Math. Program., Ser. B (2009) 116:461–479
DOI 10.1007/s10107-007-0113-9

FULL LENGTH PAPER

Epi-convergent discretizations of multistage stochastic programs via integration quadratures

Teemu Pennanen

Received: 8 March 2005 / Accepted: 6 January 2006 / Published online: 28 April 2007

Abstract This paper presents procedures for constructing numerically solvable discretizations of multistage stochastic programs that epi-converge to the original problem as the discretizations are made finer. Epi-convergence implies, in particular, that the cluster points of the first-stage solutions of the discretized problems are optimal first-stage solutions of the original problem. The discretization procedures apply to a general class of nonlinear stochastic programs where the uncertain factors are driven by time series models. Using existing routines for numerical integration allows for an easy and efficient implementation of the procedures.

Keywords Stochastic programming · Discretization · Epi-convergence · Quadrature

Mathematics Subject Classification (2000) 90C15 · 49M25 · 90C25

1 Introduction

Because of their generality, multistage stochastic programs have become popular models for dynamic decision making under uncertainty. A number of applications in economics, finance, production planning, engineering etc. can be found in the collections of Marti and Kall [30,31], Mulvey and Ziemba [53], Ruszczynski and Shapiro [47] and Marti, Ermoliev and Pflug [29].

This paper is concerned with numerical solution of the following general stochastic programming model from Rockafellar and Wets [42,43]. At each *stage* $k = 0, \dots, K$,

This work was supported by Finnish Academy under contract no. 3385.

T. Pennanen (✉)
Department of Business Technology, Helsinki School of Economics,
PL 1210, 00101 Helsinki, Finland
e-mail: pennanen@hse.fi

the decision maker observes the value of a random variable ξ_k, and makes a decision x_k depending on the observed values of $\xi_0, \ldots, \xi_k$. We will assume that each ξ_k takes values in a Borel subset Ξ_k of $\mathbb{R}^{d_k}$ and x_k is $\mathbb{R}^{n_k}$-valued. We will also assume that Ξ_0 is a singleton, so that ξ_0 and thus x_0 will be deterministic. The vector $\xi = (\xi_0, \ldots, \xi_K)$ will be modeled as a random variable in the probability space $(\Xi, \mathcal{F}, P)$, where $\Xi = \Xi_0 \times \cdots \times \Xi_K$, $\mathcal{F}$ is the Borel σ-field on Ξ and P is a probability measure on $(\Xi, \mathcal{F})$. The *multistage stochastic program* is the optimization problem

$$\underset{x \in \mathcal{N}(P)}{minimize} \quad E^P f(x(\xi), \xi), \tag{$SP(P)$}$$

where E^P denotes the expectation operator[1], f is a *convex normal integrand* on $\mathbb{R}^n \times \Xi$ and $\mathcal{N}(P)$ is the subspace of *nonanticipative* elements of $L^\infty(\Xi, \mathcal{F}, P; \mathbb{R}^n)$ where $n = n_0 + \cdots + n_K$. Recall that a function $f : \mathbb{R}^n \times \Xi \to (-\infty, +\infty]$ is a convex normal integrand if the set-valued mapping $\xi \mapsto \operatorname{epi} f(\cdot, \xi)$ is closed convex-valued and measurable; see Rockafellar and Wets [45, Chap. 14]. The set of nonanticipative elements is defined by

$$\mathcal{N}(P) = \{x \in L^\infty(\Xi, \mathcal{F}, P; \mathbb{R}^n) \mid x \text{ contains an } (\mathcal{F}_k)_{k=0}^K\text{-adapted function}\},$$

where $(\mathcal{F}_k)_{k=0}^K$ is the filtration of σ-fields

$$\mathcal{F}_k := \{B_k \times \Xi_{k+1} \times \cdots \times \Xi_K \mid B_k \in \mathcal{B}_k\},$$

where $\mathcal{B}_k$ is the Borel σ-field on $\Xi_0 \times \cdots \times \Xi_k$. Recall that a function $\tilde{x} = (\tilde{x}_0, \ldots, \tilde{x}_K)$ is said to be *adapted* to $(\mathcal{F}_k)_{k=0}^K$ if for each k, $\tilde{x}_k$ is $\mathcal{F}_k$-measurable. An $\mathcal{F}$-measurable function on Ξ is $\mathcal{F}_k$ measurable iff it depends only on $(\xi_0, \ldots, \xi_k)$.

When the stochastic process ξ is a random variable with an infinite sample space (as in most econometric models), ($SP(P)$) is an infinite-dimensional optimization problem whose solution requires discretization. Discretizations are usually obtained by replacing the original measure P by a finitely supported measure (a scenario tree) of the form

$$P^\nu = \sum_{i \in I(\nu)} p^{\nu,i} \delta_{\xi^{\nu,i}},$$

where $I(\nu)$ is a finite index set, $\delta_{\xi^{\nu,i}}$ is the unit mass at a point $\xi^{\nu,i} \in \Xi$, and $p^{\nu,i} > 0$. Then $L^\infty(\Xi, \mathcal{F}, P^\nu; \mathbb{R}^n) \cong (\mathbb{R}^n)^{I(\nu)}$, and ($SP(P^\nu)$) can be written in the finite-dimensional form

$$\underset{x \in \mathcal{N}(P^\nu)}{minimize} \quad \sum_{i \in I(\nu)} p^{\nu,i} f(x(\xi^{\nu,i}), \xi^{\nu,i}), \tag{$SP(P^\nu)$}$$

where

$$\mathcal{N}(P^\nu) = \{x \in L^\infty(\Xi, \mathcal{F}, P^\nu; \mathbb{R}^n) \mid x \text{ contains an } (\mathcal{F}_k)_{k=0}^K\text{-adapted function}\}$$

[1] when both the positive and negative parts integrate to ∞ the convention $\infty - \infty = \infty$ is to be used

$$= \{x \in L^\infty(\Xi, \mathcal{F}, P^\nu; \mathbb{R}^n) \,|\, x_k(\xi^{\nu,i}) = x_k(\xi^{\nu,j}) \text{ if } \pi_k \xi^{\nu,i} = \pi_k \xi^{\nu,j}\}$$

and π_k denotes the projection $\xi \mapsto (\xi_0, \ldots, \xi_k)$. This is a mathematical program which can in principle be solved numerically by standard solvers or special purpose algorithms designed to take advantage of problem structure. The literature on constructing discrete measures P^ν for purposes of stochastic programming is vast; see Dupačová et al. [14] for a review up to 2001 and Pennanen and Koivu [37], Dupačová et al. [15], Heitsch and Römisch [20], Casey and Sen [8] or Pflug and Hochreiter [40] for more recent developments.

Considering the large number of stochastic programming applications in practice and the wide variety of discretization approaches, surprisingly little attention has been given to the consistency properties of discretizations in the case of multistage problems with general probability distributions. Olsen [35] seems to have been the first to study this question. He gave rather general conditions for a discretization scheme to produce consistent optimal values for linear multistage stochastic programs. Lepp [27] studied nonlinear two-stage stochastic programs with relatively complete recourse. Frauendorfer [18] studied the so called barycentric approximation scheme for stochastic programs satisfying certain convexity properties with respect to random variables. Shapiro [48,49] has obtained various statistical results for random discretizations based on conditional sampling. Recently, Heitsch et al. [21] have obtained stability results for multistage stochastic programs that may be applicable to discretizations. So far, the most general class of problems has been treated in Pennanen [36] where discretizations were analyzed using *epi-convergence* (Γ-convergence), which is a general technique for studying approximations of optimization problems; see Attouch [1], Dal Maso [12], Rockafellar and Wets [45] or Braides [7] for general treatment of the subject as well as Polak [41] and Braides [7, Chap. 4] for its applications to discretizations.

The purpose of this paper is to present a class of discretization procedures that fits the framework of [36] and thus yields consistent finite-dimensional approximations of multistage stochastic programs. The procedures apply to a rather general class of multistage stochastic programs and they can be implemented quite easily using available routines for numerical integration. To our knowledge, these are the first existing discretization procedures which have been shown to yield epi-convergent discretizations of the original infinite-dimensional problem.

It should be noted that this paper is only concerned with *asymptotic* convergence properties of discretizations. Nothing is said about how should P^ν be chosen in order to guarantee that the distance of the optimal value and the solutions are within a given distance from those of the original problem. Such *quantitative* results typically require stronger assumptions that may be hard to verify in practice. Instead of taking the decision stages as fixed in ($SP(P^\nu)$), one could also study convergence properties with respect to time-discretization along the lines of Mordukhovich [32]. For this it might be more convenient to adopt a continuous-time framework such as the one in Back and Pliska [2].

The rest of this paper is organized as follows. In the next section, we recall the general assumptions from [36] that guarantee the convergence of ($SP(P^\nu)$) to ($SP(P)$) as the discretizations are made finer. Section 3, defines the class of stochastic processes that will be treated in this paper. Examples are given to illustrate the significance of this class. Section 4 describes the discretization procedures together with conditions under which they produce measures that satisfy the conditions for epi-convergence. Applications are given in Sect. 5, and the proof of the main result in Sect. 6.

2 Epi-convergent discretizations

In order to study the discretizations ($SP(P^\nu)$) via epi-convergence, it is necessary to first embed them in the original space $L^\infty(\mathcal{F}, P)$. Here and in what follows, we will use the notation

$$L^\infty(\mathcal{F}', P') := L^\infty(\Xi, \mathcal{F}', P'; \mathbb{R}^n)$$

for any sub σ-field $\mathcal{F}' \subset \mathcal{F}$ and a probability measure P' on $(\Xi, \mathcal{F})$. The original problem ($SP(P)$) is thus a minimization problem over $L^\infty(\mathcal{F}, P)$, whereas ($SP(P^\nu)$) is a minimization problem over $L^\infty(\mathcal{F}, P^\nu)$, where

$$P^\nu = \sum_{i \in I(\nu)} p^{\nu,i} \delta_{\xi^{\nu,i}}.$$

As in [36, Sect. 2], we will assume that for each P^ν there is a partition $\{\Xi^{\nu,i}\}_{i \in I(\nu)}$ of Ξ such that $P(\Xi^{\nu,i}) > 0$ and we let $\mathcal{F}^\nu$ be the σ-field generated by $\{\Xi^{\nu,i}\}_{i \in I(\nu)}$. Then $L^\infty(\mathcal{F}, P^\nu)$ is isometric to $L^\infty(\mathcal{F}^\nu, P)$ which is a subspace of $L^\infty(\mathcal{F}, P)$. Indeed, the mapping $\Pi^\nu : L^\infty(\mathcal{F}, P^\nu) \to L^\infty(\mathcal{F}^\nu, P)$ given by

$$\Pi^\nu x = \sum_{i=1}^{\nu} x(\xi^{\nu,i}) \chi_{\Xi^{\nu,i}}$$

is a bijection and

$$\|\Pi^\nu x\|_{L^\infty(\mathcal{F},P)} = \|x\|_{L^\infty(\mathcal{F},P^\nu)} \equiv \max_{i \in I(\nu)} |x(\xi^{\nu,i})|.$$

By [36, Lemma 1], the essential objective of ($SP(P^\nu)$) can be written as

$$\tilde{F}^\nu(x) = F^\nu(\Pi^\nu x),$$

where

$$F^\nu(x) = \begin{cases} E^P f(x(\xi), s^\nu(\xi))\psi^\nu(\xi) & \text{if } x \in \mathcal{N}^\nu(P), \\ +\infty & \text{otherwise,} \end{cases}$$

where $s^\nu : \Xi \to \Xi$ and $\psi^\nu : \Xi \to \mathbb{R}$ are the piecewise constant functions, defined

$$s^\nu(\xi) = \xi^{\nu,i} \quad \text{and} \quad \psi^\nu(\xi) = \frac{p^{\nu,i}}{P(\Xi^{\nu,i})} \quad \text{if } \xi \in \Xi^{\nu,i},$$

and

$$\mathcal{N}^\nu(P) = \{x \in L^\infty(\mathcal{F}, P) \,|\, x \text{ contains an } ((s^\nu)^{-1}(\mathcal{F}_k))_{k=0}^K\text{-adapted function}\}.$$

Thus, since $\mathcal{N}^\nu(P) \subset L^\infty(\mathcal{F}^\nu, P)$, problem ($SP(P^\nu)$) is equivalent to minimizing F^ν over $L^\infty(\mathcal{F}, P)$. The main result of [36] shows that, under Assumptions 1 and 2 below, the functions F^ν *epi-converge* to the essential objective of ($SP(P)$) as the discretizations are made finer. The general properties of epi-convergence then yield, in particular, the following.

Theorem 1 ([36, Corollary 3.1]) *If assumptions 1 and 2 hold and problems* ($SP(P^\nu)$) *have ϵ^ν-optimal solutions x^ν such that $\|x^\nu\|_{L^\infty(\mathcal{F},P^\nu)}$ remains bounded and $\epsilon^\nu \searrow 0$, then the optimal values of* ($SP(P^\nu)$) *converge to that of* ($SP(P)$)*, and all cluster points of $(x_0^\nu)_{\nu=1}^\infty$ are optimal first-stage solutions of* ($SP(P)$).

The first assumption concerns the discretized measures P^ν.

Assumption 1 *The sequence $(P^\nu)_{\nu=1}^\infty$ of measures is such that there exists a sequence of partitions $\{\Xi^{\nu,i}\}_{i\in I(\nu)}$ of Ξ such that*

$$\mathcal{N}^\nu(P) \subset \mathcal{N}(P), \tag{A1}$$

$$s^\nu \xrightarrow{P} I, \tag{A2}$$

$$\psi^\nu \xrightarrow{L^\infty} 1. \tag{A3}$$

The main goal of this paper is to present procedures for generating sequences of discrete measures that satisfy Assumption 1. We emphasize that it is not necessary to construct the partitions $\{\Xi^{\nu,i}\}_{i\in I(\nu)}$ explicitly. Indeed, they are not involved in the discretizations ($SP(P^\nu)$) or in Theorem 1 which only concerns the optimal values and first-stage solutions.

In stating Assumption 2, we will use terminology from Ioffe [24]. The function f has the *lower compactness property* if $f_-(x^\nu(\cdot), s^\nu(\cdot))$ is weakly precompact in L^1 whenever (x^ν) converges in $L^\infty(\mathcal{F}, P)$, (s^ν) converges in measure P and $\sup_\nu E^P f(x^\nu(\xi), s^\nu(\xi)) < \infty$. Here, $f_-(x, s) := \min\{f(x, s), 0\}$. According to [24, Remark 2], f has the lower compactness property, in particular, if there exists a nondecreasing real-valued function g on $[0, +\infty)$ and a real number b such that

$$f(x, \xi) \geq -g(|x|) - b \quad \forall x \in \mathbb{R}^n,\ \xi \in \Xi.$$

Assumption 2

1. $E^P|\xi| < \infty$.
2. *The function f is lsc and has the* lower compactness property of Ioffe [24].

3. For every feasible x, there is a uniformly bounded sequence $y^\mu \to x$ of nonanticipative, P-a.s. continuous functions such that

$$\limsup_{\nu\to\infty} E^{P^\nu} f(y^\mu(\xi),\xi) \le E^P f(y^\mu(\xi),\xi) \quad \forall \mu = 1,2,\dots,$$
$$\limsup_{\mu\to\infty} E^P f((y^\mu(\xi),\xi) \le E^P f(x(\xi),\xi).$$

Whereas Assumption 1 concerns the discretized measures P^ν, Assumption 2 can be modified so that it only concerns the original problem ($SP(P)$). Indeed, Assumption 2.3 can be replaced by the requirement that for every feasible x, there is a uniformly bounded sequence $y^\mu \to x$ of nonanticipative, P-a.s. continuous functions such that the function $\xi \mapsto f(y^\mu(\xi),\xi)$ is bounded and P-a.s. continuous and that

$$\limsup_{\mu\to\infty} E^P f((y^\mu(\xi),\xi) \le E^P f(x(\xi),\xi).$$

That this is sufficient follows from the fact that, under (A2) and (A3), $E^{P^\nu}\varphi \to E^P\varphi$ for every bounded and P-a.s. continuous function φ; see [36, Lemma 4.1].

3 Time series models with uniform innovations

The discretization technique to be presented in Sect. 4 applies to the class of stochastic programs where the stochastic processes ξ is driven by a time series model of the form

$$\xi_k = g_k(\xi_0,\dots,\xi_{k-1},\omega_k) \quad \text{for } k = 1,\dots,K, \tag{1}$$

where ξ_0 is given, $\omega_1,\dots,\omega_K$ are mutually independent random variables, with ω_k uniformly distributed in the d_k-dimensional unit cube $\Omega_k = (0,1)^{d_k}$, and $g_k : \Xi_0 \times \cdots \times \Xi_{k-1} \times \Omega_k \to \Xi_k$.

The following (multivariate generalization of) model from Shiryaev [50] is convenient especially in modeling financial time series.

Example 1 (Conditionally Gaussian processes) Consider a d-dimensional process $(\xi_t)_{t=0}^\infty$ that satisfies

$$\xi_t = \mu_t(\xi_1,\dots,\xi_{t-1}) + \sigma_t(\xi_1,\dots,\xi_{t-1})\varepsilon_t, \tag{2}$$

where $\mu_t : \mathbb{R}^{(t-1)d} \to \mathbb{R}^d$ and $\sigma_t : \mathbb{R}^{(t-1)d} \to \mathbb{R}^{d\times d}$ are given functions and ε_t are independent d-dimensional Gaussian (normally distributed) random variables with zero mean and unit variance. It follows that the conditional distribution of ξ_t, given $\xi_1,\dots,\xi_{t-1}$, is Gaussian with mean $\mu_t(\xi_1,\dots,\xi_{t-1})$ and variance $\sigma_t(\xi_1,\dots,\xi_{t-1})$ $\sigma_t(\xi_1,\dots,\xi_{t-1})^T$.

This model can be written as (1) by using the expression

$$\varepsilon_t = (\Phi^{-1}(\omega_t^1),\dots,\Phi^{-1}(\omega_t^d)),$$

where ω_t is uniformly distributed in $(0,1)^d$ and Φ is the univariate Gaussian distribution function.

In econometric models like (2), the vectors ξ_t, $t = 0, 1, \ldots$, often represent the values of a stochastic process at uniformly spaced points in time. In many stochastic programming models in practice, on the other hand, the time periods between different stages $k = 0, \ldots, K$ vary. If stage k corresponds to $t_k \neq k$, in the time units of (2), the variables in (1) would be

$$\xi_k := (\xi_{t_{k-1}+1}, \ldots, \xi_{t_k}) \quad \text{and} \quad \omega_k := (\omega_{t_{k-1}+1}, \ldots, \omega_{t_k})$$

with $\Xi_k := \mathbb{R}^{(t_k - t_{k-1})d}$ and the functions g_k would be given recursively by (2).

Conditionally Gaussian processes cover a wide variety of econometric time series models, both linear and nonlinear. When ξ is the *first difference* of another series s, i.e. $\xi_t = \Delta s_t := s_t - s_{t-1}$, (2) can be viewed as a discrete time version of a general Itô process with *drift* μ_t and *volatility* σ_t. In particular, if μ_t and σ_t are constant, one obtains a Brownian motion model for s. In economic and financial models, s is often the logarithm of a price or an index, so that ξ is the so called log return, and instead of Brownian motion, one gets a geometric Brownian motion.

When σ_t are constant but

$$\mu_t(\xi_1, \ldots, \xi_{t-1}) = c + \sum_{i=1}^{l} A_i \Delta s_{t-i}$$

for some fixed parameters $c \in \mathbb{R}^d$ and $A_i \in \mathbb{R}^{d\times d}$, one obtains the Vector Auto Regressive (VAR) model for Δs. VAR models have been used in stochastic programming in Boender et al. [6] and Kouwenberg [25]. When σ_t are invertible matrices, we can express each innovation ε_t in terms of $\xi_1, \ldots, \xi_t$. Thus (2) also covers the case where σ_t are constant invertible matrices and

$$\mu_t(\xi_1, \ldots, \xi_{t-1}) = c + \sum_{i=1}^{l} A_i \Delta s_{t-i} + \sum_{i=1}^{l} B_i \varepsilon_{t-i},$$

where $B_i \in \mathbb{R}^{d\times d}$ are fixed. Such models are called Vector Autoregressive Moving Average (VARMA) models. When, instead,

$$\mu_t(\xi_1, \ldots, \xi_{t-1}) = c + \sum_{i=1}^{l} A_i \Delta s_{t-i} + \alpha\beta^T s_{t-1}, \tag{3}$$

where $\alpha \in \mathbb{R}^{d\times d'}$ and $\beta \in \mathbb{R}^{d\times d'}$ are fixed, one obtains a Vector Equilibrium Correction (VEqC) model. VEqC models have become popular in modeling nonstationary multivariate economic time series; see for example Engle and Kranger [16] or Clements and Hendry [11]. The second term in (3) can be used to model equilibrium conditions for s. For example, if s is the logarithm of an interest rate, $A = 0$,

$\beta = 1$ and $\alpha < 0$, one obtains the Black-Karasinski mean reversion model with mean reversion level $-c/\alpha$ [5]. VEqC models have been used in stochastic programming in [6] and Hilli et al. [22].

In VAR, VEqC and VARMA-models, the drift is a linear function of ξ and the volatility σ_t is constant. These models are said to be *linear* since there ξ can be written as a linear function of the Gaussian innovations ε_t, which implies that ξ is also Gaussian. In conditionally Gaussian models where the volatility terms σ_t depend on ξ, this is no longer true, and one speaks of *nonlinear* time series models. In such models, the *conditional* distribution of ξ_t, given $\xi_0, \dots, \xi_{t-1}$, is still Gaussian, but the unconditional distribution need not be. In the simplest multivariate Auto Regressive Conditionally Heteroscedastic (ARCH) models, the variance matrices

$$\sigma_t^2 := \sigma_t(\xi_1, \dots, \xi_{t-1})\sigma_t(\xi_1, \dots, \xi_{t-1})^T$$

are given by

$$\sigma_t^2 = \sigma_0^2 + \sum_{i=1}^{l} \alpha_i (\xi_{t-i} - \mu_{t-i})(\xi_{t-i} - \mu_{t-i})^T,$$

where $\sigma_0^2 \in \mathbb{R}^{d\times d}$ and $\alpha_i \in \mathbb{R}$ are constants, whereas in GARCH-models,

$$\sigma_t^2 = \sigma_0^2 + \sum_{i=1}^{l} \alpha_i (\xi_{t-i} - \mu_{t-i})(\xi_{t-i} - \mu_{t-i})^T + \sum_{i=1}^{l} \beta_i \sigma_{t-i}^2,$$

where $\beta_i \in \mathbb{R}$. In both models, the volatility is set equal to a square matrix $\sigma_t(\xi_1, \dots, \xi_{t-1})$ that satisfies $\sigma_t(\xi_1, \dots, \xi_{t-1})\sigma_t(\xi_1, \dots, \xi_{t-1})^T = \sigma_t^2$, e.g. the Cholesky factor of σ_t^2. GARCH processes have been used in stochastic programming by Gondzio et al. [19] and Dempster et al. [13].

4 Discretization procedures

When $\xi = (\xi_0, \dots, \xi_K)$ follows the time series model (1), it is uniquely determined by $\omega = (\omega_1, \dots, \omega_K)$. Denote this mapping by G. The vector ω follows the uniform distribution $U = U_1 \times \cdots \times U_K$, where U_k is the uniform distribution on the measurable space (Ω_k, Σ_k), where $\Omega_k = (0, 1)^{d_k}$ and Σ_k is the Borel field on Ω_k. When G is Borel-measurable, we have

$$E^P \varphi(\xi) = E^U \varphi(G(\xi)),$$

for any measurable function φ on Ξ. This means that

$$P = UG^{-1}, \tag{4}$$

i.e. $P(A) = U(G^{-1}(A))$ for every $A \in \mathcal{F}$. Expression (4) suggests the following

Discretization procedure

1. Approximate each U_k, $k = 1, \ldots, K$, independently of each other by a discrete measure U_k^ν;
2. Let $U^\nu = U_1^\nu \times \cdots \times U_K^\nu$ and

$$P^\nu = U^\nu G^{-1}. \tag{5}$$

More concretely, if for $k = 1, \ldots, K$

$$U_k^\nu = \sum_{i \in I_k(\nu)} p_k^{\nu,i} \delta_{\omega_k^{\nu,i}},$$

where $I_k(\nu)$ is a finite index set, then

$$U^\nu = \sum_{i \in I(\nu)} p^{\nu,i} \delta_{\omega^{\nu,i}},$$

where

$$\begin{aligned} I(\nu) &= \{(i_1, \ldots, i_K) \mid i_k \in I_k(\nu)\}, \\ p^{\nu,i} &= p_1^{\nu,i_1} \cdots p_K^{\nu,i_K}, \\ \omega^{\nu,i} &= (\omega_1^{\nu,i_1}, \ldots, \omega_K^{\nu,i_K}), \end{aligned}$$

and (5) becomes

$$P^\nu = \sum_{i \in I(\nu)} p^{\nu,i} \delta_{\xi^{\nu,i}},$$

where

$$\xi^{\nu,i} = G(\omega_1^{\nu,i_1}, \ldots, \omega_K^{\nu,i_K}).$$

The measure U^ν can be viewed as a scenario tree with branching structure $(|I_1(\nu)|, \ldots, |I_K(\nu)|)$. Since the first k components of the mapping G do not depend on $\omega_{k+1}, \ldots, \omega_K$, the measure P^ν has a similar tree-structure.

Theorem 2 below, the main result of this paper, shows that if, for each $k = 1, \ldots, K$, one has a sequence of discrete measures U_k^ν that converge weakly to U_k, then under mild conditions on g, the sequence of measures P^ν obtained from the above procedure satisfies Assumption 1. Recall that a sequence of probability measures Q^ν is said to *converge weakly* to a probability measure Q, denoted $Q^\nu \to Q$, if

$$E^{Q^\nu} \varphi \to E^Q \varphi$$

for every bounded and continuous function φ. The marginal distribution of $(\xi_0, \ldots, \xi_k)$ will be denoted by P_k. Since $\Xi_0 = \{\xi_0\}$, P_0 is given by $P_0(\{\xi_0\}) = 1$.

Theorem 2 *Let P and P^ν be given by (4) and (5), respectively. Assume that, for $k = 1, \dots, K$,*

1. $g_k(\xi_0, \dots, \xi_{k-1}, \cdot)$ *is a bijection for every* $\xi \in \Xi$,
2. g_k *and the function* $(\xi_1, \dots, \xi_k) \mapsto g_k(\xi_0, \dots, \xi_{k-1}, \cdot)^{-1}(\xi_k)$ *are Borel-measurable,*
3. g_k *is* $P_{k-1} \times U_k$*-a.s. continuous,*
4. $U_k^\nu \to U_k$.

Then the sequence $(P^\nu)_{\nu=1}^\infty$ *satisfies Assumption 1.*

Proof See Sect. 6.

Conditions 1 and 2 essentially mean that the process ω contains the same information as ξ. In Example 2, conditions 1–3 take the form

1. $\sigma_t(\xi_1, \dots, \xi_{t-1})$ is nonsingular for every $\xi \in \Xi$,
2. μ_t and σ_t are measurable,
3. μ_t and σ_t are P_k-a.s. continuous.

Theorem 2 thus covers VARMA, VEqC and GARCH-models, in particular.

Methods for constructing measures U_k^ν that converge weakly to U_k are abundant in the literature of numerical integration. The best-known method is Monte Carlo, where $\{\omega_k^{\nu,i}\}_{i \in I_k(\nu)}$ are random samples from $(0,1)^{d_k}$ and $p_k^{\nu,i} = 1/|I_k(\nu)|$. Indeed, by Glivenko–Cantelli theorem, the corresponding measures U_k^ν converge weakly to U_k with probability one as $\nu \to \infty$. When Monte Carlo is used for constructing U_k^ν, the above discretization technique is often referred to as *conditional sampling;* see e.g. Chiralaksanakul [10] or Shapiro [48]. Combined with [36, Theorem 5], the above result implies that, under Assumption 2, *conditional sampling produces epi-convergent discretizations with probability one.*

Besides Monte Carlo, there are so called *quasi-Monte Carlo* methods that, instead of randomly throwing points into the unit cube, try to approximate the uniform distribution as well as possible in the sense of the distance ("probability metric")

$$D^*(U^\nu, U) := \sup_{C \in \mathcal{C}_0} |U^\nu(C) - U(C)|,$$

where $\mathcal{C}_0$ is the set of rectangles $C \subset (0,1)^d$ with $0 \in C$; see Niederreiter [33,34]. There exists a wide variety of methods that are able to generate sequences of discrete measures U^ν that satisfy

$$D^*(U^\nu, U) \le \gamma \frac{(\ln \nu)^d}{\nu} \quad \forall \nu = 1, 2, \dots,$$

where d is the dimension of the space and γ is a constant independent of ν. By Lucchetti et al. [28, Corollary 11], $D^*(U^\nu, U) \to 0$ is equivalent to $U^\nu \to U$; see also Römisch [46, Sect. 2.1]. The use of quasi-Monte Carlo methods in discretization of multistage stochastic programs was first proposed in Pennanen and Koivu [37], where convergence of optimal values was studied numerically. See also Chen and Womersley [9], where lattice rules (see [26,51]) were used for computing the expectation of the optimum value of the second-stage problem.

Actually, in the numerical integration literature, the number $D^*(U^\nu, U)$ is called the *star-discrepancy* of the point set $\{\omega^{\nu,i}\}_{i\in I(\nu)}$ and usually there is no reference to probability measures or weak convergence. We use the probabilistic terminology in order to emphasize the connections between the two fields. In particular, viewing star-discrepancy as a probability metric, reveals some connections of the above quadrature-based discretization technique with the techniques in Dupačová et al. [15], Heitsch and Römisch [20] and Pflug [39]; see also Pflug and Hochreiter [40]. For static problems [39,40] propose to discretize a given measure P by minimizing the Wasserstein distance

$$D^W(P^\nu, P) := \sup_{\operatorname{lip}\varphi\le 1} |E^{P^\nu}\varphi(\xi) - E^P\varphi(\xi)|,$$

where $\operatorname{lip}\varphi$ denotes the Lipschitz constant of φ. The problem of finding discrete measures that minimize a distance from a given measure can be very hard in general. Star-discrepancy (which is called Kolmogorov–Smirnov distance in [40]) has the advantage that many efficient methods, namely quasi-Monte Carlo methods, already exist for its minimization in high-dimensional spaces.

5 Applications

5.1 Static problems

When $K = 1$ and f is independent of x_1, problem ($SP(P)$) can be written as a *static* stochastic program

$$\mathop{minimize}_{x\in\mathbb{R}^n} \quad E^P f(x,\xi).$$

Theorem 3 *Assume that $E^P|\xi| < \infty$ and $P = UG^{-1}$ for some invertible, U-a.s. continuous mapping G such that both G and G^{-1} are measurable. If $P^\nu \to P$, f is lsc and has the lower compactness property, and for every feasible x, there is a sequence $y^\mu \to x$ such that*

$$\limsup_{\nu\to\infty} E^{P^\nu} f(y^\mu,\xi) \le E^P f(y^\mu,\xi) \quad \forall\mu = 1,2,\dots,$$
$$\limsup_{\mu\to\infty} E^P f(y^\mu,\xi) \le E^P f(x,\xi),$$

then the functions $E^{P^\nu} f(\cdot,\xi)$ epi-converge to $E^P f(\cdot,\xi)$.

Proof Since G^{-1} is single-valued and measurable and since $P^\nu \to P$ [4, Theorem 2.7] implies that the measures $U^\nu = P^\nu G$ are well-defined and converge weakly to U. We thus have $P^\nu = U^\nu G^{-1}$ and conditions 1–4 of Theorem 2 are satisfied. The result now follows by combining Theorem 2 and [36, Theorem 3.3].

It can be shown that, if P has a strictly positive, continuous density on a cube $(a,b) = \prod_{i=1}^d (a_i, b_i)$ ($a_i = -\infty$ and $b_i = +\infty$ are allowed) then $P = UG^{-1}$ for a *diffeomorphism* $G : (0,1)^d \to (a,b)$; see Hlawka and Mück [23]. Such a mapping G automatically has the properties asked in the above theorem. If $f(x,\cdot)$ is bounded and P-a.s. continuous for every $x \in \operatorname{dom} F$, then by the definition of weak convergence, the last condition holds with $y^\mu = x$, and, in addition to epi-convergence, we get pointwise convergence of the objectives. The above result thus covers, in particular, the applications of [38], where discretizations of *static* stochastic programs were studied.

5.2 Problems with constraint-structure

When, in the general multistage case again, f has the form

$$f(x,\xi) = \begin{cases} f_0(x,\xi) & \text{if } x \in X \text{ and } f_j(x,\xi) \le 0 \text{ for } j = 1,\dots,m, \\ +\infty & \text{otherwise,} \end{cases} \tag{6}$$

where $X \subset \mathbb{R}^n$ is convex and f_j are convex normal integrands, problem ($SP(P)$) can be written with explicit constraints as

$$\begin{aligned} &\underset{x \in \mathcal{N}(P)}{minimize}\ E^P f_0(x(\xi),\xi) \\ &subject\ to \quad f_j(x(\xi),\xi) \le 0, \quad j = 1,\dots,m, \\ &\qquad\qquad\qquad x(\xi) \in X, \\ &\qquad\qquad\qquad P\text{-a.s.} \end{aligned}$$

and ($SP(P^\nu)$) as

$$\begin{aligned} &\underset{x \in \mathcal{N}(P^\nu)}{minimize} \sum_{i \in I(\nu)} p^{\nu,i} f_0(x(\xi^{\nu,i}),\xi^{\nu,i}) \\ &subject\ to \qquad f_j(x(\xi^{\nu,i}),\xi^{\nu,i}) \le 0, \quad j = 1,\dots,m, \\ &\qquad\qquad\qquad x(\xi^{\nu,i}) \in X, \\ &\qquad\qquad\qquad \forall i \in I(\nu). \end{aligned}$$

Theorem 4 *Assume that f has the form* (6) *and*

(a) *Ξ is compact, ξ follows* (1) *and $G : \omega \mapsto \xi$ is a diffeomorphism,*
(b) *X is compact and has a nonempty interior,*
(c) *f_j are continuous on $X \times \operatorname{supp} P$,*
(d) *there exists a bounded nonanticipative function $\tilde{x}$ and an $\varepsilon > 0$ such that*

$$\tilde{x}(\xi) \in X \quad and \quad f_j(\tilde{x}(\xi),\xi) \le -\varepsilon \quad j = 1,\dots,m, \ \forall \xi \in \Xi.$$

Then, if P^ν are given by (5) *and $U_k^\nu \to U_k$, the optimal values of* ($SP(P^\nu)$) *converge to that of* ($SP(P)$)*, and if z^ν is an ϵ^ν-optimal solution of* ($SP(P^\nu)$) *and $\epsilon^\nu \searrow 0$, then all cluster points of $(z_0^\nu)_{\nu=1}^\infty$ are optimal first-stage solutions of* ($SP(P)$).

Proof By Theorem 1, it suffices to check that Assumptions 1 and 2 hold. It follows from (a) that P has the form (4), and that conditions 1–3 of Theorem 2 are satisfied. Then, by Theorem 2, $U_k^\nu \to U_k$ implies Assumption 1.

To verify Assumption 2, it suffices, by [36, Theorem 4.2], to check that P is "laminary", which according to [44, p. 304] happens in particular when P has a strictly positive density on $\Xi = \Xi_1 \times \cdots \times \Xi_K$. Since G is a diffeomorphism, the substitution formula for Lebesgue integration (see e.g. [17, Theorem 2.47]) shows that for any positive measurable functions ψ and φ on Ξ,

$$\int_{\Xi} \psi(\xi)\varphi(\xi)d\xi = \int_{(0,1)^d} \psi(G(\omega))\varphi(G(\omega))|J_G(\omega)|d\omega,$$

where J_G is the determinant of the Jacobian of G. This shows, in particular, that

$$\varphi(\xi) = \frac{1}{|J_G(G^{-1}(\xi))|}$$

is the density of $P = UG^{-1}$.

The mapping G is a diffeomorphism for example in VARMA, VEqC, and GARCH models. By truncation, such models can be modified so that Ξ becomes compact as required in condition (a) above. The compactness enters the conditions because that was assumed in [44], which is the basis of [36, Theorem 4.2]. It would be interesting to explore whether the compactness condition in [44] could be relaxed.

6 Proof of Theorem 2

The following two lemmas are adapted from Vainikko [52]. Here, $\Omega = (0, 1)^d$, Σ is the Borel field on Ω and U is the uniform distribution on (Ω, Σ).

Lemma 1 *Let $A \in \Sigma$ and $p^i > 0$ be such that $\sum_{i\in N} p^i = U(A)$. Then there is a partition $\{\Omega^i\}_{i\in N} \subset \Sigma$ of A, such that $U(\Omega^i) = p^i$.*

Proof The closed ball with radius $r \in \mathbb{R}$ and center at the origin will be denoted by $\mathbb{B}(r)$. Since $U(\text{bdry}\,\mathbb{B}(r)) = 0$, the function $r \mapsto U(A' \cap \mathbb{B}(r))$ is continuous with range $[0, U(A')]$, for every $A' \in \Sigma$. A partition having the desired properties can thus be constructed by the following procedure.

For $i \in N\{$
    Find an r_i so that $U(A \cap \mathbb{B}(r_i)) = p^i$;
    Let $\Omega^i = A \cap \mathbb{B}(r_i)$;
    Let $A := A \setminus \Omega^i$;
$\}$

The significance of weak convergence in Theorem 2 comes from the following.

Lemma 2 *For each* $\nu = 1, 2, \ldots$, *let*

$$U^\nu = \sum_{i\in I(\nu)} p^{\nu,i}\delta_{\omega^{\nu,i}}$$

be a discrete measure on Ω. *Then* $U^\nu \to U$ *if and only if there exists a sequence of partitions* $\{\Omega^{\nu,i}\}_{i\in I(\nu)} \subset \Sigma$ *of* Ω *such that*

$$s^\nu \xrightarrow{U} I, \tag{B1}$$

$$\max_{i\in I(\nu)} \left|\frac{p^{\nu,i}}{U(\Omega^{\nu,i})} - 1\right| \to 0, \tag{B2}$$

where

$$s^\nu(\omega) = \omega^{\nu,i} \quad \textit{if } \omega \in \Omega^{\nu,i}.$$

Proof Necessity: let φ be a bounded continuous function. We have $E^{U^\nu}\varphi = E^U\psi^\nu \varphi\circ s^\nu$, where ψ^ν is the step function

$$\psi^\nu(\xi) = p^{\nu,i}/U(\Omega^{\nu,i}) \quad \text{if } \xi \in \Omega^{\nu,i}.$$

It suffices to show that $\psi^\nu \xrightarrow{L^\infty} 1$ and $\varphi\circ s^\nu \xrightarrow{L^1} \varphi$. The first property is immediate from (B2). As for the second, assume for contradiction that there exists a subsequence of $(\varphi\circ s^\nu)$ that stays at a positive L^1-distance from φ. By [3, Theorem 20.5], (B1) implies that we can find a further subsequence (s^{ν^μ}) that converges U-a.s. to I. Since φ is U-a.s. continuous, it follows that $\varphi\circ s^{\nu^\mu} \to \varphi$ U-a.s. and then, by the dominated convergence theorem, $\varphi\circ s^{\nu^\mu} \xrightarrow{L^1} \varphi$.

Sufficiency: For each $\mu = 1, 2, \ldots$, let $\mathcal{P}_\mu \subset \Sigma$ be a finite partition of Ω such that

$$\max_{A\in\mathcal{P}_\mu} \operatorname{diam} A \le \frac{1}{\mu}. \tag{7}$$

Since $\mathcal{P}_\mu$ is finite, the portmanteau theorem (see e.g. [4]) guarantees that there is a ν_μ such that

$$\max_{A\in\mathcal{P}_\mu} \left|\frac{U(A)}{U^\nu(A)} - 1\right| \le \frac{1}{\mu} \tag{8}$$

for every $\nu \ge \nu_\mu$. For each ν, let μ_ν be the largest μ for which (8) is satisfied. It follows that $\mu_\nu \nearrow \infty$ as $\nu \nearrow \infty$.

For each ν and $A \in \mathcal{P}_{\mu_\nu}$, let $I_A^\nu = \{i \mid \omega^{\nu,i} \in A\}$, which is nonempty by (8). Then by Lemma 1, we can find a partition $(\Omega^{\nu,i})_{i\in I_A^\nu} \subset \Sigma$ of A such that

$$U(\Omega^{\nu,i}) = \frac{p^{\nu,i}}{U^\nu(A)}U(A) \quad \forall i \in I_A^\nu.$$

Combining this with (8), we get

$$\left|\frac{p^{\nu,i}}{U(\Omega^{\nu,i})}-1\right| = \left|\frac{U^{\nu}(A)}{U(A)}-1\right| \leq \frac{1}{\mu_\nu},$$

and by (7),

$$|\omega^{\nu,i}-\omega| \leq \frac{1}{\mu_\nu} \quad \forall \omega \in \Omega^{\nu,i}$$

for all $i \in I_A^\nu$. Combining $\{\Omega^{\nu,i}\}_{i\in I_A^\nu}$ over $A \in \mathcal{P}_{\mu_\nu}$, gives partitions of Ω that satisfy (B1) and (B2).

For each $k = 1,\dots,K$, we define the σ-fields

$$\mathcal{G}_k = \{R_k \times \Omega_{k+1} \times \cdots \times \Omega_K \mid R_k \in \mathcal{R}_k\},$$

where $\mathcal{R}_k$ denotes the Borel σ-field on $\Omega_1 \times \cdots \times \Omega_k$.

Lemma 3 *Under assumptions of Theorem 2, the mapping G is invertible, U-a.s. continuous and for every $k = 1,\dots,K$, G is $(\mathcal{G}_k,\mathcal{F}_k)$-measurable and G^{-1} is $(\mathcal{F}_k,\mathcal{G}_k)$-measurable.*

Proof By condition 1, G has an inverse given by

$$\omega_k = g_k(\xi_0,\dots,\xi_{k-1},\cdot)^{-1}(\xi_k) \quad \forall k = 1,\dots,K. \tag{9}$$

For $k = 1,\dots,K$, let G_k be the mapping that sends $(\omega_1,\dots,\omega_k)$ to $(\xi_1,\dots,\xi_k)$, so that $G = G_K$. We prove the U-a.s. continuity of G by induction on k. Let D_k be the set of discontinuities of G_k. Since $U_1(D_1) = 0$ by condition 3, it suffices to show that $(U_1\times\cdots\times U_{k-1})(D_{k-1}) = 0$ implies $(U_1\times\cdots\times U_k)(D_k) = 0$. From the expression

$$G_k(\omega_1,\dots,\omega_k) = (G_{k-1}(\omega_1,\dots,\omega_{k-1}), g_k(G_{k-1}(\omega_1,\dots,\omega_{k-1}),\omega_k)) \tag{10}$$

we get $D_k \subset D_{k-1}\cup[G_{k-1},I]^{-1}D_{g_k}$, where D_{g_k} is the set of discontinuities of g_k, and $[G_{k-1},I]$ is the mapping $(\omega_1,\dots,\omega_k) \mapsto (G_{k-1}(\omega_1,\dots,\omega_{k-1}),\omega_k)$. Consequently,

$$\begin{aligned}(U_1\times\cdots\times U_k)(D_k) &\leq (U_1\times\cdots\times U_{k-1})(D_{k-1})\\ &\quad +(U_1\times\cdots\times U_k)[G_{k-1},I]^{-1}(D_{g_k})\\ &= (U_1\times\cdots\times U_{k-1})(D_{k-1}) + (P_{k-1}\times U_k)(D_{g_k}),\end{aligned}$$

where the last term equals zero by condition 3.

Using (10) and induction on k again, it follows from the measurability of g_k in condition 2 that G_k is $(\mathcal{R}_k,\mathcal{B}_k)$-measurable. Thus, for any $B_k \in \mathcal{B}_k$ (the Borel σ-field on $\Xi_1\times\cdots\times\Xi_k$)

$$\begin{aligned}G^{-1}(B_k\times\Xi_{k+1}\times\cdots\times\Xi_K) &= \{(\omega_1,\dots,\omega_K)\,|\,G_k(\omega_1,\dots,\omega_k)\in B_k\}\\ &= G_k^{-1}(B_k)\times\Omega_{k+1}\times\cdots\times\Omega_K \in \mathcal{G}_k,\end{aligned}$$

which means that G is $(\mathcal{G}_k, \mathcal{F}_k)$-measurable. By condition 1, we have for any $R_k \in \mathcal{R}_k$ that

$$G(R_k \times \Omega_{k+1} \times \cdots \times \Omega_K) = G_k(R_k) \times \Xi_{k+1} \times \cdots \times \Xi_K.$$

Expression (9) shows that, by condition 2, G_k^{-1} is $(\mathcal{B}_k, \mathcal{R}_k)$-measurable, so the right hand side belongs to $\mathcal{F}_k$.

We will also need the following version of the so called Slutsky's theorem.

Lemma 4 *Assume that $H^\nu \xrightarrow{U} H$ and let Λ be a set containing the ranges of H and H^ν. If $G : \Lambda \to \mathbb{R}^d$ is such that the set of its discontinuity points in* rge H *has UH^{-1}-measure zero, then,*

$$G \circ H^\nu \xrightarrow{U} G \circ H.$$

Proof By [3, Theorem 20.5], it suffices to show that every subsequence of $(G \circ H^\nu)_{\nu=1}^\infty$ has a further subsequence that converges U-a.s. to $G \circ H$. By [3, Theorem 20.5] again, this follows from $H^\nu \xrightarrow{U} H$ and the continuity assumption on G.

Proof of Theorem 2 By Lemma 2, $U_k^\nu \to U_k$ implies that there exist partitions $\{\Omega_k^{\nu,i}\}_{i \in I_k(\nu)}$ satisfying (B2) and (B1). We will show that (A1)–(A3) are satisfied by the partitions $\{\Xi^{\nu,i}\}_{i \in I(\nu)}$ defined for each ν and $i \in I(\nu)$ by $\Xi^{\nu,i} = G(\Omega^{\nu,i})$, where $\Omega^{\nu,i} = \Omega_1^{\nu,i_1} \times \cdots \times \Omega_K^{\nu,i_K}$.

Note first that

$$\begin{aligned} s^\nu(\xi) &= \xi^{\nu,i} \qquad \text{if } \xi \in \Xi^{\nu,i} \\ &= G(\omega^{\nu,i}) \qquad \text{if } G^{-1}(\xi) \in \Omega^{\nu,i}, \end{aligned}$$

or in other words,

$$s^\nu = G \circ (s_1^\nu \times \cdots \times s_K^\nu) \circ G^{-1}, \tag{11}$$

where

$$s_k^\nu(\omega_k) = \omega_k^{\nu,i} \quad \text{if } \omega_k \in \Omega_k^{\nu,i}.$$

Using (11), the measurability properties in Lemma 3, and the fact that $s_1^\nu \times \cdots \times s_K^\nu$ is $(\mathcal{G}_k, \mathcal{G}_k)$-measurable, we get

$$\begin{aligned} (s^\nu)^{-1}(\mathcal{F}_k) &= G\left((s_1^\nu \times \cdots \times s_K^\nu)^{-1}\left(G^{-1}(\mathcal{F}_k)\right)\right) \\ &\subset G\left((s_1^\nu \times \cdots \times s_K^\nu)^{-1}(\mathcal{G}_k)\right) \subset G(\mathcal{G}_k) \subset \mathcal{F}_k, \end{aligned}$$

so (A1) holds.

Using the definition of P and (11), we get

$$\begin{aligned}P(|s^\nu(\xi) - \xi| > \epsilon) &= U(|s^\nu(G(\omega)) - G(\omega)| > \epsilon)\\ &= U(|G((s_1^\nu \times \cdots \times s_K^\nu)(\omega)) - G(\omega)| > \epsilon),\end{aligned}$$

so (A2) follows from Lemma 4, Lemma 3 and (B1).

To verify (A3), note first that

$$P(\Xi^{\nu,i}) = UG^{-1}\left(G(\Omega_1^{\nu,i_1} \times \cdots \times \Omega_K^{\nu,i_K})\right) = U_1(\Omega_1^{\nu,i_1}) \cdots U_K(\Omega_K^{\nu,i_K}),$$

and that for any scalars a_k

$$(a_1 \cdots a_K - 1) = \sum_{\mathcal{J} \subset \{1,\dots,K\}} \prod_{k \in \mathcal{J}} (a_k - 1).$$

Thus,

$$\begin{aligned}\|\psi^\nu - 1\|_{L^\infty} &= \max_{i \in I(\nu)} \left| \frac{p^{\nu,i}}{P(\Xi^{\nu,i})} - 1 \right|\\ &= \max_{i \in I(\nu)} \left| \frac{p_1^{\nu,i_1}}{U_1(\Omega_1^{\nu,i_1})} \cdots \frac{p_K^{\nu,i_K}}{U_K(\Omega_K^{\nu,i_K})} - 1 \right|\\ &\le \sum_{\mathcal{J} \subset \{1,\dots,K\}} \prod_{k \in \mathcal{J}} \max_{i_k \in I_k(\nu)} \left| \frac{p_k^{\nu,i_k}}{U_k(\Omega_k^{\nu,i_k})} - 1 \right|,\end{aligned}$$

so (A3) follows from (B2). □

Acknowledgments I would like to thank doctor Matti Koivu for numerous fruitful discussions at various stages of this project.

References

1. Attouch, H.: Variational Convergence for Functions and Operators. Pitman (Advanced Publishing Program), Boston (1984)
2. Back, K., Pliska, S.R.: The shadow price of information in continuous time decision problems. Stochastics **22**(2), 151–186 (1987)
3. Billingsley, P.: Probability and measure. 3rd edn. Wiley Series in Probability and Mathematical Statistics. Wiley, New York (1995)
4. Billingsley, P.: Convergence of probability measures, 2nd edn. Wiley, New York (1999)
5. Black, F., Karasinski, P.: Bond and option pricing when short rates are lognormal. Finan. Anal. J. **July–August**, 52–59 (1991)
6. Boender, C.G.E., van Aalst, P., Heemskerk, F.: Modelling and management of assets and liabilities of pension plans in The Netherlands. In: Ziemba W.T., Mulvey J.M. (eds.) Worldwide asset and liability modeling, pp. Cambridge University Press 561–580 (1998)
7. Braides, A.: Γ-convergence for beginners. Oxford Lecture Series in Mathematics and its Applications, vol. 22. Oxford University Press, Oxford

8. Casey, M., Sen, S.: The scenario generation algorithm for multistage stochastic linear programs. Math. Oper. Res. **30**(3), 615–631 (2005)
9. Chen, X., Womersley, R.S.: A parallel inexact Newton method for stochastic programs with recourse. Ann. Oper. Res. **64**, 113–141 (1996)
10. Chiralaksanakul, A.: Monte carlo methods for multi-stage stochastic programs. PhD thesis, University of Texas at Austin, Texas (2003)
11. Clements, M.P., Hendry, D.F.: Forecasting Non-Stationary Economic Time Series. MIT Press, Cambridge (1999)
12. Dal Maso, G.: An introduction to Γ-convergence, Progress in Nonlinear Differential Equations and their Applications, vol. 8. Birkhäuser Boston Inc., Boston (1993)
13. Dempster, M.A.H., Germano, M., Medova, E., Villaverde, M.: Global asset liability management. Br. Actuar. J. **9**(1), 137–195 (2003)
14. Dupačová, J., Consigli, G., Wallace, S.W.: Scenarios for multistage stochastic programs. Ann. Oper. Res. **100**, 25–53 (2001)
15. Dupačová, J., Gröwe-Kuska, N., Römisch, W.: Scenario reduction in stochastic programming. An approach using probability metrics. Math. Program. **95**(3, Ser. A), 493–511 (2003)
16. Engle, R.F., Granger, C.W.J.: Co-integration and error correction: representation, estimation, and testing. Econometrica **55**(2), 251–276 (1987)
17. Folland, G.B.: Real analysis, 2nd edn. Pure and Applied Mathematics. Wiley, New York (1999)
18. Frauendorfer, K.: Barycentric scenario trees in convex multistage stochastic programming. Math. Programm. Approximation Comput. Stoc. Programm. **75**(2, Ser. B), 277–293 (1996)
19. Gondzio, J., Kouwenberg, R., Vorst, T.: Hedging options under transaction costs and stochastic volatility. J. Econ. Dyn. Control **27**(6), 1045–1068 (2003)
20. Heitsch, H., Römisch, W.: Scenario reduction algorithms in stochastic programming. Comput. Optim. Appl. Stochastic Programming **24**(2–3), 187–206 (2003)
21. Heitsch, H., Römisch, W., Strugarek, C.: Stability of multistage stochastic programs. Preprint 255, DFG Research Center Matheon Mathematics for key technologies (2005)
22. Hilli, P., Koivu, M., Pennanen, T., Ranne, A.: A stochastic programming model for asset liability management of a finnish pension company. Ann. Oper. Res. (in press)
23. Hlawka, E., Mück, R.: Über eine Transformation von gleichverteilten Folgen. II. Computing (Arch. Elektron. Rechnen) **9**, 127–138 (1972)
24. Ioffe, A.D.: On lower semicontinuity of integral functionals. I. SIAM J. Control Optimization **15**(4), 521–538 (1977)
25. Kouwenberg, R.: Scenario generation and stochastic programming models for asset liability management. Eur. J. Oper. Res. Financ. Model. **134**(2), 279–292 (2001)
26. L'Ecuyer, P., Lemieux, C.: Variance reduction via lattice rules. Manage. Sci. **46**(2), 1214–1235 (2000)
27. Lepp, R.: Approximations to stochastic programs with complete recourse. SIAM J. Control Optim. **28**(2), 382–394 (1990)
28. Lucchetti, R., Salinetti, G., Wets, R.J.B.: Uniform convergence of probability measures: topological criteria. J. Multivariate Anal. **51**(2), 252–264 (1994)
29. Marti, K., Ermoliev, Y., Pflug, G. (eds.): Dynamic stochastic optimization, Lecture Notes in Economics and Mathematical Systems. vol. 532. Papers from the IFIP/IIASA/GAMM Workshop held in Laxenburg, March 11–14, 2002. Springer, Berlin (2004)
30. Marti, K., Kall, P. (eds.): Stochastic programming. Lecture Notes in Economics and Mathematical Systems. vol. 423. Numerical techniques and engineering applications. Springer, Berlin (1995)
31. Marti, K., Kall, P. (eds.): Stochastic programming methods and technical applications. Lecture Notes in Economics and Mathematical Systems. vol. 458. Springer-Verlag, Berlin (1998)
32. Mordukhovich, B.: On difference approximations of optimal control systems. J. Appl. Math. Mech **42**(3), 431–440 (1978)
33. Niederreiter, H.: Random number generation and quasi-Monte Carlo methods, CBMS-NSF Regional Conference Series in Applied Mathematics. vol. 63. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (1992)
34. Niederreiter, H. (ed.): Monte Carlo and quasi-Monte Carlo methods 2002. In: Proceedings of a conference held at the National Univeristy of Singapore, Republic of Singapore, November 25–28, 2002. Springer-Verlag, Berlin (2004)
35. Olsen, P.: Discretizations of multistage stochastic programming problems. Math. Programm. Stud. **6**, 111–124 (1976)

36. Pennanen, T.: Epi-convergent discretizations of multistage stochastic programs. Math. Oper. Res. **30**(1), 245–256 (2005)
37. Pennanen, T., Koivu, M.: Integration quadratures in discretization of stochastic programs. Stochastic Programming E-Print Series (2002)
38. Pennanen, T., Koivu, M.: Epi-convergent discretizations of stochastic programs via integration quadratures. Numer. Math. **100**(1), 141–163 (2005)
39. Pflug, G.: Scenario tree generation for multiperiod financial optimization by optimal discretization. Math. Program. Math. Programm. Finance **89**(2, Ser. B), 251–271 (2001)
40. 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)
41. Polak, E.: Optimization. Applied Mathematical Sciences. vol. 124. Algorithms and consistent approximations. Springer, New York (1997)
42. Rockafellar, R.T., Wets, R.J.B.: Continuous versus measurable recourse in N-stage stochastic programming. J. Math. Anal. Appl. **48**, 836–859 (1974)
43. Rockafellar, R.T., Wets, R.J.B.: Nonanticipativity and L^1-martingales in stochastic optimization problems. Math. Programm. Stud. **6**, 170–187 (1976)
44. Rockafellar, R.T., Wets, R.J.B.: Measures as Lagrange multipliers in multistage stochastic programming. J. Math. Anal. Appl. **60**(2), 301–313 (1977)
45. Rockafellar, R.T., Wets, R.J.B.: Variational analysis. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] vol. 317. Springer, Berlin (1998)
46. Römisch, W.: Stability of stochastic programming problems. In: Stochastic Programming, Handbooks in Operations Research and Management Science. vol. 10, pp. 483–554. Elsevier, Amsterdam (2003)
47. Ruszczynski, A., Shapiro, A. (eds.): Stochastic Programming, Handbooks in Operations Research and Management Science. vol. 10. Elsevier, Amsterdam (2003)
48. Shapiro, A.: Inference of statistical bounds for multistage stochastic programming problems. Math. Methods Oper. Res. **58**(1), 57–68 (2003)
49. Shapiro, A.: On complexity of multistage stochastic programs. Optimization Online (2005)
50. Shiryaev, A.N.: Essentials of stochastic finance, Advanced Series on Statistical Science & Applied Probability. vol. 3. World Scientific Publishing Co. Inc., River Edge(1999)
51. Sloan, I.H., Joe, S.: Lattice methods for multiple integration. The Clarendon Press/Oxford University Press, New York (1994)
52. Vainikko, G.M.: The convergence of the method of mechanical quadratures for integral equations with discontinuous kernels. Siberian Math. J. **12**, 29–38 (1971)
53. Ziemba, W.T., Mulvey, J.M. (eds.): Worldwide Asset and Liability Modeling, Publications of the Newton Institute, vol. 10. Cambridge University Press, Cambridge (1998)