Numer. Math. (2005) 100: 141–163
Digital Object Identifier (DOI) 10.1007/s00211-004-0571-4

Numerische Mathematik

Epi-convergent discretizations of stochastic programs via integration quadratures

Teemu Pennanen*, Matti Koivu

Department of Management Science, Helsinki School of Economics, PL 1210, 00101 Helsinki, Finland; e-mail: {pennanen, koivu}@hkkk.fi

Received January 5, 2004 / Revised version received August 22, 2004
Published online: February 4, 2005 –

Summary. The simplest and the best-known method for numerical approximation of high-dimensional integrals is the Monte Carlo method (MC), i.e. random sampling. MC has also become the most popular method for constructing numerically solvable approximations of stochastic programs. However, certain modern integration quadratures are often superior to crude MC in high-dimensional integration, so it seems natural to try to use them also in discretization of stochastic programs. This paper derives conditions that guarantee the epi-convergence of the resulting objectives to the original one. Our epi-convergence result is closely related to some of the existing ones but it is easier to apply to discretizations and it allows the feasible set to depend on the probability measure. As examples, we prove epi-convergence of quadrature-based discretizations of three different models of portfolio management and we study their behavior numerically. Besides MC, our discretizations are the only existing ones with guaranteed epi-convergence for these problem classes. In our tests, modern quadratures seem to result in faster convergence of optimal values than MC.

Mathematics Subject Classification (2000): 90C15, 49M25

1 Introduction

Let X and Ξ be complete separable metric spaces, and Σ the Borel σ-algebra on Ξ. Let P be a probability measure on (Ξ, Σ), and f an extended real-

* The work of this author was partially supported by The Finnish Foundation for Economic Education under grant no. 21599 and by Finnish Academy under contract no. 3385
Correspondence to: T. Pennanen

valued function on $X \times \Xi$, such that $f(x, \cdot)$ measurable for every $x \in X$. This paper studies numerical solution through discretization of stochastic programs (optimization problems) of the form

$$(SP) \qquad \underset{x \in X}{minimize} \quad E^P f(x) := \int_\Xi f(x, \xi) P(d\xi),$$

where the integral is interpreted as $+\infty$ when $f(x, \cdot) \notin L^1(\Xi, \Sigma, P)$. Here the decision variable x is not a function of ξ, so (SP) represents a *static* (one-stage) stochastic program. By allowing f to take on the value $+\infty$ we can incorporate constraints into the objective, which makes (SP) a very general model for static decision making problems under uncertainty. We emphasize that, unlike most studies of stochastic programs, we do not assume the feasible set

$$\operatorname{dom} E^P f(x) = \{x \in X \mid f(x, \cdot) \in L^1(\Xi, \Sigma, P)\}$$

to be known a priori. This is essential e.g. in stochastic programs without relatively complete recourse and in certain financial applications, where the determination of the feasible set is part of the problem rather than its statement; see Subsection 4.3.

In solving problems of the above form, a common approach is to replace P by a finitely supported measure of the form

$$P^\nu = \sum_{i=1}^{\nu} p_i^\nu \delta_{\xi_i^\nu},$$

where $\delta_{\xi_i^\nu}$ denotes the unit mass located at ξ_i^ν. This results in the problem

$$(SP^\nu) \qquad \underset{x \in X}{minimize} \quad E^{P^\nu} f(x) := \sum_{i=1}^{\nu} p_i^\nu f(x, \xi_i^\nu),$$

which, for moderate values of ν, is usually easier to solve than (SP). Several approaches for constructing the measures P^ν have been considered in the literature. In general, the aim is to choose P^ν so that (SP^ν) is a good approximation of (SP) and that the number ν of support points of P^ν is small enough to allow for numerical solution of (SP^ν). Note that, since f is extended real-valued and since the containment $\operatorname{dom} E^P f \subset X$ may be strict, it may happen that $\operatorname{dom} E^{P^\nu} f \neq \operatorname{dom} E^P f$.

The simplest and perhaps the most popular choice in applications is to use empirical measures of the form $P^\nu = \sum_{i=1}^{\nu} \frac{1}{\nu} \delta_{\xi_i^\nu}$, where $\{\xi_i^\nu\}_{i=1}^\nu$ is a random sample from P. Such random approximations of stochastic programs are known to be consistent as the sample size increases; see for example Artstein and Wets [2] for asymptotic analysis under very mild conditions and Shapiro [41] for more special results in the case where f is real-valued and the

feasible set is known a priori (dom $E^P f = X$). However, it is clear that a random sample can lead to a bad approximation of P, which in turn, may lead to an equally bad approximation of the optimization problem. There have been attempts to improve the accuracy of crude Monte Carlo sampling by using ideas from importance sampling technique; see Infanger [22] and Dempster and Thompson [13]. Høyland and Wallace [21] proposed to use moment matching where P^ν is constructed so that it has the first few moments of the original distribution; see also [20]. In barycentric approximation, one constructs P^ν so that, under certain convexity properties of the function $f(\cdot,\cdot)$, the optimum value of (SP^ν) provides an upper/lower bound to that of (SP); see Frauendorfer [17]. Pflug [33] proposed to construct discrete measures P^ν so that they are as close as possible to P in the sense of the so called Wasserstein-distance.

This paper studies the use of modern integration quadratures in constructing the discretizations (SP^ν). Such quadratures have the attractive feature that they have been designed to give discrete measures that approximate a given measure as well as possible. Moreover, they are just as easy to use as crude Monte Carlo and they are fast compared to methods like barycentric approximation, moment matching or that in [33]. We study the corresponding discretizations both analytically and numerically. The use of integration quadratures in solving stochastic programs have been considered in Deak [12], Lepp [28], Chen and Womersley [11] and in Pennanen and Koivu [32]. In [12], integration quadratures were not considered as feasible methods for constructing approximations (SP^ν) whereas [11,32] were mainly concerned with computational aspects without theoretical justification.

Since we are dealing with minimization problems, a natural framework for analyzing approximations is *epi-convergence;* see Attouch [3] or Rockafellar and Wets [38] for introduction to epi-convergence. Epi-convergence of the objectives is a minimal property that should be satisfied by any approximation scheme for optimization problems in order to get asymptotic convergence of optimal values and solutions. Epi-convergence of stochastic programs with respect to perturbations in the probability measure has been studied, for example, by Birge and Wets [9], Robinson and Wets [37], Dupacova and Wets [14], Kall, Ruszczyński, and Frauendorfer [23], Lucchetti and Wets [30], Artstein and Wets [1], Zervos [46], Schultz [40] and Vogel and Lachout [45]. In these studies, *weak convergence* of the approximating measures P^ν to the original measure P has been found an important property. In numerical integration, weak convergence corresponds to *consistency* which is a minimal requirement for any integration quadrature. We derive an epi-convergence result which is closely related to the ones in the above references but it is easier to apply to discretizations and it does not require the feasible set to be independent of the measure. As examples, we discretize three different models

of portfolio optimization with integration quadratures and we verify the epi-convergence of the resulting approximations. The earlier epi-convergence results do not seem applicable in these instances. In particular, in the third example, the feasible set depends on the measure in an essential way.

When using empirical measures instead of integration quadratures in approximating our test problems, one gets *almost sure* epi-convergence from the general result of [2]. In numerical tests, integration quadratures seem to result in faster convergence than empirical measures thus allowing for smaller values of ν and cheaper computations. A rigorous quantitative analysis of such phenomena would require an epigraphical analysis along the lines of Attouch and Wets [4]. Unfortunately, quantitative results for approximations of optimization problems often rely on some sort of strong convexity properties which are missing from many important problems in practice, and in particular, in two of our test problems. Some quantitative results for approximations of stochastic programs can be found in [39,42,36]. Our focus here is rather on deriving as weak conditions as possible that will allow us to deduce asymptotic epi-convergence of discretizations of as general models as possible. Even such asymptotic epi-convergence results for discretizations of stochastic programs are nontrivial, which can be seen from the lack of them for most discretization methods besides empirical approximations. This paper presents a simple deterministic approximation scheme with a theoretical justification for a rather general class of practically interesting problems.

The rest of this paper is organized as follows. Section 2 gives a brief review of modern integration quadratures and their use in generation of weakly convergent probability measures. In Section 3, we derive an epi-convergence result for $E^{P^\nu} f$. In Section 4 we combine results from Sections 2 and 3 to construct epi-convergent discretizations of some particular problems, and we study the stability of the corresponding optimal values numerically.

2 Constructing weakly convergent probability measures

Based on the importance of weak convergence in studying epi-convergence of stochastic programs, it is natural to try to choose the measures P^ν in (SP^ν) so that they converge weakly to P as $\nu \nearrow \infty$. Recall that weak convergence, which will be denoted by $P^\nu \to P$, means that

$$E^{P^\nu}\varphi \to E^P\varphi, \tag{1}$$

for all bounded and continuous functions φ; see Billingsley [8]. The literature of numerical integration is full of methods for generating such sequences. Moreover, these methods often perform much better in numerical integration than crude Monte Carlo. For very low-dimensional integrals, Gaussian

quadratures are usually most effective, but in higher dimensions, low discrepancy sequences and point sets (quasi Monte Carlo methods) often give better results. Both classes of methods are briefly reviewed below.

2.1 Low-dimensional spaces: Gaussian quadratures

Gaussian quadratures are usually very efficient in one-dimensional integration [35]. For different choices of integration limits a and b and a weight (density) function w, they yield approximations

$$\int_a^b \varphi(\eta) w(\eta) d\eta \approx \sum_{i=1}^{\nu} w_i^\nu \varphi(\eta_i^\nu), \tag{2}$$

where the quadrature points η_i^ν and weights w_i^ν are chosen so that the quadrature has an optimal order of accuracy: *a ν point Gaussian quadrature is exact for all polynomials of degree $2\nu - 1$ or less.* Given a, b and w, the values of η_i and w_i can be computed numerically. For certain choices of a, b and w, the computation of the quadrature points and weights is particularly easy.

In the case $a = -\infty$, $b = \infty$, $w(\eta) = \exp(-\eta^2)$, (2) is known as *Gauss-Hermite quadrature.* A *C*-routine for computing the the points η_i^ν and the weights w_i^ν of the Gauss-Hermite quadrature can be found in [35]. Gauss-Hermite quadrature can be used to approximate the expectation under the normal distribution P as

$$\begin{aligned} E^P \varphi &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} \varphi(\xi) d\xi \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-\eta^2} \varphi(\sqrt{2}\eta) d\eta \approx \sum_{i=1}^{\nu} \frac{1}{\sqrt{\pi}} w_i^\nu \varphi(\sqrt{2}\eta_i^\nu). \end{aligned}$$

It turns out that $w_i^\nu > 0$ and $\sum_{i=1}^{\nu} \frac{w_i}{\sqrt{\pi}} = 1$, the latter following from the fact that the quadrature is exact for constant functions. Thus,

$$P^\nu := \sum_{i=1}^{\nu} \frac{w_i^\nu}{\sqrt{\pi}} \delta_{\sqrt{2}\eta_i^\nu}$$

defines a probability measure which satisfies $E^{P^\nu} \varphi = E^P \varphi$ for all polynomials φ of order $2\nu - 1$ or less. (In particular, P^ν matches the first $2\nu - 1$ moments of the normal distribution; compare with [20]). Weierstrass approximation theorem then implies that the measures P^ν converge weakly to P as $\nu \nearrow \infty$.

Gaussian quadratures do not directly extend to higher dimensions. The generation of integration quadratures of optimal order in arbitrary dimension is an open problem [35]. If a random variable can be expressed as $\xi = G(\zeta)$,

where G is almost everywhere continuous, and $\zeta = (\zeta_1, \dots, \zeta_d)$ for independent random variables ζ_i with densities w_i, we can approximate the distribution of each ζ_i with a Gaussian quadrature to get discrete measures Q_i^ν, and then construct the measure

$$P^\nu = (Q_1^\nu, \dots, Q_d^\nu)G^{-1}; \tag{3}$$

see Theorem 1 below. Such approximations work well in dimensions 1, 2 or 3, but in higher dimensions they suffer from the "curse of dimensionality": if one approximates each Q_i with an k-point quadrature, a 10-dimensional approximation of the above form would have k^{10} quadrature points. Fortunately, there are better integration quadratures for high-dimensional spaces.

2.2 Higher dimensions: low discrepancy point sets and the method of inversion

In the scalar case, a usual thing to do is to approximate the uniform distribution on (0, 1) and to transform each point with the inverse of the distribution function of the desired distribution. This is known as the *method of inversion.* The same idea works whenever $P = QG^{-1}$, where Q is the multivariate uniform distribution and G is Q-a.s. continuous, in other words, whenever

$$\xi = G(u),$$

where u is uniformly distributed in the unit cube $(0, 1)^d$, and $G : (0, 1)^d \to \Xi$ is almost everywhere continuous. This is based on the following very useful result from Billingsley [8] where U is any metric space with Borel algebra $\mathcal{B}$.

Theorem 1 (Billingsley) *Let $G : (U, \mathcal{B}) \to (\Xi, \Sigma)$ be a measurable function and Q a probability distribution on $(U, \mathcal{B})$. Then $QG^{-1}(A) := Q(G^{-1}A)$ defines a probability measure on (Ξ, Σ), and if G is Q-a.s. continuous, then*

$$Q^\nu \to Q \implies Q^\nu G^{-1} \to QG^{-1}.$$

Given a Q-a.s. continuous G and a discrete approximation $Q^\nu = \sum_{i=1}^\nu p_i^\nu \delta_{u_i^\nu}$ of Q, Theorem 1 says that the discrete measures

$$P^\nu := Q^\nu G^{-1} = \sum_{i=1}^\nu p_i^\nu \delta_{G(u_i^\nu)}$$

converge weakly to $P = QG^{-1}$ whenever $Q^\nu \to Q$. It is then natural to try to choose discrete approximations Q^ν which are as close as possible to the uniform distribution Q. Modern methods of numerical integration do exactly this; see the books of Niederreiter [31] and Sloan and Joe [43]. Much of this theory has evolved around the following notion of distance from Q.

Definition 2 *The* star-discrepancy *of a point set* $\{u_1, \ldots, u_\nu\} \subset (0,1)^d$ *is defined as*

$$D^*(u_1, \ldots, u_\nu) = \sup_{C \in \mathcal{C}_0} |Q^\nu(C) - Q(C)|, \tag{4}$$

where

$$Q^\nu = \sum_{i=1}^{\nu} \frac{1}{\nu} \delta_{u_i},$$

and $\mathcal{C}_0$ *is the set of rectangles* $C \subset (0,1)^d$ *with one corner at the origin.*

The following is a direct consequence of Corollary 11 in Lucchetti, Salinetti and Wets [29].

Proposition 3 *For each* ν, *let* $\{u_1^\nu, \ldots, u_\nu^\nu\}$ *be a point set in the unit cube. The measures*

$$Q^\nu = \sum_{i=1}^{\nu} \frac{1}{\nu} \delta_{u_i^\nu}$$

converge weakly to the uniform distribution if and only if $D^*(u_1^\nu, \ldots, u_\nu^\nu) \to 0$.

Thus, if we can find point sets whose star-discrepancy approaches zero as $\nu \nearrow \infty$, we obtain weakly convergent discrete approximations of the uniform distribution. If $P = QG^{-1}$, we can then use the method of inversion to get weakly convergent discretizations of P. In the literature of numerical integration, many methods have been proposed that are aimed at producing point sets that have as low star-discrepancy as possible. It is thus natural to employ them in the construction of discrete measures P^ν and the corresponding approximations (SP^ν). This is what the present paper is about.

This approach to discretization of stochastic programs is close in spirit to the method proposed in Pflug [33], where the aim is to find discrete measures P^ν that are as close as possible to P in the sense of the so called *Wasserstein-distance.* In general, the problem of finding a discrete measure that minimizes a distance from a given measure can be very hard. Fortunately, in the case of star-discrepancy, many efficient methods are already available.

Example 4 (low discrepancy sequences) *Low discrepancy sequences are infinite sequences whose first* ν *points have low discrepancy for all* ν. *Examples are*

1. *Faure sequence [15]. A FORTRAN 77-routine for Faure sequence has been implemented by Fox as ACM Algorithm 647 [16].*

2. *Sobol sequence [44]. A C-routine for Sobol sequence is available in GSL (Gnu Scientific Library,* `www.gnu.org/software/gsl/gsl.html`*).*
3. *Niederreiter sequence [10]. This is also available in GSL.*

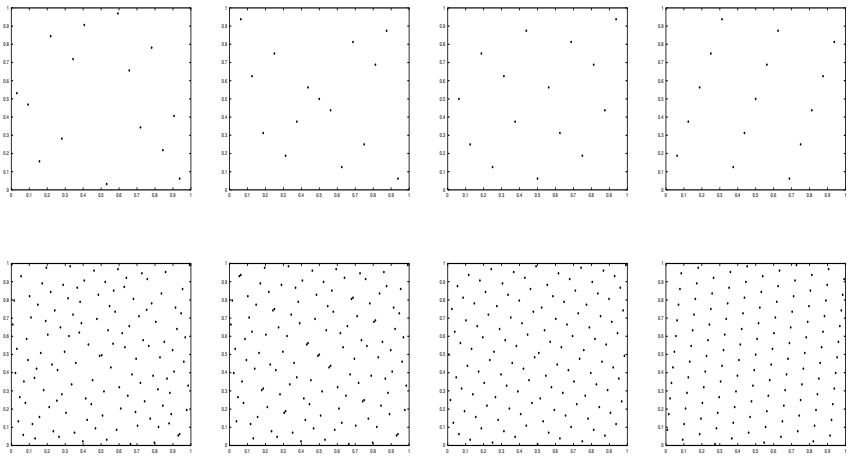
These satisfy

$$D^*(u_1, \dots, u_\nu) \le C \frac{(\log \nu)^d}{\nu} \quad \forall \nu, \tag{5}$$

for a constant C independent of ν. These examples fall in the general class of (t, s)-sequences; see [31, Chapter 4]. Figure 1 displays the first 15 and 127 points for Faure and Sobol sequences in the 2-dimensional unit cube.

In direct numerical integration, infinite low discrepancy sequences are useful in that after evaluating a ν-point quadrature, one can continue to compute the next $(\nu + 1)$-point quadrature simply by evaluating the function at one new point. In stochastic programming this advantage is lost since, in general, the solution x and thus the integrand $f(x, \cdot)$ changes every time a new point (scenario) is added to the problem. This raises the question whether it is possible to obtain more accurate quadratures if it is not required that ν points of a $(\nu + 1)$-point quadrature are the points of the ν-point quadrature. This is indeed possible.

Example 5 (low discrepancy point sets) *A set of points $\{u_1, \dots, u_\nu\}$ in the unit cube is called a* low discrepancy point set *if it has low discrepancy. Examples are*



(e) Faure (f) Sobol (g) Hammersley (h) Lattice

Fig. 1. Discretizations of the uniform distribution by 15 and 127 points

1. *Hammersley point sets [19]. Hammersley point sets can be obtained quite easily from the Halton sequence [18] that has been implemented in [16].*
2. *(t, m, s)-nets are a general class of low discrepancy point sets that are described in detail in [31, Chapter 4].*

These satisfy

$$D^*(u_1, \ldots, u_\nu) \leq C \frac{(\log \nu)^{d-1}}{\nu}, \tag{6}$$

for a constant C independent of ν. Figure 1 displays 15 and 127 Hammersley points in a 2-dimensional unit cube.

There is another class of quadratures designed to take advantage of additional regularity properties of integrands.

Example 6 (lattice rules) *Lattice rules are a general family of methods for generating point sets with low discrepancy; see for example [31, Chapter 5], [43] and L'Ecuyer and Lemieux [27]. For each d and ν there exist lattice rules that give point sets satisfying* (5)*; see [31, page 115]. This is not as good as* (6)*, but for certain classes of functions much tighter error bounds can be derived.*

Figure 1 displays lattices of 15 and 127 points produced by the so called Korobov lattice rule [26]. Parameters required by the method were provided by Pierre L'Ecuyer and Christiane Lemieux (personal communication).

It can be shown that if $Q^\nu = \sum_{i=1}^{\nu} \frac{1}{\nu} \delta_{u_i^\nu}$, then

$$\left| E^{Q^\nu} \varphi - E^Q \varphi \right| \leq V(\varphi) D^*(u_1^\nu, \ldots, u_\nu^\nu),$$

where $V(\varphi)$ is the variation of φ in the sense of Hardy and Krause; see [31, Section 2.2]. In direct integration, the above methods thus achieve the asymptotic convergence rate of ν^{-1}, whereas for pure Monte Carlo methods the rate is $\nu^{-\frac{1}{2}}$; more precisely, in Monte Carlo integration, the standard deviation of the integration error is $\sigma(\varphi)\nu^{-\frac{1}{2}}$, where $\sigma(\varphi)$ is the standard deviation of φ. It should be noticed, however, that in Monte Carlo, the error bound is independent of the dimension of the space, whereas the bounds in the above examples may depend on the dimension so that the actual error bound achieved in practice is much greater than $\sigma(\varphi)\nu^{-\frac{1}{2}}$. In numerical tests, however, low discrepancy point sets and sequences are often much more efficient than pure Monte Carlo; see for example [43]. In the tests of Section 4, one can see a similar effect in discretizations of stochastic programs.

3 Epi-convergence of the objectives

Given efficient procedures for constructing finitely supported measures P^ν that converge weakly to P, our next step is to find conditions that guarantee the epi-convergence of $E^{P^\nu} f$ to $E^P f$. Recall that the *domain* of an extended real-valued function g is the set $\operatorname{dom} g = \{x \mid g(x) < \infty\}$, and its *lower closure* is the function

$$(\operatorname{cl} g)(x) = \liminf_{y \to x} g(y).$$

A function is called *lower semicontinuous* (lsc) if it is equal to its lower closure. The *lower epi-limit* of a sequence $\{F^\nu\}$ of functions is the lsc function given by

$$(\text{e-lim inf } F^\nu)(x) = \inf_{x^\nu \to x} \liminf_{\nu \to \infty} F^\nu(x^\nu)$$

and the *upper epi-limit* is the lsc function given by

$$(\text{e-lim sup } F^\nu)(x) = \inf_{x^\nu \to x} \limsup_{\nu \to \infty} F^\nu(x^\nu).$$

If e-lim inf F^ν = e-lim sup F^ν, then the common limit, denoted e-lim F^ν, is called the *epi-limit* of $\{F^\nu\}$ and the sequence is said to *epi-converge* to it.

Epi-convergence has many important implications in studying approximations of minimization problems. The following is one of them; see Attouch [3, Section 2.2].

Theorem 7 *If* e-lim $F^\nu = F$, *then*

$$\limsup_{\nu \to \infty} \inf F^\nu \leq \inf F,$$

and if there is a convergent sequence $x^k \to x$ *such that* $x^k \in \operatorname{argmin} F^{\nu^k}$ *for some subsequence* $\{\nu^k\}_{k=1}^\infty$, *then* $x \in \operatorname{argmin} F$ *and* $\inf F^{\nu^k} \to \inf F$. *In particular, if there is a compact set* C *such that* $\operatorname{argmin} F^\nu \cap C \neq \emptyset$ *for all* ν, *then* $\inf F^\nu \to \inf F$.

Our proof of epi-convergence for (SP^ν) is based on ideas from Artstein and Wets [2], where $\{P^\nu\}_{\nu=1}^\infty$ was a sequence of empirical measures, and the main tools were the strong law of large numbers and an approximation algorithm due to Beer [5]. In our case, $\{P^\nu\}_{\nu=1}^\infty$ is a weakly convergent nonrandom sequence, and our main tools are the algorithm of Beer and Theorem 8 below. Recall that a sequence of functions $\{\varphi^\nu\}_{\nu=1}^\infty$ *converges continuously* to a function φ^0 at $\xi \in \Xi$ if

$$\varphi^\nu(\xi^\nu) \to \varphi^0(\xi),$$

whenever $\xi^\nu \to \xi$. The following is based on [1, Remark 4.3] and [40, pp. 67–68].

Theorem 8 *If $P^\nu \to P^0$, $\varphi^\nu \to \varphi^0$ continuously at P^0-almost every $\xi \in \Xi$, and if for each $\epsilon > 0$ there exists a measurable set $K_\epsilon \subset \Xi$ and a bound $b_\epsilon \in \mathbb{R}$, such that for each $\nu = 0, 1, 2, \ldots$*

1. *$|\varphi^\nu(\xi)| \le b_\epsilon$ for P^ν-almost every $\xi \in K_\epsilon$,*
2. *$\int_{\Xi \setminus K_\epsilon} |\varphi^\nu(\xi)| P^\nu(d\xi) < \epsilon$,*

then

$$\lim_{\nu \to \infty} E^{P^\nu} \varphi^\nu = E^{P^0} \varphi^0.$$

Proof. Choose an $\epsilon > 0$ and let $b_\epsilon \in \mathbb{R}$ and $K_\epsilon \subset \Xi$ be the bound and the measurable set, respectively, provided by the last condition. For each ν, let φ_ϵ^ν be the bounded function whose value at a point ξ is the projection of $\varphi^\nu(\xi)$ onto the interval $[-b_\epsilon, b_\epsilon]$. Then $|\varphi_\epsilon^\nu| \le |\varphi^\nu|$, $\varphi_\epsilon^\nu(\xi) = \varphi^\nu(\xi)$ for P^ν-almost every $\xi \in K_\epsilon$, and $\varphi_\epsilon^\nu \to \varphi_\epsilon^0$ continuously P^0-a.s. We have

$$\begin{aligned} |E^{P^\nu} \varphi^\nu - E^{P^0} \varphi^0| \le{}& |E^{P^\nu} \varphi^\nu - E^{P^\nu} \varphi_\epsilon^\nu| + |E^{P^\nu} \varphi_\epsilon^\nu - E^{P^0} \varphi_\epsilon^0| \\ &+ |E^{P^0} \varphi_\epsilon^0 - E^{P^0} \varphi^0|, \end{aligned}$$

where the second term on the right converges to zero by [7, Theorem 5.5], and for $\nu = 0, 1, \ldots,$

$$\begin{aligned} |E^{P^\nu} \varphi^\nu - E^{P^\nu} \varphi_\epsilon^\nu| &= \left| \int_{\Xi \setminus K_\epsilon} [\varphi^\nu(\xi) - \varphi_\epsilon^\nu(\xi)] P^\nu(d\xi) \right| \\ &\le 2 \int_{\Xi \setminus K_\epsilon} |\varphi^\nu(\xi)| P^\nu(d\xi) \le 2\epsilon. \end{aligned}$$

Since $\epsilon > 0$ was arbitrary, the result follows. □

In particular, if φ is P^0-a.s. continuous and bounded, then

$$E^{P^\nu} \varphi \to E^{P^0} \varphi.$$

Note that this is also implied directly by Theorem 1.

We can now state our epi-convergence result for $E^{P^\nu} f$.

Theorem 9 *Let $P^\nu \to P^0$. If for each $x \in X$,*

1. *there is an open set $N \ni x$ such that f is bounded from below on $N \times \Xi$,*
2. *$(\mathrm{cl}\, f)(x, \cdot) = f(x, \cdot)$ P^0-a.s.,*

then

$$\text{e-lim inf } E^{P^\nu} f \ge E^{P^0} f.$$

If for each $x \in \operatorname{dom} E^{P^0} f$,

3. *there is a sequence $x^\nu \to x$ such that P^ν and $\varphi^\nu := f(x^\nu, \cdot)$ satisfy the conditions of Theorem 8,*

then

$$\text{e-lim sup } E^{P^\nu} f \leq E^{P^0} f.$$

Proof. To verify the first claim, fix an $x \in X$ and let $x^\nu \to x$. According to [5] (see also the proof of [6, Theorem 7.14]), the first condition implies that there exists a sequence $\{f^k\}$ of bounded from below Lipschitz functions on $N \times \Xi$ such that $f^k \nearrow \text{cl } f$. The functions $g^k = \min\{f^k, k\}$ are also Lipschitz with $g^k \nearrow \text{cl } f$ but, in addition, they are bounded. Then

$$E^{P^\nu} f(x^\nu) \geq E^{P^\nu} g^k(x^\nu) \geq E^{P^\nu} g^k(x) - L^k d(x^\nu, x) \quad \forall \nu, k$$

where L^k is the Lipschitz constant of g^k and d is the metric on X. Since $P^\nu \to P^0$, we get

$$\liminf_{\nu\to\infty} E^{P^\nu} f(x^\nu) \geq \liminf_{\nu\to\infty} E^{P^\nu} g^k(x) = E^{P^0} g^k(x) \quad \forall k,$$

where $E^{P^0} g^k(x) \nearrow E^{P^0}(\text{cl } f)(x)$ as $k \nearrow \infty$ by the monotone convergence theorem. Since by the second condition, $E^{P^0}(\text{cl } f)(x) = E^{P^0} f(x)$, and since $x \in X$ and $x^\nu \to x$ were arbitrary, the claim follows.

For the second claim, it suffices to consider points $x \in \text{dom } E^{P^0} f$. The result then follows from the third condition and Theorem 8. □

Note that choosing $P^\nu = P^0$ for all ν, the first claim shows that under conditions 1 and 2 $E^{P^0} f$ is lsc.

Conditions 1 and 2 in Theorem 9 are strictly stronger than those in the epi-convergence result for empirical measures in [2]. The reason is that the result in [2] relies on the strong law of large numbers whereas we have used Theorem 8 which applies to any weakly convergent sequence of measures. In particular, Theorem 9 applies to all the quasi-Monte Carlo methods discussed in Section 2 as well as to the method proposed in Pflug [33], where the measures P^ν are chosen by minimizing the distance of P^ν from P^0 in the sense of the Wasserstein metric; see also Pflug and Hochreiter [34].

Theorem 9 is close to the epi-convergence results in [30, 1, 46], but it has some advantages. The results of [46] concerned real-valued functions which do not allow modeling constraints as infinite penalties. In [30, 1], the tightness-like conditions are much stronger than condition 3 above. In particular, condition 3 only asks for a measurable set K_ϵ instead of a compact one, and instead of all sequences $x^\nu \to x$, it only involves *one* sequence for each $x \in \text{dom } E^{P^0} f$. This is important since it is the tightness-like conditions that are usually hard to check for discretizations in practice. Also, requiring condition 3 to hold at all $x \in X$ would imply $\text{dom } E^{P^0} f = X$. In a sense, we have traded the stronger tightness-like conditions for conditions 1 and 2, which are often much easier to check; see Section 4. The lower-boundedness

property in condition 1 holds in many applications arising in practice. Condition 2 holds in particular if f is lsc, which was assumed in [6, Section 8.3] and [30]. According to the remark after Theorem 8, we thus have the following simplified version, which is often sufficient in applications.

Corollary 10 *Let $P^\nu \to P^0$ and assume that f is lsc. If*

1. *for each $x \in X$, there is an open set $N \ni x$ such that f is bounded from below on $N \times \Xi$,*
2. *for each $x \in \operatorname{dom} E^{P^0} f$, $f(x, \cdot)$ is P^0-a.s. continuous and bounded,*

then the functions $E^{P^\nu} f$ both pointwise and epi-converge to $E^{P^0} f$.

4 Numerical tests

4.1 Markowitz model

We start the testing with a model which can be solved exactly. Of course, discretization is unnecessary in such cases but here we get to compare the approximate solutions with the exact one. We will study the following mean-variance model

$$(MP) \qquad \underset{x\in\mathbb{R}^n}{minimize} \quad E^{P^0}(r\cdot x-\bar r\cdot x)^2$$
$$subject\ to \quad \bar r\cdot x \ge w,$$
$$\sum_{i=1}^{n} x_i \le 1,$$
$$x \in C,$$

where $x = (x_1, \dots, x_n)$ is a portfolio of assets, $r = (r_1, \dots, r_n)$ is the vector of returns (that is, r_i is the ratio of the final and initial price of asset i), $r \cdot x = \sum_{j=1}^n r_i x_i$ is the terminal wealth, w is the required level of expected wealth and C is the set of feasible portfolios. The components of the return vector r are random variables with joint distribution P^0 and expectation $\bar r$. As is well-known, the expectation in (MP) can be computed explicitly as

$$\begin{aligned} E^{P^0}(r\cdot x-\bar r\cdot x)^2 &= E^{P^0}[(r-\bar r)\cdot x]^2 \\ &= E^{P^0}[x\cdot(r-\bar r)(r-\bar r)^T x] = x\cdot Vx, \end{aligned}$$

where $V = E^{P^0}[(r-\bar r)(r-\bar r)^T]$ is the covariance matrix of r. If V and $\bar r$ are known, (MP) can then be solved without discretization with standard solvers yielding the optimal value and optimal solution.

To test the performance of integration quadratures in discretization, we will approximate problem (MP) by the discretizations

$$(MP^\nu) \qquad \underset{x\in\mathbb{R}^n}{minimize} \quad \sum_{i=1}^{\nu} p_i^\nu (r_i^\nu \cdot x - \bar{r}\cdot x)^2$$

$$\begin{aligned} subject\ to \quad \bar{r}\cdot x &\geq w,\\ \sum_{i=1}^{n} x_i &\leq 1,\\ x &\in C. \end{aligned}$$

Under mild conditions, convergence of optimal values and solutions can be guaranteed. Recall that the *support,* $\operatorname{supp} P$, of a measure P is the intersection of all closed sets of full measure. For a Borel probability measure $\operatorname{supp} P$ is well defined and unique with $P(\operatorname{supp} P) = 1$.

Proposition 11 *Assume that* $\operatorname{supp} P^0$ *is bounded, C is closed, and that the measures*

$$P^\nu = \sum_{i=1}^{\nu} p_i^\nu \delta_{r_i^\nu}$$

converge weakly to P^0 *and satisfy* $\operatorname{supp} P^\nu \subset \operatorname{supp} P^0$. *If the feasible set is bounded, then the optimal values of* (MP^ν) *converge to that of* (MP) *and the cluster points of the solutions of* (MP^ν) *are solutions of* (MP).

Proof. This fits the format of (SP) with $\xi = r$ and

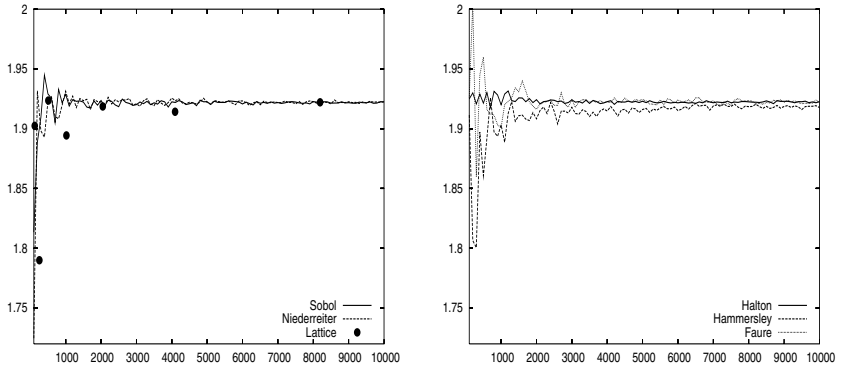
$$f(x, r) = (r\cdot x - \bar{r}\cdot x)^2 + \delta_{C'}(x),$$

where $C' = \left\{x \in C \;\middle|\; \bar{r}\cdot x \geq w,\ \sum_{i=1}^n x_i \leq 1\right\}$. So by Theorem 7, it suffices to verify the conditions of Corollary 10. Lower semicontinuity and condition 1 are clear. Since $\operatorname{supp} P^\nu \subset \operatorname{supp} P^0$, we can assume that $\Xi = \operatorname{supp} P^0$, and then condition 2 holds by boundedness of $\operatorname{supp} P^0$. □

In our test, the number of assets $n = 10$ and

$$r = \bar{r} + \sqrt{12}L(u - \frac{1}{2}e),$$

where u is uniformly distributed in the 10-dimensional unit cube, L is a 10×10 matrix and e is a vector of ones. Then $\operatorname{supp} P^0$ is bounded, r has mean $\bar{r}$ and variance $V = LL^T$. We can then solve (MP) exactly by standard QP-solvers and the discretizations (MP^ν) are easily generated by the integration quadratures described in Section 2.2. Note that the objective of



(a) Lattice rule, Niederreiter and Sobol. (b) Halton, Hammersley and Faure.

Fig. 2. Optimal values of (MP^ν) as a function of ν

(MP^ν) can be written as $x \cdot V^\nu x$, where $V^\nu = E^{P^\nu}[(r - \bar{r})(r - \bar{r})^T] = \sum_{i=1}^{\nu} p_i^\nu (r_i^\nu - \bar{r})(r_i^\nu - \bar{r})^T$, so (MP^ν) can also be solved with a QP-solver. We chose $C = \mathbb{R}^n_+$, which means that "short selling" is prohibited.

With our choices of $\bar{r}$ and V, the optimal value in the original problem (MP) turned out to be 1.9221. Figure 2 shows the development of the optimal value of (MP^ν) as a function of the number of quadrature points $\nu = 100, 200, \ldots, 10000$ for six quadratures. In our implementation, the number of points in the lattice rule is restricted to powers of 2. Each quadrature produces discretizations whose optimal value seems to converge toward the exact value 1.9221. The objective values corresponding to Halton sequence seem to behave most stably whereas Hammersley exhibits slowest convergence.

For comparison, we discretized the problem also with Monte Carlo sampling. Almost sure epi-convergence of such discretizations have been established under quite general conditions in [2]. These conditions are strictly weaker than those in Theorem 9, which guarantees *sure* (not just almost sure) epi-convergence. For each $\nu = 100, 200, \ldots, 10000$, we generated 250 discretizations and computed the average and the 90% confidence interval of the corresponding optimal values. In other words, 25 out of the 250 discretizations obtained with Monte Carlo fell outside this interval. The results are displayed in Figure 3. The optimal values obtained with Sobol sequence are repeated for reference. The average of the Monte Carlo values seems to converge towards the correct value but the convergence of the confidence interval seems slow.

Figure 4(a) displays the logarithmic error in the optimal value for Sobol discretizations as a function of $\ln \nu$. Figure 4(b) does the same for Monte

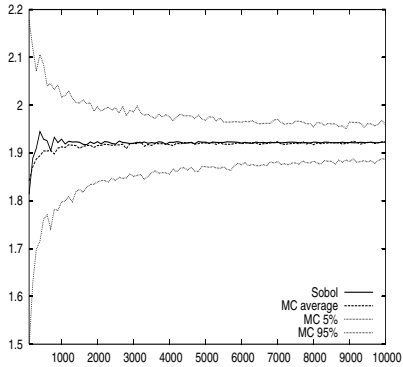
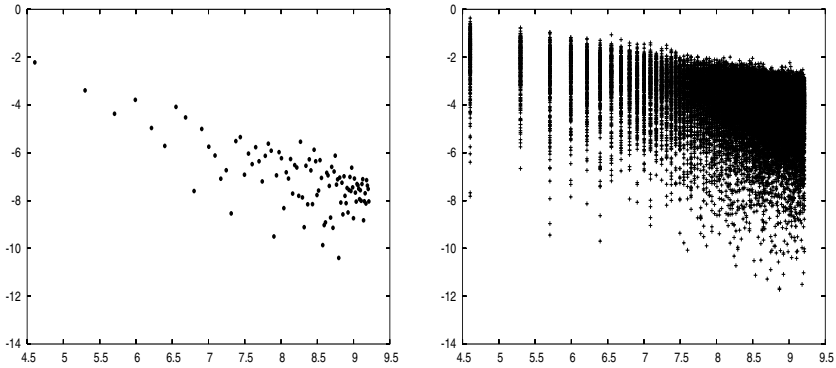


Fig. 3. Average and 90% confidence intervals for Monte Carlo



(a) Sobol (b) Monte Carlo

Fig. 4. Log of objective errors as a function of $\ln \nu$

Carlo. Interestingly, in both cases, there seems to be an upper bound on the log-error which is linear in $\ln \nu$. For Sobol, the slope is roughly -1 whereas for MC it seems to be close to $-1/2$. This kind of quantitative behavior is not explained by the general results of Section 3, but since, in this particular example, the objective is strongly convex, the quantitative results of [39,42, 36] may apply. However, it is not at all clear how one should quantitatively compare the performance of a deterministic discretization method with a random one. One may be able to give convergence rate results for a sequence of deterministic approximations of a specially structured stochastic program as the number of quadrature points is increased, but the same cannot be done for Monte Carlo. Indeed, no matter how large a sample is taken, Monte Carlo can lead to an arbitrarily bad approximation of the original problem. For

Monte Carlo, one may be able to estimate confidence intervals (or "statistical bounds") for the approximate optimal values, but such intervals have to do only with the randomness of the approximation method, not with the actual problem to be solved. The above way of computing confidence intervals and averages from 250 independent sample approximations does not seem very practical since it involves the solution of 250 optimization problems.

4.2 Utility maximization

The objective in the Markowitz model penalizes for exceeding the expected wealth $\bar{r} \cdot x$. When the distribution of r is symmetric, this does not matter, but in practice, the distribution of r is usually nonsymmetric since $r \geq 0$. The following utility maximization problem still makes sense

$$(UP) \qquad \underset{x \in \mathbb{R}^n}{maximize} \quad E^{P^0} u(r \cdot x)$$

$$subject\ to \quad \sum_{i=1}^{n} x_i \leq w_0,$$

$$x \in C.$$

Here x, r and C are as in the previous example and u measures the utility from terminal wealth.

In general, (UP) cannot be solved analytically, so we will consider the discretizations

$$(UP^\nu) \qquad \underset{x \in \mathbb{R}^n}{maximize} \quad \sum_{i=1}^{\nu} p_i^\nu u(r_i^\nu \cdot x)$$

$$subject\ to \quad \sum_{i=1}^{n} x_i \leq w_0,$$

$$x \in C.$$

Proposition 12 *Assume* $\operatorname{supp} P^0 \subset \mathbb{R}^n_+$, *$u$ is continuous and bounded on $\mathbb{R}_+$, C is closed and contained in $\mathbb{R}^n_+$ (short selling is not allowed) and that the measures*

$$P^\nu = \sum_{i=1}^{\nu} p_i^\nu \delta_{r_i^\nu}$$

converge weakly to P^0 and satisfy $\operatorname{supp} P^\nu \subset \mathbb{R}^n_+$. *Then the optimal values of (UP^ν) converge to that of (UP) and the cluster points of the solutions of (UP^ν) are solutions of (UP).*

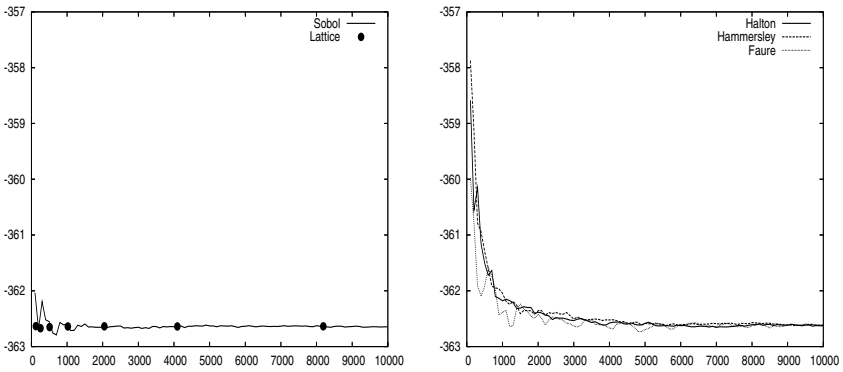
Proof. This fits the format of (SP) with $\Xi = \mathbb{R}^n_+$, $\xi = r$, and

$$f(x, r) = -u(r \cdot x) + \delta_{C'}(x),$$

where $C' = \left\{ x \in C \;\middle|\; \sum_{i=1}^n x_i \leq w_0 \right\}$. So by Theorem 7, it suffices to verify the conditions of Corollary 10. Since u is continuous and C is closed, f is lsc. Condition 1 follows from the boundedness of u on $\mathbb{R}_+$, and the fact that $r \cdot x \in \mathbb{R}_+$ for all $r \geq 0$ and $x \in C \subset \mathbb{R}_+$. Condition 2 follows from the boundedness and continuity of u on $\mathbb{R}_+$. □

Note that many familiar utility functions, like the exponential utility, are bounded on $\mathbb{R}_+$. More general utility functions are easily modified to be bounded on $\mathbb{R}_+$ in a way that does not affect computations in practice.

In our test, the number of assets $n = 30$, r follows a multivariate log-normal distribution, $u(w) = -\exp(-w)$ and $C = \mathbb{R}^n_+$. Figure 5 shows the development of the optimal value of (UP^ν) as a function of the number of quadrature points $\nu = 100, 200, \ldots, 10000$ for five quadratures. Again, the quadratures seem to converge to a common value, but this time, Halton, Hammersley and Faure exhibit slowest convergence whereas Sobol and lattice rule seem to work best. In our implementation of the Niederreiter sequence, the maximum dimension is 12, so it cannot be used here. Figure 6 depicts the development of the average and the 90% confidence interval for the optimal values obtained with 250 Monte Carlo samples for each value of ν. The values obtained with Sobol are shown for reference. The average seems to converge to the same value as the optimal values obtained by the quadratures but the confidence interval narrows down very slowly.



(a) Lattice rule and Sobol.

(b) Halton, Hammersley and Faure.

Fig. 5. Optimal values of (UP^ν) as a function of ν

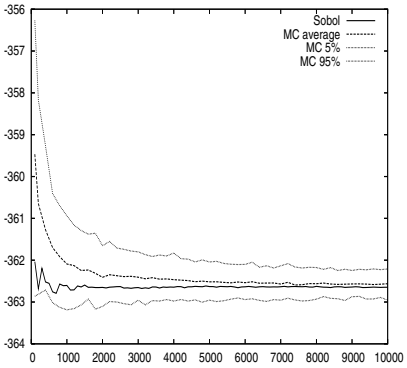


Fig. 6. Average and 90% confidence intervals for Monte Carlo

4.3 Super-replication of contingent claims

Consider the problem

$$(PP) \qquad \underset{V,\theta}{minimize} \quad V$$

$$\begin{aligned} subject\ to \quad S_0 \cdot \theta &\leq V, \\ S \cdot \theta &\geq F, \quad P^0\text{-}a.s. \\ \theta &\in C, \end{aligned}$$

where V is the wealth invested in a portfolio $\theta = (\theta_1, \ldots, \theta_J)$ of assets that have prices $S_0 = (S_0^1, \ldots, S_0^J)$ at the beginning and $S = (S^1, \ldots, S^J)$ at the end of a holding period and F is a cash-flow at the end of the holding period. S and F are random variables with joint distribution P^0. (PP) can be interpreted as a pricing problem of a seller of the cash-flow F; see for example King [24] and Korf [25, Section 7]. The seller tries to find the least amount of initial wealth that can be used to buy a portfolio that is almost surely worth at least F at the end of the holding period. The optimum value of (PP) is called the seller's price of F and an optimal portfolio θ is called a seller's hedge.

(PP) is a semi-infinite linear programming problem and, in general, impossible to solve analytically. Replacing P^0 by a discrete measure $P^\nu = \sum_{i=1}^{\nu} p_i^\nu \delta_{(S_i^\nu, F_i^\nu)}$ with $p_i^\nu > 0$, for all $i = 1, \ldots \nu$ gives the problem

$$(PP^\nu) \qquad \underset{V,\theta}{minimize} \quad V$$

$$\begin{aligned} subject\ to \quad S_0 \cdot \theta &\leq V, \\ S_i^\nu \cdot \theta &\geq F_i^\nu, \quad i = 1, \ldots, \nu, \\ \theta &\in C, \end{aligned}$$

which is an LP problem for which many solvers are available.

Note that, in this example, the feasible region depends on the measure, and dom $E^{P^0} f$ is impossible to characterize explicitly in general. This problem does not fit the frameworks of [30,1,46,41].

Proposition 13 *Assume that the points* $\{(S_i^\nu, F_i^\nu)\}_{i=1}^\nu$ *are all contained in* supp P^0 *and that for some* $\{p_i^\nu\}_{i=1}^\nu$, $\nu = 0, 1, 2, \ldots$, *with* $p_i^\nu > 0$, *for all* $i = 1, \ldots \nu$, *the measures*

$$P^\nu = \sum_{i=1}^{\nu} p_i^\nu \delta_{(S_i^\nu, F_i^\nu)}$$

converge weakly to P^0. *If the feasible set is bounded, then the optimal values to* (PP^ν) *converge to the seller's price of* F *and the cluster points of the solutions of* (PP^ν) *are seller's hedges for* F.

Proof. This can be written as (SP) with $x = (V, \theta)$, $\xi = (S, F)$ and

$$f(V, \theta, S, F) = V + \delta_C(\theta) + \delta_{C_0}(V, \theta) + \delta_{C_1}(\theta, S, F),$$

where

$$C_0 = \{(V, \theta) \mid S_0 \cdot \theta \leq V\}$$

and

$$C_1 = \{(\theta, S, F) \mid S \cdot \theta \geq F\}.$$

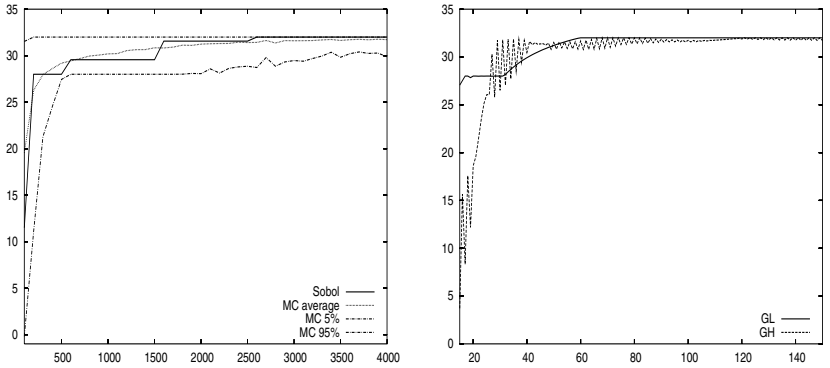
Since C_0 and C_1 are closed, f is lsc. It is also clear that condition 1 of Corollary 10 holds. To verify condition 2, note first that for each $(V, \theta) \in$ dom $E^{P^0} f$, $f(V, \theta, \cdot)$ is the constant function V on the set

$$C_1(\theta) = \{(S, F) \mid S \cdot \theta \geq F\}$$

which is of full measure. Since $C_1(\theta)$ is closed, we must have supp $P^0 \subset C_1(\theta)$ for every $(V, \theta) \in$ dom $E^{P^0} f$. Thus, condition 2 holds if we let $\Xi =$ supp P^0, which is legitimate since supp $P^\nu \subset$ supp P^0. □

In our test, the set of assets consists of cash, SP500 index and 28 European call and put options on the index with maturity of 17 calendar days. In this case, the value of S is fully determined by the value of the index at the maturity which is assumed to be log-normally distributed. The cash-flow F is taken to be that of a call option with the same maturity but different strike than any other call included in S.

Figure 7(a) displays the objective values obtained with Sobol along with the averages and 90% confidence intervals obtained with Monte Carlo from 250 samples for each value of $\nu = 100, 110, \ldots, 4000$.



(a) MC and Sobol. (b) Gaussian quadratures.

Fig. 7. Optimal values of (PP^{ν}) as a function of ν

Our random variable being one-dimensional in the current problem suggests using Gaussian quadratures. The use of Gauss-Hermite quadrature for discretizing the normal distribution has been already described in Subsection 2.1. We will also utilize Gauss-Legendre quadrature that gives convergent discretizations of the one-dimensional uniform distribution on $(0, 1)$. From this we obtain discretizations of the normal distribution by mapping each point through the inverse of the normal distribution function. The results are shown in Figure 7(b). With $\nu = 60$, the optimal values obtained with Gauss-Legendre quadrature have converged to the same value as the optimal values obtained with Sobol after 2500 points. Gauss-Hermite is almost as good but it results in slight oscillations.

References

1. Artstein, Z., Wets, R.J.-B.: Stability results for stochastic programs and sensors, allowing for discontinuous objective functions. SIAM J. Optim. **4**(3), 537–550 (1994)
2. Artstein, Z., Wets, R.J.-B.: Consistency of minimizers and the SLLN for stochastic programs. J. Convex Anal. **2**(1–2), 1–17 (1995)
3. Attouch, H.: Variational convergence for functions and operators. Pitman (Advanced Publishing Program), Boston, MA, 1984
4. Attouch, H., Wets, R.J.-B.: Quantitative stability of variational systems. I. The epigraphical distance. Trans. Amer. Math. Soc. **328**(2), 695–729 (1991)
5. Beer, G.: A geometric algorithm for approximating semicontinuous function. J. Approx. Theory **49**(1), 31–40 (1987)
6. Bertsekas, D.P., Shreve, S.E.: Stochastic optimal control, volume 139 of Mathematics in Science and Engineering. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1978. The discrete time case
7. Billingsley, P.: Convergence of probability measures. John Wiley & Sons Inc., New York, 1968

8. Billingsley, P.: Convergence of probability measures. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication
9. Birge, J.R., Wets, R.J.-B.: Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse. Math. Programming Stud., Stochastic programming 84. I. (27), 54–102 (1986)
10. Bratley, P., Fox, B.L., Niederreiter, H.: Implementation and tests of low-discrepancy sequences. ACM Trans. Model. Comput. Simul. **2**(3), 195–213 (1992)
11. Chen, X., Womersley, R.S.: A parallel inexact Newton method for stochastic programs with recourse. Ann. Oper. Res. **64**, 113–141 (1996); Stochastic programming, algorithms and models (Lillehammer, 1994)
12. Deák, I.: Multidimensional integration and stochastic programming. In Numerical techniques for stochastic optimization, volume 10 of Springer Ser. Comput. Math. Springer, Berlin, 1988, pp. 187–200
13. Dempster, M.A.H., Thompson, R.T.: EVPI-based importance sampling solution procedures for multistage stochastic linear programmes on parallel MIMD architectures. Ann. Oper. Res. **90**, 161–184 (1999)
14. Dupačová, J., Wets, R.J.-B.: Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. Ann. Statist. **16**(4), 1517–1549 (1988)
15. Faure, H.: Discrépance de suites associées à un système de numération (en dimension s). Acta Arith. **41**(4), 337–351 (1982)
16. Fox, B.L.: Algorithm 647: Implementation and relative efficiency of quasirandom sequence generators. ACM Transactions on Mathematical Software **12**(4), 362–376 (1986)
17. Frauendorfer, K.: Stochastic two-stage programming, volume 392 of Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, Berlin, 1992 Habilitationsschrift, University of Zürich, Zürich, 1992
18. Halton, J.H.: On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numer. Math. **2**, 84–90 (1960)
19. Hammersley, J.M.: Monte Carlo methods for solving multivariable problems. Ann. New York Acad. Sci. **86**, 844–874 (1960) 1960
20. Høyland, K., Kaut, M., Wallace, S.W.: Heuristic for moment-matching scenario generation. Comput. Optim. Appl. **24**(2–3), 169–185 (2003)
21. Høyland, K., Wallace, S.W.: Generating scenario trees for multistage decision problems. Managament Science **47**(2), 295–307 (2001)
22. Infanger, G.: Monte Carlo (importance) sampling within a Benders decomposition algorithm for stochastic linear programs. Ann. Oper. Res. **39**(1–4), 69–95 (1993, 1992)
23. Kall, P., Ruszczyński, A., Frauendorfer, K.: Approximation techniques in stochastic programming. In Numerical techniques for stochastic optimization, volume 10 of Springer Ser. Comput. Math. Springer, Berlin, 1988, pp. 33–64
24. King, A.J.: Duality and martingales: a stochastic programming perspective on contingent claims. Math. Program. **91**(3, Ser. B), 543–562 (2002); ISMP 2000, Part 1 (Atlanta, GA)
25. Korf, L.A.: Stochastic programming duality: $\mathcal{L}^\infty$ multipliers for unbounded constraints with an application to mathematical finance. Math. Program. **99**(2, Ser. A), 241–259 (2004)
26. Korobov, N.M.: Approximate evaluation of repeated integrals. Dokl. Akad. Nauk SSSR **124**, 1207–1210 (1959)
27. L'Ecuyer, P., Lemieux, C.: Variance reduction via lattice rules. Management Science **46**(2), 1214–1235 (2000)

28. Lepp, R.: Approximations to stochastic programs with complete recourse. SIAM J. Control Optim. **28**(2), 382–394 (1990)
29. Lucchetti, R., Salinetti, G., Wets, R.J.-B.: Uniform convergence of probability measures: topological criteria. J. Multivariate Anal. **51**(2), 252–264 (1994)
30. Lucchetti, R., Wets, R.J.-B.: Convergence of minima of integral functionals, with applications to optimal control and stochastic optimization. Statist. Decisions **11**(1), 69–84 (1993)
31. Niederreiter, H.: Random number generation and quasi-Monte Carlo methods, volume 63 of CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992
32. Pennanen, T., Koivu, M.: Integration quadratures in discretization of stochastic programs. Stochastic Programming E-Print Series, 2002
33. Pflug, G.: Scenario tree generation for multiperiod financial optimization by optimal discretization. Math. Program., **89**(2, Ser. B), 251–271 (2001); Mathematical programming and finance
34. Pflug, G., Hochreiter, R.: Scenario generation for stochastic multi-stage decision processes as facility location problems. Technical report, Department of Statistics and Decision Support Systems, University of Vienna, 2003
35. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical recipes in C. Cambridge University Press, Cambridge, second edition, 1992 The art of scientific computing
36. Svetlozar T.: Rachev and Werner Römisch. Quantitative stability in stochastic programming: the method of probability metrics. Math. Oper. Res. **27**(4), 792–818 (2002)
37. Robinson, S.M., Wets, R.J.-B.: Stability in two-stage stochastic programming. SIAM J. Control Optim. **25**(6), 1409–1416 (1987)
38. Rockafellar, R.T., Wets, R.J.-B.: Variational analysis, volume 317 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1998
39. Römisch, W., Schultz, R.: Stability of solutions for stochastic programs with complete recourse. Math. Oper. Res. **18**(3), 590–609 (1993)
40. Schultz, R.: Some aspects of stability in stochastic programming. Ann. Oper Res. **100**, 55–84 (2001), 2000. Research in stochastic programming (Vancouver, BC, 1998)
41. Shapiro, A.: Stochastic programming by monte carlo simulation methods. Stochastic Programming E-Print Series, 2000
42. Shapiro, A.: Quantitative stability in stochastic programming. Math. Programming **67**(1, Ser. A), 99–108 (1994)
43. Sloan, I.H., Joe, S.: Lattice methods for multiple integration. Oxford Science Publications. The Clarendon Press Oxford University Press, New York, 1994
44. Sobol', I.M.: The distribution of points in a cube and the approximate evaluation of integrals. U.S.S.R. Comput. Math. And Math. Phys. (4), 86–112 (1967)
45. Vogel, S., Lachout, P.: On continuous convergence and epi-convergence of random functions. II. Sufficient conditions and applications. Kybernetika (Prague) **39**(1), 99–118 (2003)
46. Zervos, M.: On the epiconvergence of stochastic optimization problems. Math. Oper. Res. **24**(2), 495–508 (1999)