly determine the value of the finite random time

$$N^*(\epsilon) := \inf\left\{n \in \mathbb{N} \,|\, \forall k \geq n : \|\hat{f}_k - f\|_\infty \leq (1+\epsilon) b_k \sigma(\widetilde{X})\right\}$$

or the related last exit time

$$\tau^*(\epsilon) := \sup\left\{n \in \mathbb{N} \,|\, \|\hat{f}_n - f\|_\infty > (1+\epsilon) b_n \sigma(\widetilde{X})\right\}$$

(if $\hat{f}_1$ and $\hat{f}_2$ are not identical to f, then $N^*(\epsilon) = \tau^*(\epsilon) + 1$), as this depends on the particular realisation of the underlying random sequence $\{\xi_i\}$. Yet, the last exit time $\tau^*(\epsilon)$ may be linked to the counting variable

$$J^*(\epsilon) := \left|\left\{n \in \mathbb{N} : \ \|\hat{f}_n - f\|_\infty > (1+\epsilon) b_n \sigma(\widetilde{X})\right\}\right|$$

by $\tau^*(\epsilon) \geq J^*(\epsilon)$, of which we know that $\mathbb{E}_{\mathbb{P}}[(J^*)^\lambda] = \infty$ for any $\lambda > 0$ if f is a real-valued object, cf. Slivka (1969). This might be taken as a strong indication that a similar result also holds in the Banach space case, telling us that the *asymptotic rate* only holds for very large N.

3.2.1 Rate of Convergence of Optimal Values

Once having assertion (15), the rate of convergence of the optimal value $\{\hat{f}_N^*\}$ to f^* is easily obtained by recalling the Lipschitz continuity of the continuous minimum value function $\vartheta(\psi) = \inf_{x \in \mathcal{X}} \psi(x)$, with $\hat{f}_N^* = \vartheta(\hat{f}_N)$ and $f^* = \vartheta(f)$. We thus have the following result, in analogy to Theorem 2.

Theorem 4 *Suppose that assumptions (A1)–(A2) hold. Then,*

$$\forall N \in \mathbb{N} : \quad |\hat{f}_N^* - f^*| \leq \|\hat{f}_N - f\|_\infty.$$

In particular, it holds that $\{\hat{f}_N^*\}$ converges to f^*, $\mathbb{P}$-almost surely, at a rate of $\boldsymbol{O}(b_N)$.

3.2.2 Rate of Convergence of Optimal Solutions

Next, we proceed with analysing the rate of convergence of optimal solutions in the almost sure sense. Considering the space $C(\mathcal{X})$ of continuous functions on $\mathcal{X}$, we note first of all that if the random function h only satisfies the moment and Lipschitz conditions (A1) and (A2), respectively, then a slower rate of almost sure convergence can be obtained under the regularity conditions (B1) and (B2), as the following proposition shows.

Proposition 1 *Suppose that assumptions (A1)–(A2) and (B1)–(B2) hold. Then, there exists a finite random variable $N^* \in \mathbb{N}$ such that*

$$\forall N \geq N^* : \quad \|\hat{x}_N^* - x^*\|^2 \leq \frac{2}{\alpha}\|\hat{f}_N - f\|_\infty, \tag{16}$$

$\mathbb{P}$-almost surely. In particular, it holds that $\{\hat{x}_N^*\}$ converges to x^*, $\mathbb{P}$-almost surely, at a rate of $\boldsymbol{O}(\sqrt{b_N})$.

Proof By assumptions (A1)–(A2) and (B1), $\{\hat{x}_N^*\}$ converges to x^*, $\mathbb{P}$-almost surely, for $N \to \infty$ (e.g., Shapiro et al (2014), Theorems 5.3 and 7.53). This implies that $\hat{x}_N^* \in V$ holds $\mathbb{P}$-almost surely for $N \geq N^*$, for some finite random $N^* \in \mathbb{N}$. Hence, the second-order growth condition (B2) at x^* with $\alpha > 0$ yields

$$\begin{aligned}
\|\hat{x}_N^* - x^*\|^2 &\leq \frac{1}{\alpha}\big(f(\hat{x}_N^*) - f(x^*)\big) \\
&\leq \frac{1}{\alpha}\big(f(\hat{x}_N^*) - \hat{f}_N(\hat{x}_N^*) + \hat{f}_N(\hat{x}_N^*) - f(x^*)\big) \\
&\leq \frac{1}{\alpha}\big(f(\hat{x}_N^*) - \hat{f}_N(\hat{x}_N^*) + \hat{f}_N(x^*) - f(x^*)\big) \\
&\leq \frac{1}{\alpha}\big(\big|\hat{f}_N(\hat{x}_N^*) - f(\hat{x}_N^*)\big| + \big|\hat{f}_N(x^*) - f(x^*)\big|\big) \\
&\leq \frac{2}{\alpha}\|\hat{f}_N - f\|_\infty,
\end{aligned}$$

where $\hat{f}_N(\hat{x}_N^*)$ has been added and subtracted from the first line to the second and $\hat{f}_N(\hat{x}_N^*) \leq \hat{f}_N(x^*)$ has been used from the second line to the third. This proves (16), and the remaining assertion then follows from Lemma 1. □

To achieve a faster rate of almost sure convergence, stronger assumptions on h and the gradient $\nabla_x h$ in the subspace $C^1(\mathcal{X})$ are required, as described in Section 3.1.2 for convergence in distribution. Specifically, if, in addition to assumptions (A1) and (A2), we assume that h is also continuously differentiable on $\mathcal{X}$, i.e. assumption (A3) holds, then f is an element of the Banach space $C^1(\mathcal{X})$ and $\hat{f}_N$ is $C^1(\mathcal{X})$-valued. Consequently, on condition that the moment and Lipschitz assumptions of $\nabla_x h$ in (A1') and (A2'), respectively, are also fulfilled, $\widetilde{X}$ satisfies the compact LIL in $C^1(\mathcal{X})$ and we can state the following, cf. Lemma 1.

Lemma 2 *Suppose that assumptions (A1)–(A3) and (A1')–(A2') hold. Then, for any $\epsilon > 0$, there exists a finite random variable $N^* = N^*(\epsilon) \in \mathbb{N}$ such that*

$$\forall N \geq N^* : \quad \|\hat{f}_N - f\|_{1,\infty} \leq (1+\epsilon) b_N \sigma(\widetilde{X}),$$

$\mathbb{P}$-almost surely. Here, $\sigma(\widetilde{X})$ is given in general form by (4) for the $C^1(\mathcal{X})$-valued random variable $\widetilde{X}$.

Moreover, we further consider the regularity assumptions (B1) and (B2) on the original problem (1), where we marginally strengthen the latter according to:

(B2') The function f satisfies the second-order growth condition at x^*, i.e. there exists $\alpha > 0$ and a neighbourhood V of x^* such that

$$f(x) \geq f(x^*) + \alpha \|x - x^*\|^2, \quad \forall x \in \mathcal{X} \cap V.$$

Further, V can be chosen such that $\mathcal{X} \cap V$ is star-shaped with center x^*.

We are then able to derive the following result on the speed of convergence of optimal solutions of the SAA problem. Note that this result holds in parallel to Theorem 3 in the almost sure case.

Theorem 5 *Suppose that assumptions (A1)–(A3), (A1')–(A2'), (B1) and (B2') hold. Then, there exists a finite random variable $N^* \in \mathbb{N}$ such that*

$$\forall N \geq N^* : \quad \|\hat{x}_N^* - x^*\| \leq \frac{1}{\alpha} \|\hat{f}_N - f\|_{1,\infty}. \tag{17}$$

$\mathbb{P}$-almost surely. In particular, it holds that $\{\hat{x}_N^*\}$ converges to x^*, $\mathbb{P}$-almost surely, at a rate of $\boldsymbol{O}(b_N)$.

Proof Again, by assumptions (A1)–(A2) and (B1), $\{\hat{x}_N^*\}$ converges to x^*, $\mathbb{P}$-almost surely, for $N \to \infty$. This implies that $\hat{x}_N^* \in V$ holds $\mathbb{P}$-almost surely for $N \geq N^*$, for some finite random $N^* \in \mathbb{N}$. Hence, the second-order growth condition (B2') at x^* with $\alpha > 0$ yields

$$\begin{aligned}
\|\hat{x}_N^* - x^*\|^2 &\leq \frac{1}{\alpha}\big(f(\hat{x}_N^*) - f(x^*)\big) \\
&\leq \frac{1}{\alpha}\big(f(\hat{x}_N^*) - \hat{f}_N(\hat{x}_N^*) + \hat{f}_N(\hat{x}_N^*) - f(x^*)\big) \\
&\leq \frac{1}{\alpha}\big(f(\hat{x}_N^*) - \hat{f}_N(\hat{x}_N^*) + \hat{f}_N(x^*) - f(x^*)\big) \\
&\leq \frac{1}{\alpha}\big(f(\hat{x}_N^*) - \hat{f}_N(\hat{x}_N^*) - \big(f(x^*) - \hat{f}_N(x^*)\big)\big),
\end{aligned}$$

and therefore

$$\|\hat{x}_N^* - x^*\| \leq \frac{\big|f(\hat{x}_N^*) - \hat{f}_N(\hat{x}_N^*) - \big(f(x^*) - \hat{f}_N(x^*)\big)\big|}{\alpha \|\hat{x}_N^* - x^*\|}.$$

Since $(f - \hat{f}_N)$ is assumed to be differentiable on $\mathcal{X}$, $\mathbb{P}$-almost surely, and $\mathcal{X} \cap V$ is star-shaped with centre x^*, it further holds by the mean value theorem (e.g., Dieudonné (1960), Theorem 8.5.4) that

$$\begin{aligned}
&\frac{\big|f(\hat{x}_N^*) - \hat{f}_N(\hat{x}_N^*) - \big(f(x^*) - \hat{f}_N(x^*)\big)\big|}{\alpha \|\hat{x}_N^* - x^*\|} \\
&\quad \leq \frac{1}{\alpha} \sup_{0 \leq t \leq 1} \big\|\nabla\big(f(\hat{x}_N^* + t(x^* - \hat{x}_N^*)) - \hat{f}_N(\hat{x}_N^* + t(x^* - \hat{x}_N^*))\big)\big\| \\
&\quad \leq \frac{1}{\alpha} \sup_{x \in \mathcal{X} \cap V} \big\|\nabla\big(f(x) - \hat{f}_N(x)\big)\big\|.
\end{aligned}$$

Thus, by definition of the norm $\|\cdot\|_{1,\infty}$, the latter then provides inequality (17), and applying Lemma 2 yields the statement on the rate of convergence. □

Note that the results established in Proposition 1 and Theorem 5 require fewer assumptions on the objective function f than the corresponding Theorem 3 on convergence in distribution, while providing almost sure convergence instead of convergence in distribution. This becomes most notable in that the former results are able to dispense with assumptions (B3) and (B4), while these are necessary for the second order Hadamard directional derivative ϑ_f'' in the latter. In particular, we are thus able to deal with an optimal solution on the boundary of the feasible set $\mathcal{X}$ without requiring any regularity condition for $\mathcal{X}$.

It is to be expected from the above analysis that improved almost sure convergence rates for the difference of the optimal values might be obtained in a similar manner as for convergence in distribution by the second order delta method under analogous assumptions. We leave this question for future research, and instead focus on rates of convergence in mean in the following.

3.3 Rates of Convergence in Mean

By recalling that the $C(\mathcal{X})$-valued random variable $\widetilde{X} = h(\cdot, \xi) - \mathbb{E}_{\mathbb{P}}[h(\cdot, \xi)]$ satisfies the compact LIL under assumptions (A1) and (A2), we can apply Kuelbs's equivalence (6) (cf. also Kuelbs (1977), Theorem 4.1) to obtain

$$\mathbb{E}_{\mathbb{P}}\left[\left\|\sum_{i=1}^{N} \widetilde{X}_i\right\|_{\infty}\right] = \mathbf{o}(a_N).$$

This, in turn, directly leads to the following proposition.

Proposition 2 *Suppose that assumptions (A1)–(A2) hold. Then,*

$$\lim_{N\to\infty} \mathbb{E}_{\mathbb{P}}\left[\frac{\|\hat{f}_N - f\|_{\infty}}{b_N}\right] = 0, \tag{18}$$

i.e. $\mathbb{E}_{\mathbb{P}}[\|\hat{f}_N - f\|_{\infty}] = \boldsymbol{o}(b_N)$*, and in particular* $\{\hat{f}_N\}$ *converges to* f *in* $L_1(C(\mathcal{X}))$ *at a rate of* $\boldsymbol{o}(b_N)$*.*

Let us emphasise that Proposition 2 constitutes an important and novel result which can be directly obtained from the compact LIL without any further technicalities. To the best of our knowledge, only the convergence in mean of $\{\hat{f}_N\}$ to f, i.e. $\mathbb{E}_{\mathbb{P}}[\|\hat{f}_N - f\|_{\infty}] \to 0$ as $N \to \infty$, was known thus far, albeit without specifying any rate. Such a result may be obtained, for instance, by convergence in distribution of $\{\hat{f}_N\}$ to f and additional assumptions like (A1) and (A2), yielding uniform integrability of the sequence $\{\hat{f}_N\}$. Deducing uniform integrability for an upscaled sequence like $\{(\hat{f}_N - f)/b_N\}$, however, is not possible in such a way, which renders the above result (18) and its implication even more noteworthy.

3.3.1 Rate of Convergence of Optimal Values and Biasedness

By the Lipschitz continuity of the minimum value function $\vartheta(\psi) = \inf_{x\in\mathcal{X}} \psi(x)$, $\psi \in C(\mathcal{X})$, and Proposition 2, we immediately arrive at the corresponding result for the convergence of optimal values.

Theorem 6 *Suppose that assumptions (A1)–(A2) hold. Then,* $\{\hat{f}_N^*\}$ *converges to* f^* *in* L_1*, and* $\mathbb{E}_{\mathbb{P}}[|\hat{f}_N^* - f^*|] = \boldsymbol{o}(b_N)$*. In particular, one has that the bias of* $\hat{f}_N^*$ *vanishes at the same rate, i.e.* $|\mathbb{E}_{\mathbb{P}}[\hat{f}_N^*] - f^*| = \boldsymbol{o}(b_N)$*.*

As we have seen, Theorem 6 states that $\hat{f}_N^*$ is an asymptotically unbiased estimator of f^* and that the bias $\mathbb{E}[\hat{f}_N^*] - f^*$ is of order $\mathbf{o}(b_N)$. In contrast to classical results, cf. Shapiro et al (2014), pp. 185, these results on the bias do not need the additional strong assumption of uniform integrability of the sequence $\{\sqrt{N}(\hat{f}_N^* - f^*)\}$. Instead, one deduces here directly that $\{\sqrt{N}/\sqrt{2\,\mathrm{LLog}(N)}(\hat{f}_N^* - f^*)\}$ is uniformly integrable (as it is convergent in L_1, see, e.g., Bauer (2001), Theorem 21.4).

We also want to stress the fact that Theorem 6 puts non-technical discussions about the asymptotic bias of the optimal value on a sound theoretical basis, cf. Homem-de-Mello and Bayraksan (2014), and especially their discussion following Example 8 in Section 2.2.

Remark 3 The well-known fact that $\mathbb{E}_{\mathbb{P}}[\hat{f}_N^*] \leq \mathbb{E}_{\mathbb{P}}[\hat{f}_{N+1}^*] \leq f^*$ for any $N \in \mathbb{N}$, cf. Mak et al (1999), can be combined with the above proposition to obtain that for any $\epsilon > 0$, there exists an $N^* = N^*(\epsilon)$ such that

$$\forall N \geq N^* : \quad \mathbb{E}_{\mathbb{P}}\big[\hat{f}_N^*\big] \leq \mathbb{E}_{\mathbb{P}}\big[\hat{f}_{N+1}^*\big] \leq f^* \leq \mathbb{E}_{\mathbb{P}}\big[\hat{f}_N^*\big] + \epsilon b_N, \tag{19}$$

which brackets the unknown optimal value f^* in a interval of known size.

Remark 4 Under the additional assumption that $f^* > 0$, one can obtain further insight into the speed at which $\mathbb{E}_{\mathbb{P}}[\hat{f}_N^*]$ approaches f^*. For this purpose, let us first observe that

$$\begin{aligned}\hat{f}_{N+1}^* \leq \hat{f}_{N+1}(\hat{x}_N^*) &= \frac{1}{N+1}\sum_{i=1}^{N+1} h(\hat{x}_N^*, \xi_i) \\ &= \frac{N}{N+1}\hat{f}_N^* + \frac{1}{N+1}h(\hat{x}_N^*, \xi_{N+1}).\end{aligned}$$

Taking expectations on both sides and using the fact that $\mathbb{E}_{\mathbb{P}}[\hat{f}_N^*] > 0$ for sufficiently large N (as $f^* > 0$), we then arrive at

$$\mathbb{E}_{\mathbb{P}}\big[\hat{f}_{N+1}^*\big] \leq \mathbb{E}_{\mathbb{P}}\big[\hat{f}_N^*\big] + \frac{c_1}{N+1},$$

with the constant $c_1 := \mathbb{E}_{\mathbb{P}}[\|h(\cdot, \xi_N)\|_\infty]$. In summary, we thus have derived an upper bound for the difference of subsequent expected minimum function values, showing that these expected values grow at most at a logarithmic speed.

It is also of importance to consider the second moment of $\hat{f}_N^*$, e.g., for constructing confidence intervals for f^* or for bounding the optimality gap, cf. Mak et al (1999), Section 3. To obtain such results, a version of the CLT is usually invoked to estimate $\mathbb{E}_{\mathbb{P}}[\hat{f}_N^*]$, which is only valid under the assumption that $\mathbb{E}_{\mathbb{P}}[(\hat{f}_N^*)^2] < \infty$. However, while the latter is often (implicitly) assumed and not treated explicitly, e.g. Mak et al (1999), formula (6), only Homem-de-Mello and Bayraksan (2014), Section 4.1, seems to carefully consider the finiteness of the second moment of $\hat{f}_N^*$. To the best of our knowledge, there is no result known on the asymptotic behaviour of $\mathbb{E}[(\hat{f}_N^*)^2]$. The following proposition closes this gap by providing an asymptotic rate on the standard deviation $\mathrm{std}(\hat{f}_N^*)$ of $\hat{f}_N^*$.

Proposition 3 *Suppose that assumptions (A1)–(A2) hold. Then*

$$\operatorname{std}\left(\hat{f}_N^*\right) \leq \|\hat{f}_N^*\|_{L_2} < \infty. \tag{20}$$

Further, if in addition assumptions (B1)–(B2) are satisfied and if there exists $\gamma_\delta := \mathbb{E}_{\mathbb{P}}[G(\xi)^{2+\delta}] < \infty$ *for some* $\delta > 0$*, then, with* $p = 2 + 4/\delta$*, it holds*

$$\operatorname{std}\left(\hat{f}_N^*\right) = \boldsymbol{o}\big(b_N^{1/p}\big). \tag{21}$$

Proof Let us start by considering the inequality

$$\|\hat{f}_N^*\|_{L_2} \leq \|\hat{f}_N(\hat{x}_N^*) - \hat{f}_N(x^*)\|_{L_2} + \|\hat{f}_N(x^*)\|_{L_2}.$$

Using the Lipschitz continuity of h, we have for the first term on the right-hand side that

$$\begin{aligned}\big\|\hat{f}_N(\hat{x}_N^*) - \hat{f}_N(x^*)\big\|_{L_2} &\leq \big\|\frac{1}{N}\sum_{i=1}^N G(\xi_i)\,\|\hat{x}_N^* - x^*\|\big\|_{L_2} \\ &\leq \operatorname{diam}(\mathcal{X})\,\big\|\frac{1}{N}\sum_{i=1}^N G(\xi_i)\big\|_{L_2},\end{aligned}$$

where $\operatorname{diam}(\mathcal{X}) := \sup\{\|x_1 - x_2\| : x_1, x_2 \in \mathcal{X}\}$ denotes the finite diameter of $\mathcal{X}$. For the second term, we easily get

$$\|\hat{f}_N(x^*)\|_{L_2}^2 = \mathbb{V}\mathrm{ar}\left(\hat{f}_N(x^*)\right) + \mathbb{E}_{\mathbb{P}}\big[\hat{f}_N(x^*)\big]^2 = \frac{1}{N}\,\mathbb{V}\mathrm{ar}\left(h(x^*,\xi)\right) + (f^*)^2,$$

such that we obtain assertion (20) under the respective assumptions (A1)–(A2).

To prove (21), we use the subadditivity of the standard deviation to get

$$\operatorname{std}\left(\hat{f}_N^*\right) \leq \operatorname{std}\left(\hat{f}_N(\hat{x}_N^*) - \hat{f}_N(x^*)\right) + \operatorname{std}\left(\hat{f}_N(x^*)\right). \tag{22}$$

For the first term on the right-hand side of inequality (22) we proceed as above, but eventually apply the generalised Hölder inequality to obtain

$$\begin{aligned}\operatorname{std}\left(\hat{f}_N(\hat{x}_N^*) - \hat{f}_N(x^*)\right) &\leq \big\|\hat{f}_N(\hat{x}_N^*) - \hat{f}_N(x^*)\big\|_{L_2} \\ &\leq \big\|\frac{1}{N}\sum_{i=1}^N G(\xi_i)\,\|\hat{x}_N^* - x^*\|\big\|_{L_2} \\ &\leq \big\|\frac{1}{N}\sum_{i=1}^N G(\xi_i)\big\|_{L_{2+\delta}}\,\big\|\|\hat{x}_N^* - x^*\|\big\|_{L_p}.\end{aligned}$$

The first factor of the latter expression can then be bounded according to the assumption by

$$\big\|\frac{1}{N}\sum_{i=1}^N G(\xi_i)\big\|_{L_{2+\delta}}^{2+\delta} \leq \gamma_\delta,$$

while for the second factor it holds $\big\|\|\hat{x}_N^* - x^*\|\big\|_{L_p} = \mathbf{o}(b_N^{1/p})$, due to Theorem 7 (as proved independently in the next subsection). For the second term on the right-hand side of inequality (22), it holds $\operatorname{std}(\hat{f}_N(x^*)) = 1/\sqrt{N}\,\operatorname{std}(h(x^*,\xi))$, as seen above. Hence, in summary, we obtain $\operatorname{std}(\hat{f}_N^*) = \mathbf{o}(b_N^{1/p})$, which proves (21). □

An immediate consequence of this proposition is the important insight that the standard deviation of $\hat{f}_N^*$ converges to zero for N to infinity. This implies that the mean squared error of $\hat{f}_N^*$ converges to zero at a known rate, thus implying L_2-convergence of $\{\hat{f}_N^*\}$ to f^*.

A further implication of the above proposition is related to the optimality gap: the upper bound of the usual confidence set to bound the optimality gap converges to zero for sufficiently large N, see Section 4.2.1 for more details.

3.3.2 Rate of Convergence of Optimal Solutions

Finally, if assumptions (A1)–(A2) are met together with (B1)–(B2), then convergence of optimal solutions $\{\hat{x}_N^*\}$ to x^* in any L_p, $1 \leq p < \infty$, is easily obtained.

Proposition 4 *Suppose that assumptions (A1)–(A2) and (B1)–(B2) hold. Then, $\{\hat{x}_N^*\}$ converges to x^* in L_p, $1 \leq p < \infty$, i.e.*

$$\mathbb{E}_{\mathbb{P}}\left[\|\hat{x}_N^* - x^*\|^p\right] \to 0, \quad \text{as } N \to \infty.$$

In particular, this implies that $\hat{x}_N^*$ is an asymptotically unbiased estimator for x^* and that the mean squared error $\mathbb{E}_{\mathbb{P}}[\|\hat{x}_N^* - x^*\|^2]$ vanishes asymptotically.

Proof From Proposition 1, we know that $\{\hat{x}_N^*\}$ converges to x^*, $\mathbb{P}$-almost surely, i.e. for each $1 \leq p < \infty$, we have $\|\hat{x}_N^* - x^*\|^p \to 0$, $\mathbb{P}$-almost surely. Further, due to compactness of $\mathcal{X}$, we have $\|\hat{x}_N^* - x^*\|^p \leq \operatorname{diam}(\mathcal{X})^p$. The main statement now follows directly from Lebesgue's dominated convergence theorem (e.g., Serfling (1980), Theorem 1.3.7). The remaining statements are easy consequences. □

Remark 5 The above proposition relies on the initially made assumption that the set $\mathcal{X}$ is compact. Considering unbounded $\mathcal{X}$, it is quite easy to construct a counterexample to the above result. More specifically, one can construct a uniformly convex quadratic objective function, where optimal solutions still converge almost surely but not in mean.

To further derive the corresponding rates for the convergence of $\{\hat{x}_N^*\}$ to x^* in L_p, we additionally require the following lemma. It quantifies the probability that $\hat{x}_N^*$ lies outside the set V of the second-order growth condition (B2), in terms of the rate b_N.

Lemma 3 *Suppose that assumptions (A1)–(A2) and (B1)–(B2) hold. Then, there exists a $\delta > 0$ (depending on V), such that for all $x \in \mathcal{X}$,*

$$f(x) < f(x^*) + \delta \quad \Rightarrow \quad x \in V.$$

Further, it holds

$$\mathbb{P}(\hat{x}_N^* \notin V) = \boldsymbol{o}(b_N).$$

Proof We prove the first statement by contradiction, assuming that there exists no such δ. Then we can find a sequence $\{\delta_N\}$ which converges monotonically to 0, together with a sequence $\{x_N\} \in \mathcal{X}\backslash V$ with $f(x_N) < f(x^*) + \delta_N$. As $\mathcal{X}\backslash V$ is compact, the sequence has a least one cluster point $\bar{x} \neq x^*$ with $f(\bar{x}) \leq f(x^*)$. This, however, yields the contradiction to the uniqueness of x^*, as assumed by (B1).

Now, let us consider the following chain of inequalities

$$\begin{aligned}\mathbb{P}(\hat{x}_N^* \notin V) &\leq \mathbb{P}\big(f(\hat{x}_N^*) - f(x^*) \geq \delta\big) \\ &= \mathbb{P}\big(f(\hat{x}_N^*) - \hat{f}_N(\hat{x}_N^*) + \hat{f}_N(\hat{x}_N^*) - f(x^*) \geq \delta\big) \\ &\leq \mathbb{P}\big(|f(\hat{x}_N^*) - \hat{f}_N(\hat{x}_N^*)| + |\hat{f}_N(x^*) - f(x^*)| \geq \delta\big) \\ &\leq \mathbb{P}\big(2\|f - \hat{f}_N\|_\infty \geq \delta\big) \\ &\leq \frac{2\,\mathbb{E}_\mathbb{P}\big[\|f - \hat{f}_N\|_\infty\big]}{\delta},\end{aligned}$$

where we have used Markov's inequality in the last step. Proposition 2 now yields the claim. □

Finally, we are now in position to state the following result on rates of convergence in L_p for optimal solutions.

Theorem 7 *Suppose that assumptions (A1)–(A2) and (B1)–(B2) hold. Then, $\{\hat{x}_N^*\}$ converges to x^* in L_1 at a rate of $\boldsymbol{o}(\sqrt{b_N})$ and in L_p, $2 \leq p < \infty$, at a rate of $\boldsymbol{o}(b_N)$, i.e.*

$$\mathbb{E}_\mathbb{P}\left[\frac{\|\hat{x}_N^* - x^*\|}{\sqrt{b_N}}\right] \to 0, \quad \text{and} \quad \mathbb{E}_\mathbb{P}\left[\frac{\|\hat{x}_N^* - x^*\|^p}{b_N}\right] \to 0,$$

respectively, as $N \to \infty$.

Moreover, if assumptions (A1)–(A3), (A1')–(A2'), (B1) and (B2') are satisfied, then the rate for convergence in L_1 is $\boldsymbol{o}(b_N)$.

Proof Under the assumptions (A1)–(A2) and (B1)–(B2), we only need to prove the statement for $p = 2$. The case $p > 2$ follows from the case $p = 2$ using $\|\hat{x}_N^* - x^*\| \leq \operatorname{diam}(\mathcal{X})$; the case $p = 1$ follows directly from Hölder's inequality. Accordingly, in analogy to the proof of Proposition 1, we have

$$\begin{aligned}\mathbb{E}_\mathbb{P}\left[\frac{\|\hat{x}_N^* - x^*\|^2}{b_N}\right] &= \mathbb{E}_\mathbb{P}\left[\frac{\|\hat{x}_N^* - x^*\|^2}{b_N}\mathbf{1}_{\{\hat{x}_N^* \in V\}}\right] + \mathbb{E}_\mathbb{P}\left[\frac{\|\hat{x}_N^* - x^*\|^2}{b_N}\mathbf{1}_{\{\hat{x}_N^* \notin V\}}\right] \\ &\leq \frac{2}{\alpha}\mathbb{E}_\mathbb{P}\left[\frac{\|\hat{f}_N - f\|_\infty}{b_N}\mathbf{1}_{\{\hat{x}_N^* \in V\}}\right] + \mathbb{E}_\mathbb{P}\left[\frac{\|\hat{x}_N^* - x^*\|^2}{b_N}\mathbf{1}_{\{\hat{x}_N^* \notin V\}}\right] \\ &\leq \frac{2}{\alpha}\mathbb{E}_\mathbb{P}\left[\frac{\|\hat{f}_N - f\|_\infty}{b_N}\right] + \mathbb{E}_\mathbb{P}\left[\frac{\|\hat{x}_N^* - x^*\|^2}{b_N}\mathbf{1}_{\{\hat{x}_N^* \notin V\}}\right].\end{aligned}$$

The first term of the latter expression already shows the proposed rate according to Proposition 2. For the second term, we use

$$\mathbb{E}_\mathbb{P}\left[\frac{\|\hat{x}_N^* - x^*\|^2}{b_N}\mathbf{1}_{\{\hat{x}_N^* \notin V\}}\right] \leq \operatorname{diam}(\mathcal{X})^2\frac{\mathbb{P}(\hat{x}_N^* \notin V)}{b_N}$$

which, together with Lemma 3, shows the claim for convergence in L_2.

Finally, assuming (A1)–(A3), (A1')–(A2'), (B1) and (B2'), the stronger rate of $\mathbf{o}(b_N)$ can be obtained analogously for convergence in L_1, cf. Theorem 5. □

4 Further Implications

In addition to the previous section on rates of convergence that hold almost surely and in mean, we now derive some further results from our analysis of the LIL in Banach spaces. Specifically, by exploiting the obtained rates of convergence in mean, we first infer in Section 4.1 rates of convergence in probability for the sequences of optimal estimators as well as (slow) rates of error probabilities under considerably mild conditions. This is opposed to other approaches yielding (fast) exponential rates of convergence but relying on a strong exponential moment condition (or boundedness condition). In Section 4.2, we then provide novel insights into the size of the optimality gap and show, most importantly, that the confidence set for the optimality gap ultimately converges to zero at a known rate in the almost sure sense. We also reconsider more traditional confidence sets for the optimal value and the optimal solution, and discuss their validity and potential to form universal confidence sets.

4.1 Convergence in Probability

From the well-known fact that almost sure convergence implies convergence in probability, all convergence rates obtained in Section 3.2 also hold in probability. However, slightly better convergence results can be obtained by making use of the rates of convergence in mean (or equivalently, by equivalence (5) of the compact LIL), see Section 4.1.1. By referring to related results from the literature on the LIL in Banach spaces, Section 4.1.2 provides some further insights into the asymptotic behaviour of error probabilities. The main difference between the first and the second subsection is that the former considers rates for the size of the deviation corridor (i.e. inside the probability), whereas the latter is concerned with rates of a fixed deviation probability (i.e. outside the probability).

4.1.1 Rates of Convergence in Probability

Applying the results from Section 3.3 immediately yields the following result.

Proposition 5 *Suppose that assumptions (A1)–(A2) hold and let $\delta > 0$ be arbitrary. Then,*

$$\mathbb{P}\left(\frac{|\hat{f}_N^* - f^*|}{b_N} > \delta \right) \to 0, \quad \text{as } N \to \infty.$$

Further, if in addition assumptions (B1)–(B2) are satisfied, then we have

$$\mathbb{P}\left(\frac{\|\hat{x}_N^* - x^*\|^2}{b_N} > \delta \right) \to 0, \quad \text{as } N \to \infty.$$

Finally, if assumptions (A1)–(A3), (A1')–(A2'), (B1) and (B2') are satisfied, then it holds

$$\mathbb{P}\left(\frac{\|\hat{x}_N^* - x^*\|}{b_N} > \delta \right) \to 0, \quad \text{as } N \to \infty. \tag{23}$$

Proof The results follow straightforwardly from Proposition 2, and Theorems 6 and 7. □

Note that, by Theorems 2 and 3 on the asymptotic distribution of $\{\sqrt{N}(\hat{f}_N^* - f^*)\}$ and $\{\sqrt{N}(\hat{x}_N^* - x^*)\}$, it immediately follows under the respective assumptions that the sequences are also bounded in probability as N tends to infinity. Also, considering the case of optimal solutions under the assumptions (A1)–(A3), (A1')–(A2'), (B1) and (B2'), it is already possible to infer from assertion (23) of the above proposition that $\{(\hat{x}_N^* - x^*)/b_N\}$ is bounded in probability, thus providing a slightly weaker rate under weaker assumptions.

Remark 6 In addition to Proposition 5, the rate of convergence in probability obtained from the compact LIL may be further characterised in terms of sums of probabilities, see, e.g., Li (1991) or Li et al (2007). In particular, given assumptions (A1) and (A2), it follows from Corollary 2.1 in Li (1991) that the sequence $\{(\hat{f}_N - f)/b_N\}$ must also satisfy for all $\delta > \sigma(\widetilde{\mathcal{X}})$ that

$$\sum_{N=1}^{\infty} \frac{\mathrm{LLog}(N)}{N} \mathbb{P}\left(\frac{\|\hat{f}_N - f\|_\infty}{b_N} \geq \delta \right) < \infty,$$

and

$$\mathbb{P}\left(\sup_{k \geq N} \frac{\|\hat{f}_k - f\|_\infty}{b_k} \geq \delta \right) = \mathbf{o}\left(\frac{1}{\mathrm{LLog}(N)} \right), \quad \text{as } N \to \infty,$$

where $\sigma(\widetilde{\mathcal{X}})$ is given by (14) for the $C(\mathcal{X})$-valued random variable $\widetilde{\mathcal{X}}$. Under the relevant assumptions, these characterisations of the rate of convergence may then be transferred to the respective optimal estimators.

Note that under strong exponential moment conditions on $\|\widetilde{X}\|$ and further weak requirements, it is also possible to derive exponential rates of convergence in probability by a large deviation principle, cf. Theorem 2.3 in de Acosta (1992). For every closed set F of $C(\mathcal{X})$, it then holds that

$$\limsup_{N \to \infty} \frac{N}{a_N^2} \log \mathbb{P}\left(\frac{\hat{f}_N - f}{b_N} \in F \right) \leq - \inf_{x \in F} I(x),$$

where I denotes the corresponding rate function of the Hilbert space $H_{\widetilde{X}}$ associated to the covariance of $\widetilde{X}$.

4.1.2 Rates of Error Probabilities

Related rates of error probabilities for the difference in objective function values, in optimal values, and in optimal solutions can also be derived from Section 3.3 on rates of convergence in mean. To this end, reconsider equality (18) under assumptions (A1)–(A2), implying that for any $\epsilon > 0$ there exists a deterministic $N^* = N^*(\epsilon) \in \mathbb{N}$ such that

$$\forall N \geq N^* : \quad \frac{1}{b_N} \mathbb{E}_{\mathbb{P}}\left[\|\hat{f}_N - f\|_\infty \right] \leq \epsilon. \tag{24}$$

In consequence of this inequality, we are then able to formulate the following probabilistic estimates for the differences in objective function values, where we distinguish between the case when no further moment conditions on the random variable $\widetilde{X} = h(\cdot, \xi) - \mathbb{E}_{\mathbb{P}}[h(\cdot, \xi)]$ are available (to apply Markov's inequality) and the case when higher moment conditions on $\widetilde{X}$ are satisfied (to use an inequality by Einmahl and Li (2008)).

Theorem 8 *Suppose that assumptions (A1)–(A2) hold and let $\delta > 0$. Then, the following statements hold:*

a) For any $\epsilon > 0$, there exists an $N^* = N^*(\epsilon) \in \mathbb{N}$ such that

$$\forall N \geq N^* : \quad \mathbb{P}\Big(\|\hat{f}_N - f\|_\infty \geq \delta\Big) \leq \frac{\epsilon}{\delta} b_N. \tag{25}$$

b) If $\mathbb{E}_{\mathbb{P}}[\|\widetilde{X}\|^s] < \infty$ for $s > 2$ then there exists an $N^* = N^*(\delta) \in \mathbb{N}$ such that for all $N \geq N^*$:

$$\mathbb{P}\Big(\|\hat{f}_N - f\|_\infty \geq \delta\Big) \leq \exp\left\{-\frac{N\delta^2}{12\sigma^2(\widetilde{X})}\right\} + \frac{c_2}{N^{s-1}\big(\frac{\delta}{2}\big)^s}\mathbb{E}_{\mathbb{P}}\Big[\|\widetilde{X}\|^s\Big], \tag{26}$$

where $\sigma(\widetilde{X})$ is given by (14) for the $C(\mathcal{X})$-valued random variable $\widetilde{X}$ and c_2 is a positive constant.

Proof Assertion a) follows directly from Markov's inequality and inequality (24).

To show b), we first observe that according to inequality (24), for $\delta > 0$ and some arbitrary but fixed $0 < \eta \leq 1$ there exists an $N^* = N^*(\delta, \eta) \in \mathbb{N}$ such that for all $N \geq N^*$,

$$\begin{aligned}\mathbb{P}\Big(\|\hat{f}_N - f\|_\infty \geq \delta\Big) &\leq \mathbb{P}\left(\|\hat{f}_N - f\|_\infty \geq (1+\eta)\mathbb{E}_{\mathbb{P}}\big[\|\hat{f}_N - f\|_\infty\big] + \frac{\delta}{2}\right) \\ &\leq \mathbb{P}\left(\max_{1\leq k\leq N}\|\hat{f}_k - f\|_\infty \geq (1+\eta)\mathbb{E}_{\mathbb{P}}\big[\|\hat{f}_N - f\|_\infty\big] + \frac{\delta}{2}\right),\end{aligned}$$

where the second inequality follows from $\max_{1\leq k\leq N}\|\hat{f}_k - f\| \geq \|\hat{f}_N - f\|$. By applying Theorem 4 of Einmahl and Li (2008) (with $\delta = 1$ and $t = \delta/2$) on the $C(\mathcal{X})$-valued random variables $\widetilde{X}_i/N$ under the moment condition $\mathbb{E}_{\mathbb{P}}[\|\widetilde{X}\|^s] < \infty$, we then obtain $\forall N \geq N^*$:

$$\begin{aligned}&\mathbb{P}\left(\max_{1\leq k\leq N}\|\hat{f}_k - f\|_\infty \geq (1+\eta)\mathbb{E}_{\mathbb{P}}[\|\hat{f}_N - f\|_\infty] + \frac{\delta}{2}\right) \\ &\qquad\leq \exp\left\{-\frac{N\delta^2}{12\sigma^2(\widetilde{X})}\right\} + \frac{c_2}{N^{s-1}\big(\frac{\delta}{2}\big)^s}\mathbb{E}_{\mathbb{P}}\left[\|\widetilde{X}\|^s\right],\end{aligned}$$

with the specified constants $\sigma(\widetilde{X})$ and c_2. □

Interestingly, while the error probability of $\hat{f}_N$ with respect to f in (25) has essentially the usual rate b_N, the rate is of order $1/N^{s-1}$ in (26) provided that $s < \infty$. However, it has to be noted that in both cases the exact number N^* needed for the validity of both estimates is not known. Moreover, given $N \geq N^*$, both inequalities imply that for sufficiently small values δ, the condition $N \gg 1/\delta^2$ is sufficient to obtain reasonably small probabilities for errors larger than δ.

Remark 7 Given Theorem 4, both estimates (25) and (26) in Theorem 8 can further be used to infer rates in error probability for the absolute error of the optimal values, i.e. $\mathbb{P}(|\hat{f}_N^* - f^*| \geq \delta)$. Moreover, Markov's inequality can be applied to obtain similar rates for the error probability of the optimal solutions $\mathbb{P}(\|\hat{x}_N^* - x^*\| \geq \delta)$, based on Theorem 7.

4.2 Confidence Sets for Optimal Values and Solutions

4.2.1 Bounding the Optimality Gap

In what follows, we reconsider the idea of the optimality gap to derive a confidence interval for f^*, presumably first considered by Mak et al (1999). Given our results of Section 3.3, especially Theorem 6 and Proposition 3, we are able to improve known results on the optimality gap and to state that the size of the corresponding confidence set converges to zero with known rate, almost surely.

Definition 1 The *optimality gaps* of a point $\bar{x} \in \mathcal{X}$ with respect to problems (1) and (2) are defined as

$$\Gamma(\bar{x}) := f(\bar{x}) - f^* \quad \text{and} \quad \widehat{\Gamma}_N(\bar{x}) := \hat{f}_N(\bar{x}) - \hat{f}_N^*,$$

respectively.

Based on the results of Sections 3.2 and 3.3, some important properties of the nonnegative estimator $\widehat{\Gamma}_N(\bar{x})$ for $\Gamma(\bar{x})$ can be derived.

Proposition 6 *Suppose that assumptions (A1)–(A2) hold, and let $\bar{x} \in \mathcal{X}$ be fixed. Then, it holds:*

a) $\widehat{\Gamma}_N(\bar{x}) - \Gamma(\bar{x}) = \boldsymbol{O}(b_N)$, $\mathbb{P}$*-almost surely, and* $\mathbb{E}_{\mathbb{P}}[|\widehat{\Gamma}_N(\bar{x}) - \Gamma(\bar{x})|] = \boldsymbol{o}(b_N)$.

b) $0 \leq \mathbb{E}_{\mathbb{P}}[\widehat{\Gamma}_N(\bar{x})] - \Gamma(\bar{x}) = \boldsymbol{o}(b_N)$.

Further, if in addition assumptions (B1)–(B2) are satisfied and if there exists $\gamma_\delta := \mathbb{E}_{\mathbb{P}}[G(\xi)^{2+\delta}] < \infty$ *for some* $\delta > 0$*, then, with* $p = 2 + 4/\delta$*, it holds*

c) $\operatorname{std}(\widehat{\Gamma}_N(\bar{x})) \leq \frac{1}{\sqrt{N}} \operatorname{std}(h(\bar{x}, \xi)) + \operatorname{std}(\hat{f}_N^*) = \boldsymbol{o}(b_N^{1/p})$.

Proof To show the statements, let us rearrange $\widehat{\Gamma}_N(\bar{x}) - \Gamma(\bar{x})$ as

$$\widehat{\Gamma}_N(\bar{x}) - \Gamma(\bar{x}) = \big(\hat{f}_N(\bar{x}) - f(\bar{x})\big) + \big(f^* - \hat{f}_N^*\big).$$

By this representation, the first part of statement a) is a direct consequence of Lemma 1 and Theorem 4, and the second part follows analogously by Proposition 2 and Theorem 6. The first inequality of statement b) follows from $\mathbb{E}_{\mathbb{P}}[\hat{f}_N^*] \leq f^*$, while the rate of the bias follows from the second part of statement a). Finally, the last statement is due to the subadditivity of the standard deviation and the second part of Proposition 3, under the additionally made assumptions. □

The main idea for bounding the optimality gap for a given candidate point $\bar{x}$ was introduced in Mak et al (1999), Section 3.2: for a given $\epsilon > 0$ find a (random) upper bound $u_N = u_N(\epsilon)$ for $\mathbb{E}_{\mathbb{P}}[\widehat{\Gamma}_N(\bar{x})]$ with

$$\mathbb{P}\Big(u_N \geq \mathbb{E}_{\mathbb{P}}\big[\widehat{\Gamma}_N(\bar{x})\big]\Big) \geq 1 - \epsilon,$$

since then Proposition 6b) implies that

$$\mathbb{P}\big(f(\bar{x}) \leq f^* + u_N\big) \geq 1 - \epsilon,$$

providing a performance guarantee for the candidate point $\bar{x}$ with high probability. Mak et al (1999) suggest to find u_N by means of the CLT: sample M i.i.d.

realisations $\widehat{\Gamma}_{j,N}(\bar{x})$ of the random variable $\widehat{\Gamma}_N(\bar{x})$ by independent batches with length N each, then estimate

$$\mu_{\widehat{\Gamma}_N} := \mathbb{E}_{\mathbb{P}}\big[\widehat{\Gamma}_N(\bar{x})\big], \quad \text{and} \quad \sigma_{\widehat{\Gamma}_N} := \operatorname{std}\big(\widehat{\Gamma}_N(\bar{x})\big)$$

by the classical estimators

$$\hat{\mu}_{\widehat{\Gamma}_N,M} := \frac{1}{M}\sum_{j=1}^{M}\widehat{\Gamma}_{j,N}(\bar{x}), \quad \text{and} \quad \hat{\sigma}^2_{\widehat{\Gamma}_N,M} := \frac{1}{M}\sum_{j=1}^{M}\big(\widehat{\Gamma}_{j,N}(\bar{x}) - \hat{\mu}_{\widehat{\Gamma}_N,M}\big)^2,$$

respectively, and set

$$\hat{u}_{N,M} := \hat{\mu}_{\widehat{\Gamma}_N,M} + \frac{z_\epsilon}{\sqrt{M}}\hat{\sigma}_{\widehat{\Gamma}_N,M}.$$

Here, z_ϵ is the corresponding $(1-\epsilon)$-quantile of the t-distribution with $M-1$ degrees of freedom. If the CLT holds for $\widehat{\Gamma}_N(\bar{x})$ and if M is chosen large enough, the upper bound u_N can be approximately computed by asymptotic normality in a standard manner. More formally, if the CLT holds for $\widehat{\Gamma}_N(\bar{x})$, then

$$\forall N: \quad \mathbb{P}\Big(\hat{u}_{N,M} \geq \mathbb{E}_{\mathbb{P}}\big[\widehat{\Gamma}_N(\bar{x})\big]\Big) \to 1-\epsilon, \quad \text{as } M \to \infty.$$

For the CLT to hold, it is required that the random variable $\widehat{\Gamma}_N(\bar{x})$ has a finite second moment, a property which is guaranteed, for example, by Proposition 6c).

Interestingly, concerning the asymptotic behaviour of $\hat{u}_{N,M}$, we have not been able to identify any investigations concerning the asymptotic behaviour of $\hat{u}_{N,M}$. However, based on the results obtained in this exposition, especially Proposition 6, we are able to characterise the asymptotic behaviour more precisely.

Proposition 7 *Suppose that assumptions (A1)–(A2) hold, and let $\bar{x} \in \mathcal{X}$ be fixed. Then, for any fixed $M > 1$ we have*

$$\hat{u}_{N,M} = \Gamma(\bar{x}) + \boldsymbol{O}(b_N),$$

$\mathbb{P}$*-almost surely.*

Proof From Proposition 6a), we immediately have that $\hat{\mu}_{\widehat{\Gamma}_N,M} = \Gamma(\bar{x}) + \mathbf{O}(b_N)$, $\mathbb{P}$-almost surely, as this holds for each term of the sum in the definition of $\hat{\mu}_{\widehat{\Gamma}_N,M}$. Similarly, let us consider

$$\big|\widehat{\Gamma}_{j,N}(\bar{x}) - \hat{\mu}_{\widehat{\Gamma}_N,M}\big| = \big|\big(\widehat{\Gamma}_{j,N}(\bar{x}) - \Gamma(\bar{x})\big) - \big(\hat{\mu}_{\widehat{\Gamma}_N,M} - \Gamma(\bar{x})\big)\big|.$$

For the first term on the right-hand side, we have by Proposition 6a) that $|\widehat{\Gamma}_{j,N}(\bar{x}) - \Gamma(\bar{x})| = \mathbf{O}(b_N)$, $\mathbb{P}$-almost surely, while for the second term we have already obtained an almost sure asymptotic rate of $\mathbf{O}(b_N)$. Combining these results proves the statement. □

This shows that even for a fixed M (as typically suggested in the SAA literature), an arbitrarily exact upper bound $\hat{u}_{N,M}$ can be found. Thus, the choice of M seems to be mainly important for the quality of the normal approximation, but not for the size of the uncertainty set. Nevertheless, let us remark that a more careful analysis of the almost sure asymptotic behaviour of $\hat{\mu}_{\widehat{\Gamma}_N,M}$ will yield that the above rate can be further improved to include rates in M as well. These

kind of estimates can be obtained, e.g., by applying a LIL in M (under some slightly stronger assumption on the existence of fourth moments) to $\widehat{\Gamma}_{j,N}(\bar{x})$ and $(\widehat{\Gamma}_{j,N}(\bar{x}) - \hat{\mu}_{\widehat{\Gamma}_N,M})^2$.

The above idea can be taken one step further to obtain a proper (i.e. consistent and degenerate) confidence interval for f^*: instead of fixing a candidate point $\bar{x}$, an independent estimate $\bar{x}_N$ on a further (independent) batch based on N samples could be computed. Then, conditional on $\bar{x}_N$, the above analysis remains completely valid, with the exception that now we also have to consider the asymptotic behaviour of $\Gamma(\bar{x}_N)$. Using Lipschitz continuity of f and the rate of the almost sure convergence of $\bar{x}_N$ to x^*, it holds that $\Gamma(\bar{x}_N) = \mathbf{O}(b_N^{1/2})$ or $\Gamma(\bar{x}_N) = \mathbf{O}(b_N)$, depending on the specific assumptions made. In summary, this leads to

$$\hat{u}_{N,M} = \mathbf{O}(b_N^{1/2}), \quad \text{or} \quad \hat{u}_{N,M} = \mathbf{O}(b_N),$$

each $\mathbb{P}$-almost surely. The final step to obtain a consistent confidence set is the replacement of $1/\sqrt{M}$ by $1/b_M$, which replaces the upper bound by some slightly larger upper bound. According to previous considerations, this then leads to a 100% coverage in the limit, see also the subsequent section for similar constructions.

4.2.2 Confidence Sets

Let us briefly discuss the possibility to derive confidence sets for the optimal value f^* and an optimal solution x^* (provided the latter is unique) by other methods, where we mainly follow ideas by Lai (1976), Pflug (2003) and Vogel (2008b). To avoid a lengthy discussion of measurability issues, we focus on random convex compact sets. As notation sometimes differs among authors, we first recall the following definitions to avoid any ambiguity.

Definition 2 Let $\{C_N\}$ be a sequence of random convex compact subsets of $\mathbb{R}^l$. For an unknown fixed vector $q \in \mathbb{R}^l$, the sequence $\{C_N\}$ of random sets is called

(i) *consistent* if $\mathbb{P}(q \in C_N) \to 1$ for $N \to \infty$,
(ii) *degenerate* if $\operatorname{diam}(C_N) \to 0$, $\mathbb{P}$-almost surely, as $N \to \infty$, and
(iii) a *proper confidence sequence* if it is consistent and degenerate.

It is further called

(iv) an *ultimate ϵ-level confidence sequence* if

$$\mathbb{P}\big(\{\forall N \in \mathbb{N} : \, q \in C_N\}\big) \geq 1 - \epsilon, \text{ and}$$

(v) a *universal ϵ-level confidence sequence* if

$$\forall N \in \mathbb{N} : \quad \mathbb{P}(q \in C_N) \geq 1 - \epsilon.$$

Of course, the quantities of interest in our context are the optimal value f^* and the optimal solution x^*. Let us also recall in this context that it is well-known that in case a CLT holds, the classical confidence sets on this basis must fail the condition for an ultimate confidence if a LIL also holds. Therefore, it is reasonable to aim instead for universal confidence sets in the sense of Pflug and Vogel.

Unfortunately, the approach followed in this paper does not seem to be able to yield explicit estimates which could be exploited for the construction of universal confidence sequences. Nevertheless, the existence of a tail behaviour function in the sense of Pflug (2003) is guaranteed by the following argumentation. According to Theorem 6 and by applying the Markov inequality, we obtain for all $\delta > 0$ that

$$\sup_{N\in\mathbb{N}} \mathbb{P}\left(\frac{|\hat{f}_N^* - f^*|}{b_N} \geq \delta\right) \leq \frac{1}{\delta} \sup_{N\in\mathbb{N}} \frac{\mathbb{E}_{\mathbb{P}}\big[|\hat{f}_N^* - f^*|\big]}{b_N} = \frac{1}{\delta} c_{f^*},$$

with unknown constant $c_{f^*} := \sup_{N\in\mathbb{N}} \mathbb{E}_{\mathbb{P}}[|\hat{f}_N^* - f^*|]/b_N < \infty$. If some upper bound on c_{f^*} is known, a universal confidence set can be constructed from this inequality along the lines described by Pflug (2003) and Vogel (2008b).

In any case, based on the results from the previous section, it is straightforward to show that proper confidence sequences can be easily obtained.

Corollary 1 *Suppose that assumptions (A1)–(A2) hold and let $\delta > 0$ be arbitrary. Then,*

$$C_N = \big\{z \in \mathbb{R} : |z - \hat{f}_N^*| \leq \delta b_N\big\}$$

yields a proper confidence sequence for f^*.

Proof This follows directly from the first statement of Proposition 5. □

Two main comments are in order. First, a similarly sized proper confidence sequence can be easily obtained from the CLT approach under the same assumptions; for the above corollary though, the CLT has not been used. Second, in contrast to the approach via the CLT, here no approximate estimate of the coverage probability of C_N is available, while after all no kind of variance estimate is necessary for its construction.

Corollary 2 *Suppose that assumptions (A1)–(A2) and (B1)–(B2) hold and let $\delta > 0$ be arbitrary. Then,*

$$C_N = \big\{z \in \mathbb{R}^n : \|z - \hat{x}_N^*\| \leq \sqrt{\delta b_N}\big\} \tag{27}$$

yields a proper confidence sequence for x^*. Under the assumptions (A1)–(A3), (A1')–(A2'), (B1) and (B2'), it further holds that

$$C_N = \big\{z \in \mathbb{R}^n : \|z - \hat{x}_N^*\| \leq \delta b_N\big\} \tag{28}$$

also yields a proper confidence sequence for x^*.

Proof Again, this follows directly from the second and the third statement of Proposition 5. □

Note that, for sufficiently large N, the confidence set (28) is much smaller than the set (27), as $\delta b_N < \sqrt{\delta b_N}$ for N large enough.

The main novelty of the latter results lies in the fact that they allow to derive a proper confidence sequence for the optimal solution x^* under quite weak assumptions. These assumptions are indeed weaker than those which lead to confidence sets of x^* via asymptotic normality, cf. Theorem 3. Again, let us point out that although under asymptotic normality approximate coverage probabilities are available, no exact knowledge of the coverage probability is available here in either case.

Remark 8 Suppose δ is chosen such that $\delta > \sigma(\widetilde{X})$ for f^* or $\delta > \frac{2}{\alpha}\sigma(\widetilde{X})$ for x^*, where $\alpha > 0$ denotes the constant of either the second-order growth condition (B2) or of (B2'). Then, ultimately, on each sample path, the confidence set C_N covers f^* or x^* from some random N^* onwards, see, e.g., Serfling (1980), Section 1.10, for a discussion. These upper estimate for $\sigma(\widetilde{X})$ might be derived, for instance, via (14) or from the boundedness of h. As already pointed out above, this behaviour of the confidence sequence is in contrast to the classical confidence sets provided by asymptotic normality; here, the quantities f^* or x^* drop out of the confidence interval infinitely often on each sample path.

In summary, we thus have seen that the approach followed here is not able to yield universal confidence sets (for instance, by deriving explicit tail behaviour of the estimators), but is able to provide proper confidence sequences.

5 Numerical Illustration

In this section we illustrate the main results of our analysis by means of the well-known newsvendor problem. To this end, we consider the problem in its most simple version:

$$\min_{x\in[0,x_u]} \quad \mathbb{E}_{\mathbb{P}}\big[cx - r\min(x,\xi)\big], \tag{29}$$

where ξ denotes the random demand for a certain good (newspaper), c the costs associated with keeping the good in stock, r the price at which the good can be sold, and x_u the maximum amount of goods that can be stored. The objective function which is to be minimised represents the expected negative revenues from deciding to keep x goods in stock. For a more detailed treatment of the newsvendor problem including several visualisations, we refer to the thorough review by Homem-de-Mello and Bayraksan (2014) and the references therein.

For our numerical experiments, we set the parameters as $c = 2$, $r = 5$, and $x_u = 100$, and assume that ξ is distributed according to a lognormal distribution with parameters $\mu_{LN} = 0$ and $\sigma_{LN} = 1$, that is $\xi \sim LN(0,1)$. The optimal solution to (29) is then given by the 60%-quantile of the lognormal distribution, i.e. by $x^* = F^{-1}_{LN(0,1)}(0.6) \approx 1.288330$ with optimal value $f^* \approx -3.753092$ (which can also be calculated analytically in our specific case). To approximatively solve problem (29) by the SAA approach, we choose $N = 1, 2, 4, 8, \ldots, N^{\max}$ with $N^{\max} = 2^{24}$, and set the number of batches (independent repetitions to solve the SAA problem) to $M = 2^{15}$.

5.1 Illustrations on the CLT and the LIL

5.1.1 Convergence in Distribution

Let us start with an illustration of the asymptotic distribution of the sequences $\{\sqrt{N}(\hat{f}^*_N - f^*)\}$ and $\{\sqrt{N}(\hat{x}^*_N - x^*)\}$, for which we have plotted in Figure 1 the quantities $\sqrt{N}(\hat{f}^*_N - f^*)$ and $\sqrt{N}(\hat{x}^*_N - x^*)$ for a small and a large N each.

As expected by Theorems 2 and 3, it can be observed from Figure 1 that the distributions of $\sqrt{N}(\hat{f}^*_N - f^*)$ and $\sqrt{N}(\hat{x}^*_N - x^*)$ look already quite normal for small

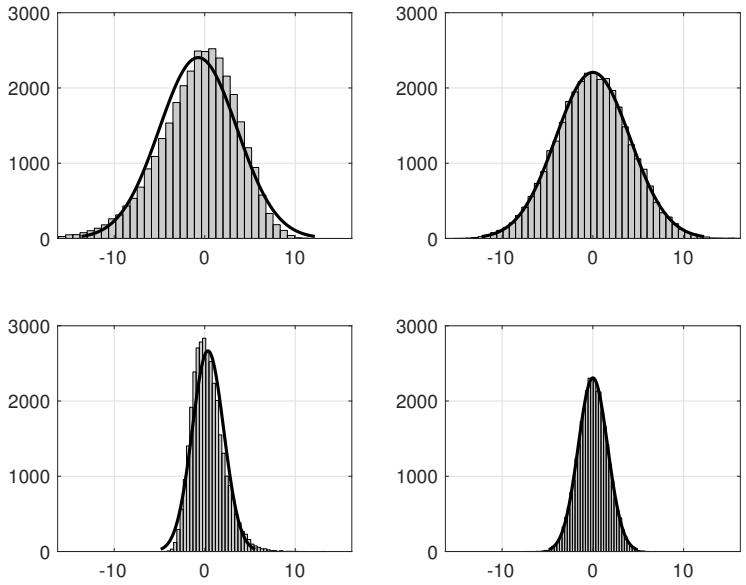

Fig. 1: Distribution of $\sqrt{N}(\hat{f}_N^* - f^*)$ (upper half) and of $\sqrt{N}(\hat{x}_N^* - x^*)$ (lower half) for $N = 32$ (left) and $N = N^{\max}$ (right). Note that in the considered example approximate normality can already be obtained for small N.

N and very close to normal for large N. By contrast, however, as $\hat{f}_N^* \leq \hat{f}_N(x^*)$, Theorem 3 also tells us that we cannot expect that $\{N(\hat{f}_N^* - \hat{f}_N(x^*))\}$ converges in distribution to a normal distribution, cf. Figure 2.

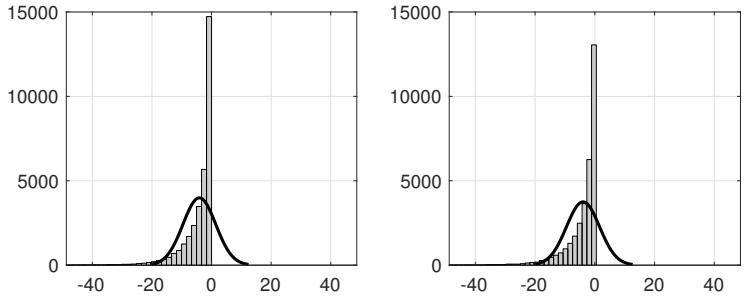

Fig. 2: Distribution of $N(\hat{f}_N^* - \hat{f}_N(x^*))$ for $N = 32$ (left) and $N = N^{\max}$ (right). Note that convergence in distribution to a non-normal distribution can be observed empirically.

5.1.2 Almost Sure Convergence

Next, we illustrate the behaviour of $\{\hat{f}_N^*\}$ and $\{\hat{x}_N^*\}$ in the almost sure sense. To this end, we have plotted in Figure 3 the first 500 sample paths (out of M) of

the upscaled quantities $(\hat{f}_N^* - f^*)/b_N$ and $(\hat{x}_N^* - x^*)/b_N$ for the grid of different sample sizes N. Note that the remaining paths behave very similarly.

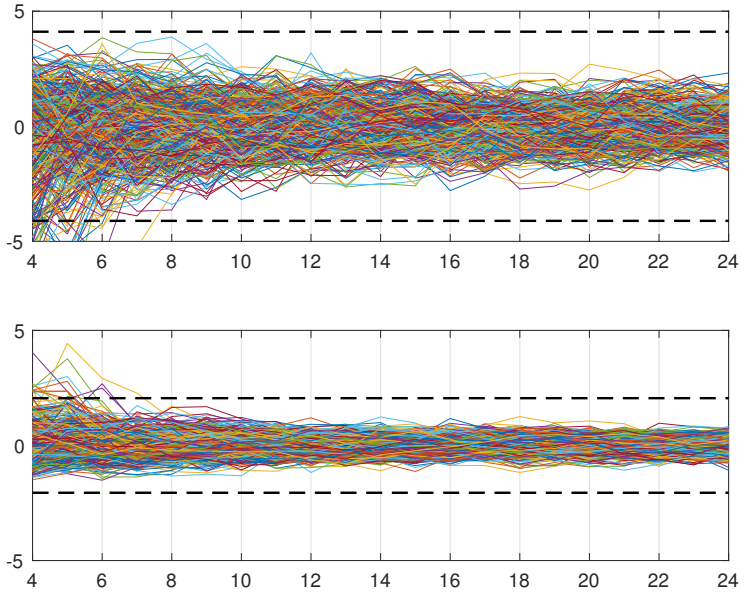

Fig. 3: Plots of the first 500 sample paths of $(\hat{f}_N^* - f^*)/b_N$ (top) and $(\hat{x}_N^* - x^*)/b_N$ (bottom) for the different sample sizes N (in $\log_2$-scale). Note that for f^* the standard deviation of $h(x^*, \xi)$ may be used as a good upper bound, while no such estimate is readily available for x^*, due to the unknown α in the second-order growth condition (B2'). Therefore, a conservative estimate of $\alpha = 2$ has been used.

In accordance with theory, cf. Theorems 4 and 5, it can indeed be observed from Figure 3 that most upscaled sample paths remain within a band of width equal to the standard deviation. Note that for the case of $\hat{f}_N^*$, we have been able to use the standard deviation of $h(x^*, \xi)$ as an upper bound, cf. Theorem 4 and Remark 8. However, as the positive constant α in the second-order growth condition (B2') is usually not known in the case of $\hat{x}_N^*$ (and no readily available estimate is available), we have taken an estimate of α corresponding to function $\hat{f}_N$, leading to $\alpha \approx 2$. Considering Figure 3, we note that the almost-sure speed of convergence is of course already implied by the fact that the upscaled sequences stay bounded. By closer inspection, we can further observe that we cannot expect a better convergence rate as the confidence band is almost completely covered by each path.

Finally, in accordance with the construction of the confidence sets in Section 4.2.2 and Remark 8, we can see from Figure 4 that the probability of falling outside these confidence bands drops to zero.

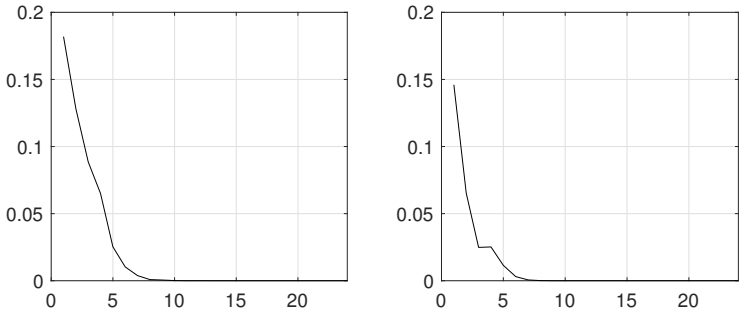

Fig. 4: Plots of $\mathbb{P}((\hat{f}_N^* - f^*)/b_N \notin C^f)$ (left) and $\mathbb{P}((\hat{x}_N^* - x^*)/b_N \notin C^x)$ (right) for the different sample sizes N (in $\log_2$-scale), where the fixed confidence bands C^f and C^x are constructed according to Figure 3. Note that these probabilities converge to zero as $N \to \infty$, as expected.

5.1.3 Convergence in Mean

We illustrate the convergence in mean of the sequences $\{\hat{f}_N^*\}$ and $\{\hat{x}_N^*\}$ by considering the average of $\hat{f}_N^*$ and $\hat{x}_N^*$ for the different sample sizes N over the M batches and together with the standard error of the corresponding estimator. The results obtained are presented in Figure 5, where we again have plotted the upscaled quantities for better visibility.

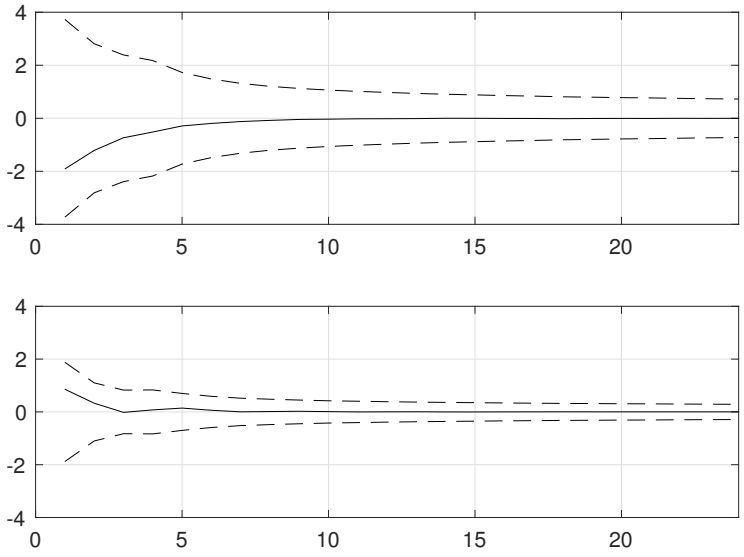

Fig. 5: Plots of the averages of $\hat{f}_N^*$ (top) and $\hat{x}_N^*$ (bottom) over M batches, centred at f^* and x^*, respectively, and upscaled by $1/b_N$, for the different sample sizes N (in $\log_2$-scale). The dashed lines represent the corresponding sample standard deviations, again upscaled by $1/b_N$. Note that the negative bias of $\hat{f}_N^*$ is clearly visible but vanishes asymptotically, as suggested by its order $\mathbf{o}(b_N)$.

From Figure 5, a few interesting insights can be gained. First, the negative bias of $\hat{f}_N^*$ can be visually identified in the top panel. Second, in full accordance with Theorems 6 and 7, we observe that the upscaled sequences $\{(\hat{f}_N^* - f^*)/b_N\}$ and $\{(\hat{x}_N^* - x^*)/b_N\}$ still converge to zero, confirming the convergence order $\mathbf{o}(b_N)$. Third, it seems that the corresponding standard deviations of both estimators, upscaled by $1/b_N$, remain at least bounded – a much better behaviour than could be expected from assertion (21) (where a scale of $1/b_N^{1/p}$ is considered).

5.2 Further Illustrations

5.2.1 Convergence in Probability

Considering the last mode of convergence to be discussed, convergence in probability, Figure 6 depicts how fast the deviation probabilities in Proposition 5 converge to zero.

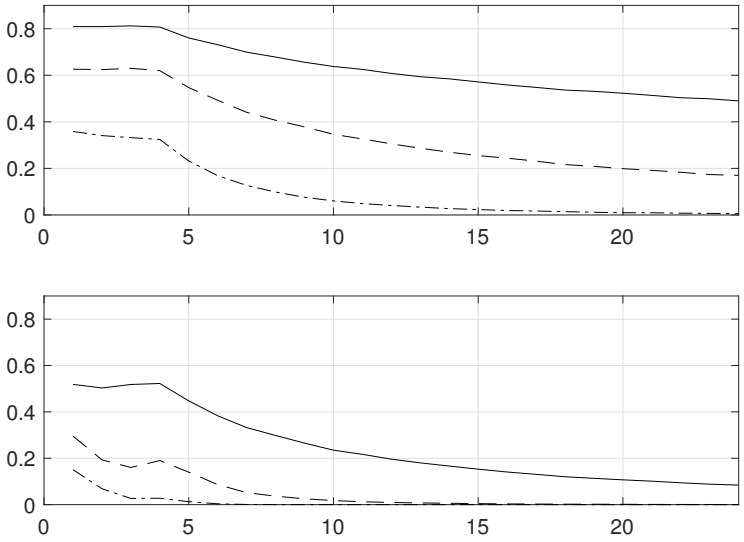

Fig. 6: Plots of $\mathbb{P}(|\hat{f}_N^* - f^*|/b_N > \delta)$ (top) and $\mathbb{P}(\|\hat{x}_N^* - x^*\|/b_N > \delta)$ (bottom) for the different sample sizes N (in $\log_2$-scale) and $\delta \in \{1/2, 1, 2\}$. Note that the convergence of these probabilities to zero is in accordance with Proposition 5.

It can clearly be observed that $\mathbb{P}(|\hat{f}_N^* - f^*|/b_N > \delta) \to 0$ and $\mathbb{P}(\|\hat{x}_N^* - x^*\|/b_N > \delta) \to 0$ holds in accordance with Proposition 5.

5.2.2 Estimation of the Optimality Gap

As a final illustration and application of our results, let us investigate the behaviour of the optimality gap. For this purpose, Figure 7 illustrates the behaviour of $\mu_{\widehat{\Gamma}_N}$, $\sigma_{\widehat{\Gamma}_N}$ and $\hat{u}_{N,M}$ for different sample sizes N.

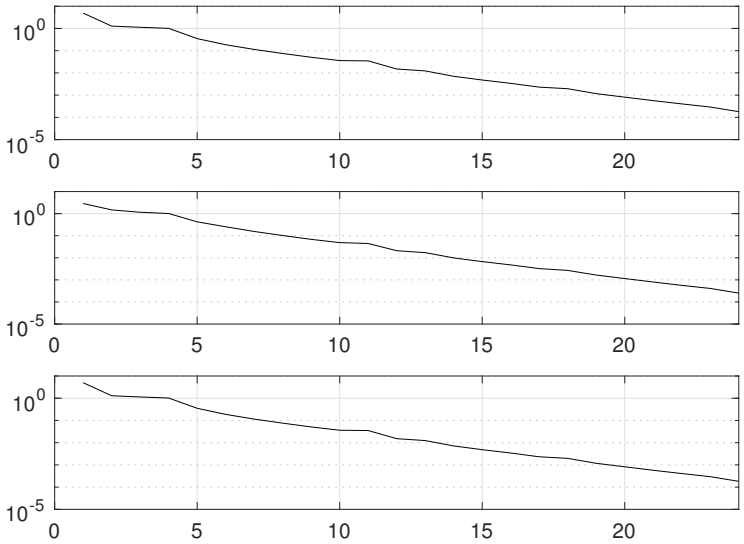

Fig. 7: Plots of $\mu_{\widehat{\Gamma}_N}/b_N$ (top), $\sigma_{\widehat{\Gamma}_N}/b_N$ (middle) and $\hat{u}_{N,M}/b_N$ (bottom) (in $\log_{10}$-scale) for different sample sizes N (in $\log_2$-scale), obtained by averaging over M batches. Note that all upscaled sequences converge to zero in our specific example, as expected.

First, we can observe that the upscaled sequences $\{\mu_{\widehat{\Gamma}_N}/b_N\}$ and $\{\sigma_{\widehat{\Gamma}_N}/b_N\}$ appear to converge to zero at a rate of at least $\mathbf{o}(b_N)$, as indicated by Proposition 6b). Second, for the example considered, it can further be observed that the scale of $\sigma_{\widehat{\Gamma}_N}$ is comparable to the one of $\mu_{\widehat{\Gamma}_N}$. For already small M, we thus have $\hat{u}_{N,M} \approx \hat{\mu}_{\Gamma_N}$. In general, if M is chosen large enough, it is expected that this is always the case, i.e. it generally holds that $\hat{u}_{N,M} \approx \hat{\mu}_{\Gamma_N}$, as the second term in the definition of $\hat{u}_{N,M}$ vanishes.

6 Conclusion

In this paper, we have derived rates of convergence almost surely and in mean for optimal estimators in the SAA approach, a matter which has not been investigated so far. Both rates can essentially be quantified as $\sqrt{\mathrm{LLog}(N)}/\sqrt{N}$ and may be inferred under rather mild assumptions by applying a version of the LIL in Banach spaces, similar to the case of the functional CLT that allows to derive asymptotic distributions and related convergence rates for the optimal estimators. On the basis of the obtained convergence results in mean, we have been able to quantify the asymptotic bias and the mean squared errors of the optimal estimators. Moreover, from the rates of convergence in mean, we have derived convergence in probability for the deviation of the optimal estimators from their respective counterparts and rates of error probabilities that are rather weak but do not rely on the strong exponential moment conditions as in other approaches. We have also analysed the idea of constructing confidence sets for optimal values and solutions by bounding the optimality gap and by more traditional methods. Finally, we have provided

a numerical illustration of our results by considering the well-known newsvendor problem.

Acknowledgements This work was supported by the Engineering and Physical Sciences Research Council [grant no. EP/M50662X/1] of the UK. The authors want to thank two anonymous referees and the guest editor for their highly valuable comments which have helped to significantly improve the paper both in terms of clarity and exposition.

References

Albiac F, Kalton NJ (2006) Topics in Banach Space Theory. Springer, New York, NY

Aliprantis CD, Border K (2006) Infinite Dimensional Analysis – A Hitchhiker's Guide. Springer, Berlin Heidelberg

Araujo A, Giné E (1980) The Central Limit Theorem for Real and Banach Valued Random Variables. Wiley, New York, NY

Bates C, White H (1985) A unified theory of consistent estimation for parametric models. Econometric Theory 1(2):151–178

Bauer H (2001) Measure and Integration Theory. De Gruyter, Berlin

Bonnans JF, Shapiro A (2000) Perturbation Analysis of Optimization Problems. Springer, New York, NY

Dai L, Chen CH, Birge JR (2000) Convergence properties of two-stage stochastic programming. Journal of Optimization Theory and Applications 106(3):489–509

Danskin JM (1966) The theory of max-min, with applications. SIAM Journal on Applied Mathematics 14(4):641–664

de Acosta A (1992) Moderate deviations and associated Laplace approximations for sums of independent random vectors. Transactions of the American Mathematical Society 329(1):357–375

Dieudonné J (1960) Foundations of Modern Analysis. Academic Press, New York, NY

Dupačová J, Wets R (1988) Asymptotic Behavior of Statistical Estimators and of Optimal Solutions of Stochastic Optimization Problems. The Annals of Statistics 16(4):1517–1549

Einmahl U, Li D (2008) Characterization of LIL behavior in Banach space. Transactions of the American Mathematical Society 360(12):6677–6693

Fernique X (1970) Intégrabilité des vecteurs gaussiens. Comptes rendus de l'Académie des Sciences Paris 270(25):1698–1699

Goodman V, Kuelbs J, Zinn J (1981) Some Results on the LIL in Banach Space with Applications to Weighted Empirical Processes. The Annals of Probability 9(5):713–752

Hartman P, Wintner A (1941) On the Law of the Iterated Logarithm. American Journal of Mathematics 63(1):169–176

He X, Wang G (1995) Law of the iterated logarithm and invariance principle for M-estimators. Proceedings of the American Mathematical Society 123(2):563–573

Homem-de-Mello T (2003) Variable-sample methods for stochastic optimization. ACM Transactions on Modeling and Computer Simulation 13(2):108–133

Homem-de-Mello T (2008) On Rates of Convergence for Stochastic Optimization Problems Under Non-Independent and Identically Distributed Sampling. SIAM Journal on Optimization 19(2):524–551

Homem-de-Mello T, Bayraksan G (2014) Monte Carlo sampling-based methods for stochastic optimization. Surveys in Operations Research and Management Science 19(1):56–85

Kaniovski YM, King AJ, Wets R (1995) Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems. Annals of Operations Research 56(1):189–208

King AJ, Rockafellar RT (1993) Asymptotic Theory for Solutions in Statistical Estimation and Stochastic Programming. Mathematics of Operations Research 18(1):148–162

Kuelbs J (1976a) A Strong Convergence Theorem for Banach Space Valued Random Variables. The Annals of Probability 4(5):744–771

Kuelbs J (1976b) A Counterexample for Banach space valued random variables. The Annals of Probability 4(4):684–689

Kuelbs J (1977) Kolmogorov's law of the iterated logarithm for Banach space valued random variables. Illinois Journal of Mathematics 21(4):784–800

Lai TL (1976) On Confidence Sequences. The Annals of Statistics 4(2):265–280

Ledoux M, Talagrand M (1988) Characterization of the Law of the Iterated Logarithm in Banach Space. The Annals of Probability 16(3):1242–1264

Ledoux M, Talagrand M (1991) Probability in Banach Spaces: Isoperimetry and Processes. Springer, Berlin Heidelberg

Li D (1991) Convergence Rates of Law of Iterated Logarithm for B-valued Random Variables. Science in China A(34):395–404

Li D, Rosalsky A, Volodin A (2007) On the Relationship Between the Baum-Katz-Spitzer Complete Convergence Theorem and the Law of the Iterated Logarithm. Acta Mathematica Sinica, English Series 23(4):599–612

Mak WK, Morton DP, Wood RK (1999) Monte Carlo bounding techniques for determining solution quality in stochastic programs. Operations Research Letters 24(1-2):47–56

Pflug GC (2003) Stochastic optimization and statistical inference. In: Ruszczyński A, Shapiro A (eds) Stochastic Programming, Handbooks in Operations Research and Management Science, vol 10, Elsevier, Amsterdam, pp 427–482

Pisier G (1975) Le théorème de la limite centrale et la loi du logarithme itérée dans les espaces de Banach. Séminaire Maurey-Schwartz 1975-1976, and exposés III et IV

Pisier G, Zinn J (1978) On the limit theorems for random variables with values in the spaces L_p $(2 \leq p < \infty)$. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 41(4):289–304

Robinson SM (1996) Analysis of sample-path optimization. Mathematics of Operations Research 21(3):513–528

Rockafellar RT, Wets RJB (1998) Variational Analysis. Springer, Berlin Heidelberg

Serfling RJ (1980) Approximation Theorems of Mathematical Statistics. Wiley, New York, NY

Shapiro A (1989) Asymptotic Properties of Statistical Estimators in Stochastic Programming. Annals of Statistics 17(2):841–858

Shapiro A (1990) On differential stability in stochastic programming. Mathematical Programming 47(1-3):107–116

Shapiro A (1991) Asymptotic analysis of stochastic programs. Annals of Operations Research 30(1):169–186
Shapiro A (2000) Statistical inference of stochastic optimization problems. In: Uryasev SP (ed) Probabilistic Constrained Optimization, Nonconvex Optimization and Its Applications, vol 49, Springer, Boston, MA, pp 282–304
Shapiro A (2003) Monte Carlo Sampling Methods. In: Ruszczyński A, Shapiro A (eds) Stochastic Programming, Handbooks in Operations Research and Management Science, vol 10, Elsevier, Amsterdam, pp 353–425
Shapiro A, Homem-de-Mello T (2000) On the Rate of Convergence of Optimal Solutions of Monte Carlo Approximations of Stochastic Programs. SIAM Journal on Optimization 11(1):70—86
Shapiro A, Dentcheva D, Ruszczyński A (2014) Lectures on Stochastic Programming: Modeling and Theory. SIAM, Philadelphia, PA
Slivka J (1969) On the law of the iterated logarithm. Proceedings of the National Academy of Sciences of the USA 63(2):289–291
Strassen V (1964) An invariance principle for the law of the iterated logarithm. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 3(3):211–226
Strassen V (1966) A converse to the law of the iterated logarithm. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 4(4):265–268
Vogel S (1988) Stability results for stochastic programming problems. Optimization 19(2):269–288
Vogel S (1992) On stability in multiobjective programming – A stochastic approach. Mathematical Programming 56(1-3):91–119
Vogel S (2008a) Confidence sets and convergence of random functions. In: Tammer C, Heyde F (eds) Festschrift in Celebration of Prof. Dr. Wilfried Grecksch's 60th Birthday, Shaker, Aachen
Vogel S (2008b) Universal Confidence Sets for Solutions of Optimization Problems. SIAM Journal on Optimization 19(3):1467–1488
Vogel S (2017) Random approximations in multiobjective optimization. Mathematical Programming 164(1-2):29–53