MATHEMATICS OF OPERATIONS RESEARCH
Vol. 30, No. 1, February 2005, pp. 245–256
ISSN 0364-765X | EISSN 1526-5471 | 05 | 3001 | 0245
DOI 10.1287/moor.1040.0114

Epi-Convergent Discretizations of Multistage Stochastic Programs

Teemu Pennanen
Department of Management Science, Helsinki School of Economics, PL 1210,
00101 Helsinki, Finland, pennanen@hkkk.fi

In many dynamic stochastic optimization problems in practice, the uncertain factors are best modeled as random variables with an infinite support. This results in infinite-dimensional optimization problems that can rarely be solved directly. Therefore, the random variables (stochastic processes) are often approximated by finitely supported ones (scenario trees), which result in finite-dimensional optimization problems that are more likely to be solvable by available optimization tools. This paper presents conditions under which such finite-dimensional optimization problems can be shown to epi-converge to the original infinite-dimensional problem. Epi-convergence implies the convergence of optimal values and solutions as the discretizations are made finer. Our convergence result applies to a general class of convex problems where neither linearity nor complete recourse are assumed.

Key words: multistage stochastic program; discretization; epi-convergence
MSC2000 subject classification: Primary: 90C15, 49M25, 90C25
OR/MS subject classification: Primary: Programming/stochastic
History: Received October 12, 2003; revised February 18, 2004, and April 29, 2004.

---

1. Introduction. This paper is concerned with numerical solution of multistage decision problems where, at each *stage* $k = 0, \dots, K$, the decision maker observes the value of a random variable ξ_k, and makes a decision x_k depending on the observed values of $\xi_0, \dots, \xi_k$. In many applications, the first stage decision x_0 is deterministic, which corresponds to ξ_0 being constant. A sequence of decisions $x = (x_0, \dots, x_K)$ together with a realization of $\xi = (\xi_0, \dots, \xi_K)$ determines a "cost" $f(x, \xi)$. The objective is to find a decision rule $x(\xi)$ that minimizes the expectation of $f(x(\xi), \xi)$. Many decision problems in practice can be cast in this general framework; see for example Ziemba and Mulvey [40], Marti and Kall [17], Föllmer and Schied [11, Part II] and Ruszczynski and Shapiro [32].

We will assume that the random variable ξ_k takes values in a Borel subset Ξ_k of $\mathbb{R}^{d_k}$ and the decision x_k is $\mathbb{R}^{n_k}$-valued. The vector ξ 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 $(\Xi, \mathcal{F})$. For $k = 0, \dots, K$, $\mathcal{B}_k$ denotes the Borel σ-field on $\Xi_0 \times \cdots \times \Xi_k$, and π_k the projection of $\Xi_0 \times \cdots \times \Xi_K$ on $\Xi_0 \times \cdots \times \Xi_k$. The σ-fields

$$\mathcal{F}_k := \pi_k^{-1}(\mathcal{B}_k) = \{B_k \times \Xi_{k+1} \times \cdots \times \Xi_K \mid B_k \in \mathcal{B}_k\}$$

define a filtration $\mathcal{F}_0 \subset \cdots \subset \mathcal{F}_K = \mathcal{F}$ that describes the information available to the decision maker at each stage k. Indeed, the requirement that the function x_k: $\Xi \to \mathbb{R}^{n_k}$ depends only on the values of $\xi_0, \dots, \xi_k$ means that it is $\mathcal{F}_k$-measurable (when $\mathbb{R}^{n_k}$ is endowed with its Borel structure). A function $x = (x_0, \dots, x_K)$ is said to be *nonanticipative* or *adapted to the filtration* $(\mathcal{F}_k)_{k=0}^K$ if x_k is $\mathcal{F}_k$-measurable, for $k = 0, \dots, K$.

The fact that the expectation $E^P f(x(\xi), \xi)$ is not affected if we alter x on a set of P-measure zero, suggests taking as decision variables equivalence classes of functions that agree P-almost surely. More specifically, we will restrict the decision variables x to be elements of the Lebesque space $X(P) := L^\infty(\Xi, \mathcal{F}, P; \mathbb{R}^n)$, where $n = n_0 + \cdots + n_K$. For a filtration $(\mathcal{G}_k)_{k=0}^K$, an element $x = (x_0, \dots, x_K)$ of $X(P)$ will be called $(\mathcal{G}_k)_{k=0}^K$-*adapted* if it contains a function which is adapted to $(\mathcal{G}_k)_{k=0}^K$. The set of $(\mathcal{F}_k)_{k=0}^K$-adapted elements of $X(P)$ will be denoted by $\mathcal{N}(P)$, i.e.,

$$\mathcal{N}(P) := \{x \in X(P) \mid x \text{ contains an } (\mathcal{F}_k)_{k=0}^K\text{-adapted function}\}.$$

Our decision problem will be modeled as the *multistage stochastic program*

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

where f is a convex normal integrand on $\mathbb{R}^n\times\Xi$. Recall that a function $f\colon \mathbb{R}^n\times\Xi\to(-\infty,+\infty]$ is called a *convex normal integrand* if $f(\cdot,\xi)$ is convex and lower semicontinuous for every $\xi\in\Xi$ and if f is $\mathcal{B}(\mathbb{R}^n)\times\mathcal{F}$-measurable, where $\mathcal{B}(\mathbb{R}^n)$ is the Borel σ-field on $\mathbb{R}^n$. Problem ($SP(P)$) and its dual has been studied in a series of papers by Rockafellar and Wets [26, 27, 28, 29]. It should be noted that by allowing f to take on the value $+\infty$, explicit constraints can be incorporated into the objective by infinite penalties. This makes ($SP(P)$) a very general model for decision problems.

Unless P is finitely supported, $X(P)$ is an infinite-dimensional space and ($SP(P)$) cannot be solved directly except in some special cases. Infinitely supported measures come up quite naturally, for example, in various financial applications; see, e.g., Shiryaev [34]. In practice, such infinite-dimensional problems are often discretized by replacing P by a finitely supported measure of the form

$$P^\nu=\sum_{i=1}^{\nu} p^{\nu,i}\delta_{\xi^{\nu,i}},$$

where $\delta_{\xi^{\nu,i}}$ is the point mass at a point $\xi^{\nu,i}\in\Xi$, and the set $\{\xi^{\nu,i},p^{\nu,i}\}_{i=1}^{\nu}$ of *scenarios* and associated probabilities is believed to somehow describe the underlying uncertainties. For simplicity, we will assume throughout that $p^{\nu,i}>0$. Then $X(P^\nu)\cong(\mathbb{R}^n)^\nu$ and ($SP(P^\nu)$) can be written in the finite-dimensional form

$$\underset{x\in\mathcal{N}(P^\nu)}{\text{minimize}}\; \sum_{i=1}^{\nu} p^{\nu,i} f(x(\xi^{\nu,i}),\xi^{\nu,i}), \tag{$SP(P^\nu)$}$$

where

$$\begin{aligned}\mathcal{N}(P^\nu)&=\{x\in X(P^\nu)\mid x \text{ contains an } (\mathcal{F}_k)_{k=0}^K\text{-adapted function}\}\\&=\big\{x\in X(P^\nu)\mid x_k(\xi^{\nu,i})=x_k(\xi^{\nu,j}) \text{ if } \pi_k\xi^{\nu,i}=\pi_k\xi^{\nu,j}\big\}.\end{aligned}$$

This is a mathematical program that can in principle be solved numerically by standard solvers or special purpose algorithms designed to take advantage of problem structure.

It is natural to ask whether ($SP(P^\nu)$) can really be considered an approximation of ($SP(P)$). We will study this question under the general framework of *epi-convergence*; see for example Attouch [3], Rockafellar and Wets [30, Chapter 7] or Braides [6]. Epi-convergence is by now widely recognized as the right framework for studying approximations of optimization problems and it has been successfully applied to discretizations of various infinite-dimensional optimization problems such as finite element and finite difference approximations in numerical analysis of ordinary and partial differential equations. Our aim here is to find conditions under which the optimal values and solutions of ($SP(P^\nu)$) converge to those of ($SP(P)$) as the discretizations are made finer by increasing the number ν of scenarios.

In one-stage (static) stochastic programs, where the decision variables are not functions of ξ, the situation is much simpler. In the present notation, a one-stage problem is obtained when $K=1$, ξ_0 is constant, and f is independent of x_1. Then the decision variables both in ($SP(P)$) and ($SP(P^\nu)$) can be viewed as elements of $\mathbb{R}^n$ instead of $X(P)$ and $X(P^\nu)$. One-stage problems have been studied by many authors and conditions have been found that guarantee the epi-convergence of the objectives when the underlying probability measure is approximated; see for example Birge and Wets [5], Robinson and Wets [24], Dupačová and Wets [8], Lucchetti and Wets [16], Artstein and Wets [1, 2], Zervos [39], Schultz [33], Vogel and Lachout [37], Pennanen and Koivu [21] and their references.

For general multistage problems, the situation is not as good. Here, the analysis is complicated due to the fact that the underlying decision space depends on the probability measure: ($SP(P)$) is a minimization problem over the space $X(P)$ whereas ($SP(P^\nu)$) is over $X(P^\nu)$. One does not have that $X(P^\nu) \subset X(P)$, and it is not even clear how to interpret the variables of ($SP(P^\nu)$). Discretizations of multistage stochastic programs have been studied by various techniques in the linear case by Olsen [18] and Casey and Sen [7], in the two-stage case by Lepp [15], and in the case of concave-convex value functions by Frauendorfer [12]. Wang [38] gave conditions for epi-convergence of certain perturbations of linear and linear-quadratic stochastic programs, but his results do not apply to the above discretizations. In this paper, we give conditions that imply that the essential objectives of appropriate reformulations of ($SP(P^\nu)$) epi-converge to the essential objective of ($SP(P)$) with respect to the weak*-topology of $X(P)$. The general properties of epi-convergence then allow us to deduce the convergence of optimal values and of optimal first stage solutions.

This paper is concerned only with the behavior of ($SP(P^\nu)$) as the number of scenarios tends to infinity. In particular, nothing is said about how many scenarios are required to get good approximations of the original problem, or how "far" ($SP(P^\nu)$) is from ($SP(P)$) for given P^ν. Quantitative results for approximations of optimization problems usually require stronger assumptions on the given problem. Our aim here is to give as general conditions as possible that will guarantee the asymptotic consistency of discretizations which is a minimal requirement for any approximation scheme. Quantitative results for certain class of perturbations of multistage stochastic programs can be found in Fiedler and Römisch [9]. Empirical tests have shown that the number of scenarios required to get convergence of optimal values and optimal solutions grows rapidly with the number of stages K.

The rest of this paper is organized as follows. We will first reformulate, in §2, problems ($SP(P^\nu)$) as problems of minimizing certain functions F^ν over the original space $X(P)$. In §3, we give conditions that guarantee the epi-convergence of these functions F^ν to the essential objective of ($SP(P)$) with respect to the weak*-topology of $X(P)$. This yields results on the convergence of optimal values and optimal solutions of ($SP(P^\nu)$). Our conditions concern both the problem being discretized and the discrete measures P^ν. In §4, we show that the class of stochastic programs analyzed in Rockafellar and Wets [28] satisfies the conditions we pose on the problem. Methods for constructing discrete measures that satisfy our conditions, along with some numerical tests, have been already presented in Pennanen and Koivu [20]. That those methods really satisfy the conditions given in this paper (Assumption 3.1 below) will be verified in Pennanen [19].

2. Reformulations. In order to analyze the discretizations ($SP(P^\nu)$) through the theory of epi-convergence, we will first express both ($SP(P)$) and ($SP(P^\nu)$) as problems of minimizing certain functions F and F^ν, respectively, over $X(P)$. For ($SP(P)$), we simply set

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

Under mild conditions, F is convex and lower semicontinuous (lsc) in the weak*-topology which is defined as the weakest topology which makes continuous all functions of the form

$$x \mapsto E^P[x(\xi) \cdot x^*(\xi)],$$

where $x^* \in L^1(\Xi, \mathcal{F}, P; \mathbb{R}^n)$. Indeed, the subspace $\mathcal{N}(P)$ is weak*-closed, and the function $x \mapsto E^P f(x(\xi), \xi)$ is weak*-lsc provided there exists an $x^* \in L^1(\Xi, \mathcal{F}, P; \mathbb{R}^n)$ such that $E^P f^*(x^*(\xi), \xi) < \infty$, where $f^*(\cdot, \xi)$ denotes the conjugate of $f(\cdot, \xi)$ (Rockafellar [25, Corollary 3D]).

As to ($SP(P^\nu)$), we will first embed the spaces $X(P^\nu)$ in $X(P)$. To this end, we assume that, for each $\nu = 1, 2, \ldots,$ there is a partition $\{\Xi^{\nu,i}\}_{i=1}^{\nu}$ of Ξ into ν sets (one for each

$\xi^{\nu,i}$) with $\Xi^{\nu,i} \in \mathcal{F}$ and $P(\Xi^{\nu,i}) > 0$. These partitions allow us to define finite-dimensional subspaces of $SP(P)$ on which one can define a minimization problem that is equivalent to ($SP(P^\nu)$). Our convergence analysis will require the existence of such partitions, but note that, the partitions are not needed in the formulation of the discretizations ($SP(P^\nu)$). The idea of using partitions in studying approximations of stochastic programs is quite natural and has been used for example in Olsen [18], Kall et al. [14], Lepp [15], Frauendorfer [12], Casey and Sen [7].

Let $\mathcal{F}^\nu$ be the σ-field generated by the νth partition, and define the finite-dimensional subspace

$$X^\nu(P) := \{x \in X(P) \mid x \text{ contains an } \mathcal{F}^\nu\text{-measurable function}\}$$

of $X(P)$. This is the set of equivalence classes of functions equivalent to step functions of the form

$$x = \sum_{i=1}^{\nu} z^{\nu,i} \chi_{\Xi^{\nu,i}},$$

where $z^{\nu,i} = (z_0^{\nu,i}, \ldots, z_K^{\nu,i}) \in \mathbb{R}^n$. The restriction to $X^\nu(P)$ of the linear mapping A^ν: $X(P) \to X(P^\nu)$,

$$(A^\nu x)(\xi^{\nu,i}) := \frac{1}{P(\Xi^{\nu,i})} \int_{\Xi^{\nu,i}} x(\xi)\, P(d\xi)$$

defines a continuous bijection from $X^\nu(P)$ to $X(P^\nu)$. The inverse Π^ν: $X(P^\nu) \to X^\nu(P)$ of this bijection maps points $z \in X(P^\nu)$ to their *prolongations*

$$\Pi^\nu z = \sum_{i=1}^{\nu} z^{\nu,i} \chi_{\Xi^{\nu,i}} \in X^\nu(P).$$

It follows that ($SP(P^\nu)$) is equivalent to the problem of minimizing over $X(P)$ the function

$$F^\nu(x) = \begin{cases} \tilde{F}^\nu(A^\nu x) & \text{if } x \in X^\nu(P), \\ +\infty & \text{otherwise,} \end{cases}$$

where $\tilde{F}^\nu$ denotes the essential objective of ($SP(P^\nu)$). Indeed, a z solves ($SP(P^\nu)$) if and only if $\Pi^\nu z$ minimizes F^ν. Note that F^ν is lsc since it is the composition of a continuous mapping with a lsc function.

Lemma 2.1. *We have*

$$F^\nu(x) = \begin{cases} E^P f(x(\xi), s^\nu(\xi))\psi^\nu(\xi) & \text{if } x \in \mathcal{N}^\nu(P), \\ +\infty & \text{otherwise,} \end{cases} \tag{1}$$

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

$$s^\nu(\xi) = \xi^{\nu,i} \quad \text{and} \quad \psi^\nu(\xi) = \frac{p^{\nu,i}}{P(\Xi^{\nu,i})} \qquad \text{if } \xi \in \Xi^{\nu,i}, \tag{2}$$

and $\mathcal{N}^\nu(P)$ is the set of $((s^\nu)^{-1}(\mathcal{F}_k))_{k=0}^K$-adapted elements of $X(P)$.

Proof. Note first that, by the Doob-Dynkin lemma (see, e.g., Rao [23], p. 4),

$$\begin{aligned} \mathcal{N}^\nu(P) &= \{x \in X(P) \mid x \ni \tilde{x} \circ s^\nu \text{ for some } (\mathcal{F}_k)_{k=1}^K\text{-adapted function } \tilde{x}\} \\ &= \{x \in X^\nu(P) \mid (A^\nu x)(\xi^{\nu,i}) = \tilde{x}(\xi^{\nu,i}) \quad \forall\, i = 1, \ldots, \nu \\ &\qquad \text{for some } (\mathcal{F}_k)_{k=1}^K\text{-adapted function } \tilde{x}\} \\ &= \{x \in X^\nu(P) \mid A^\nu x \in \mathcal{N}(P^\nu)\}. \end{aligned} \tag{3}$$

We thus have that

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

where $E^{P^\nu} f(A^\nu x(\xi), \xi) = E^P f(x(\xi), s^\nu(\xi))\psi^\nu(\xi)$, for any $x \in X^\nu(P)$. $\square$

3. Epi-convergence. From now on, unless otherwise specified, we equip $X(P)$ with the weak*-topology. The *lower epi-limit* of a sequence $\{F^\nu\}_{i=1}^\infty$ of functions is the lsc function defined 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 defined by

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

If $\text{e-lim inf}\, F^\nu = \text{e-lim sup}\, F^\nu$, then the common limit, denoted $\text{e-lim}\, F^\nu$, is called the *epi-limit* of $\{F^\nu\}_{i=1}^\infty$ and the sequence is said to *epi-converge* to it.

Epi-convergence has many important implications for approximations of minimization problems. The following theorem, where the set $\{x \mid F(x) \le \inf F + \epsilon\}$ of ϵ-minimizers of a function F will be denoted by $\epsilon\text{-}\arg\min F$, lists some of them; see, e.g., Attouch [3, §2.2].

Theorem 3.1. *Let $\{\epsilon^\nu\}_{\nu=1}^\infty$ be a sequence of nonnegative real numbers converging to zero. If $\{F^\nu\}_{\nu=1}^\infty$ epi-converges to F, then*

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

and if there is a sequence $x^\mu \to x$ such that $x^\mu \in \epsilon^{\nu^\mu}\text{-}\arg\min F^{\nu^\mu}$ for some subsequence $\{\nu^\mu\}_{\mu=1}^\infty$, then $x \in \arg\min F$ and $\inf F^{\nu^\mu} \to \inf F$. In particular, if there is a compact set C such that $\epsilon^\nu\text{-}\arg\min F^\nu \cap C \neq \varnothing$ for all ν, then $\inf F^\nu \to \inf F$.

Our epi-convergence result for the discretizations will rely on the following.

Assumption 3.1. *The sequence $\{P^\nu\}_{\nu=1}^\infty$ of scenario trees is such that there exists a sequence of partitions $\{\Xi^{\nu,i}\}_{i=1}^\nu$ such that*

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

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

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

where $\mathcal{N}^\nu(P)$, s^ν, and ψ^ν are as in Lemma 2.1.

By (A2) we mean that the functions s^ν converge in probability to the identity function, i.e., $P(|s^\nu(\xi) - \xi| \ge \epsilon) \to 0$ for every $\epsilon > 0$. Assumption (A3) means that

$$\max_{i=1,\dots,\nu} \left| \frac{p^{\nu,i}}{P(\Xi^{\nu,i})} - 1 \right| \to 0,$$

so it may be seen as a relaxed version of $P(\Xi^{\nu,i}) = p^{\nu,i}$. Assumption (A3) implies that, eventually, $P(\Xi^{\nu,i}) > 0$, so that the reformulations of ($SP(P^\nu)$) in §2 are valid.

Assumption (A1) means that the sets $\mathcal{N}^\nu(P)$ of $((s^\nu)^{-1}(\mathcal{F}_k))_{k=1}^K$-adapted elements in Lemma 2.1 are all contained in $\mathcal{N}(P)$. Since $\mathcal{F}_k = \pi_k^{-1}(\mathcal{B}_k)$, we can write (A1) as $(\pi_k s^\nu)^{-1}(\mathcal{B}_k) \subset \pi_k^{-1}(\mathcal{B}_k)$, which means that the function $\pi_k s^\nu$ only depends on $\pi_k \xi$. It can thus be written as

$$\pi_k(\Xi^{\nu,i}) \cap \pi_k(\Xi^{\nu,j}) \neq \varnothing \quad \Longrightarrow \quad \pi_k \xi^{\nu,i} = \pi_k \xi^{\nu,j},$$

which is the defining feature of a "grid" in Olsen [18, Definition 1.1]. When $K = 1$ and Ξ_0 is a singleton, as in Lepp [15], (A1) is trivially satisfied. More interesting is the following example from Olsen [18, Example 1.2].

Example 3.1. Assume that $P = P_0 \times \cdots \times P_K$ and that

$$P_k^\nu = \sum_{i=1}^{\nu_k(\nu)} p_k^{\nu,i} \delta_{\xi_k^{\nu,i}}$$

is a discretization of P_k and $\{\Xi_k^{\nu,i}\}_{i=1}^{\nu_k(\nu)}$ is a partition of Ξ_k. Define a discretization of P by

$$P^\nu = P_0^\nu \times \cdots \times P_K^\nu$$

and a partition of Ξ by

$$\Xi^{\nu,i} := \Xi_0^{\nu,i_0} \times \cdots \times \Xi_K^{\nu,i_K}.$$

Then, (A1) is satisfied.[1]

Situations where $P = P_0 \times \cdots \times P_K$ are of course rather special, but with an appropriate change of variables, many practically interesting models can be reduced into such a form; see for example Shiryaev [34, Chapter II] and Pennanen and Koivu [20]. A simple technique for generating scenario trees P^ν that satisfy Assumption 3.1 have been already presented in Pennanen and Koivu [20]; see Pennanen [19] for a detailed analysis.

Lemma 3.1. *Under* (A2), $\varphi \circ s^\nu$ *converges to* φ, *both in the* L^1*- and weak*-topologies, for any bounded and* P*-a.s. continuous function* φ *on* Ξ.

Proof. It suffices to show that every subsequence of $\{\varphi \circ s^\nu\}_{\nu=1}^\infty$ has a subsequence converging to φ. Assumption (A2) implies that every subsequence of $\{s^\nu\}_{\nu=1}^\infty$ has a subsequence $\{s^{\nu^\mu}\}_{\mu=1}^\infty$ that converges P-a.s. to I; see for example Folland [10, Theorem 2.30]. Since φ is P-a.s. continuous it follows that $|\varphi(s^{\nu^\mu}(\xi)) - \varphi(\xi)|$ converges to zero P-a.s., and then, by the dominated convergence theorem, the boundedness of φ implies $\|\varphi \circ s^{\nu^\mu} - \varphi\|_{L^1} \to 0$. Weak*-convergence follows from L^1-convergence together with the boundedness of φ. Indeed, for any $\phi \in L^1$, the measure $|\phi|P$ is absolutely continuous with respect to P, and it follows that every subsequence of $\{s^\nu\}_{\nu=1}^\infty$ has a subsequence $\{s^{\nu^\mu}\}_{\mu=1}^\infty$ such that $|\varphi(s^{\nu^\mu}(\xi)) - \varphi(\xi)|$ converges to zero $|\phi|P$-a.e. Consequently,

$$|E^P[\varphi(s^{\nu^\mu}(\xi)) - \varphi(\xi)] \cdot \phi(\xi)| \le E^{|\phi|P} |\varphi(s^{\nu^\mu}(\xi)) - \varphi(\xi)| \to 0,$$

by the boundedness of φ. □

Following Ioffe [13], we will say that f has the *lower compactness property* if $f_-(x^\nu(\cdot), s^\nu(\cdot))$ is weakly precompact in L^1 whenever $\{x^\nu\}$ converges in $X(P)$, $\{s^\nu\}$ converges in measure P, and $\sup_\nu E^P f(x^\nu(\xi), s^\nu(\xi)) < \infty$. Here, $f_-(x, s) := \min\{f(x, s), 0\}$. More specific conditions implying the lower compactness property can be found in Ioffe [13, §3]. In particular, f has the lower compactness property if there exist real numbers a and b such that

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

We are now ready to prove our main result.

Theorem 3.2. *Assume that Assumption* 3.1 *holds and that* $E^P|\xi| \in \mathbb{R}$.

(1) *If* f *is lsc and has the lower compactness property, then* $\text{e-lim inf}\, F^\nu \ge F$.

(2) *If for every* $x \in \operatorname{dom} F$, *there is a uniformly bounded sequence* $y^\mu \to x$ *of nonanticipative,* P*-a.s. continuous functions such that*

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

$$\limsup_{\mu\to\infty} F(y^\mu) \le F(x),$$

then $\text{e-lim sup}\, F^\nu \le F$.

In particular, if both conditions (1) *and* (2) *hold, then* $\text{e-lim}\, F^\nu = F$.

[1] Here $i = (i_0, \ldots, i_K)$ is a multi-index and the number of scenarios $\xi^{\nu,i}$ and sets $\Xi^{\nu,i}$, for each ν, is $\nu_0(\nu) \times \cdots \times \nu_K(\nu)$. For simplicity, we do not use the multi-index notation in our general analysis but everything that is said applies to that case as well.

Proof. To verify the first claim, it suffices to show that $\liminf F^\nu(x^\nu) \geq F(x)$ whenever $x^\nu \to x$. The only challenging cases are the ones where there exists a subsequence for which the objective values are bounded from above. By passing to a subsequence if necessary, we can thus assume that $\sup_\nu F^\nu(x^\nu) < \infty$. Then, by Lemma 2.1, $x^\nu \in \mathcal{N}^\nu(P)$ and

$$F^\nu(x^\nu) = E^P f(x^\nu(\xi), s^\nu(\xi))\psi^\nu(\xi). \tag{4}$$

We will now apply the lower semicontinuity result of Ioffe [13]. Let $M = X(P)$ and $L = L^0 \times L^\infty$, where L^0 denotes the linear space of integrable $\mathbb{R}^{d_0+\cdots+d_K}$-valued functions with the topology of convergence in probability P. These spaces satisfy hypotheses (H_1) and (H_2) of Ioffe [13], and since $E^P|\xi| < \infty$, we have that both (s^ν, ψ^ν) and $(I, 1)$ belong to L. Since $\psi^\nu \geq 0$, we can write (4) as

$$F^\nu(x^\nu) = E^P \tilde{f}(x^\nu(\xi), s^\nu(\xi), \psi^\nu(\xi)),$$

where

$$\tilde{f}(x, s, \psi) = \begin{cases} f(x, s)\psi & \text{if } \psi \geq 0, \\ +\infty & \text{otherwise.} \end{cases}$$

The advantage of this form is that $\tilde{f}_-(x, s, \psi) \geq f_-(x, s)\psi$, which shows that the lower compactness property of f implies that $\tilde{f}$ is lower compact on $M \times L$ (see, e.g., Ioffe [13, Proposition 1]). Then, since $x^\nu \to x$ in M and, by (A2) and (A3), $(s^\nu, \psi^\nu) \to (I, 1)$ in L, Ioffe [13, Theorem 1][2] gives

$$\liminf_{\nu\to\infty} F^\nu(x^\nu) \geq E^P f(x(\xi), \xi).$$

Since $x^\nu \in \mathcal{N}^\nu(P)$, where $\mathcal{N}^\nu(P) \subset \mathcal{N}(P)$ by (A1), and since $\mathcal{N}(P)$ is closed, we must have $x \in \mathcal{N}(P)$, so the right-hand side equals $F(x)$. This completes the proof of the first claim.

The bound in the second claim holds trivially for $x \notin \operatorname{dom} F$, so let $x \in \operatorname{dom} F$ and consider the sequence $\{y^\mu\}_{\mu=1}^\infty$ given by the second condition. Let $y^{\mu,\nu} \in X(P)$ be the equivalence class of functions corresponding to

$$y^\mu \circ s^\nu = \sum_{i=1}^{\nu} y^\mu(\xi^{\nu,i})\chi_{\Xi^{\nu,i}}.$$

By Lemma 3.1,

$$y^{\mu,\nu} \to y^\mu, \tag{5}$$

as $\nu \to \infty$. For any Borel set $C \subset \mathbb{R}^{n_k}$, the nonanticipativity of y^μ implies

$$(y^\mu_k \circ s^\nu)^{-1}(C) = (s^\nu)^{-1}((y^\mu_k)^{-1}(C)) \in (s^\nu)^{-1}(\mathcal{F}_k),$$

so $y^{\mu,\nu} \in \mathcal{N}^\nu(P)$ and by Lemma 2.1,

$$F^\nu(y^{\mu,\nu}) = E^P f(y^\mu \circ s^\nu(\xi), s^\nu(\xi))\psi^\nu(\xi) = E^{P^\nu} f(y^\mu(\xi), \xi).$$

Thus, by the first inequality in the second condition,

$$\limsup_{\nu\to\infty} F^\nu(y^{\mu,\nu}) \leq E^P f(y^\mu(\xi), \xi) = F(y^\mu). \tag{6}$$

The proof is now completed by a diagonalization argument. Since the sequence y^μ is uniformly bounded, there exists an $M \in \mathbb{R}$ such that $y^{\mu,\nu} \in B := \{y \in X(P) \mid \|y\|_{L^\infty} \leq M\}$

[2] The assumption that the measure space is nonatomic (made on the first page of Ioffe [13]) is not needed for the proof of Ioffe [13, Theorem 1]. (The author is grateful to Professor Alexander Ioffe for pointing this out.)

for all μ and ν. By the Banach-Alaoglu theorem, B is weak*-compact so, by Rudin [31, Theorem 3.16], we can find a metric d that induces the weak*-topology on B. By Attouch [3, Corollary 1.16], there exists a mapping $\nu \mapsto \mu(\nu) \in \mathbb{N}$ such that $\mu(\nu) \nearrow \infty$ as $\nu \to \infty$, and

$$\limsup_{\nu\to\infty} \max\{F^\nu(y^{\mu(\nu),\nu}) - F(x), d(y^{\mu(\nu),\nu}, x)\}$$
$$\leq \limsup_{\mu\to\infty} \limsup_{\nu\to\infty} \max\{F^\nu(y^{\mu,\nu}) - F(x), d(y^{\mu,\nu}, x)\}.$$

By (5) and (6), the right-hand side is less than or equal to

$$\limsup_{\mu\to\infty} \max\{F(y^\mu) - F(x), d(y^\mu, x)\},$$

which is zero by the second inequality in the second condition. Thus, $y^{\mu(\nu),\nu} \to x$ and $\limsup F^\nu(y^{\mu(\nu),\nu}) \leq F(x)$, which proves the second claim. □

REMARK 3.1. (1) The lower semicontinuity property in condition (1) holds in most applications in practice, and it has been assumed also in the analysis of Bertsekas and Shreve [4, §8.3] and Lucchetti and Wets [16]. Section 4 describes situations where conditions (1) and (2) of Theorem 3.2 are automatically satisfied whenever Assumption 3.1 holds. Procedures for generating discrete measures for which Assumption 3.1 can be guaranteed will be presented in Pennanen [19]; see also Pennanen and Koivu [20].

(2) Instead of (A1), it would suffice that the outer limit $\limsup \mathcal{N}^\nu(P)$ of the sets $\mathcal{N}^\nu(P)$ is contained in $\mathcal{N}(P)$. Such a condition can also be expressed in terms of the σ-fields $(s^\nu)^{-1}(\mathcal{F}_k)$ and $\mathcal{F}_k$; see Piccinini [22]. We have chosen to use (A1) for simplicity since it holds for all the practical constructions we have in mind; see Pennanen [19].

(3) The sets $\mathcal{N}^\nu(P)$ are said to converge to $\mathcal{N}(P)$ in the sense of Painleve-Kuratowski if, in addition to $\limsup \mathcal{N}^\nu(P) \subset \mathcal{N}(P)$, one has $\mathcal{N}(P) \subset \liminf \mathcal{N}^\nu(P)$. Taking $f \equiv 0$ in the above proof shows that the sets $\mathcal{N}^\nu(P)$ converge to $\mathcal{N}(P)$ in the sense of Painleve-Kuratowski if

(a) $(s^\nu)^{-1}(\mathcal{F}_k) \subset \mathcal{F}_k$,

(b) $s^\nu \xrightarrow{P} I$,

(c) for every $x \in \mathcal{N}(P)$, there exists a uniformly bounded sequence $y^\mu \to x$ of P-a.s. continuous functions in $\mathcal{N}(P)$.

Combining Theorem 3.2 with Theorem 3.1, we obtain conditions for the convergence of the optimal values and the prolongations $\Pi^\nu z^\nu$ of the solutions z^ν of ($SP(P^\nu)$). Note that when Ξ_0 is a singleton (which is often the case in practice), the first-stage decision (which is what one is usually most interested in) is deterministic and the prolongation $(\Pi^\nu z^\nu)_0$ of the first-stage solution of ($SP(P^\nu)$) is the constant function z_0^ν. In this case, Theorem 3.2 yields the following.

COROLLARY 3.1. *Assume that the conditions of Theorem 3.2 hold and that problems* ($SP(P^\nu)$) *have* ϵ^ν*-optimal solutions* z^ν *such that* $\epsilon^\nu \searrow 0$ *and* $\max_{i=1,\dots,\nu} |z^{\nu,i}|$ *remains bounded. Then the optimal values of* ($SP(P^\nu)$) *converge to that of* ($SP(P)$), *and if* Ξ_0 *is a singleton, all cluster points of* $\{z_0^\nu\}_{\nu=1}^\infty$ *are optimal first-stage solutions of* ($SP(P)$).

PROOF. In terms of F^ν, we have $\Pi^\nu z^\nu \in \epsilon_\nu\text{-}\arg\min F^\nu$. The boundedness condition implies that the sequence $\{\Pi^\nu z^\nu\}_{\nu=1}^\infty$ is bounded in $X(P)$, so it is weakly relatively compact by the Banach-Alaoglu theorem. The conclusions now follow from Theorems 3.2 and 3.1 and the observation that any cluster point $\bar{z}_0$ of $\{z_0^\nu\}_{\nu=1}^\infty$ can be expressed as $\bar{z}_0 = \bar{x}_0$ for a cluster point $\bar{x}$ of $\{\Pi^\nu z^\nu\}_{\nu=1}^\infty$. Indeed, if $\{\nu^\mu\}_{\mu=1}^\infty$ is a subsequence such that $z_0^{\nu^\mu} \to \bar{z}_0$, we can, by weak compactness, find a subsequence of $\{(A^{\nu^\mu})^{-1} z^{\nu^\mu}\}_{\mu=1}^\infty$ that converges to a point $\bar{x}$. Since Ξ_0 is a singleton, $((A^{\nu^\mu})^{-1} z^{\nu^\mu})_0$ is the constant function $z_0^{\nu^\mu}$ for each μ, and we must have $\bar{x}_0 = \bar{z}_0$. □

If $K = 1$, Ξ_0 is a singleton and f is independent of x_1, ($SP(P)$) becomes a one-stage stochastic program, and the decision variables both in ($SP(P)$) and ($SP(P^\nu)$) can be viewed as elements of $\mathbb{R}^{n_0}$. In this case, Theorem 3.2 gives the following.

Corollary 3.2. *Consider the one-stage case and assume that Ξ is compact, $P^\nu \to P$, f is lsc and has the lower compactness property, and that for every $x \in \operatorname{dom} F$, there is a sequence $y^\mu \to x$ such that*

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

$$\limsup_{\mu\to\infty} F(y^\mu) \le F(x).$$

Then $\operatorname{e-lim} F^\nu = F$.

Proof. When Ξ is compact, $E^P|\xi| \in \mathbb{R}$ holds automatically. Then also, the weak convergence of P^ν to P implies that there exists a sequence of partitions such that (A2) and (A3) hold; see Vainikko [35]. Since $K = 1$, (A1) is trivially satisfied. The claim now follows from Theorem 3.2. □

Corollary 3.2 is close to Pennanen and Koivu [21, Corollary 10], but it applies to more general f and adds the compactness assumption on Ξ. It still allows implicit constraints of the form $f(x, \cdot) \in L^1(\Xi, \mathcal{F}, P)$ unlike most existing epi-convergence results for one-stage problems.

4. Weak convergence and continuous recourse. A sequence $\{P^\nu\}_{\nu=1}^\infty$ of probability measures is said to *converge weakly* to a measure P, denoted $P^\nu \to P$, if for every bounded and continuous function φ,

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

Weak convergence implies the convergence of expectations, not only of bounded and continuous functions, but also of P-a.s. continuous functions φ such that $E^{P^\nu}|\varphi(\xi)| \to E^P|\varphi(\xi)|$; (van der Vaart and Wellner [36, 1.11.3 Theorem, p. 69]).

Lemma 4.1. *Under* (A2) *and* (A3), $P^\nu \to P$.

Proof. In terms of the functions s^ν and ψ^ν in Lemma 2.1, we have

$$E^{P^\nu} \varphi = E^P \psi^\nu \varphi \circ s^\nu,$$

where $\psi^\nu \xrightarrow{L^\infty} \psi$ by (A3) and $\varphi \circ s^\nu \xrightarrow{L^1} \varphi$ by (A2) and Lemma 3.1. □

By the above lemma, condition (2) of Theorem 3.2 will hold under Assumption 3.1 if the feasible points of ($SP(P)$) can be approximated by feasible points which are sufficiently continuous functions of ξ. Such properties of ($SP(P)$) have been studied in Rockafellar and Wets [26, 28]. The problems studied in Rockafellar and Wets [28] correspond to functions f of the form

$$f(x, \xi) = \begin{cases} f_0(x, \xi) & \text{if } x \in X \text{ and } f_i(x, \xi) \le 0 \text{ for } i = 1, \ldots, m, \\ +\infty & \text{otherwise,} \end{cases} \tag{7}$$

where

(a) $X \subset \mathbb{R}^n$ is convex and compact with nonempty interior,
(b) f_i are convex in x and continuous on $X \times \operatorname{supp} P$.

Given $S \subset \Xi$, for each $k = 0, \ldots, K - 1$, let

$$\Lambda_k^S(\xi_0, \ldots, \xi_k) = \{(\xi_{k+1}, \ldots, \xi_K) \mid (\xi_0, \ldots, \xi_k, \xi_{k+1}, \ldots, \xi_K) \in S\}.$$

Following Rockafellar and Wets [26], we will say that a probability measure P is *laminary* if

(i) if $S \subset \operatorname{supp} P$ is such that $S \in \mathcal{F}$, $P(S) = 1$ and $\pi_k S \in \mathcal{B}_k$, then

$$\operatorname{cl} \Lambda_k^S(\xi_0, \ldots, \xi_k) = \Lambda_k^{\operatorname{supp} P}(\xi_0, \ldots, \xi_k) \quad \text{for } P\text{-a.e. } (\xi_0, \ldots, \xi_k) \in \pi_k S;$$

(ii) the set $\{(\xi_0,\ldots,\xi_k) \mid \Lambda_k^{\operatorname{supp} P}(\xi_0,\ldots,\xi_k) \cap V \neq \varnothing\}$ is open relative to $\pi_k(\operatorname{supp} P)$ for every open V.

Here, $\operatorname{supp} P$ denotes the *support*, of P which is defined as the intersection of all closed sets of full measure. It can be shown that, since Ξ has a countable base of open sets and P is a Borel measure, $\operatorname{supp} P$ is well defined and $P(\operatorname{supp} P) = 1$. As noted in Rockafellar and Wets [26, p. 840], a measure P is laminary in particular when it has a strictly positive density with respect to a product measure $P_0 \times \cdots \times P_K$, where P_k is a probability measure on Ξ_k.

As in Rockafellar and Wets [28], we will say that the problem ($SP(P)$) corresponding to (7) is *strictly feasible* if there is a bounded nonanticipative function $\tilde{x}$ and an $\varepsilon > 0$ such that

$$\tilde{x}(\xi) \in X \quad \text{and} \quad f_i(\xi, \tilde{x}(\xi)) \leq -\varepsilon \qquad \text{for } i = 1, \ldots, m \text{ and all } \xi \in \operatorname{supp} P.$$

It is clear that problem ($SP(P)$) is not affected if we replace Ξ by $\operatorname{supp} P$.

THEOREM 4.1. *Assume that f is of the form* (7), (a) *and* (b) *hold, P is laminary, $\Xi = \operatorname{supp} P$ is compact, and ($SP(P)$) is strictly feasible and Assumption 3.1 is satisfied. Then conditions* (1) *and* (2) *of Theorem 3.2 hold. Furthermore, the optimal values of ($SP(P^\nu)$) converge to that of ($SP(P)$), and if Ξ_0 is a singleton and z^ν is an ϵ^ν-optimal solution of ($SP(P^\nu)$), then all cluster points of $\{z_0^\nu\}_{\nu=1}^\infty$ are optimal first-stage solutions of ($SP(P)$).*

PROOF. Since f_i are continuous and $X \times \Xi$ is compact, the function f is lsc and bounded from below. The lower boundedness implies the lower compactness property of f (see, e.g., Ioffe [13, Proposition 1]), so the first condition holds.

To verify the second condition, let $x \in \operatorname{dom} F$. According to Rockafellar and Wets [28, Proposition 1], there exists a continuous nonanticipative function y^μ which is feasible in ($SP(P)$) and has

$$P(A) < 1/\mu,$$

where $A = \{\xi \in \Xi \mid |x(\xi) - y^\mu(\xi)| > \delta\}$. It follows that

$$\begin{aligned} \|y^\mu - x\|_{L^1} &= \int_A |y^\mu(\xi) - x(\xi)| P(d\xi) + \int_{\Xi\setminus A} |y^\mu(\xi) - x(\xi)| P(d\xi) \\ &\leq \|y^\mu - x\|_{L^\infty} P(A) + \delta P(\Xi\setminus A) \leq [(\operatorname{diam} X) + 1]\,\delta. \end{aligned}$$

Looking into the proof of Rockafellar and Wets [28, Proposition 1] shows also that y^μ is the function $\tilde{x}$ constructed in Rockafellar and Wets [26, pp. 852–853], and it satisfies $F(y^\mu) < F(x) + 2\delta$. Furthermore, since y^μ is feasible,

$$f(y^\mu(\cdot), \cdot) = f_0(y^\mu(\cdot), \cdot) \quad P\text{-a.s.},$$

where the right-hand side is continuous and bounded by the continuity of y^μ and f_0 and the boundedness of X and Ξ. The first inequality in condition (2) thus holds as an equality by Lemma 4.1. We have thus shown that condition (2) holds, except that instead of weak*-convergence of y^μ to x, we have convergence in the L^1-norm. However, since y^μ are uniformly bounded, by boundedness of X, the L^1-convergence implies the weak*-convergence, as in the proof of Lemma 3.1.

To verify the last claim, it suffices, by Corollary 3.1, to note that since Ξ is compact, $E^P|\xi| \in \mathbb{R}$ holds automatically, and since X is compact, problems ($SP(P^\nu)$) have optimal solutions such that $\max_{i=1,\ldots,\nu} |z^{\nu,i}|$ remains bounded. □

The conditions in Theorem 4.1 hold in many situations arising in practice, but they are only sufficient for the conditions of Theorem 3.2 to hold. Indeed, they imply the existence of *continuous* functions that converge in the L^1-norm, while P-a.s. continuity and weak*-convergence suffice in Theorem 3.2. It would be interesting to explore whether such properties could be obtained for problem classes other than (7).

Acknowledgments. The author would like to thank professors Roger J-B Wets, Alexander Ioffe, and Michael Casey as well as Matti Lassas and Matti Koivu for their advice and fruitful discussions at various stages of this project. This work was supported by Finnish Academy under Contract 3385.

References

[1] Artstein, Z., R. J-B. Wets. 1994. Stability results for stochastic programs and sensors, allowing for discontinuous objective functions. *SIAM J. Optim.* **4**(3) 537–550.
[2] Artstein, Z., R. J-B. Wets. 1995. Consistency of minimizers and the SLLN for stochastic programs. *J. Convex Anal.* **2**(1–2) 1–17.
[3] Attouch, H. 1984. *Variational Convergence for Functions and Operators*. Pitman (Advanced Publishing Program), Boston, MA.
[4] Bertsekas, D. P., S. E. Shreve. 1978. *Stochastic Optimal Control: The Discrete Time Case*, Vol. 139, *Mathematics in Science and Engineering*. Academic Press Inc., New York.
[5] Birge, J. R., R. J-B. Wets. 1986. Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse. *Stochastic Programming* 84.I. *Math. Programming Stud.* **27** 54–102.
[6] Braides, Andrea. 2002. *Γ-Convergence for Beginners*, Vol. 22. *Oxford Lecture Series in Mathematics and Its Applications*. Oxford University Press, Oxford, U.K.
[7] Casey, M., S. Sen. 2005. The scenario generation algorithm for multistage stochastic linear programs. *Math. Oper. Res.* Forthcoming.
[8] Dupačová, J., R. J-B. Wets. 1988. Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *Ann. Statist.* **16**(4) 1517–1549.
[9] Fiedler, Olga, Werner Römisch. 1995. Stability in multistage stochastic programming. *Ann. Oper. Res.* **56** 79–93.
[10] Folland, G. B. 1984. *Real Analysis: Modern Techniques and Their Applications*, *Pure and Applied Mathematics*. John Wiley & Sons Inc., New York.
[11] Föllmer, H., A. Schied. 2002. *Stochastic Finance: An Introduction in Discrete Time*, Vol. 27, *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin.
[12] Frauendorfer, K. 1996. Barycentric scenario trees in convex multistage stochastic programming. *Math. Programming Ser. B* **75**(2) 277–293.
[13] Ioffe, A. D. 1977. On lower semicontinuity of integral functionals. I. *SIAM J. Control Optim.* **15**(4) 521–538.
[14] Kall, P., A. Ruszczyński, K. Frauendorfer. 1988. Approximation techniques in stochastic programming. Yu. Ermoliev, ed. *Numerical Techniques for Stochastic Optimization*, Vol. 10, *Springer Series in Computational Mathematics*. Springer, Berlin, Germany, 33–64.
[15] Lepp, R. 1990. Approximations to stochastic programs with complete recourse. *SIAM J. Control Optim.* **28**(2) 382–394.
[16] Lucchetti, R., R. J-B. Wets. 1993. Convergence of minima of integral functionals, with applications to optimal control and stochastic optimization. *Statist. Decisions* **11**(1) 69–84.
[17] Marti, Kurt, Peter Kall, eds. 1998. *Stochastic Programming Methods and Technical Applications*, Vol. 458, *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, Berlin.
[18] Olsen, P. 1976. Discretizations of multistage stochastic programming problems. *Math. Programming Stud.* **6** 111–124.
[19] Pennanen, T. 2005. Epi-convergent discretizations of multistage stochastic programs via integration quadratures. *Numerische Mathematik*. Forthcoming.
[20] Pennanen, T., M. Koivu. 2002. Integration quadratures in discretization of stochastic programs. *Stochastic Programming E-Print Series*.
[21] Pennanen, T., M. Koivu. 2003. Epi-convergent discretizations of stochastic programs via integration quadratures. *Numerische Mathematik*. Forthcoming.
[22] Piccinini, L. 1998. Convergence of nonmonotone sequence of sub-σ-fields and convergence of associated subspaces $L^p(\mathcal{B}_n)$ $(p \in [1, +\infty])$. *J. Math. Anal. Appl.* **225**(1) 73–90.
[23] Rao, M. M. 1981. *Foundations of Stochastic Analysis, Probability and Mathematical Statistics.* Academic Press Inc., New York.
[24] Robinson, S. M., R. J-B. Wets. 1987. Stability in two-stage stochastic programming. *SIAM J. Control Optim.* **25**(6) 1409–1416.
[25] Rockafellar, R. T. 1976. Integral functionals, normal integrands and measurable selections. *Nonlinear Operators and the Calculus of Variations, Lecture Notes in Mathematics*, Vol. 543. Springer, Berlin, 157–207.
[26] Rockafellar, R. T., R. J-B. Wets. 1974. Continuous versus measurable recourse in N-stage stochastic programming. *J. Math. Anal. Appl.* **48** 836–859.
[27] Rockafellar, R. T., R. J-B. Wets. 1976. Nonanticipativity and L^1-martingales in stochastic optimization problems. *Math. Programming Stud.* **6** 170–187.

[28] Rockafellar, R. T., R. J-B. Wets. 1977. Measures as Lagrange multipliers in multistage stochastic programming. *J. Math. Anal. Appl.* **60**(2) 301–313.

[29] Rockafellar, R. T., R. J-B. Wets. 1978. The optimal recourse problem in discrete time: L^1-multipliers for inequality constraints. *SIAM J. Control Optim.* **16**(1) 16–36.

[30] Rockafellar, R. T., R. J-B. Wets. 1998. *Variational Analysis*, Vol. 317, *Grundlehren der Mathematischen Wissenschaften* (*Fundamental Principles of Mathematical Sciences*), Springer-Verlag, Berlin, Germany.

[31] Rudin, Walter. 1973. *Functional Analysis*. McGraw-Hill Series in Higher Mathematics, McGraw-Hill Book Co., New York.

[32] Ruszczynski, A., A. Shapiro, eds. 2003. *Stochastic Programming*, Vol. 10. *Handbooks in Operations Research and Management Science*, Elsevier, Amsterdam, The Netherlands.

[33] Schultz, R. 2001. Some aspects of stability in stochastic programming. *Ann. Oper. Res.* **100** 55–84.

[34] Shiryaev, Albert N. 1999. *Essentials of Stochastic Finance. Advanced Series on Statistical Science & Applied Probability*, Vol. 3. World Scientific Publishing Co. Inc., River Edge, NJ. Translated from the Russian manuscript by N. Kruzhilin.

[35] Vainikko, G. M. 1971. The convergence of the method of mechanical quadratures for integral equations with discontinuous kernels. *Siberian Math. J.* **12** 29–38.

[36] van der Vaart, Aad W., Jon A. Wellner. 1996. *Weak Convergence and Empirical Processes*. Springer Series in Statistics, Springer-Verlag, New York.

[37] Vogel, Silvia, Petr Lachout. 2003. On continuous convergence and epi-convergence of random functions. II. Sufficient conditions and applications. *Kybernetika* (*Prague*) **39**(1) 99–118.

[38] Wang, Jin De. 1995. Stability of multistage stochastic programming. *Ann. Oper. Res.* **56** 313–322.

[39] Zervos, M. 1999. On the epiconvergence of stochastic optimization problems. *Math. Oper. Res.* **24**(2) 495–508.

[40] Ziemba, W. T., J. M. Mulvey, eds. 1998. *Worldwide Asset and Liability Modeling*, Vol. 10, *Publications of the Newton Institute*. Cambridge University Press, Cambridge, UK.