SIAM J. OPTIM.
Vol. 29, No. 1, pp. 290–317

GENERIC CONSISTENCY FOR APPROXIMATE STOCHASTIC PROGRAMMING AND STATISTICAL PROBLEMS*

CHRISTIAN HESS† AND RAFFAELLO SERI‡

Abstract. In stochastic programming, statistics, or econometrics, the aim is in general the optimization of a criterion function that depends on a decision variable θ and reads as an expectation with respect to a probability $\mathbb{P}$. When this function cannot be computed in closed form, it is customary to approximate it through an empirical mean function based on a random sample. On the other hand, several other methods have been proposed, such as quasi–Monte Carlo integration and numerical integration rules. In this paper, we propose a general approach for approximating such a function, in the sense of epigraphical convergence, using a sequence of functions of simpler type which can be expressed as expectations with respect to probability measures $\mathbb{P}_n$ that, in some sense, approximate $\mathbb{P}$. The main difference with the existing results lies in the fact that our main theorem does not impose conditions directly on the approximating probabilities but only on some integrals with respect to them. In addition, the $\mathbb{P}_n$'s can be transition probabilities, i.e., are allowed to depend on a further parameter, ξ, whose value results from deterministic or stochastic operations, depending on the underlying model. This framework allows us to deal with a large variety of approximation procedures such as Monte Carlo, quasi–Monte Carlo, numerical integration, quantization, several variations on Monte Carlo sampling, and some density approximation algorithms. As by-products, we discuss convergence results for stochastic programming and statistical inference based on dependent data, for programming with estimated parameters, and for robust optimization; we also provide a general result about the consistency of the bootstrap for M-estimators.

Key words. stochastic programming, approximation methods, statistical inference, epigraphical convergence, transition probabilities, robust optimization

AMS subject classifications. 90C15, 90C59, 60F15, 60B10, 62F40, 90C34

DOI. 10.1137/17M1156769

1. Introduction.

In stochastic programming, statistics, or econometrics, one often looks for the solution of optimization problems of the following form (see, e.g., [9, Page 332]):

$$\inf_{\theta\in\Theta} \mathbb{E}_{\mathbb{P}}\, g(\cdot,\theta) = \inf_{\theta\in\Theta} \int_{\mathbb{R}^q} g(y,\theta)\mathbb{P}(dy), \tag{1.1}$$

where Θ is a Borel subset of $\mathbb{R}^p$ and $\mathbb{P}$ is a probability measure defined on $\mathbf{Y} = \mathbb{R}^q$ endowed with its Borel σ-field $\mathcal{B}(\mathbf{Y})$ (but more general spaces can be considered).[1] Most of the time, the mean functional $\mathbb{E}_{\mathbb{P}}\, g(\cdot,\theta)$ cannot be explicitly or easily calculated. Fortunately, there are situations where it is possible to approximate problem (1.1) by a sequence of more tractable problems where $\mathbb{P}$ is replaced with a probability $\mathbb{P}_n$. In this approximation process, it is expected that the optimization problem relative to $\mathbb{P}_n$, namely

$$\inf_{\theta\in\Theta} \mathbb{E}_{\mathbb{P}_n}\, g(\cdot,\theta) = \inf_{\theta\in\Theta} \int_{\mathbb{R}^q} g(y,\theta)\mathbb{P}_n(dy), \tag{1.2}$$

*Received by the editors November 15, 2017; accepted for publication (in revised form) October 8, 2018; published electronically January 24, 2019.
http://www.siam.org/journals/siopt/29-1/M115676.html
Funding: This work was funded by the Fog Research Institute under contract no. FRI-454.
†Département MIDO, Université Paris Dauphine, PSL Research University, Paris, France (christian.hess@dauphine.fr).
‡Dipartimento di Economia, Università degli Studi dell'Insubria, Varese, Italy (raffaello.seri@uninsubria.it).
[1]As usual, $g(\cdot,\theta)$ denotes the map $y \mapsto g(y,\theta)$.

is simpler than (1.1) and that the sequence $\mathbb{P}_n$ converges to $\mathbb{P}$ in some sense, for example, in the sense of narrow (or weak[2]) convergence. Often, the function to be minimized can be expressed as the expectation

$$\mathbb{E}_{\mathbb{Q}}\, g(Y,\theta) = \int_\Omega g(Y(\omega),\theta)\mathbb{Q}(d\omega), \tag{1.3}$$

where Y is an $\mathbb{R}^q$-valued random variable defined on a probability space $(\Omega, \mathcal{A}, \mathbb{Q})$. In such a case, $\mathbb{P} = \mathbb{Q}_Y$ is the image measure[3] of $\mathbb{Q}$ by Y. If a sample of copies of the random variable Y, say $(Y_i)_{i=1,\dots,n}$, is available, the expectation can be approximated through the empirical mean. The corresponding approximating probabilities $\mathbb{P}_n$ are the transition probabilities given by

$$\mathbb{P}_n(\omega, dy) = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i(\omega)}(dy), \qquad \omega \in \Omega, \qquad n \geq 1, \tag{1.4}$$

where $\delta_{Y_i(\omega)}$ stands for the Dirac measure at $Y_i(\omega)$. This is the basis of the so-called M-estimation used in statistics and econometrics (see, e.g., [40, Chapter III]), that is, estimation obtained by optimizing a function of the sample with respect to one or several parameters.

Thus, it is natural to look for conditions under which the solution of the approximating problem (1.2) converges almost surely to the solution of the original problem, namely, problem (1.1). To this aim, it is common to assume that $(Y_i)_{i=1,\dots,n}$ is a sample of independent and identically distributed (i.i.d.) copies of the random variable Y. By the Glivenko–Cantelli theorem, the sequence of approximating objective functions converges $\mathbb{Q}$-almost surely on Ω and uniformly on Θ to the original objective function as n goes to infinity:

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^n g(Y_i(\omega),\theta) - \mathbb{E}_{\mathbb{Q}}\, g(Y,\theta)\right| \to 0, \qquad \mathbb{Q}\text{-a.s.}$$

Under suitable technical assumptions, this implies the convergence of minimizers.

In the present paper, we shall rather make use of epigraphical convergence (epi-convergence for short) for sequences of functions defined on a metric space Θ (and depending on a random parameter). The motivation is that epigraphical convergence holds under weaker assumptions than uniform convergence, but is equally well-suited for approximating the value and the solution of a minimization problem. In particular, under suitable compactness assumptions on the parameter space, epi-convergence entails the convergence of infima and minimizers (see Appendix A or, e.g., [6, 19, 33] for a more detailed presentation). In fact, unlike uniform convergence, epi-convergence is a one-sided concept. A symmetric notion, called hypographical convergence, enjoys similar properties with respect to maximization problems. This immediately follows from the fact that hypo-convergence of a sequence $(f_n)_{n\geq 1}$ of extended real-valued functions is equivalent to epi-convergence of the sequence $(-f_n)$.

If we take $(Y_i)_{i=1,\dots,n}$ as an i.i.d. sequence and $-g$ as the logarithm of a density, problem (1.2) becomes a maximum likelihood estimation (MLE) problem. This has been covered in [33], where it was shown that almost-sure hypographical convergence of the sequence of the likelihood functionals implies almost-sure convergence of the

[2]As is known, *weak** convergence would be more appropriate.
[3]Recall that $\mathbb{Q}_Y$ is defined by $\mathbb{Q}_Y(B) = \mathbb{Q}(Y^{-1}(B))$ for all $B \in \mathcal{B}(\mathbf{Y})$.

sequence of maximum likelihood estimators to the true value, say θ_0, of the parameter. The same structure is shared by the Monte Carlo approximation approach to stochastic programming, but in this case g is a more general function (see, e.g., [9, Chapter 9] or [64]).[4]

A first novelty of our approach is that our main results (Theorems 3.1 and 3.2) do not impose conditions directly on the probabilities $\mathbb{P}_n$ that approximate $\mathbb{P}$, but only on some integrals with respect to them. A second novelty is that the $\mathbb{P}_n$'s are allowed to be transition probabilities, i.e., to depend on a random parameter. As we shall see, this is an important issue that arises quite naturally in applications. Another advantage of our approach lies in the fact that it is not restricted to the i.i.d. case. Indeed, it is possible to derive approximation results for problem (1.1) using more general sequences of random observations, such as pairwise i.i.d., ergodic stationary, and asymptotically mean stationary sequences, in which some kind of dependence is present.

For example, if the observations $(Y_i)_{i=1,\ldots,n}$ display some dependence, pseudo-maximum likelihood estimation (PMLE) is obtained, supposing that $-g$ is the logarithm of the marginal density of the process (see [16, section 2.5]). The framework can also accommodate conditional maximum likelihood estimation, provided $-g$ is replaced by the logarithm of the conditional density of Y_i given a short section of the past of $(Y_i)_{i=1,\ldots,n}$, say $Y_{i-1}, \ldots, Y_{i-k}$ for some positive integer k.

At this point, let us say a word about the adjective "generic," which appears in the title. By this, we mean that our model can be applied to a wide category of problems sharing the same structure. Further, the model is versatile in that it can take into account many variants, as will be illustrated by applications. In particular, the use of epi-convergence and, on the probabilistic side, of transition probabilities permits very general results. On the other hand, our first main result (Theorem 3.1) can be regarded as a generic strong law of large numbers (SLLN) for integrands and can be applied to derive an SLLN for random sets (see, e.g., [16]).

Results related to ours can be found in the literature, for example, on the epi-convergence of discretizations for various optimization problems or integration quadratures (see, e.g., [54, 55, 52, 53]). In those works, however, the approximating probabilities $\mathbb{P}_n$ are not transition probabilities, which prevents the researcher from dealing with a large class of situations. In fact, the consideration of transition probabilities is sometimes necessary in order to deal with subtle probabilistic issues, as shown in subsection 4.5.

A first version of Theorem 3.1 was given in [17] with an application to ergodic theory. In the present paper, we also deal with the convergence of minimizers and provide a larger variety of applications, for which precise statements and proofs are provided.

The paper is organized as follows. In section 2, we introduce the general framework of our approach and illustrate it with a few introductory examples. The main results are contained in section 3. Theorem 3.1 provides a sufficient condition under which the sequence of approximating objective functions epi-converges to the original one. The convergence of minimizers is addressed by Theorem 3.2. These results are illustrated by various applications presented in section 4 and Appendix B. Several of them, such as Monte Carlo and quasi–Monte Carlo methods, have been extensively used in stochastic programming. Others, such as numerical integration rules and quantization, have been proposed more recently. Some others, such as variants of

[4] The case in which the space Θ is discrete has been covered in [42, 18].

Monte Carlo methods and density approximation techniques, seem to be new. We also present convergence results for stochastic programming and statistical inference based on dependent data, for programming with estimated parameters (subsection 4.3) and for robust optimization (subsection 4.6). In addition, we provide a general result in connection with the strong consistency of the bootstrap for M-estimators (subsection 4.4) and another dealing with U-statistics (subsection 4.5). The proofs are deferred to section 5. A short presentation of epi-convergence is provided in Appendix A. Further applications are briefly presented in Appendix B.

2. Approximation methods: A general framework. Consider a function $g : \mathbf{Y} \times \Theta \to \overline{\mathbb{R}}$, where $\mathbf{Y}$ is a metric space endowed with its Borel σ-field $\mathcal{B}(\mathbf{Y})$ and Θ is a separable[5] metric space. As in the introduction, we consider problem (1.1), and we assume that it admits a unique solution. As already mentioned, we are interested in conditions under which the sequence of solutions of problem (1.2), the approximating problem, converges to the solution of (1.1) as n tends to infinity. This is why our main goal is to derive a.s. epi-convergence results for the above sequence. In a lot of models, the $\mathbb{P}_n$'s are transition probabilities, i.e., they depend on a parameter. Consequently, it is necessary to introduce a second probability space $(\Xi, \mathcal{X}, \mu)$, whose generic element is denoted by ξ. The role that the latter space can play is illustrated in the following short examples.

Example 2.1. Consider the case of quasi–Monte Carlo (QMC) integration with respect to the Lebesgue measure on the unit hypercube. Let f be a real-valued function defined on $\mathbf{Y}$. Suppose we have a configuration $(y_i)_{i=1,\ldots,n}$ of QMC points. Then it is possible to define the empirical probability $\mathbb{P}_n$ by

$$\mathbb{P}_n(B) = \frac{1}{n} \sum_{i=1}^{n} \delta_{y_i}(B), \qquad B \in \mathcal{B}(\mathbf{Y}), \tag{2.1}$$

where $(y_i)_{i \geq 1}$ is a low discrepancy sequence and $\delta_y(\cdot)$ is a Dirac delta. The integral $\int_{\mathbf{Y}} f(y)\mathbb{P}(dy)$ can be approximated by

$$\int_{\mathbf{Y}} f(y)\mathbb{P}_n(dy) = \frac{1}{n} \sum_{i=1}^{n} f(y_i).$$

Here, no parameter such as ξ is needed.

Example 2.2. Consider Monte Carlo integration through a sample of i.i.d. random variables $(Y_i)_{i=1,\ldots,n}$ defined on the measurable space $(\Omega, \mathcal{A})$. In this case, $\mathbb{P}_n$ is given by (1.4). Here, the parameter ξ is present and equal to ω, that is, the measurable space $(\Xi, \mathcal{X})$ is equal to $(\Omega, \mathcal{A})$.

Example 2.3. Consider now a more sophisticated Monte Carlo integration procedure. Suppose we also draw a sample of nonnegative weights $(W_{i,n})_{i=1,\ldots,n}$ defined on a measurable space $(\Lambda, \mathcal{L})$ and summing to 1, i.e., $\sum_{i=1}^{n} W_{i,n}(\lambda) = 1$ for all $\lambda \in \Lambda$. These weights can be (and in most cases are) independent of the sample $(Y_i)_{i=1,\ldots,n}$. Thus, it is possible to take $(\Xi, \mathcal{X}) = (\Lambda \times \Omega, \mathcal{L} \otimes \mathcal{A})$, $\xi = (\lambda, \omega)$, and $\mathbb{P}_n$ defined by the more general formula

$$\mathbb{P}_n(\xi, B) = \sum_{i=1}^{n} W_{i,n}(\lambda) \cdot \delta_{Y_i(\omega)}(B), \qquad \xi = (\lambda, \omega) \in \Lambda \times \Omega, \qquad B \in \mathcal{B}(\mathbf{Y}), \qquad n \geq 1. \tag{2.2}$$

This is, for example, the bootstrap case, which will be examined in subsection 4.4.

[5]That is, it contains a dense countable subset.

All the other examples described in section 4 and Appendix B can be embedded in this framework. In a precise probabilistic setting, each $\mathbb{P}_n$ is defined as the following mapping:

$$\text{(2.3)} \qquad \begin{array}{rrcl} \mathbb{P}_n : & \Xi \times \mathcal{B}(\mathbf{Y}) & \to & [0,1] \\ & (\xi, B) & \mapsto & \mathbb{P}_n(\xi, B) \end{array}$$

such that

- for any $\xi \in \Xi$, $\mathbb{P}_n(\xi, \cdot)$ is a probability measure on $(\mathbf{Y}, \mathcal{B}(\mathbf{Y}))$;
- for any $B \in \mathcal{B}(\mathbf{Y})$, $\mathbb{P}_n(\cdot, B)$ is $\mathcal{X}$-measurable.

Thus, $\mathbb{P}_n$ is a *transition probability* on $\Xi \times \mathcal{B}(\mathbf{Y})$ (see, e.g., [59, Chapter III] for the basic properties of this object). For those objects, the following type of convergence is used. Given a positive measure μ on $(\Xi, \mathcal{X})$, a sequence $(\mathbb{P}_n)$ of transition probabilities is said to *converge* to the probability $\mathbb{P}$, as n goes to $+\infty$, if $\mathbb{P}_n(\xi, \cdot)$ converges narrowly to $\mathbb{P}$ for μ-almost all $\xi \in \Xi$. This sort of convergence (or variants of it) has been used in several fields, especially in the theory of Young measures (see, e.g., [8, 69]).[6] When $\mathbb{P}_n$ is a transition probability, the integral appearing in (1.2) becomes

$$\text{(2.4)} \qquad \mathbb{E}_{\mathbb{P}_n(\xi,\cdot)}\, g(\cdot, \theta) = \int_{\mathbf{Y}} g(y, \theta) \mathbb{P}_n(\xi, dy).$$

3. Main results. In this section, we state our main results on almost-sure epi-convergence for the objective functions of approximate stochastic programming problems and on the convergence of minimizers. We consider a separable metric space (Θ, d) endowed with its Borel σ-field $\mathcal{B}(\Theta)$, a metric space $(\mathbf{Y}, \rho)$ endowed with its Borel σ-field $\mathcal{B}(\mathbf{Y})$, and a probability measure $\mathbb{P}$ defined on $(\mathbf{Y}, \mathcal{B}(\mathbf{Y}))$. An extended real-valued function $g : \mathbf{Y} \times \Theta \to \overline{\mathbb{R}}$ is called an *integrand*[7] if it is $\mathcal{B}(\mathbf{Y}) \otimes \mathcal{B}(\Theta)$-measurable.[8] It is said to be a *normal integrand* if $g(y, \cdot)$ is also lower semicontinuous (l.s.c.) for $\mathbb{P}$-almost all $y \in \mathbf{Y}$ and is said to be *k-Lipschitz* on Θ if, for $\mathbb{P}$-almost all $y \in \mathbf{Y}$ and for all $\theta, \theta' \in \Theta$,

$$|g(y, \theta) - g(y, \theta')| \leq k d(\theta, \theta').$$

It is said to be *nonnegative* if, for $\mathbb{P}$-almost every $y \in \mathbf{Y}$, $g(y, \cdot)$ takes on its values in $[0, +\infty]$. Given a normal integrand g and an integer $k \geq 1$, the *Lipschitz approximation* of order k of g (with respect to θ) is the function g^k defined by

$$g^k(y, \theta) = \inf_{\theta' \in \Theta} \{g(y, \theta') + kd(\theta, \theta')\}, \qquad y \in \mathbf{Y}, \qquad \theta \in \Theta,$$

where k is a superscript. Like g, g^k is a normal integrand (see, e.g., [33, Proposition 4.4]). Finally, let Θ_0 denote a dense countable subset of Θ and assume the following hold.

(A1) g is a nonnegative normal integrand on $\mathbf{Y} \times \Theta$ and the function $\theta \mapsto \mathbb{E}_{\mathbb{P}}\, g(\cdot, \theta)$ is not identically $+\infty$ on Θ.

(A2) $(\Xi, \mathcal{X}, \mu)$ is a probability space and $(\mathbb{P}_n)$ is a sequence of transition probabilities defined on $\Xi \times \mathcal{B}(\mathbf{Y})$ (see (2.3)).

[6] In that theory, the limit probability $\mathbb{P}$ may depend on ξ, but in the present paper it does not, i.e., $\mathbb{P}$ is a probability with the usual meaning.

[7] Etymologically, "integrand" means "what is to be integrated."

[8] For real-valued functions, this corresponds exactly to the definition of a *measurable random real function* in [27, Definition 1, Page 157].

Remark 3.1. As to the notation, when there is no risk of ambiguity, we simply denote $\mathbb{E}_{\mathbb{P}}\, g(\cdot,\theta)$ by $\mathbb{E}\, g(\cdot,\theta)$. The expectation of the measurable function $y \mapsto g(y,\theta)$ with respect to the transition probability $\mathbb{P}_n$, namely,

$$\mathbb{E}_{\mathbb{P}_n(\xi,\cdot)}\, g(\cdot,\theta) = \int_{\mathbf{Y}} g(y,\theta)\, \mathbb{P}_n(\xi,dy), \qquad \theta\in\Theta, \qquad \xi\in\Xi,$$

will be often abbreviated in $[\mathbb{E}_n g(\cdot,\theta)](\xi)$, where the notation clearly displays the dependence on n, θ, ξ, and the integration is performed with respect to y.

(A3) For each $\theta\in\Theta_0$, there exists a μ-negligible set $N_1(\theta)\subseteq\Xi$ such that

$$\limsup_{n\to+\infty}[\mathbb{E}_n(g(\cdot,\theta))](\xi) \le \mathbb{E}g(\cdot,\theta) \qquad \forall\xi\in\Xi\setminus N_1(\theta). \tag{3.1}$$

(A4) For each positive integer k and for $\theta\in\Theta_0$, there exists a μ-negligible set $N_2(\theta,k)\subseteq\Xi$ such that

$$\liminf_{n\to+\infty}[\mathbb{E}_n(g^k(\cdot,\theta))](\xi) \ge \mathbb{E}g^k(\cdot,\theta) \qquad \forall\xi\in\Xi\setminus N_2(\theta,k). \tag{3.2}$$

Our main result reads as follows.

THEOREM 3.1. *Let (Θ,d) be a separable metric space endowed with its Borel σ-field $\mathcal{B}(\Theta)$, $(\mathbf{Y},\rho)$ be a metric space endowed with its Borel σ-field $\mathcal{B}(\mathbf{Y})$, and g be a nonnegative normal integrand on $\mathbf{Y}\times\Theta$. If assumptions* (A1) *to* (A4) *hold, then one can find a μ-negligible subset N of Ξ such that, for every $\xi\in\Xi\setminus N$ and $\theta\in\Theta$, the sequence of functions*

$$\theta\mapsto[\mathbb{E}_n g(\cdot,\theta)](\xi), \qquad n\ge 1, \tag{3.3}$$

epi-converges on Θ to

$$\theta\mapsto\mathbb{E}\, g(\cdot,\theta). \tag{3.4}$$

This is also denoted by

$$\mathbb{E}\, g(\cdot,\theta) = \text{epi-lim}_{n\to+\infty}[\mathbb{E}_n g(\cdot,\theta)](\xi), \qquad \theta\in\Theta. \tag{3.5}$$

Remark 3.2.

(i) In assumption (A1), the nonnegativity condition can be relaxed (see, e.g., [33] or [16]).

(ii) Assumption (A3) needs only to be checked for those $\theta\in\Theta_0$ that are members of the *domain* of $\theta\mapsto\mathbb{E}_{\mathbb{P}}\, g(\cdot,\theta)$, that is, of the set

$$D_0 = \{\theta\in\Theta_0 : \mathbb{E}_{\mathbb{P}}\, g(\cdot,\theta) < +\infty\},$$

which is nonempty by (A1).

(iii) As we shall see, assumption (A3) (resp., (A4)) allows one to prove the epi-lim sup (resp., epi-lim inf) part of epi-convergence.

Remark 3.3. (i) Often, assumptions (A3) and (A4) can be checked in the form of an SLLN, namely,

$$\lim_{n\to+\infty}[\mathbb{E}_n\,(g(\cdot,\theta))](\xi) = \mathbb{E}\, g(\cdot,\theta) \qquad \forall\xi\in\Xi\setminus N_1(\theta), \tag{3.6}$$

$$\lim_{n\to+\infty}[\mathbb{E}_n(g^k(\cdot,\theta))](\xi) = \mathbb{E}g^k(\cdot,\theta) \qquad \forall k\ge 1, \qquad \forall\xi\in\Xi\setminus N_2(\theta,k). \tag{3.7}$$

For example, these relationships hold true if the approximating functions can be expressed as

(3.8)

$$[\mathbb{E}_n\, g(\cdot,\theta)](\omega) = \int_{\mathbf{Y}} g(y,\theta)\, \mathbb{P}_n(\omega, dy) = \frac{1}{n}\sum_{i=1}^{n} g(Y_i(\omega),\theta), \qquad \omega\in\Omega,\ \theta\in\Theta,\ n\geq 1,$$

where the transition probabilities $\mathbb{P}_n$ are given by (1.4). Indeed, if $Y_1,\dots,Y_n$ are i.i.d. and g is a nonnegative integrand, then, for all $\theta\in\Theta$, the random variables $g(Y_1,\theta),\dots,g(Y_n,\theta)$ are i.i.d. and nonnegative too. A similar argument can be used for g^k, because the measurability properties of g are inherited by g^k through the Lipschitz approximation operation (see [33, Proposition 4.4]). It follows that the identical distribution and independence properties of g are transmitted to g^k. Obviously, if g is nonnegative, so is g^k.

(ii) The case of pairwise i.i.d. sequences can also be treated by appealing to the Etemadi SLLN (see [24]). The case of ergodic stationary sequences can be dealt with by using the Birkhoff ergodic theorem (see [11, Corollary 6.23]). The extension to the case of asymptotically mean stationary sequences is obtained using [35, Theorem 3] (see [16, section 2.5] for a statistical application).

Thus, Theorem 3.1 can be seen as a device that transforms an SLLN-like convergence result into an SLLN-like epi-convergence result, which in turn can be used to approximate optimization problems. The SLLN approach and its extensions allow us to solve a large variety of problems. However, this is not the only available method. In particular, when purely deterministic algorithms are used, it is no longer possible to invoke an SLLN-like result to show the validity of assumptions (A3) and (A4), and a recourse to other methods is necessary. An example of this kind, dealing with robust optimization, is treated in subsection 4.6. Others are briefly described in Appendix B.3. The following remark presents a weaker form of assumption (A4), namely, assumption (A4$_w$), that can be useful in such situations. More precisely, it provides another way to prove the epi-lower limit part of epi-convergence, as shown at the end of the proof of Corollary 4.7.

Remark 3.4. Define the functions ψ_n by

$$\psi_n(\theta,\xi) = [\mathbb{E}_n(g(\cdot,\theta))](\xi), \qquad \theta\in\Theta, \qquad \xi\in\Xi, \qquad n\geq 1,$$

and consider assumption (A4$_w$).

(A4$_w$) For each $k\geq 1$ and $\theta\in\Theta_0$, there exist an integer $m(k)\geq 1$ and a μ-negligible subset $N_2(\theta,k)\subseteq\Xi$ such that

- (i) $\lim_{k\to+\infty} m(k) = +\infty$,
- (ii) $\liminf_{n\to+\infty} \psi_n^k(\theta,\xi) \geq \mathbb{E}\, g^{m(k)}(\cdot,\theta)$ for all $\xi\in\Xi\setminus N_2(\theta,k)$, where the Lipschitz approximation on the left-hand side is performed with respect to θ, namely,

$$\psi_n^k(\theta,\xi) = \inf_{\theta'\in\Theta}[\psi_n(\theta',\xi) + kd(\theta',\theta)].$$

Assumption (A4$_w$) is implied by (A4), because $\psi_n^k(\theta,\xi)\geq[\mathbb{E}_n(g^k(\cdot,\theta))](\xi)$ for all $n,\ k,\ \theta,\ \xi$. Simple examples show that the converse implication is not true.

We end this section with an application of Theorem 3.1 to the convergence of infima and minimizers for the sequence of objective functions given by (3.3).

THEOREM 3.2. *Assume that* (A1) *and* (A2) *hold, and that the sequence given by* (3.3) *epi-converges to the function given by* (3.4) *for μ-almost all $\xi\in\Xi$ as n goes to infinity. Then, the following three statements hold true.*

(a) *If, for each $n \geq 1$, $\tilde{\varepsilon}_n$ is a positive $\mathcal{X}$-measurable function defined on Ξ, then there exists a sequence of $\mathcal{X}/\mathcal{B}(\Theta)$-measurable functions $\tilde{\theta}_n : \Xi \to \Theta$ such that, for μ-almost all $\xi \in \Xi$, $\tilde{\theta}_n(\xi)$ is an $\tilde{\varepsilon}_n(\xi)$-minimizer of $\theta \mapsto [\mathbb{E}_n g(\cdot, \theta)](\xi)$, that is,*

$$\mathbb{E}_{\mathbb{P}_n(\xi,\cdot)}\, g(\cdot, \tilde{\theta}_n(\xi)) \leq \inf_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}_n(\xi,\cdot)}\, g(\cdot, \theta) + \tilde{\varepsilon}_n(\xi). \tag{3.9}$$

(b) *Assume that one can find a μ-negligible subset $N \subseteq \Xi$ such that, for all $\xi \in \Xi \setminus N$, the sequence $(\tilde{\theta}_n(\xi))_{n \geq 1}$ admits a cluster point $\tilde{\theta}(\xi)$ and $\lim_{n \to +\infty} \tilde{\varepsilon}_n(\xi) = 0$. Then, for all $\xi \in \Xi \setminus N$, $\tilde{\theta}(\xi)$ is a minimizer of $\theta \mapsto \mathbb{E} g(\cdot, \theta)$ and one has*

$$\mathbb{E} g(\cdot, \tilde{\theta}(\xi)) = \limsup_{n \to +\infty} \mathbb{E} g(\cdot, \tilde{\theta}_n(\xi)).$$

(c) *If the metric space Θ is compact and if problem* (1.1) *admits a unique solution θ^*, then for μ-almost all $\xi \in \Xi$ one has*

$$\lim_{n \to \infty} \tilde{\theta}_n(\xi) = \theta^* \qquad \textit{and} \qquad \mathbb{E} g(\cdot, \theta^*) = \lim_{n \to +\infty} \mathbb{E} g(\cdot, \tilde{\theta}_n(\xi)). \tag{3.10}$$

Remark 3.5.
(i) In (c), the uniqueness of the minimizer is assumed to hold only for $\theta \mapsto \mathbb{E} g(\cdot, \theta)$, but no uniqueness requirement is required for the $\tilde{\theta}_n(\xi)$'s (even if $\tilde{\theta}_n(\xi)$ is an exact minimizer of $\theta \mapsto \mathbb{E}_n g(\cdot, \theta)$ for each $n \geq 1$). On the other hand, the convergence properties in (3.10) hold for each sequence $(\tilde{\theta}_n(\xi))_{n \geq 1}$ satisfying the required conditions.
(ii) Statement (b) is more general than (c). However, in order to get simpler statements, we shall mainly refer to statement (c).
(iii) Obviously, if the transition probabilities $\mathbb{P}_n$'s do not depend on the parameter ξ, the conclusions of Theorems 3.1 and 3.2 still hold and do not involve ξ. This situation is encountered, for example, when one has recourse to deterministic numerical integration rules (see Appendix B.3).

4. Applications. The following examples present situations where Theorems 3.1 and 3.2 can be applied. Most of them involve various sorts of randomness that are treated using SLLN-like results, but an example on robust optimization (subsection 4.6) shows another available method. In some cases, precise statements are provided as corollaries of the main results. In others, only a short description is given. Further applications are briefly presented in Appendix B.

4.1. Monte Carlo approximation of stochastic programs. A first case arises, as in Example 2.2, when the approximating measure $\mathbb{P}_n$ is specified by the empirical distribution associated with a sample $(Y_i)_{i=1,\ldots,n}$ of random variables. In this case, $\mathbb{P}_n$ is a transition probability and one has $\Xi = \Omega$. As already mentioned, under suitable conditions on the sequence $(Y_n)_{n \geq 1}$, assumptions (A3) and (A4) hold in the form of an SLLN.

As mentioned in Remark 3.3(ii), this permits recovery of not only the i.i.d. case, as in [5], but also the pairwise i.i.d. case [32, 33] and the stationary ergodic case [70, 44, 71, 16]. Moreover, the case of asymptotically mean stationary ergodic sequences $(Y_i)_{i=1,\ldots,n}$ (see [45] or [29] for the definition) is dealt with in [17]. This result can be useful in order to establish strong consistency of estimators and SLLN for random closed sets, reasoning as in [16]. SLLN-like results that do not require integrability of the functions, but only quasi-integrability, are provided in [35]. These epi-convergence results can be used for approximating optimization problems. This has been considered, e.g., in [5], but several extensions are possible. Sometimes it is of interest to use deterministic weights different from $\frac{1}{n}$. In this case, one can have recourse to SLLN for weighted sums (this kind of SLLN has been derived in [15, 47]). An application of Theorems 3.1 and 3.2 in connection with the consistency of bootstrap estimators is given in subsection 4.4. As shown by Shapiro in [64], the Monte Carlo sampling method is quite efficient for solving large-scale stochastic programming problems. In [60], besides the approximation of stochastic programming problems, Römisch also examines the question of their stability, namely, how the infimum and the optimal solution vary when the original probability measure is perturbed. (See also Appendix B.1 for other applications in the framework of Monte Carlo techniques.)

4.2. Nonparametric simulated approximation. Let $Y : \Omega \to \mathbf{Y} = \mathbb{R}^q$ be a random variable defined on $(\Omega, \mathcal{A}, \mathbb{Q})$ and let $(Y_i)_{i\geq 1}$ be an i.i.d. sequence of random variables defined on the same space and having the same distribution as Y. As in Example 2.2, it is assumed that $\Xi = \Omega$. Further, the distribution $\mathbb{P} = \mathbb{Q}_Y$ of Y is assumed to have a density f with respect to the q-dimensional Lebesgue measure. Given a separable metric space Θ and a $\mathcal{B}(\mathbf{Y}) \otimes \mathcal{B}(\Theta)$-measurable integrand $u : \mathbf{Y} \times \Theta \to \overline{\mathbb{R}}$, consider the minimization problem

$$\inf_{\theta\in\Theta} \mathbb{P}(B(\theta)) = \inf_{\theta\in\Theta} \int_{B(\theta)} f(y)\, dy, \tag{4.1}$$

where $B(\theta) = \{y \in \mathbf{Y} : u(y, \theta) \leq 0\}$. This type of problem arises, for example, in portfolio management (see, e.g., [58]).

We first need the following result that describes a situation where the function $\theta \to \mathbb{P}(B(\theta))$ is l.s.c. Recall that the l.s.c. property is required by (A1).

PROPOSITION 4.1. *Assume the following two conditions hold.*

(a) *For $\mathbb{P}$-almost all $y \in \mathbf{Y}$, $u(y, \cdot)$ is upper semicontinuous (u.s.c.).*

(b) *For every $\theta \in \Theta$ the set $D(\theta) = \{y \in \mathbf{Y} : u(y, \theta) = 0\}$ satisfies $\mathbb{P}(D(\theta)) = 0$. Then, the function $\theta \to \mathbb{P}(B(\theta))$ is l.s.c.*

Remark 4.1. If in addition it is assumed that $u(\cdot, \theta)$ is u.s.c. for all $\theta \in \Theta$, then $\mathbb{P}(B(\theta)) > 0$ for all θ such that $B'(\theta) \cap \operatorname{supp}(\mathbb{P}) \neq \emptyset$, where

$$B'(\theta) = \{y \in \mathbf{Y} : u(y, \theta) < 0\}$$

and $\operatorname{supp}(\mathbb{P})$ is the *support* of $\mathbb{P}$, that is, the smallest closed subset of $\mathbf{Y}$ with full $\mathbb{P}$-measure.

In case the density is unknown or exceedingly complex, it is possible to approximate $\mathbb{P}$ by the same empirical measures $\mathbb{P}_n$ as in (1.4). Indeed, the SLLN can be applied to the sequence $(Y_i)_{i\geq 1}$ and implies that, for each $B \in \mathcal{B}(Y)$, one can find a $\mathbb{Q}$-negligible subset of Ω such that

$$\lim_{n\to+\infty} \mathbb{P}_n(\omega, B) = \lim_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} 1_B(Y_i(\omega)) = \mathbb{P}(B) \qquad \text{a.s.} \tag{4.2}$$

Thus, one can think of approximating problem (4.1) by

$$\inf_{\theta\in\Theta} [\mathbb{E}_n 1_{B(\theta)}](\omega) = \inf_{\theta\in\Theta} \frac{1}{n} \sum_{i=1}^{n} 1_{B(\theta)}(Y_i(\omega)). \tag{4.3}$$

This is indeed possible by applying Theorem 3.1 to the integrand g defined on $\mathbf{Y} \times \Theta$ by

$$g(y, \theta) = \mathbf{1}_{B(\theta)}(y). \tag{4.4}$$

However, this approximation is generally discontinuous with respect to θ and can cause problems for numerical optimization algorithms.[9] An alternative is to replace the empirical distribution function with a smoother estimator. A common choice is given by a nonparametric estimator of f involving a *kernel* $K(\cdot)$, namely, a probability density function, such as the Gaussian density. Then, the density $f(\cdot)$ can be approximated by the mixture

$$f_n(\omega, y) = \frac{1}{n} \sum_{i=1}^{n} \frac{K(\frac{y - Y_i(\omega)}{h_n})}{h_n^q}, \qquad \omega \in \Omega, \qquad y \in \mathbf{Y}, \qquad n \geq 1. \tag{4.5}$$

The positive parameter h_n, whose choice generally depends on the sample size, is called the *bandwidth* or *smoothing parameter*. It permits the control of the degree of smoothness of the approximating distribution. The bandwidth h_n can also depend on the data (i.e., on ω), but we will not consider this case here. The theory of kernel density estimators is reviewed in [72, 66]. As to the $\mathbb{Q}$-almost-sure convergence of $f_n(\omega, \cdot)$ to f, we shall consider the L^1-distance on the space of all probability densities on $\mathbf{Y}$. Also consider a sequence $(h_n)_{n \geq 1}$ of bandwidths satisfying the conditions

$$\lim_{n \to +\infty} h_n = 0 \qquad \text{and} \qquad \lim_{n \to +\infty} n\, h_n^q = +\infty. \tag{4.6}$$

In order to state the epi-convergence result, we consider the integrand g defined in (4.4).

COROLLARY 4.2. *Assume conditions* (a) *and* (b) *of Proposition* 4.1 *and condition* (4.6) *hold. Further, assume that* $K : \mathbf{Y} \to \mathbb{R}_+$ *is measurable and satisfies* $\int_{\mathbf{Y}} K(y)\, dy = 1$. *Then, there exists a* $\mathbb{Q}$*-negligible subset* N *of* Ω *such that for all* $\omega \in \Omega \setminus N$ *the sequence of functions*

$$\theta \mapsto [\mathbb{E}_n\, g(\cdot, \theta)](\omega) = \int_{\mathbf{Y}} g(y, \theta) f_n(\omega, y)\, dy = \frac{1}{n} \sum_{i=1}^{n} \int_{B(\theta)} \frac{K(\frac{y - Y_i(\omega)}{h_n})}{h_n^q}\, dy$$

epi-converges to

$$\theta \mapsto [\mathbb{E}\, g(\cdot, \theta)] = \int_{\mathbf{Y}} g(y, \theta) f(y)\, dy = P(B(\theta))$$

as n *goes to infinity.*

Remark 4.2. Condition (4.6) is a well-known necessary and sufficient condition for $f_n(\omega, \cdot)$ to be $\mathbb{Q}$-a.s. convergent to $f(\cdot)$ in the L^1-metric on $\mathbf{Y} = \mathbb{R}^q$ (see, e.g., [22, Theorem 1]).

For example, consider the approximation of (4.1) through the kernel estimator (4.5) and suppose that $K(\cdot)$ is given by the density of a q-dimensional standard Gaussian distribution, and let $\Phi_q(B; m, \Sigma)$ be the probability of the set B under a

[9] An approximation of this kind is known in econometrics as a *crude frequency simulator* (see [46]). Related numerical problems are documented, e.g., in [31, Pages 98–99].

Gaussian distribution with mean $m = Y(\omega)$ and variance matrix $\Sigma = h_n \cdot I_q$, where I_q is the $q \times q$ identity matrix. Then, (4.1) can be approximated through

$$\inf_{\theta\in\Theta} \frac{1}{n} \sum_{i=1}^{n} \Phi_q(B(\theta); Y_i(\omega), \Sigma). \tag{4.7}$$

Convergence of this kind of approximation has been studied in [25]. The following corollary, which deals with the convergence of minimizers, is an immediate consequence of Theorem 3.2. The details are left to the reader.

Corollary 4.3. *Assume the same assumptions as in Corollary* 4.2 *and the following additional conditions.*

(i) *The metric space Θ is compact, and the function in* (4.1) *admits a unique minimizer θ^*.*

(ii) *For each $n \geq 1$, let $\tilde{\varepsilon}_n$ be a positive $\mathcal{A}$-measurable function defined on Ω and such that for $\mathbb{Q}$-almost all $\omega \in \Omega$ the sequence $(\tilde{\varepsilon}_n(\omega))_{n\geq 1}$ converges to* 0 *as n goes to $+\infty$.*

Then, the following two properties hold.

(a) *For each $n \geq 1$, there exists a sequence of $\mathcal{A}/\mathcal{B}(\Theta)$-measurable functions $\tilde{\theta}_n : \Omega \to \Theta$ such that, for $\mathbb{Q}$-almost all $\omega \in \Omega$, $\tilde{\theta}_n(\omega)$ is an $\tilde{\varepsilon}_n(\omega)$-minimizer of $\theta \mapsto [\mathbb{E}_n g(\cdot, \theta)](\omega)$.*

(b) *For each sequence $(\tilde{\theta}_n)_{n\geq 1}$ as in* (a), *one has $\lim_{n\to\infty} \tilde{\theta}_n(\omega) = \theta^*$ for $\mathbb{Q}$-almost all $\omega \in \Omega$.*

4.3. Programming with estimated parameters. Let $(\Omega, \mathcal{A}, \mathbb{Q})$ be a probability space, $\mathbf{Y}$ be a Borel subset of $\mathbb{R}^q$, and $Y : \Omega \to \mathbf{Y}$ be a random variable defined on that space. The probability $\mathbb{P} = \mathbb{Q}_Y$ often depends on an unknown parameter $\eta \in \mathbf{H}$, where $\mathbf{H}$ is a Borel subset of $\mathbb{R}^m$. In such a case, we use the notation $\mathbb{P}^\eta$, where η is a superscript. This sort of problem arises when one seeks the solution of a program in which unknown parameters are replaced with estimators. For example, one could consider a portfolio choice program where the unknown quantities are the asset moments that have to be estimated from data.

Let ν be a σ-finite positive measure[10] on $(\mathbf{Y}, \mathcal{B}(\mathbf{Y}))$ and assume that $\mathbb{P}^\eta$ admits a density f with respect to ν, that is,

$$\mathbb{P}^\eta(B) = \int_B f(y,\eta)\,\nu(dy), \qquad \eta \in \mathbf{H}, \qquad B \in \mathcal{B}(\mathbf{Y}).$$

In addition, f is assumed to be $\mathcal{B}(\mathbf{Y}) \otimes \mathcal{B}(\mathbf{H})$-measurable. Consequently, given g satisfying assumption (A1), one has

$$\mathbb{E}_{\mathbb{P}^\eta} g(\cdot,\theta) = \int_{\mathbf{Y}} g(y,\theta)\, f(y,\eta)\,\nu(dy), \qquad \eta \in \mathbf{H}, \qquad \theta \in \Theta. \tag{4.8}$$

Further, if for each $n \geq 1$ there exists an estimator S_n of η, i.e., an $\mathcal{A}/\mathcal{B}(\mathbf{H})$-measurable map $S_n : \Omega \to \mathbf{H}$, we consider the transition probabilities defined by

$$\mathbb{P}_n(\omega, B) = \int_B f(y, S_n(\omega))\,\nu(dy), \qquad B \in \mathcal{B}(\mathbf{Y}), \qquad \omega \in \Omega, \qquad n \geq 1.$$

[10]The introduction of a σ-additive measure allows the case of infinite discrete probability distributions to be included.

The corresponding expectations are given by
(4.9)

$$[\mathbb{E}_n g(\cdot,\theta)](\omega) = \int_{\mathbf{Y}} g(y,\theta)\, f(y, S_n(\omega))\, \nu(dy), \qquad \theta \in \Theta, \qquad \omega \in \Omega, \qquad n \geq 1.$$

The following result presents a situation where the minimization of (4.8) can be reached approximately for $\mathbb{Q}$-almost all ω by minimizing (4.9). We denote by η_0 the true (unknown) value of the parameter.

Corollary 4.4. *Assume the following conditions hold.*
(i) $\mathbf{H}$ *is a Borel subset of* $\mathbb{R}^m$*, and* Θ_0 *is a dense countable subset of* Θ*.*
(ii) g *satisfies* (A1)*, and for all* $\theta \in \Theta_0$ *the function* $y \mapsto g(y,\theta)$ *is bounded* ν*-almost everywhere on* $\mathbf{Y}$*.*
(iii) *For* $\mathbb{Q}$*-almost all* $\omega \in \Omega$*, one has*

$$\lim_{n\to+\infty} \int_{\mathbf{Y}} |f(y,\eta_0) - f(y, S_n(\omega))|\, \nu(dy) = 0. \tag{4.10}$$

Under the above conditions, for $\mathbb{Q}$*-almost all* $\omega \in \Omega$ *the expectation in* (4.9) *epi-converges on* Θ *to the expectation in* (4.8)*. Further, if* Θ *is compact and if the problem*

$$\min_{\theta\in\Theta} \mathbb{E}_{\mathbb{P}^{\eta_0}}\, g(\cdot,\theta) = \min_{\theta\in\Theta} \int_{\mathbf{Y}} g(y,\theta)\, f(y,\eta_0)\, \nu(dy)$$

admits a unique solution θ^**, then Theorem* 3.2 *is applicable as well (as in Corollary* 4.3*).*

Remark 4.3. Condition (iii) of Corollary 4.4 simply requires S_n to be consistent for η_0 in the L^1-metric. It will be satisfied if the three conditions below hold.
(a) For $\mathbb{Q}$-almost all $\omega \in \Omega$, one has $\eta_0 = \lim_{n\to+\infty} S_n(\omega)$, that is, S_n is strongly consistent.
(b) For ν-almost all $y \in \mathbf{Y}$, $\eta \mapsto f(y,\eta)$ is continuous on $\mathbf{H}$.
(c) There exists a positive integrable function h defined on $\mathbf{Y}$ such that

$$|f(y,\eta)| \leq h(y), \qquad y \in \mathbf{Y}, \qquad \eta \in \mathbf{H}.$$

Indeed, it is a straightforward application of the Lebesgue dominated convergence theorem. More generally, using Hölder's inequality, a simple extension of Corollary 4.4 can be given when L^1-convergence in (4.10) is replaced with L^p-convergence ($p > 1$) and $g(\cdot,\theta) \in L^q$ for all $\theta \in \Theta$, where $1/p + 1/q = 1$.

Example 4.1. Assume that $\mathbf{Y}$ is a closed convex subset of $\mathbb{R}^q$ and let $Y : \Omega \to \mathbf{Y}$ still denote a random variable defined on $(\Omega, \mathcal{A}, \mathbb{Q})$ whose expectation $\mathbb{E}Y$ exists (i.e., all its components are finite). Further, let $(Y_i)_{i\geq 1}$ be a sequence of random variables defined on the same space with the same distribution as Y. Consider a continuous map $\psi : \mathbf{Y} \to \mathbf{H}$ and assume that the unknown parameter is $\eta_0 = \psi(\mathbb{E}Y)$. Also consider the random variables S_n defined by

$$S_n(\omega) = \psi\left(\frac{1}{n}\sum_{i=1}^{n} Y_i(\omega)\right), \qquad \omega \in \Omega, \qquad n \geq 1.$$

If the sequence $(Y_i)_{i\geq 1}$ is assumed to be i.i.d., pairwise i.i.d., or stationary ergodic, then we can invoke the classical SLLN, the Etemadi SLLN, or the Birkhoff ergodic theorem, respectively, to show the existence of a $\mathbb{P}^{\eta_0}$-negligible set N of Ω such that $\eta_0 = \psi(\mathbb{E}Y) = \lim_{n\to+\infty} S_n(\omega)$ for all $\omega \in \Omega \setminus N$. This kind of reasoning is possible in any case where an SLLN-like result can be shown to hold.

4.4. Approximations involving the consistency of bootstrap estimators. The classical bootstrap algorithm employed in statistics corresponds to the algorithm described in Example 2.3, where $(n\,W_{i,n})_{i=1,\dots,n}$ is a random vector having the multinomial distribution with all probabilities equal to $1/n$ (see, e.g., [4, Pages 49–50] and Appendix B.1 for examples). So, Y still denotes a real-valued random variable defined on the probability space $(\Omega, \mathcal{A}, \mathbb{Q})$, $(Y_i)_{i=1,\dots,n}$ an i.i.d. sample defined on the same space, and $(n\,W_{i,n})_{i=1,\dots,n}$ a random vector defined on the probability space $(\Lambda, \mathcal{L}, \mathbb{L})$ with generic element λ. It is assumed that the random vectors $(Y_i)_{i=1,\dots,n}$ and $(n\,W_{i,n})_{i=1,\dots,n}$ are independent, so that Ξ is given as the product of two probability spaces, namely,

$$(\Xi, \mathcal{X}, \mu) = (\Lambda \times \Omega, \mathcal{L} \otimes \mathcal{A}, \mathbb{L} \otimes \mathbb{Q})$$

and $\xi = (\lambda, \omega)$. The approximating transition probabilities are given by (2.2). It is not hard to check that, given a normal integrand $g : \mathbf{Y} \times \Theta \to \overline{\mathbb{R}}_+$, assumptions (A3) and (A4) can be verified through the bootstrap SLLN as in [4] (e.g., Theorem 2.1) or [3]. Therefore, Theorem 3.1 allows us to prove epi-convergence of the sequence of approximating objective functions

$$\theta \mapsto \int_{\mathbf{Y}} g(y,\theta)\,\mathbb{P}_n((\lambda,\omega), dy) = \sum_{i=1}^{n} W_{i,n}(\lambda) \cdot g(Y_i(\omega), \theta) \tag{4.11}$$

to

$$\theta \mapsto \mathbb{E}_{\mathbb{L}\otimes\mathbb{Q}}\, g(Y,\theta) = \int_{\Lambda\times\Omega} g(Y(\omega),\theta)\,(\mathbb{L}\otimes\mathbb{Q})(d(\lambda,\omega)) = \mathbb{E}_{\mathbb{Q}}\, g(Y,\theta) \tag{4.12}$$

for $(\mathbb{L}\otimes\mathbb{Q})$-almost any $(\lambda,\omega) \in \Lambda \times \Omega$ as n goes to $+\infty$. This means that, for almost all random samples $(Y_i)_{i=1,\dots,n}$ and for almost all resampling from $(Y_i)_{i=1,\dots,n}$, the sequence of approximate (bootstrapped) objective functions epi-converges to the original one.

The following result is an immediate consequence of this fact and of Theorem 3.2. It presents a situation where any sequence of (exact or approximate) minimizers associated with the functions given by (4.11) converges to the unique minimizer of (4.12).

COROLLARY 4.5. *Assume that Θ is compact and that the function given by* (4.12) *admits a unique minimizer θ^*. Let $\tilde{\varepsilon}_n : \Xi \to (0, +\infty)$ and $\tilde{\theta}_n : \Xi \to \Theta$ be sequences*[11] *like those in Theorem* 3.2, *where the functions $\theta \mapsto [\mathbb{E}_n g(\cdot,\theta)](\xi)$ are given by* (4.11). *Then for μ-almost all $\xi \in \Xi$, the sequence $(\tilde{\theta}_n(\xi))_{n\geq 1}$ converges to θ^* in Θ as n goes to $+\infty$.*

4.5. Epi-convergence of *U*-statistics. Let $Y : \Omega \to \mathbf{Y}$ be a random variable defined on $(\Omega, \mathcal{A}, \mathbb{Q})$ with distribution $\mathbb{P} = \mathbb{Q}_Y$. Here, we assume that $\mathbf{Y} = \mathbb{R}$. Consider an i.i.d. sequence $(Y_i)_{i\geq 1}$ defined on the same space and a separable metric space (Θ, d). For a given normal integrand $h : \mathbf{Y}^2 \times \Theta \to \overline{\mathbb{R}}$, we define

$$\begin{aligned} V_n(\omega,\theta) &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} h(Y_i(\omega), Y_j(\omega), \theta) \\ &= \int_{\mathbf{Y}^2} h(y_1,y_2,\theta)\,\mathbb{P}_n(\omega, dy_1 \otimes dy_2), \qquad \omega \in \Omega, \qquad \theta \in \Theta, \end{aligned} \tag{4.13}$$

[11] The existence of $\tilde{\theta}_n$ is guaranteed by Theorem 3.2(a).

where

$$\mathbb{P}_n(\omega, B) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{(Y_i(\omega),Y_j(\omega))}(B), \qquad \omega \in \Omega, \qquad B \in \mathcal{B}(\mathbf{Y}) \otimes \mathcal{B}(\mathbf{Y}) = \mathcal{B}(\mathbf{Y}^2).$$

V_n is a V-statistic or *von Mises statistic* (see, e.g., [62, section 5.1.2 and Chapter 6]) depending on a parameter $\theta \in \Theta$. These statistics arise, for example, in the theory of U- and V-*processes* (see, e.g., [20, Page 67]). We would like to find conditions ensuring that, for $\mathbb{Q}$-almost all $\omega \in \Omega$, $V_n(\omega, \cdot)$ in (4.13) converges epigraphically to $V(\cdot)$ defined by

$$V(\theta) = \mathbb{E}_{\mathbb{Q}}\, h(Y_1, Y_2, \theta), \qquad \theta \in \Theta,$$

where Y_1 and Y_2 are two independent copies of Y. Equivalently, we have

$$V(\theta) = \int_{\Omega} h(Y_1(\omega), Y_2(\omega), \theta)\, \mathbb{Q}(d\omega) = \int_{\mathbf{Y}^2} h(y_1, y_2, \theta)\, \mathbb{P}(dy_1)\mathbb{P}(dy_2). \tag{4.14}$$

Combining our Theorem 3.1 and [1, Theorem U(i), (ii)], it is possible to give a simple epi-convergence result. As in [1], we assume that h is *bounded by a* $\mathbb{P}$-*integrable product* (BIP).

COROLLARY 4.6. *Let h be a nonnegative normal integrand on $\mathbf{Y}^2 \times \Theta$ and assume the following condition.*

(BIP) *There exists a measurable function $h_0 : \mathbf{Y} \times \Theta \to \mathbb{R}$ such that, for all $\theta \in \Theta$,*
- (a) *$h_0(\cdot, \theta)$ is $\mathbb{P}$-integrable,*
- (b) *$h(y_1, y_2, \theta) \le h_0(y_1, \theta)\, h_0(y_2, \theta)$, $y_1, y_2 \in \mathbf{Y}$.*

Also assume one of the following two conditions:
- (j) *$\mathbb{P}$ is discrete,*
- (jj) *one can find a dense countable subset Θ_0 such that, for $\mathbb{P} \otimes \mathbb{P}$-almost all $(y_1^0, y_2^0) \in \mathbf{Y}^2$, the family of functions $\mathcal{F} = \{(y_1, y_2) \mapsto h(y_1, y_2, \theta) : \theta \in \Theta_0\}$ is equi-continuous at (y_1^0, y_2^0), i.e., for each $\varepsilon > 0$ there exists $\eta > 0$ such that $\rho(y_1, y_1^0) < \eta$ and $\rho(y_2, y_2^0) < \eta$ together imply $|h(y_1, y_2, \theta) - h(y_1^0, y_2^0, \theta)| < \varepsilon$ for all $\theta \in \Theta_0$.*

Under the above conditions, one can find a $\mathbb{Q}$-negligible set $N \subseteq \Omega$ satisfying

$$V(\cdot) = \text{epi-lim}_{n\to+\infty} V_n(\omega, \cdot)$$

on Θ for all $\omega \in \Omega \setminus N$.

Remark 4.4. (i) As pointed out by Aaronson et al. in [1, Page 2853], there is a difference in stating that N is $\mathbb{Q}$-negligible[12] (as can be established using their Theorem U) or that N is $\mathbb{Q} \otimes \mathbb{Q}$-negligible (as obtained from the Birkhoff ergodic theorem), since $\mathbb{Q}$ can even be singular with respect to $\mathbb{Q} \otimes \mathbb{Q}$, the limit integration measure. Example 4.1 in [1] presents a case that clarifies this difference. Only a probabilistic result, such as our Theorem 3.1, can allow the derivation of $\mathbb{Q}$-almost-sure epi-convergence and help identify the measure with respect to which N is negligible.

(ii) In the above result, we have assumed that $(Y_n)_{n\ge1}$ is an i.i.d. sequence, but an easy extension is possible. Indeed, since Theorem U of [1] holds for an ergodic, stationary process, our conclusion remains valid in this case, provided the function V appearing in (4.14) is suitably modified.

(iii) Condition (jj) of Corollary 4.6 is quite natural. Indeed, if h does not depend on θ, this condition reduces to Theorem U(ii) of [1], i.e., $(y_1, y_2) \mapsto h(y_1, y_2)$ is continuous at (y_1^0, y_2^0) for $\mathbb{P} \otimes \mathbb{P}$-almost all $(y_1^0, y_2^0) \in \mathbf{Y}^2$.

[12] Equivalently, $\mathbb{Q}$ can be seen as the diagonal measure on $\Omega \times \Omega$.

As to the convergence of minimizers, the precise statement and proof follow the same lines as the previous examples and are left to the reader.

4.6. An approximation result for robust optimization problems. As in [55], our main theorem can be applied when the sequence of functions whose epi-convergence is under scrutiny has a domain that varies with n. Interesting examples of this kind are provided by robust optimization, as considered in [7, 63, 36], or semi-infinite programming, as in [57, 37]. In order to provide a precise statement of the problem, it is necessary to define the $\mathbb{P}$-essential intersection for a random set.

Consider a map $F : \mathbf{Y} \to 2^{\Theta}$, whose values are subsets of Θ. The *graph* of F is denoted by $\mathrm{Gr}(F)$ and defined by

$$\mathrm{Gr}(F) = \{(y, \theta) \in \mathbf{Y} \times \Theta : \theta \in F(y)\}.$$

F is called a *random set* if $\mathrm{Gr}(F) \in \mathcal{B}(\mathbf{Y}) \otimes \mathcal{B}(\Theta)$. For general references on random sets see, e.g., [48] or [34]. The $\mathbb{P}$-*essential intersection*[13] of F is the subset denoted by $\wedge_{\mathbb{P}}(F)$ and defined by

$$\wedge_{\mathbb{P}}(F) = \bigcup_{N \in \mathcal{N}} \bigcap_{y \in \mathrm{supp}(\mathbb{P}) \setminus N} F(y),$$

where $\mathcal{N}$ denotes the set of all negligible sets of $(\mathbf{Y}, \mathcal{B}(\mathbf{Y}), \mathbb{P})$, and $\mathrm{supp}(\mathbb{P})$ is the *support* of $\mathbb{P}$. Thus, $\theta \in \wedge_{\mathbb{P}}(F)$ if and only if $\theta \in F(y)$ for $\mathbb{P}$-almost all $y \in \mathbf{Y}$. From this, it follows that $\wedge_{\mathbb{P}}(F)$ is closed if and only if $F(y)$ is closed for $\mathbb{P}$-almost all $y \in \mathbf{Y}$. For references on essential intersection see, e.g., [38, Page IV-34] or [36, section 3].

Essential intersection allows the problem of robust optimization to be stated precisely, as follows:

$$\min_{\theta \in \Theta} f_0(\theta) \quad \text{subject to} \quad f(y, \theta) \leq 0, \qquad \mathbb{P}\text{-a.s.}, \tag{4.15}$$

where $f_0 : \Theta \to \mathbb{R}_+$ and $f : \mathbf{Y} \times \Theta \to \mathbb{R}$ are assumed to be l.s.c. The above constraint can be expressed in terms of essential intersection, which gives an equivalent formulation of (4.15):

$$\min_{\theta \in \Theta} f_0(\theta) \quad \text{subject to} \quad \theta \in \wedge_{\mathbb{P}}(H), \tag{4.16}$$

where the (closed valued) random set H is defined by

$$H(y) = \{\theta \in \Theta : f(y, \theta) \leq 0\}. \tag{4.17}$$

Remark 4.5. This formulation is different from others that are often encountered in the literature. Alternative formulations, such as

$$\min_{\theta \in \Theta} f_0(\theta) \quad \text{subject to} \quad \theta \in \{\theta' \in \Theta : f(y, \theta') \leq 0 \ \forall y \in \mathrm{supp}(\mathbb{P})\} \tag{4.18}$$

are not precise enough, because they are correct only when the set $\mathcal{N}$ of negligible sets of $(\mathbf{Y}, \mathcal{B}(\mathbf{Y}), \mathbb{P})$ reduces to the singleton $\{\emptyset\}$. In fact, apart from this special case, the constraint set in (4.18) may be empty, whereas $\wedge_{\mathbb{P}}(H)$ is not.

[13] This object is also called *continuous intersection* in reference to the case where $\mathbb{P}$ is atomless.

The *indicator function* of a subset C of Θ is denoted by $\chi(\cdot, C)$ and defined by

$$\chi(\theta, C) = 0 \text{ if } \theta \in C, \qquad \chi(\theta, C) = +\infty \text{ if } \theta \notin C, \qquad \theta \in \Theta.$$

This function allows us to give another characterization for essential intersection. Indeed, it readily follows from the definitions that

$$\chi(\theta, \wedge_{\mathbb{P}}(H)) = \mathbb{E}_{\mathbb{P}}\, \chi(\theta, H(\cdot)) = \int_{\mathbf{Y}} \chi(\theta, H(y))\, \mathbb{P}(dy), \qquad \theta \in \Theta.$$

Consequently, problem (4.16) can be rewritten as

$$\min_{\theta\in\Theta}(f_0(\theta) + c(\theta)), \tag{4.19}$$

where $c(\theta)$ is defined by

$$c(\theta) = \chi(\theta, \wedge_{\mathbb{P}}(H)). \tag{4.20}$$

Thus, minimizing $f_0(\theta)$ under the constraint $\theta \in \wedge_{\mathbb{P}}(H)$ amounts to minimizing $f_0(\theta)+c(\theta)$ without any constraint.

Let $(\Xi, \mathcal{X}, \mu)$ be a probability space. Corollary 4.7 presents a simple approximation procedure for problem (4.19), where $\mathbb{P}$ is approximated narrowly by a sequence $(\mathbb{P}_n)_{n\geq 1}$ of transition probabilities defined on $\Xi \times \mathcal{B}(\mathbf{Y})$. For each $n \geq 1$ and $\xi \in \Xi$, we set

$$c_n(\theta, \xi) = \chi(\theta, \wedge_{\mathbb{P}_n(\xi,\cdot)}(H)) = \mathbb{E}_{\mathbb{P}_n(\xi,\cdot)}\chi(\theta, H(\cdot)) \tag{4.21}$$

and we consider the minimization problem

$$\min_{\theta\in\Theta}(f_0(\theta) + c_n(\theta, \xi)) \tag{4.22}$$

or, equivalently,

$$\min_{\theta\in\Theta} f_0(\theta) \quad \text{subject to} \quad \theta \in \wedge_{\mathbb{P}_n(\xi,\cdot)}(H).$$

COROLLARY 4.7. *Assume that $\mathbf{Y}$ and Θ are closed subsets of $\mathbb{R}^q$ and $\mathbb{R}^p$ respectively, and that the following five conditions hold.*

- (C_0) *The set of constraints of problem* (4.15) (*or* (4.16)) *is nonempty; namely, there exists at least one $\theta \in \Theta$ such that $f(y, \theta) \leq 0$ for $\mathbb{P}$-almost all $y \in \mathbf{Y}$.*
- (C_1) *$f_0 : \Theta \to \mathbb{R}_+$ and $f : \mathbf{Y} \times \Theta \to \mathbb{R}$ are l.s.c.*
- (C_2) *For μ-almost all $\xi \in \Xi$, the sequence $(\mathbb{P}_n(\xi, \cdot))_{n\geq 1}$ converges narrowly to $\mathbb{P}$.*
- (C_3) *For each $n \geq 1$ and for μ-almost all $\xi \in \Xi$, $\mathbb{P}_n(\xi, \cdot)$ is absolutely continuous with respect to $\mathbb{P}_{n+1}(\xi, \cdot)$ (which is denoted by $\mathbb{P}_n(\xi, \cdot) \ll \mathbb{P}_{n+1}(\xi, \cdot)$).*
- (C_4) *For each $\theta \in \wedge_{\mathbb{P}}(H)$ and for μ-almost all $\xi \in \Xi$, there exists an integer $m = m(\theta, \xi)$ such that $\theta \in \wedge_{\mathbb{P}_n}(H)$ for all $n \geq m$.*

Under the above conditions, for μ-almost all $\xi \in \Xi$ the sequence $(f_0(\cdot) + c_n(\cdot, \xi))_{n\geq 1}$ epi-converges to $f_0(\cdot) + c(\cdot)$ on Θ as n goes to infinity.

Remark 4.6.

(i) Condition (C_3) will be satisfied if $\operatorname{supp} \mathbb{P}_n \subseteq \operatorname{supp} \mathbb{P}_{n+1}$ for all $n \geq 1$. This holds, for example, if for each $n \geq 1$ and μ-almost all $\xi \in \Xi$ the approximating probability $\mathbb{P}_n(\xi, \cdot)$ is supported by a finite subset of $\mathbf{Y}$, say $I_n = I_n(\xi) = \{y_1^n(\xi), \ldots, y_{k_n}^n(\xi)\}$, with $I_n \subset I_{n+1}$ and $k_n < k_{n+1}$. We thus have

$$\mathbb{P}_n(\xi, dy) = \sum_{i=1}^{k_n} w_{i,n}(\xi)\, \delta_{y_i^n(\xi)}(dy),$$

where the $w_{i,n}(\xi)$'s $(i = 1, \ldots, k_n)$ are positive weights summing to 1.

(ii) Problem (4.15) (or (4.16)) may have an infinite number of constraints. However, in many situations, such as the one just mentioned, $\mathbb{P}_n(\xi, \cdot)$ has a finite support, so that the number of constraints of problem (4.22) is finite, which makes the approximation process more tractable.

(iii) Condition (C_4) means that if θ satisfies the constraints of problem (4.16), then it also satisfies the constraints of all the approximating problems for n large enough. This is, of course, a minimal requirement.

As to the convergence of minimizers, we can state the following corollary.

COROLLARY 4.8. *Assume the same hypotheses as in Corollary* 4.7*, that* Θ *is compact, and that problem* (4.19) *admits a unique solution* θ^**. For each* $n \geq 1$*, let* $\tilde{\varepsilon}_n$ *be a positive* $\mathcal{X}$*-measurable function defined on* Ξ *and such that for* μ*-almost all* $\xi \in \Xi$ *the sequence* $(\tilde{\varepsilon}_n(\xi))_{n\geq 1}$ *converges to* 0 *as* n *goes to* $+\infty$*.*

Then, the following two properties hold.

(a) *For each* $n \geq 1$*, there exists a sequence of* $\mathcal{X}/\mathcal{B}(\Theta)$*-measurable functions* $\tilde{\theta}_n : \Xi \to \Theta$ *such that, for* μ*-almost all* $\xi \in \Xi$*,* $\tilde{\theta}_n(\xi)$ *is an* $\tilde{\varepsilon}_n(\xi)$*-minimizer of* $\theta \mapsto f_0(\theta) + c_n(\theta, \xi)$*.*

(b) *For each sequence* $(\tilde{\theta}_n)_{n\geq 1}$ *as in* (a)*, one has* $\lim_{n\to\infty} \tilde{\theta}_n(\xi) = \theta^*$ *for* μ*-almost all* $\xi \in \Xi$*.*

We refer the reader to [36], where similar results are proved for asymptotic mean stationary and pairwise i.i.d. sequences using a different approach.

5. Proofs.

Proof of Theorem 3.1. For any $n \geq 1$, define on $\Xi \times \Theta$ the function

$$h_n(\xi, \theta) = [\mathbb{E}_n g(\cdot, \theta)](\xi).$$

To show (3.5), it is enough to prove the following two inequalities:

$$\text{epi-lim inf}_{n\to+\infty} h_n(\xi, \theta) \geq \mathbb{E} g(\cdot, \theta) \qquad \forall \xi \in \Xi \setminus N_1, \qquad \forall \theta \in \Theta, \tag{5.1}$$

$$\text{epi-lim sup}_{n\to+\infty} h_n(\xi, \theta) \leq \mathbb{E} g(\cdot, \theta) \qquad \forall \xi \in \Xi \setminus N_2, \qquad \forall \theta \in \Theta, \tag{5.2}$$

where N_1 and N_2 are some negligible subsets of Ξ that will be specified below. As stated in Theorem 3.1, Θ_0 denotes a dense countable subset of Θ. For any $\xi \in \Xi$ and for any fixed integer $k \geq 1$, the Lipschitz approximation of order k of $h_n(\xi, \cdot)$ is defined by

$$h_n^k(\xi, \theta) = \inf_{\theta' \in X} \{h_n(\xi, \theta') + k d(\theta, \theta')\} \qquad \forall \theta \in \Theta.$$

Since the expectation of the infimum is not greater than the infimum of the expectation, we easily obtain

$$h_n^k(\xi, \theta) \geq [\mathbb{E}_n g^k(\cdot, \theta)](\xi). \tag{5.3}$$

An appeal to Proposition 4.4 in [33] shows that g^k and h_n^k are $\widehat{\mathcal{X}} \otimes \mathcal{B}(\Theta)$-measurable.[14] Consequently, for any $\theta \in \Theta_0$ and $k \geq 1$, we can apply assumption (A4) to the sequence $([\mathbb{E}_n g^k(\cdot, \theta)](\xi))_{n\geq 1}$. This proves the existence of a negligible subset $N_1(\theta, k)$ such that, for any $\xi \in \Xi \setminus N_1(\theta, k)$,

$$\liminf_n h_n^k(\xi, \theta) \geq [\mathbb{E} g^k(\cdot, \theta)](\xi). \tag{5.4}$$

[14] Here $(\Xi, \widehat{\mathcal{X}})$ denotes the *universal completion* of $(\Xi, \mathcal{X})$ (see, e.g., [21, Page 31]).

Set $N_1 = \bigcup_{\theta\in\Theta_0}\bigcup_{k\geq 1} N_1(\theta, k)$. Inequality (5.4) is valid for $\xi \in \Xi \setminus N_1$, $k \geq 1$, and $\theta \in \Theta_0$. Moreover, it remains valid for any $\theta \in \Theta$ because each side of (5.4) defines a Lipschitz function of θ, with Lipschitz constant k. Then, taking the supremum with respect to k on both sides of (5.4) and using (A.6) together with the monotone convergence theorem, we obtain (5.1).

To prove (5.2), it is useful to define the function φ by $\varphi(\theta) = \mathbb{E}g(\cdot, \theta)$ and, for any $k \geq 1$, the function φ^k by $\varphi^k(\theta) = \inf\{\varphi(\theta') + kd(\theta, \theta') : \theta' \in \Theta\}$, where $\theta \in \Theta$. Observe first that assumption (A1) implies the finiteness of φ^k on Θ. Further, for any $\theta \in \Theta_0$, $p \geq 1$, and $k \geq 1$, one can find $\theta' = \theta'(\theta, p, k) \in \Theta$ such that $\varphi(\theta') + kd(\theta, \theta') \leq \varphi^k(\theta) + \frac{1}{p}$. Hence, for each $\theta \in \Theta_0$ and $k \geq 1$, the following equality holds true:

$$\varphi^k(\theta) = \inf\{\varphi(\theta'(\theta, p, k)) + kd(\theta, \theta'(\theta, p, k)) : p \geq 1\}. \tag{5.5}$$

Further, applying assumption (A3) to the sequence $([\mathbb{E}_n g(\cdot, \theta'(\theta, p, k))](\xi))_{n\geq 1}$, we can see that, for every $\theta \in \Theta_0$, $k \geq 1$, and $p \geq 1$, there exists a μ-negligible subset $N_2(\theta, p, k)$ such that, for every $\xi \in \Xi \setminus N_2(\theta, p, k)$,

$$\limsup_n [\mathbb{E}_n g(\cdot, \theta'(\theta, p, k))](\xi) \leq \varphi(\theta'(\theta, p, k)). \tag{5.6}$$

Set $N_2 = \bigcup_{\theta\in\Theta_0}\bigcup_{p\geq 1}\bigcup_{k\geq 1} N_2(\theta, p, k)$ and consider $\xi \in \Xi \setminus N_2$. For any $\theta \in \Theta_0$ and $k \geq 1$, we have

$$\limsup_n h_n^k(\xi, \theta) \leq \inf_{\theta'\in\Theta} \limsup_{n\to+\infty} [h_n(\xi, \theta') + kd(\theta, \theta')].$$

Restricting the infimum to the subset $\{\theta'(\theta, p, k) : p \geq 1\}$ and using (5.6) and (5.5), we obtain

$$\limsup_n h_n^k(\xi, \theta) \leq \inf_{p\geq 1} [\varphi(\theta'(\theta, p, k), \xi) + kd(\theta, \theta'(\theta, p, k))] = \varphi^k(\theta).$$

So, we have proved, for each $k \geq 1$ and $\xi \in \Xi \setminus N_2$,

$$\limsup_n h_n^k(\xi, \theta) \leq \varphi^k(\theta) \qquad \forall \theta \in \Theta_0. \tag{5.7}$$

Then, again invoking the Lipschitz property, we conclude that (5.7) remains valid for all $\theta \in \Theta$. Finally, taking the supremum on k in both sides of (5.7) and using (A.7), we get (5.2). ∎

Proof of Theorem 3.2. By Theorem 3.1, for μ-almost all $\xi \in \Xi$, the sequence given by (3.3) epi-converges to the function given by (3.4) as n goes to infinity. As to statement (a), using measurable selection arguments, it is not difficult to show that for each $n \geq 1$ there exists a $\mathcal{X}/\mathcal{B}(\Theta)$-measurable function $\tilde{\theta}_n : \Xi \to \Theta$ such that

$$[\mathbb{E}_n g(\cdot, \tilde{\theta}_n(\xi))](\xi) \leq \inf_{\theta\in\Theta} [\mathbb{E}_n g(\cdot, \theta)] + \tilde{\varepsilon}_n(\xi) \tag{5.8}$$

for μ-almost all $\xi \in \Xi$ (see, e.g., [12, 32, 33]). Thus, for each $n \geq 1$ and μ-almost all $\xi \in \Xi$, $\tilde{\theta}_n(\xi)$ is an $\tilde{\varepsilon}_n(\xi)$-minimizer of $\theta \mapsto [\mathbb{E}_n g(\cdot, \theta)](\xi)$ on Θ. To prove statement (b), fix any $\xi \in \Xi$ for which statement (a) holds. Then, the conclusion easily follows from Proposition A.3(a) applied to the sequence $\theta_n = \tilde{\theta}_n(\xi)$ and to $\theta_\infty = \theta^*$. For statement (c), use Proposition A.3(b) and Remark A.1(i) ∎

Proof of Proposition 4.1. Consider $\theta \in \Theta$ and a sequence (θ_n) converging to θ. Let $y \in B(\theta)$, or, equivalently, let y be such that $u(y,\theta) \leq 0$. In view of condition (b), we have

$$\mathbb{P}(B(\theta)) = \mathbb{P}(B(\theta) \setminus D(\theta)) = \mathbb{P}(\{y \in \mathbf{Y} : u(y,\theta) < 0\}).$$

It is thus possible to assume that $u(y,\theta) < 0$. By condition (a), we have

$$\limsup_{n\to+\infty} u(y,\theta_n) \leq u(y,\theta),$$

which implies $u(y,\theta_n) \leq u(y,\theta) < 0$ for all $n \geq m$ for m large enough. Thus, the following inclusion is valid up to a $\mathbb{P}$-negligible subset of $\mathbf{Y}$:

$$B(\theta) \subseteq \bigcup_{m\geq 1} \bigcap_{n\geq m} B(\theta_n).$$

It follows that

$$P(B(\theta)) \leq \liminf_{n\to+\infty} P(B(\theta_n)),$$

which proves the desired result. □

Proof of Corollary 4.2. For each integer $k \geq 1$, consider the Lipschitz approximation of g of order k given by

$$g^k(y,\theta) = \inf_{\theta'\in\Theta} \{1_{B(\theta')}(y) + kd(\theta,\theta')\}, \qquad (y,\theta) \in \mathbf{Y} \times \Theta.$$

For each $n \geq 1$, the empirical measure $\mathbb{P}_n$ is given by $\mathbb{P}_n(\omega, dy) = f_n(\omega, y)\, dy$, where f_n is defined by (4.5) and dy corresponds to the q-dimensional Lebesgue measure. Thus, assumption (A2) is satisfied. The validity of assumption (A1) obviously follows from Proposition 4.1, that is, g is a finite normal integrand. As to assumption (A3), observe that by Theorem 1 of [22], one has, for $\mathbb{Q}$-almost all $\omega \in \Omega$,

$$\lim_{n\to+\infty} \int_{\mathbf{R}^q} |f_n(\omega,y) - f(y)|\, dy = 0. \tag{5.9}$$

On the other hand, one has, for all $\omega \in \Omega$,

$$\begin{aligned} |[\mathbb{E}_n g(\cdot,\theta)](\omega) - \mathbb{E}g(\cdot,\theta)| &= \left| \int_{\mathbf{Y}} g(y,\theta) f_n(\omega,y)\, dy - \int_{\mathbf{Y}} g(y,\theta) f(y)\, dy \right| \\ &\leq \int_{\mathbf{Y}} g(y,\theta) |f_n(\omega,y) - f(y)|\, dy. \end{aligned}$$

Since the only possible values of g are 0 and 1, the above inequality and (5.9) show that assumption (A3) holds. Similarly, for all $k \geq 1$ the values of g^k are contained in $[0,1]$, from which it follows that assumption (A4) is satisfied. It only remains to apply Theorem 3.1. □

Proof of Corollary 4.4. As in the proof of Corollary 4.2 we get, for every $\theta \in \Theta$,

$$\left|[\mathbb{E}_n g(\cdot,\theta)](\omega) - \mathbb{E}_{\mathbb{P}^{\eta_0}} g(\cdot,\theta)\right| \leq \int_{\mathbf{Y}} g(y,\theta) \left| f(y, S_n(\omega)) - f(y,\eta) \right| \nu(dy).$$

In view of the boundedness condition on g and condition (iii), this shows that assumption (A3) holds. Replacing g with g^k proves that assumption (A4) holds too, because the boundedness condition is inherited by g^k from g. □

Proof of Corollary 4.6. Here, $\mathbf{Y}$ is replaced with $\mathbf{Y}^2$, and the function g of Theorem 3.1 is replaced with h, the normal integrand introduced in subsection 4.5. It is not difficult to check assumptions (A1) and (A2). Let us explain why assumptions (A3) and (A4) are satisfied. Theorem U of [1] can be applied to both h and h^k, where h^k is defined by

$$h^k(y_1,y_2,\theta)=\inf_{\theta'\in\Theta}[h(y_1,y_2,\theta)+kd(\theta,\theta')],\qquad y_1,y_2\in\mathbf{Y},\qquad \theta\in\Theta,\qquad k\geq 1.$$

Observe that, like h, h^k is bounded by a $\mathbb{P}$-integrable product, because $0\leq h^k\leq h$. Further, if condition (j) is satisfied, it is enough to invoke Theorem U(i) of [1]. If condition (jj) is satisfied, then it is readily checked that the function $h^k(\cdot,\cdot,\theta)$ is continuous at $\mathbb{P}\otimes\mathbb{P}$-almost every point of $\mathbf{Y}^2$ for all $k\geq 1$ and $\theta\in\Theta$. Thus, it only remains to appeal to Theorem U(ii) of [1]. ◻

Proof of Corollary 4.7. Let N be a negligible subset of Ξ such that conditions (C_2)–(C_4) hold for all $\xi\in\Xi\setminus N$. Without loss of generality, we can fix $\xi\in\Xi\setminus N$ and set $\mathbb{P}_n=\mathbb{P}_n(\xi,\cdot)$, that is, it suffices to consider the simpler case where the $\mathbb{P}_n$'s do not depend on ξ.

First observe that problem (4.19) (resp., (4.22)) admits the following equivalent formulation:

$$\min_{\theta\in\Theta}\mathbb{E}_{\mathbb{P}}\,g(\cdot,\theta),\qquad \Big(\text{resp., }\min_{\theta\in\Theta}\mathbb{E}_{\mathbb{P}_n}\,g(\cdot,\theta)\Big),\tag{5.10}$$

where g is given by

$$g(y,\theta)=f_0(\theta)+\chi(\theta,H(y))$$

for all $(y,\theta)\in\mathbf{Y}\times\Theta$ and H is given by (4.17). In view of (4.20) and (4.21), this yields

$$\mathbb{E}g(\cdot,\theta)=f_0(\theta)+c(\theta)\qquad\text{and}\qquad \mathbb{E}_n g(\cdot,\theta)=f_0(\theta)+c_n(\theta),\qquad \theta\in\Theta,\qquad n\geq 1.$$

Let us show that it is possible to apply Theorem 3.1; namely, that assumptions (A1)–(A4) are satisfied. By (C_0) and the closedness of the values of the random set H, (A1) is satisfied with g defined as above. Assumption (A2) is clearly satisfied too. As for (A3), it is enough to show that, for all $\theta\in\Theta$,

$$\limsup_{n\to+\infty}c_n(\theta)\leq c(\theta),\tag{5.11}$$

where $c(\theta)$ and $c_n(\theta)$ are defined by (4.20) and (4.21), respectively. Obviously, it suffices to prove (5.11) when $c(\theta)$ is finite (whence equal to 0), which amounts to $\theta\in\wedge_{\mathbb{P}}(H)$. By (C_4) it is possible to find $m\geq 1$ such that $\theta\in\wedge_{\mathbb{P}_n}(H)$ for all $n\geq m$. Since the c_n's are indicator functions, we deduce $c_n(\theta)=0$ for all $n\geq m$, from which (5.11) follows. As to (A4), it is more convenient to prove $(\mathrm{A4}_w)$ (see Remark 3.4). Let us first prove that for all $\theta\in\Theta$

$$\liminf_{n\to+\infty}c_n(\theta)\geq c(\theta).\tag{5.12}$$

The l.s.c. property of $y\mapsto f(y,\theta)$ implies that $y\mapsto\chi(\theta,H(y))$ is l.s.c. By (C_2) we can use the definition of narrow convergence, namely,

$$\int_{\mathbf{Y}}\varphi(y)\mathbb{P}(dy)=\lim_{n\to+\infty}\int_{\mathbf{Y}}\varphi(y)\mathbb{P}_n(dy)$$

for every bounded continuous function $\varphi : \mathbf{Y} \to \mathbb{R}$, and the fact that any extended real-valued l.s.c. function can be expressed as the supremum of an increasing sequence of bounded continuous functions. Applying this to $y \mapsto \chi(\theta, H(y))$ gives (5.12). From (5.11) and (5.12), we deduce

$$c(\theta) = \lim_{n\to+\infty} c_n(\theta). \tag{5.13}$$

Also observe that condition (C_3) implies $\wedge_{\mathbb{P}_{n+1}}(H) \subseteq \wedge_{\mathbb{P}_n}(H)$ for all $n \geq 1$ and that (5.13) implies

$$\wedge_{\mathbb{P}}(H) = \bigcap_{n\geq 1} \wedge_{\mathbb{P}_n}(H).$$

Thus, $(\wedge_{\mathbb{P}_n}(H))_{n\geq 1}$ is a nonincreasing sequence of closed subsets of Θ, whose intersection is $\wedge_{\mathbb{P}}(H)$. Since the closed balls of Θ are compact, it is readily seen that, for all $\theta \in \Theta$,

$$d(\theta, \wedge_{\mathbb{P}}(H)) = \lim_{n\to+\infty} d(\theta, \wedge_{\mathbb{P}_n}(H)) = \sup_{n\geq 1} d(\theta, \wedge_{\mathbb{P}_n}(H)), \tag{5.14}$$

where $\theta \mapsto d(\theta, C) = \inf_{\theta' \in \Theta} d(\theta, \theta')$ denotes the *distance function* of the subset C of Θ. Now, consider for all integers $n, k \geq 1$ the Lipschitz approximation of order k of c_n, namely,

$$c_n^k(\theta) = \inf_{\theta' \in \Theta} \left[\chi(\theta', \wedge_{\mathbb{P}_n}(H)) + kd(\theta', \theta)\right] = kd(\theta, \wedge_{\mathbb{P}_n}(H)).$$

Similarly, we have

$$c^k(\theta) = kd(\theta, \wedge_{\mathbb{P}}(H)).$$

In view of (5.14) it follows that, for all $k \geq 1$ and $\theta \in \Theta$,

$$c^k(\theta) = \lim_{n\to+\infty} c_n^k(\theta). \tag{5.15}$$

Further, since the infimum of the sum is not greater than the sum of the infima, one has, for all integers $n, k \geq 1$, and $\theta \in \Theta$,

$$(f_0 + c_n)^k(\theta) \geq f_0^{k/2}(\theta) + c_n^{k/2}(\theta).$$

Taking the lim inf on each side yields

$$\liminf_{n\to+\infty} (f_0 + c_n)^k(\theta) \geq f_0^{k/2}(\theta) + \liminf_{n\to+\infty} c_n^{k/2}(\theta) = f_0^{k/2}(\theta) + c^{k/2}(\theta), \tag{5.16}$$

where the last equality follows from (5.15). This is not exactly (A4$_w$), but a variant of it (the objective functions have been split into two terms). However, taking the supremum over k on each side of (5.16) yields

$$\text{epi-}\liminf\,(f_0(\theta) + c_n(\theta)) \geq \sup_{k\geq 1}(f_0^{k/2}(\theta) + c^{k/2}(\theta)) = f_0(\theta) + c(\theta),$$

where the supremum on the right-hand side is also a limit (see Proposition A.2). ☐

Appendix A. Necessary facts about epigraphical convergence. For a more complete treatment of epigraphical convergence, we refer the reader to [6] or [19].

Let (Θ, d) be a metric space and let $\varphi : \Theta \to \overline{\mathbb{R}} = [-\infty, +\infty]$ be a function from Θ into the extended reals. Its *epigraph* (or *upper graph*) is defined by

$$\text{Epi}(\varphi) = \{(\theta, \lambda) \in \Theta \times \mathbb{R} : \varphi(\theta) \leq \lambda\}. \tag{A.1}$$

Given a sequence[15] $(\varphi_n)_{n\geq 1}$ of functions from Θ into $\overline{\mathbb{R}}$, its *epigraphical lower limit* and *epigraphical upper limit* are respectively denoted by epi-lim inf φ_n and epi-lim sup φ_n, and are defined for all $\theta \in \Theta$ by

$$\text{epi-lim inf}\, \varphi_n(\theta) = \sup_{r>0} \liminf_{n\to+\infty} \inf_{\theta' \in \mathsf{B}(\theta,r)} \varphi_n(\theta'), \tag{A.2}$$

$$\text{epi-lim sup}\, \varphi_n(\theta) = \sup_{r>0} \limsup_{n\to+\infty} \inf_{\theta' \in \mathsf{B}(\theta,r)} \varphi_n(\theta'), \tag{A.3}$$

where $\mathsf{B}(\theta, r)$ denotes the open ball of radius r centered at θ. These functions are often referred to as the *lower epi-limit* and the *upper epi-limit* of $(\varphi_n)_{n\geq 1}$. Both functions are l.s.c., and one has, for all $\theta \in \Theta$,

$$\text{epi-lim inf}\, \varphi_n(\theta) \leq \text{epi-lim sup}\, \varphi_n(\theta). \tag{A.4}$$

When the equality holds in (A.4) at some point $\theta \in \Theta$, the sequence (φ_n) is said to be *epigraphically convergent* at θ. Obviously, this holds if and only if the following two inequalities are satisfied:

$$\text{epi-lim sup}\, \varphi_n(\theta) \leq \varphi(\theta) \leq \text{epi-lim inf}\, \varphi_n(\theta). \tag{A.5}$$

If this holds for all $\theta \in \Theta$, the common value defines a function φ called the *epigraphical limit* (*epi-limit*) of the sequence (φ_n). This is denoted by $\varphi = \text{epi-lim}\varphi_n$ and the sequence (φ_n) is said to *epi-converge* to φ on Θ.

Epi-convergence can be conveniently characterized by means of Lipschitz approximations. This is a decisive argument of the proofs in section 5. Given an l.s.c. function $\varphi : \Theta \to \overline{\mathbb{R}}$ and an integer $k \geq 1$, the *Lipschitz approximation* of order k of φ is defined by

$$\varphi^k(\theta) = \inf_{\theta' \in \Theta} \{\varphi(\theta') + kd(\theta, \theta')\}, \qquad k \geq 1.$$

Its main properties are listed in the following proposition (see [33, Proposition 3.3]).

Proposition A.1. *Let $\varphi : \Theta \to \overline{\mathbb{R}}$ be an l.s.c. function nonidentically equal to $+\infty$. Suppose that there exist $a > 0$, $b \in \mathbb{R}$, and $\theta_0 \in \Theta$ such that, for all $\theta \in \Theta$, $\varphi(\theta) + ad(\theta, \theta_0) + b \geq 0$. Then, the following three properties are satisfied:*

(a) *$\forall k > a$ and $\forall \theta \in \Theta$, $\varphi^k(\theta) + ad(\theta, \theta_0) + b \geq 0$;*

(b) *$\forall k \geq 1$, $\varphi^k < +\infty$ on Θ and φ^k is a Lipschitz function of constant k;*

(c) *$\forall \theta \in \Theta$, the sequence $(\varphi^k(\theta))_{k\geq 1}$ is increasing and $\varphi(\theta) = \sup_{k\geq 1} \varphi^k(\theta)$.*

The role of Lipschitz approximations for characterizing epi-convergence is made clear in the following result (see [33, Proposition 3.4]).

Proposition A.2. *Let $\varphi_n : \Theta \to \overline{\mathbb{R}}$ be a sequence of functions satisfying the following condition: there exist $a > 0$, $b \in \mathbb{R}$, and $\theta_0 \in \Theta$ such that, for every $n \geq 1$ and $\theta \in \Theta$, $\varphi_n(\theta) + ad(\theta, \theta_0) + b \geq 0$. Then, for all $\theta \in \Theta$,*

$$\text{epi-lim inf}\varphi_n(\theta) = \sup_{k\geq 1} \liminf_{n\to+\infty} \varphi_n^k(\theta), \tag{A.6}$$

$$\text{epi-lim sup}\varphi_n(\theta) = \sup_{k\geq 1} \limsup_{n\to+\infty} \varphi_n^k(\theta). \tag{A.7}$$

[15] Or, for short, (φ_n).

It is also necessary to recall the *variational properties* of epi-convergence, namely, the properties related to the convergence of infima and minimizers. For any function $\varphi : \Theta \to \overline{\mathbb{R}}$ such that $\inf_{\theta\in\Theta} \varphi(\theta) > -\infty$, the *set of exact minimizers* of φ on Θ is denoted by $\text{Argmin}(\varphi)$ and defined by

$$\text{Argmin}(\varphi) = \Big\{\theta_0 \in \Theta : \varphi(\theta_0) = \inf_{\theta\in\Theta} \varphi(\theta)\Big\}.$$

More generally, given $\varepsilon > 0$, the *set of ε-approximate minimizers* (or, for short, ε-minimizers) of φ is defined by

$$\varepsilon\text{-Argmin}(\varphi) = \Big\{\theta_\varepsilon \in \Theta : \varphi(\theta_\varepsilon) \le \inf_{\theta\in\Theta} \varphi(\theta) + \varepsilon\Big\}.$$

Clearly, the set of exact minimizers of φ corresponds to $\varepsilon = 0$. The following result concerns the convergence of infima and minimizers of an epi-convergent sequence of functions (see Theorem 7.4 and Corollary 7.20 of [19]).

PROPOSITION A.3. *Assume that the sequence (φ_n) epi-converges to a function φ in Θ. For every $n \ge 1$, let θ_n be a minimizer of φ_n (or, more generally, an ε_n-minimizer, where (ε_n) is a sequence of positive reals converging to 0).*

(a) *If θ_∞ is a cluster point of (θ_n), then θ_∞ is a minimizer of φ on Θ and $\varphi(\theta_\infty) = \limsup_{n\to+\infty} \varphi_n(\theta_n)$.*

(b) *If the sequence (θ_n) converges to $\theta_\infty \in \Theta$, then θ_∞ is a minimizer of φ on Θ and one has*

$$\varphi(\theta_\infty) = \lim_{n\to\infty} \varphi_n(\theta_n) = \lim_{n\to\infty} \inf_{\theta\in\Theta} \varphi(\theta).$$

Remark A.1. (i) The convergence of (θ_n) to θ_∞ will hold if the metric space Θ is compact and if θ_∞ is the unique minimizer of φ on Θ. Indeed, the compactness of Θ implies the existence of a subsequence (θ_{n_k}) converging to some point θ_∞, which is a minimizer of φ by statement (a). Since θ_∞ is the unique minimizer of φ, it follows that the whole sequence (θ_n) converges to θ_∞.

(ii) More generally, this convergence holds if, instead of assuming the compactness of Θ, we assume that the sequence (φ_n) is *equi-coercive* (see [19] for the definition and examples). However, in order to get simpler statements, only the compactness hypothesis on Θ has been used in this paper.

Appendix B. Further applications.

B.1. Variations on Monte Carlo. Monte Carlo sampling can be modified in order to obtain more general algorithms. For this purpose, it is convenient to assume that $(\Xi, \mathcal{X})$ is given by the product of two measurable spaces, say $(\Lambda, \mathcal{L})$ and $(\Omega, \mathcal{A})$, whose generic elements are denoted by λ and ω, so that $(\Xi, \mathcal{X}) = (\Lambda \times \Omega, \mathcal{L} \otimes \mathcal{A})$ and $\xi = (\lambda, \omega)$. This framework allows λ and ω to be stochastically independent (when the probability on $(\Xi, \mathcal{X})$ is the product of probabilities defined on $(\Lambda, \mathcal{L})$ and $(\Omega, \mathcal{A})$), stochastically dependent (when this condition is not satisfied), or even deterministically dependent (for example, when λ and ω are equal). The general structure of $\mathbb{P}_n$ is given by (2.2), where the weights $W_{i,n}$ can be chosen as described in (a) and (b) below.

(a) Consider the case in which both the nodes and the weights of numerical integration result from random drawings. We can take the same formula as (2.2), where $(W_{i,n})_{i=1,\dots,n}$ is a random vector such that $\sum_{i=1}^n W_{i,n} = 1$ and $W_{i,n} \ge 0$ for $i = 1, \dots, n$. For example, this holds if $(W_{i,n})_{i=1,\dots,n}$ has the Dirichlet distribution (see [26]). Asymptotic theory in this case has been worked out in [61]. The most prominent example in this area is the classical bootstrap algorithm employed in statistics. As already seen, it is obtained when $(n\, W_{i,n})_{i=1,\dots,n}$ is a random vector having the multinomial distribution with all probabilities equal to $1/n$ (see, e.g., [4, Pages 49–50]). A review of several bootstrap methods with special emphasis on the representation in terms of $W_{i,n}$ can be found in [14]. Laws of large numbers have been worked out in [4, 3].

(b) Monte Carlo methods can be made more robust using weights that penalize extreme values of Y_i. For example, consider the simple case where the points $(Y_i)_{i=1,\dots,n}$ come from a distribution defined on the real line. Arrange the $(Y_i)_{i=1,\dots,n}$ in increasing order so as to constitute the *order statistics* of the sample, that is, the set of ordered values $(Y_{(i)})_{i=1,\dots,n}$, with $Y_{(i)} \le Y_{(i+1)}$ for $i = 1, \dots, n-1$. Then, *trimming* gives null weights to the $k = k_n$ largest and smallest observations,

$$W_{i,n} = \begin{cases} 0, & i = 1, \dots, k, \\ \frac{1}{n-2k}, & i = k+1, \dots, n-k, \\ 0, & i = n-k+1, \dots, n, \end{cases}$$

and yields

$$\mathbb{P}_n(\omega, \cdot) = \sum_{i=1}^{n} w_{i,n} \cdot \delta_{Y_{(i)}(\omega)}(\cdot) = \frac{1}{n-2k} \cdot \sum_{i=k+1}^{n-k} \delta_{Y_{(i)}(\omega)}(\cdot)$$

(see the law of large numbers in [2]). A further alternative is *Winsorizing*, where $w_{i,n}$ is equal to 0 for any i such that $Y_i \notin [-a_n, a_n]$, where $(a_n)_{n\ge1}$ is a given sequence of positive reals. In all these cases, the weights are deterministic functions of the sample $(Y_i)_{i=1,\dots,n}$. A full treatment of these topics can be found in [65, Chapter 7].

B.2. General density approximations. Some other direct approximations of the density exist in the literature. The *Edgeworth expansion*, known to work reasonably well for approximately normal distributions, is usually introduced as a refinement of the central limit theorem, but can be also of interest in more general settings (see [10], or [13, 67] for the multidimensional case). The *saddlepoint approximations* (see, e.g., [41]) are more complex to deal with, but often much more accurate in the tails of the distributions. These methods yield approximations of the densities that depend on parameters whose value can be increased so as to obtain a greater accuracy. These parameters can be either known or estimated from a sample (as in subsection 4.3). They can be used whenever a reference distribution exists and the interest lies in small deviations with respect to it, such as in quantitative finance.

Remark B.1. In some of the previous examples, even if $\mathbb{P}$ is a probability measure, the $\mathbb{P}_n$'s may be only signed measures. This happens with certain kinds of nonparametric simulated approximation and some other kinds of approximations, such as Edgeworth expansions. A method that can be used to solve some of these problems is given in [28]. One can consult [68], for references on kernel and orthogonal series estimators taking on negative values, and for a nonparametric estimator that does not integrate to 1. The above methods can be combined: for example, one could approximate an unknown density using Edgeworth expansion, and then approximate the integral with respect to this density using numerical integration.

B.3. Deterministic algorithms. As already observed, when the approximation scheme is purely deterministic, it is no longer possible to invoke SLLN-like results (see Remark 3.3). In this section, we present some possible applications in this case. A few hybrid examples are also sketched to show the flexibility of the method.

Example B.1 (quasi–Monte Carlo). If one uses quasi–Monte Carlo (QMC) integration, as in Example 2.1, the random sample is replaced by a low discrepancy point set $(y_i)_{i=1,\ldots,n}$ and $\mathbb{P}_n$ is given by (2.1). Here, the empirical distribution is not random, so there is no parameter such as ξ or ω. QMC integration methods are in general more efficient than Monte Carlo ones (they have a faster convergence rate). However, QMC methods require more stringent hypotheses on the behavior of the function. Indeed, while any Lebesgue integrable function can be integrated using Monte Carlo algorithms, the use of low discrepancy point sets requires the function to be Riemann integrable (recall that a real-valued function is Riemann integrable if and only if it is almost everywhere continuous). The reader is referred to [23] for an excellent recent survey on QMC.

On the other hand, hybrid techniques appealing to a deterministic sequence perturbed by stochastic mechanisms can be considered, such as scrambled nets (see [50, 49]). In [43], a randomized QMC method is applied to stochastic programming. This case involves a transition probability $\mathbb{P}_n(\xi, \cdot)$, where the random element ξ does not arise from sampling from the Y's but from the "scrambling" mechanism. Here, ξ simply reduces to λ.

Example B.2 (numerical integration rules). Faster convergence rates can be obtained using numerical integration rules. In this case, the approximating measure reads as

$$\mathbb{P}_n(B) = \sum_{i=1}^{n} w_{i,n} \cdot \delta_{y_{i,n}}(B) \tag{B.1}$$

for $B \in \mathcal{B}(\mathbf{Y})$, where the sequence of nodes $(y_{i,n})_{i=1,\ldots,n}$ and the sequence of weights $(w_{i,n})_{i=1,\ldots,n}$ (often constrained to respect $\sum_{i=1}^{n} w_{i,n} = 1$) are chosen so as to optimize a certain measure of accuracy. For example, an n-point Gaussian quadrature on the real line is obtained by imposing the condition that the integrals with respect to $\mathbb{P}$ and $\mathbb{P}_n$ coincide for all polynomials of degree $2n - 1$ or less. In this case, $\mathbb{P}_n$ is just a probability, not a transition probability. The use of these methods for the approximation of stochastic programming problems has been proposed in [54, 52, 55]. This can be solved with the help of epi-convergence following the same lines as in the previous examples.

Example B.3 (quantization). The idea of quantization is to replace the original probability measure $\mathbb{P}$ with a discrete one supported by n points, say $\mathbb{P}_n$, so that $\mathbb{P}$ and $\mathbb{P}_n$ are close together with respect to a certain distance. Consider again a random variable $Y : \Omega \to \mathbf{Y}$ defined on $(\Omega, \mathcal{A}, \mathbb{Q})$, and $\mathbb{P} = \mathbb{Q}_Y$, the image measure of $\mathbb{Q}$ by Y. For each nonnegative integer n and $0 < r < +\infty$, the n-level L^r-quantization problem for Y amounts to minimizing

$$\mathbb{E}_{\mathbb{Q}} \min_{b \in \beta} \|Y - b\|^r = \int_{\mathbf{Y}} \min_{b \in \beta} \|y - b\|^r \ \mathbb{P}(dy)$$

on the set $[\mathbf{Y}]_n = \{\beta \subseteq \mathbf{Y} : \#\beta \leq n\}$, namely, over all subsets β whose cardinality is not greater than n. The optimal solution $\mathbb{P}_n$, that is, the discrete probability closest to $\mathbb{P}$, can be looked for on the set $M([\mathbf{Y}]_n)$ of all probability measures on $[\mathbf{Y}]_n$. Here, $\mathbb{P}_n$ is given by formula (B.1), where the points constitute a grid $\Gamma = \{y_{1,n}, \ldots, y_{n,n}\}$

and the weights are chosen so that

$$w_{i,n} = \mathbb{P}_n(Y = y_{i,n}) = \mathbb{P}(Y \in C_i)$$

and C_i is the Voronoi tessel of $y_{i,n} \in \Gamma$. These techniques were first applied in information theory (see, e.g., [30]), but they have been used profitably in numerical integration, as shown in [51]. Two related techniques can be mentioned here: *moment matching*, in which a discrete probability measure $\mathbb{P}_n$ is built in such a way that its first moments match the respective moments of $\mathbb{P}$ (see [39]), and *optimal discretization*, in which a distance $d(\mathbb{P}, \mathbb{P}_n)$ is minimized for $\mathbb{P}_n$ belonging to the class of finitely supported distributions (see [56]). Once $\mathbb{P}_n$ has been determined by one of the above techniques, our results can be applied to any normal integrand g satisfying the required conditions.

Acknowledgments. We are glad to thank Prof. Ruszczynski and two anonymous referees for their remarks, which helped us to improve the first version of our work. We are also grateful to Prof. Christine Choirat for valuable discussions at an early stage of the preparation of the paper.

REFERENCES

[1] J. Aaronson, R. Burton, H. Dehling, D. Gilat, T. Hill, and B. Weiss, *Strong laws for L- and U-statistics*, Trans. Amer. Math. Soc., 348 (1996), pp. 2845–2866.

[2] J. Aaronson and H. Nakada, *Trimmed sums for non-negative, mixing stationary processes*, Stochastic Process. Appl., 104 (2003), pp. 173–192.

[3] S. E. Ahmed, D. Li, A. Rosalsky, and A. I. Volodin, *Almost sure lim sup behavior of bootstrapped means with applications to pairwise i.i.d. sequences and stationary ergodic sequences*, J. Statist. Plann. Inference, 98 (2001), pp. 1–14.

[4] E. Arenal-Gutiérrez, C. Matrán, and J. A. Cuesta-Albertos, *On the unconditional strong law of large numbers for the bootstrap mean*, Statist. Probab. Lett., 27 (1996), pp. 49–60.

[5] Z. Artstein and R. J.-B. Wets, *Consistency of minimizers and the SLLN for stochastic programs*, J. Convex Anal., 2 (1995), pp. 1–17.

[6] H. Attouch, *Variational Convergence for Functions and Operators*, Appl. Math. Ser., Pitman (Advanced Publishing Program), Boston, MA, 1984.

[7] D. Bai, T. J. Carpenter, and J. M. Mulvey, *Making a case for robust models*, Manag. Sci., 43 (1997), pp. 895–907.

[8] E. J. Balder, *Lectures on Young measures theory and its applications in economics*, Rend. Istit. Mat. Univ. Trieste, XXXI, Suppl. 1 (2000), pp. 1–69.

[9] J. R. Birge and F. Louveaux, *Introduction to Stochastic Programming*, 2nd ed., Springer Ser. Oper. Res. Financ. Eng., Springer, New York, 2011.

[10] S. Blinnikov and R. Moessner, *Expansions for nearly Gaussian distributions*, Astron. Astrophys. Suppl. Ser., 130 (1998), pp. 193–205.

[11] L. Breiman, *Probability*, Classics Appl. Math. 7, SIAM, Philadelphia, PA, 1992.

[12] L. D. Brown and R. Purves, *Measurable selections of extrema*, Ann. Statist., 1 (1973), pp. 902–912.

[13] J. M. Chambers, *On methods of asymptotic approximation for multivariate distributions*, Biometrika, 54 (1967), pp. 367–383.

[14] S. Chatterjee, *Another look at the jackknife: Further examples of generalized bootstrap*, Statist. Probab. Lett., 40 (1998), pp. 307–319.

[15] X. Chen, L.-X. Zhu, and K.-T. Fang, *Almost sure convergence of weighted sums*, Statist. Sinica, 6 (1996), pp. 499–507.

[16] C. Choirat, C. Hess, and R. Seri, *A functional version of the Birkhoff ergodic theorem for a normal integrand: A variational approach*, Ann. Probab., 31 (2003), pp. 63–92.

[17] C. Choirat, C. Hess, and R. Seri, *Approximation of stochastic programming problems*, in Monte Carlo and Quasi–Monte Carlo Methods 2004, Springer, Berlin, 2006, pp. 45–59.

[18] C. Choirat and R. Seri, *Estimation in discrete parameter models*, Statist. Sci., 27 (2012), pp. 278–293.

[19] G. Dal Maso, *An Introduction to Γ-Convergence*, Progr. Nonlinear Differential Equations Appl. 8, Birkhäuser, Boston, MA, 1993.

[20] H. Dehling and W. Philipp, *Empirical process techniques for dependent data*, in Empirical Process Techniques for Dependent Data, Birkhäuser, Boston, MA, 2002, pp. 3–113.

[21] C. Dellacherie and P.-A. Meyer, *Probabilities and Potential*, North-Holland Math. Stud. 29, North-Holland, Amsterdam, 1979.

[22] L. Devroye, *The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates*, Ann. Statist., 11 (1983), pp. 896–904.

[23] J. Dick, F. Kuo, and I. Sloan, *High-dimensional integration: The quasi–Monte Carlo way*, Acta Numer., 11 (2013), pp. 133–288.

[24] N. Etemadi, *An elementary proof of the strong law of large numbers*, Z. Wahrscheinlichkeitstheorie verw. Gebiete, 55 (1981), pp. 119–122.

[25] J.-D. Fermanian and B. Salanié, *A nonparametric simulated maximum likelihood estimation method*, Econometric Theory, 20 (2004), pp. 701–734.

[26] S. Gao, J. Zhang, and T. Zhou, *Law of large numbers for sample mean of random weighting estimate*, Inform. Sci., 155 (2003), pp. 151–156.

[27] I. I. Gikhman and A. V. Skorokhod, *Introduction to the Theory of Random Processes*, W. B. Saunders Co., Philadelphia, PA, 1969.

[28] I. K. Glad, N. L. Hjort, and N. G. Ushakov, *Correction of density estimators that are not densities*, Scand. J. Statist., 30 (2003), pp. 415–427.

[29] R. M. Gray and J. C. Kieffer, *Asymptotically mean stationary measures*, Ann. Probab., 8 (1980), pp. 962–973.

[30] R. M. Gray and D. L. Neuhoff, *Quantization*, IEEE Trans. Inf. Theory, 44 (1998), pp. 2325–2383.

[31] V. Hajivassiliou, D. McFadden, and P. Ruud, *Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results*, J. Econometrics, 72 (1996), pp. 85–134.

[32] C. Hess, *Epi-convergence of Sequences of Normal Integrands and Strong Consistency of the Maximum Likelihood Estimator*, Cah. Math. Décis. 9121, Université Paris Dauphine, Paris, 1991.

[33] C. Hess, *Epi-convergence of sequences of normal integrands and strong consistency of the maximum likelihood estimator*, Ann. Statist., 24 (1996), pp. 1298–1315.

[34] C. Hess, *Set-valued integration and set-valued probability theory: An overview*, in Handbook of Measure Theory, E. Pap, ed., North Holland, Amsterdam, 2002, pp. 617–673.

[35] C. Hess, R. Seri, and C. Choirat, *Ergodic theorems for extended real-valued random variables*, Stochastic Process. Appl., 120 (2010), pp. 1908–1919.

[36] C. Hess, R. Seri, and C. Choirat, *Essential intersection and approximation results for robust optimization*, J. Nonlinear Convex Anal., 15 (2014), pp. 974–1002.

[37] R. Hettich and K. Kortanek, *Semi-infinite programming: Theory, methods, and applications*, SIAM Rev., 35 (1993), pp. 380–429.

[38] J.-B. Hiriart-Urruty, *Contributions à la programmation mathématique: cas déterministe et stochastique*, Ph.D. thesis, Université de Clermont-Ferrand II, Clermont, France, 1977.

[39] K. Høyland, M. Kaut, and S. W. Wallace, *A heuristic for moment-matching scenario generation*, Comput. Optim. Appl., 24 (2003), pp. 169–185.

[40] P. J. Huber, *Robust Statistical Procedures*, 2nd ed., CBMS-NSF Reg. Conf. Ser. in Appl. Math. 68, SIAM, Philadelphia, PA, 1996.

[41] J. L. Jensen, *Saddlepoint Approximations*, Oxford Statist. Sci. Ser. 16, Clarendon Press, New York, 1995.

[42] A. J. Kleywegt, A. Shapiro, and T. Homem-de Mello, *The sample average approximation method for stochastic discrete optimization*, SIAM J. Optim., 12 (2002), pp. 479–502.

[43] M. Koivu, *Variance reduction in sample approximations of stochastic programs*, Math. Program., 103 (2005), pp. 463–485.

[44] L. A. Korf and R. J.-B. Wets, *Random LSC functions: An ergodic theorem*, Math. Oper. Res., 26 (2001), pp. 421–445.

[45] U. Krengel, *Ergodic Theorems*, De Gruyter Stud. Math. 6, Walter de Gruyter & Co., Berlin, 1985.

[46] S. R. Lerman and C. F. Manski, *On the use of simulated frequencies to approximate choice probabilities*, in Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, MA, 1981, pp. 305–319.

[47] H.-Y. Liang and C. Su, *Strong laws for weighted sums of random elements*, Statist. Sinica, 10 (2000), pp. 1011–1019.

[48] I. S. Molchanov, *Theory of Random Sets*, Probab. Appl. 19, Springer, London, 2005.

[49] A. B. Owen, *Monte Carlo variance of scrambled net quadrature*, SIAM J. Numer. Anal., 34 (1997), pp. 1884–1910.
[50] A. B. Owen, *Scrambled net variance for integrals of smooth functions*, Ann. Statist., 25 (1997), pp. 1541–1562.
[51] G. Pagès, *A space quantization method for numerical integration*, J. Comput. Appl. Math., 89 (1998), pp. 1–38.
[52] T. Pennanen, *Epi-convergent discretizations of multistage stochastic programs*, Math. Oper. Res., 30 (2005), pp. 245–256.
[53] T. Pennanen, *Epi-convergent discretizations of multistage stochastic programs via integration quadratures*, Math. Program., 116 (2009), pp. 461–479.
[54] T. Pennanen and M. Koivu, *Integration quadratures in discretizations of stochastic programs*, Stoch. Program. E-Print Ser. 2002 (2002), No. 2001-11, https://doi.org/10.18452/8275.
[55] T. Pennanen and M. Koivu, *Epi-convergent discretizations of stochastic programs via integration quadratures*, Numer. Math., 100 (2005), pp. 141–163.
[56] G. C. Pflug, *Scenario tree generation for multiperiod financial optimization by optimal discretization*, Math. Program., 89 (2001), pp. 251–271.
[57] E. Polak, *On the mathematical foundations of nondifferentiable optimization in engineering design*, SIAM Rev., 29 (1987), pp. 21–89.
[58] G. Redaelli, *Vector stochastic optimization problems*, in Generalized Convexity and Generalized Monotonicity, Lecture Notes in Econ. and Math. Systems 502, Springer, Berlin, 2001, pp. 362–380.
[59] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, 3rd ed., Springer, Berlin, 1999.
[60] W. Römisch, *Stability of stochastic programming problems*, in Stochastic Programming, Handb. Oper. Res. Manag. Sci. 10, Springer, Berlin, 2003, pp. 483–554.
[61] A. Rosalsky and M. Sreehari, *On the limiting behavior of randomly weighted partial sums*, Statist. Probab. Lett., 40 (1998), pp. 403–410.
[62] R. Serfling, *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York, 1980.
[63] R. Seri and C. Choirat, *Approximation of robust and chance-constrained programs*, J. Optim. Theory Appl., 158 (2013), pp. 590–614.
[64] A. Shapiro, *Monte Carlo sampling methods*, in Stochastic Programming, Handb. Oper. Res. Manag. Sci. 10, Springer, 2003, pp. 353–425.
[65] G. R. Shorack, *Probability for Statisticians*, Springer Texts Statist., Springer, New York, 2000.
[66] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Monogr. Statist. Appl. Probab., Chapman & Hall, London, 1986.
[67] I. M. Skovgaard, *On multivariate Edgeworth expansions*, Int. Stat. Rev., 54 (1986), pp. 169–186.
[68] G. R. Terrell and D. W. Scott, *On improving convergence rates for nonnegative kernel density estimators*, Ann. Statist., 8 (1980), pp. 1160–1163.
[69] M. Valadier, *Young measures*, in Methods of Nonconvex Analysis, Lecture Notes in Math. 1446, Springer, Berlin, 1990, pp. 152–188.
[70] M. Valadier, *Conditional expectation and ergodic theorem for a positive integrand*, J. Nonlinear Convex Anal., 1 (2000), pp. 233–244.
[71] M. Valadier, *What differentiates stationary stochastic processes from ergodic ones: A survey*, Sūrikaisekikenkyūsho Kōkyūroku, 1215 (2001), pp. 33–52.
[72] M. Wand and M. Jones, *Kernel Smoothing*, Chapman & Hall, London, 1995.