Reflections on Output Analysis for Multistage Stochastic Linear Programs

Jitka Dupačová

Charles University Prague, Department of Probability and Mathematical Statistics, Sokolovská 83, CZ-186 75 Prague *dupacova@karlin.mff.cuni.cz*

Abstract. When solving a dynamic decision problem under uncertainty it is essential to choose or to build a suitable model taking into account the nature of the real-life problem, the character and availability of the input data, etc. There exist hints when to use stochastic dynamic programming models or multiperiod and multistage stochastic programs. Still, it is difficult to provide a general recipe. We refer to recent papers [1, 15] which characterize the main features and basic requirements of these models and indicate the cases which allow for multimodeling and comparisons or for exploitation of different approaches within one decision problem.

For both approaches, solution procedures are mostly based on an approximation scheme and it is important to relate the optimal value and optimal solutions of an approximating problem and the underlying one. It is interesting to recognize that methods of output analysis for stochastic dynamic programs were developed already in the eighties, cf. [25] and references ibidem. Regarding the solution method – the backward recursion connected with the principle of optimality – special emphasis was put on properties of discretization of state and control spaces.

We shall focus on multistage stochastic linear programs with recourse and with already given horizon and stages, that result by approximation of the underlying probability distribution. It turns out that generalization of various results well-known for two-stage stochastic linear programs to the multistage problems is not straightforward and it requires various additional assumptions, e.g., the interstage independence. We shall discuss possible generalizations of output analysis methods as delineated in [10].

1 Preliminaries

Let us consider a frequent framework for stochastic programs with recourse

$$\text{minimize } F(\boldsymbol{x}, P) := E_P f_0(\boldsymbol{x}, \omega) \text{ on a set } \mathcal{X} \tag{1}$$

where P is a *known probability distribution* on $(\Omega, \mathcal{B})$, $\Omega \subset R^m$ *which does not depend on* $\boldsymbol{x}$, E_P is the corresponding expectation operator, $\mathcal{X} \subset R^n$ is a nonempty closed set which does not depend on P and the random objective f_0 is a (usually quite complicated) function $f_0 : \mathcal{X} \times \Omega \rightarrow R^1$. For instance, the values $f_0(\boldsymbol{x}, \omega)$ in two-stage stochastic programs are obtained as optimal values of certain second-stage mathematical programs whereas for T-stage stochastic program, $f_0(\boldsymbol{x}, \omega)$ is an optimal value of a $T-1$-stage stochastic program.

We refer to the objective function in (1) as the *expectation functional* and to (1) as the *expectation-type stochastic program*; its objective function $F(\boldsymbol{x}, P)$ is linear in P. We will assume for simplicity that *all infima are attained,* and can be thus replaced by minima, and that *all expectations exist.* We denote

– $\varphi(P)$ the optimal value of (1),
– $\mathcal{X}^*(P)$ the set of optimal solutions of (1), not necessarily a singleton,
– $\boldsymbol{x}^*(P)$ the unique optimal solution of (1) in case $\mathcal{X}^*(P)$ is a singleton.

Because of incomplete information and also for the sake of numerical tractability one mostly solves an approximating, scenario-based stochastic program instead of the underlying "true" decision problem. However, the obtained output (the optimal value and optimal solutions of the approximating stochastic program) should be used to replace the sought solution of the "true" problem only after a careful analysis. An expert may create sensible scenarios and scenario trees relying on his/her experience and belief, however, methods of output analysis have to be tailored to the structure of the problem and they should also reflect the source, character and precision of the input data. We may compare scenario-generation to estimation and the output analysis to hypotheses testing. Methods of output analysis, cf. [10], will be surveyed from the point of view of their applicability to multistage stochastic linear programs with recourse.

2 Multistage stochastic linear programs with recourse

In the general *T-stage stochastic program* we think of a stochastic data process

$$\omega = (\omega_1, \ldots, \omega_{T-1}) \quad \text{or} \quad \omega = (\omega_1, \ldots, \omega_T)$$

whose realizations are (multidimensional) data trajectories and of a vector decision process

$$\boldsymbol{x} = (\boldsymbol{x}_1, \ldots \boldsymbol{x}_T),$$

a measurable function of ω. The sequence of decisions and observations is

$$\boldsymbol{x}_1, \omega_1, \boldsymbol{x}_2(\boldsymbol{x}_1, \omega_1), \omega_2, \ldots, \boldsymbol{x}_T(\boldsymbol{x}_1, \omega_1, \ldots, \omega_{T-1}). \tag{2}$$

Realizations of ω_T, i.e., those behind the horizon, do not affect the decision process, they may however contribute to the overall costs. Thus, the decision process may be affected by the *probability distribution* of ω_T. The decision process is *nonanticipative* in that sense that a sequence of decisions is built along each of the considered data trajectories in such a way that decisions based on the same part of trajectory, on the same history, are identical. It means that decisions taken at any stage of the process do neither depend on future *realizations* of random elements nor on future decisions, whereas the past information as well as the knowledge of the probability distribution P of ω are exploited. We denote by $\omega^{t-1,\bullet} := (\omega_1, \ldots, \omega_{t-1})$ the part of the stochastic data process that precedes the stage t and, similarly, by $\boldsymbol{x}^{t-1,\bullet} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1})$ the sequence of decisions at stages $1, \ldots, t-1$. Thus the decision at stage t is $\boldsymbol{x}_t = \boldsymbol{x}_t(\boldsymbol{x}^{t-1,\bullet}, \omega^{t-1,\bullet})$, or more precisely, $\boldsymbol{x}_t = \boldsymbol{x}_t(\boldsymbol{x}^{t-1,\bullet}, \omega^{t-1,\bullet}, P)$. We denote $P_t = P(\omega_t)$, $t = 1, \ldots, T-1$, the marginal probability distributions, $\mathcal{S}_t$, $t = 1, \ldots, T-1$, their supports and $P_t(\omega^{t-1,\bullet}) = P(\omega_t | \omega^{t-1,\bullet})$, $t = 2, \ldots, T-1$, the conditional probability distributions, with supports $\mathcal{S}_t(\omega^{t-1,\bullet})$.

The first-stage decisions consist of all decisions that have to be selected before further information is revealed whereas the second-stage decisions are allowed to adapt to this information, etc. In each of the stages, the decision is limited by constraints that may depend on the previous decisions and observations. *Stages do not necessarily refer to time periods, they correspond to steps in the decision process.*

Consider now the following T-stage stochastic linear program

Minimize

$$\boldsymbol{c}_1^\top \boldsymbol{x}_1 + E_{P_1}\left\{\varphi_1(\boldsymbol{x}_1, \omega_1)\right\} \tag{3}$$

subject to

$$\boldsymbol{A}_1 \boldsymbol{x}_1 = \boldsymbol{b}_1$$

$$\boldsymbol{l}_1 \leq \ \boldsymbol{x}_1 \leq \ \boldsymbol{u}_1,$$

where the functions $\varphi_{t-1}, t = 2, \ldots, T$, are defined recursively as

$\varphi_{t-1}(\boldsymbol{x}^{t-1,\bullet}, \omega^{t-1,\bullet}) =$

$$\inf_{\boldsymbol{x}_t} \left[\boldsymbol{c}_t(\omega^{t-1,\bullet})^\top \boldsymbol{x}_t + E_{P_t(\omega^{t-1,\bullet})}\left\{\varphi_t(\boldsymbol{x}^{t-1,\bullet}, \boldsymbol{x}_t, \omega^{t-1,\bullet}, \omega_t)\right\}\right] \tag{4}$$

subject to

$$\sum_{\tau=1}^{t-1} \boldsymbol{B}_{t\tau}(\omega^{t-1,\bullet})\, \boldsymbol{x}_\tau + \boldsymbol{A}_t(\omega^{t-1,\bullet})\, \boldsymbol{x}_t = \boldsymbol{b}_t(\omega^{t-1,\bullet}),$$

$$\boldsymbol{l}_t(\omega^{t-1,\bullet}) \leq \boldsymbol{x}_t \leq \boldsymbol{u}_t(\omega^{t-1,\bullet})$$

and $\varphi_T \equiv \ 0$ or it is an explicitly given function of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T, \omega_1, \ldots, \omega_T$ if contribution of ω_T is taken into account. Constraints involving random elements hold almost surely.

We assume that $\boldsymbol{A}_t$ are (m_t, n_t) matrices and that the remaining vectors and matrices are of consistent dimensions. To simplify the exposition we shall work with problems (4) which have the *staircase structure*, i.e., with $\boldsymbol{B}_{t\tau} \equiv 0$ for $\tau < t-1$ and we put $\boldsymbol{B}_{t,t-1} = \boldsymbol{B}_t$; the generalization to the general case is straightforward. For the first stage, known values of all elements of $\boldsymbol{b}_1, \boldsymbol{c}_1, \boldsymbol{A}_1, \boldsymbol{l}_1, \boldsymbol{u}_1$ are assumed. According to our assumption, an optimal solution of (4) exists for all t and all considered histories $\boldsymbol{x}^{t-1,\bullet}, \omega^{t-1,\bullet}$ – the case of the *relatively complete recourse.* In the case of the *fixed* relatively complete recourse, matrices $\boldsymbol{A}_t\ \forall t$ do not have any random elements.

The main decision variable is $\boldsymbol{x}_1$ that corresponds to the first stage and the first-stage problem (3) may be related to the general expectation -type stochastic program (1).

Many early papers on multistage stochastic linear programming with recourse were devoted to description and analysis of the corresponding expectation-type problem (1). The motivation came from the already known results for two-stage stochastic linear programs summarized, e.g. in [2]. The key questions were the description of the set $\mathcal{X}$ on which the expectation functional $F(\boldsymbol{x}, P)$ is finite, properties of the random objective $f_0(\boldsymbol{x}, \omega)$ and properties of its expectation $F(\boldsymbol{x}, P)$, such as convexity. This was done under various assumptions about the structure of the problem and properties of ω. For instance, convexity of the resulting deterministic program was proved already in [37] for problems involving only interstage independent random right-hand sides, [38] extends these convexity results to fixed recourse problems and [29] allows for interstage dependence of right-hand sides and coefficients of the objective function. Of course, under our simplifying assumption that all expectations exist and all minima are attained, such results are quite straightforward. A special result applies to discrete probability distributions P concentrated on a finite number of atoms. In this case, the set $\mathcal{X}$ is convex polyhedral and the expectation functional $F(\boldsymbol{x}, P)$ is convex piecewise linear; cf. [38]. These results for multistage stochastic linear programs with recourse and further references may be found in survey papers, e.g. [6] and books [2, 14].

For purposes of applications one mostly approximates the true probability distribution P of ω by a discrete probability distribution concentrated on a finite number of atoms $\omega^1, \dots, \omega^S$, which may be done, e.g., by sampling or by discretization. Accordingly, the supports of conditional probability distributions of ω_t conditioned by past realizations of $\omega_1, \dots \omega_{t-1}$ and the supports of marginal probability distributions of the components $\omega_t\ \forall t$ are finite sets. For disjoint sets of indices $\mathcal{K}_t,\ t = 2, \dots, T$, let us list as $\tilde{\omega}_{k_t},\ k_t \in \mathcal{K}_t$ all possible realizations of $\omega^{t-1,\bullet}$ and denote by the same subscripts the corresponding values of the t-stage coefficients. The total number of scenarios S equals the number of elements of $\mathcal{K}_T$. Each scenario $\omega^s = \{\omega_1^s, \dots, \omega_{T-1}^s\}$ thus generates a sequence of coefficients $\{\boldsymbol{c}_{k_2}, \dots, \boldsymbol{c}_{k_T}\}$, $\{\boldsymbol{A}_{k_2}, \dots, \boldsymbol{A}_{k_T}\}$, $\{\boldsymbol{B}_{k_2}, \dots, \boldsymbol{B}_{k_T}\}$, $\{\boldsymbol{b}_{k_2}, \dots, \boldsymbol{b}_{k_T}\}$, $\{\boldsymbol{l}_{k_2}, \dots, \boldsymbol{l}_{k_T}\}$, $\{\boldsymbol{u}_{k_2}, \dots, \boldsymbol{u}_{k_T}\}$. A specific organization of data in the form of the scenario tree means that each value $\tilde{\omega}_{k_{t+1}}$ of $\omega^{t\bullet}$ has a unique ancestor $\tilde{\omega}_{k_t}$ (the value of the corresponding $\omega^{t-1,\bullet}$); we denote it by subscript $a(k_{t+1})$. This allows to rewrite the *T-stage stochastic linear program with recourse in the following arborescent form*:

Minimize

$$\boldsymbol{c}_1^\top \boldsymbol{x}_1 + \sum_{k_2 \in \mathcal{K}_2} p_{k_2} \boldsymbol{c}_{k_2}^\top \boldsymbol{x}_{k_2} + \sum_{k_3 \in \mathcal{K}_3} p_{k_3} \boldsymbol{c}_{k_3}^\top \boldsymbol{x}_{k_3} + \ldots + \sum_{k_T \in \mathcal{K}_T} p_{k_T} \boldsymbol{c}_{k_T}^\top \boldsymbol{x}_{k_T} \quad (5)$$

subject to

$$\begin{array}{llllll}
\boldsymbol{A}_1 \boldsymbol{x}_1 & & & & = \boldsymbol{b}_1 & \\
\boldsymbol{B}_{k_2} \boldsymbol{x}_1 + & \boldsymbol{A}_{k_2} \boldsymbol{x}_{k_2} & & & = \boldsymbol{b}_{k_2}, & k_2 \in \mathcal{K}_2 \\
 & \boldsymbol{B}_{k_3} \boldsymbol{x}_{a(k_3)} & + \boldsymbol{A}_{k_3} \boldsymbol{x}_{k_3} & & = \boldsymbol{b}_{k_3}, & k_3 \in \mathcal{K}_3 \\
 & \ddots & \ddots & & \vdots & \\
 & & \boldsymbol{B}_{k_T} \boldsymbol{x}_{a(k_T)} & + \boldsymbol{A}_{k_T} \boldsymbol{x}_{k_T} & = \boldsymbol{b}_{k_T}, & k_T \in \mathcal{K}_T
\end{array}$$

$$\boldsymbol{l}_1 \le \boldsymbol{x}_1 \le \boldsymbol{u}_1, \, \boldsymbol{l}_{k_t} \le \boldsymbol{x}_{k_t} \le \boldsymbol{u}_{k_t}, \qquad k_t \in \mathcal{K}_t, \, t = 2, \ldots, T. \quad (6)$$

We adopt the natural choice $\mathcal{K}_t = \{K_{t-1} + 1, \ldots, K_t\}$, $t = 2, \ldots T$, with $K_1 = 1$. The problem is thus based on $S = K_T - K_{T-1}$ scenarios ω^s which generate sequences $(\boldsymbol{c}_{k_t}, \boldsymbol{A}_{k_t}, \boldsymbol{B}_{k_t}, \boldsymbol{b}_{k_t}, \boldsymbol{l}_{k_t}, \boldsymbol{u}_{k_t}, \, t = 2, \ldots T)$ of realizations of coefficients for all stages, and on *path probabilities* $p_{k_t} > 0 \, \forall k_t$, $\sum_{k_t \in \mathcal{K}_t} p_{k_t} = 1$, $t = 2, \ldots T$, of partial sequences of these coefficients, hence, probabilities of realizations of $\omega^{t-1,\bullet} \, \forall t$. The path probabilities p_{k_t} for $t > 2$ may be obtained by stepwise multiplication of the marginal probabilities p_{k_2} by the conditional arc (transition) probabilities, say, $\pi_{k_{\tau-1} k_\tau}$, $\tau = 3, \ldots, t$. Probabilities p^s of individual scenarios ω^s are equal to the corresponding path probabilities p_{k_T}.

The nonanticipativity constraints are included in an implicit form. Decomposition of (5)–(6) along scenarios is possible but it requires that the nonanticipativity constraints are spelled out in an explicit way. Given scenario ω^s denote by $\boldsymbol{c}(\omega^s)$ the vector composed of all corresponding coefficients, say, $\boldsymbol{c}_1, \boldsymbol{c}_{k_t}, t = 2, \ldots, T$, in the objective function (5), by $\boldsymbol{A}(\omega^s)$ the matrix of all coefficients of system of constraints (6) for scenario ω^s, and, similarly, by $\boldsymbol{b}(\omega^s)$, $\boldsymbol{l}(\omega^s)$, $\boldsymbol{u}(\omega^s)$ the vectors composed of right-hand sides in (6) and bounds of the box constraints for scenario ω^s. Disregarding the nonanticipativity constraints we replace the multistage stochastic linear program (5)–(6) by

$$\text{minimize} \sum_{s=1}^{S} p^s \boldsymbol{c}(\omega^s)^\top \boldsymbol{x}(\omega^s) \quad (7)$$

subject to

$$\boldsymbol{A}(\omega^s) \boldsymbol{x}(\omega^s) = \boldsymbol{b}(\omega^s), \; s = 1, \ldots S$$

and the box constraints

$$\boldsymbol{l}(\omega^s) \le \boldsymbol{x}(\omega^s) \le \boldsymbol{u}(\omega^s), \; s = 1, \ldots, S.$$

This is already an ordinary large linear program. The components of its optimal solutions, say, $\boldsymbol{x}^*(\omega^s), s = 1, \ldots, S$, depend on the underlying scenarios ω^s, they are not nonanticipative. To recover nonanticipativity, we must add constraints $\boldsymbol{x}^*(\omega^s) = \boldsymbol{x}^*(\omega^{s'})\, \forall s, s'$ to get scenario independent first-stage decisions and, moreover, similar constraints to guarantee that the t-stage decisions based on the same history are equal.

Besides the formulation of goals and constraints and identification of the driving random process ω, building a scenario-based multistage stochastic program requires specification of the horizon and stages, cf. [1, 11] and generation of the input in the form of scenario tree; see [12] and references ibidem, such as [3, 19, 21, 30]. Contrary to stochastic dynamic programs with discrete time, the number of stages is relatively small. On the other hand, formulations of the scenario-based multistage stochastic programming problems are not connected with any prescribed solution technique and it is possible to avoid special requirements such as the Markov structure of the problem. However, possibilities of drawing conclusions about the optimal value $\varphi(P)$ and the optimal solutions of the "true" stochastic program (1) using the results of the approximating scenario-based program depend essentially on the structure of the solved problem as well as on the origin of scenarios. Generally speaking, the output can hardly be more precise than the input and it is easier to answer questions concerning precision of the obtained optimal values than those concerning the sets of optimal solutions. There are results valid for two-stage stochastic programs that do not carry over to the multistage models, and we shall try to detect some of them. As we shall see, interstage independence or Markov property of the process ω play an important role in generalizations of output analysis results.

3 Output analysis

We accept that the true probability distribution P has been replaced by another probability distribution $\hat{P}$ obtained by using a simplified theoretical model and/or by sampling, discretization and simulation techniques. Both for interpretation of results and for designing numerical methods one should answer questions about the difference of the optimal values $\varphi(P), \varphi(\hat{P})$ and about the distance of the sets of optimal solutions $\mathcal{X}^*(P)$ and $\mathcal{X}^*(\hat{P})$. Regarding the origin of the approximating probability distribution one may rely on results of asymptotic statistics or exploit techniques known from parametric optimization.

3.1 Asymptotic Inference

Assume that the true probability distribution P in (1) can be well approximated by an infinite sequence of probability distributions based on an increasing level of information about P. This can be modeled in the following way: Consider a *sample space* $(Z, \mathcal{F}, \mu)$ with an increasing sequence of σ-fields $(\mathcal{F}^\nu)_{\nu=1}^\infty$ contained in $\mathcal{F}$. A sample ζ leads to a sequence of $\mathcal{F}^\nu$-measurable probability distributions $\{P^\nu(\bullet, \zeta), \nu = 1, 2, \ldots\}$ on $(\Omega, \mathcal{B})$ that are based on the information collected up to ν. The optimal value $\varphi(P^\nu)$ and the optimal solutions of the approximating stochastic program

$$\min_{\boldsymbol{x} \in \mathcal{X}} F(\boldsymbol{x}, P^\nu) = \min_{\boldsymbol{x} \in \mathcal{X}} E_{P^\nu} f_0(\boldsymbol{x}, \omega) \tag{8}$$

based on $P^\nu(\bullet, \zeta)$ depend on the used sample path ζ and the best one can get are results valid for *almost all sample paths* ζ, i.e., μ-a.s. The probability distributions P^ν will be called the *wide-sense empirical* probability distributions and the same designation will be used also for the approximating stochastic programs (8), their optimal values and optimal solutions. This helps to distinguish among general asymptotic results and those valid for the empirical probability distributions. In the latter case, the sample path $\zeta = \{\omega^1, \omega^2, \ldots\}$ is obtained by simple random sampling from $(\Omega, \mathcal{B}, P)$, $\mu = P^\infty$ and the *empirical stochastic program,* called also the *sample average approximation problem* is

$$\min_{\boldsymbol{x} \in \mathcal{X}} F(\boldsymbol{x}, P^\nu) = \min_{\boldsymbol{x} \in \mathcal{X}} \frac{1}{\nu} \sum_{j=1}^{\nu} f_0(\boldsymbol{x}, \omega^j). \tag{9}$$

Consistency Results. Under the assumption that $P^\nu \to P$ weakly (μ-a.s.) and that $f_0(\boldsymbol{x}, \bullet)$ is a continuous bounded function of ω for every $\boldsymbol{x} \in \mathcal{X}$, the pointwise convergence of the expected value objectives $F(\boldsymbol{x}, P^\nu) \to F(\boldsymbol{x}, P)$ μ-a.s. $\forall \boldsymbol{x} \in \mathcal{X}$ follows directly from the definition of weak convergence. If $\mathcal{X}$ is *compact* and the convergence of the expectations is *uniform* on $\mathcal{X}$ we get immediately (μ-a.s.) convergence of the optimal values

$$\varphi(P^\nu) \to \varphi(P).$$

If, moreover, $\mathcal{X}$ is *convex* and $f_0(\bullet, \omega)$ is *strictly convex* on $\mathcal{X}$ it is easy to get (μ-a.s.) convergence of the (unique) optimal solutions $\boldsymbol{x}^*(P^\nu)$ of $\min_{\boldsymbol{x} \in \mathcal{X}} F(\boldsymbol{x}, P^\nu)$ to the unique optimal solution $\boldsymbol{x}^*(P)$ of the underlying problem (1) and some rates of the convergence. Notice that merely the pointwise convergence of the empirical expectations does not imply consistency of the optimal values.

A more general approach is via epi-convergence of lower semicontinuous (lsc) functions, cf. [16]. The main step is to prove that the approximating objective functions $F(\boldsymbol{x}, P^\nu)$ epi-converge to the true objective function in (1), which in turn implies the following consistency result (cf. Theorem 3.9 of [16]).

Proposition 1. *Assume that $F(\boldsymbol{x}, P)$ is the μ-a.s. epi-limit of $F(\boldsymbol{x}, P^\nu)$ for $\nu \to \infty$. Then μ-a.s.*

$$\limsup_{\nu \to \infty} \varphi(P^\nu) \le \varphi(P)$$

and any cluster point $\hat{\boldsymbol{x}}$ *of any sequence* $\{\boldsymbol{x}^\nu, \nu = 1, 2, \ldots\}$ *with* $\boldsymbol{x}^\nu \in \mathcal{X}^*(P^\nu)$ *belongs* μ*-a.s. to* $\mathcal{X}^*(P)$.

In particular, if there is a compact set $\mathcal{D} \subset R^n$ *such that* μ*-a.s.,* $\mathcal{X}^*(P^\nu) \cap \mathcal{D} \neq \emptyset$ *for* $\nu = 1, 2, \ldots$ *and* $\boldsymbol{x}^* \in \mathcal{X}^*(P) \cap \mathcal{D}$ *then there exists a measurable selection* $\boldsymbol{x}^\nu$ *of* $\mathcal{X}^*(P^\nu)$ *such that* $\boldsymbol{x}^* = \lim_{\nu\to\infty} \boldsymbol{x}^\nu$ *for* μ*-almost all* ζ *and also* $\varphi(P) = \lim_{\nu\to\infty} \varphi(P^\nu)$ μ*-a.s.*

In the *convex* case, i.e. for convex function $f_0(\bullet, \omega)$, convex set $\mathcal{X}$, and for empirical probability distributions P^ν, epi-convergence of $F(\boldsymbol{x}, P^\nu)$ to $F(\boldsymbol{x}, P)$ follows from the Strong Law of Large Numbers for sums of random closed sets, see e.g. [24], and $F(\boldsymbol{x}, P^\nu)$ is an unbiased estimator of $F(\boldsymbol{x}, P)$ for any fixed $\boldsymbol{x} \in \mathcal{X}$. Moreover, for a discrete probability distribution P and for approximation of (1) by the empirical stochastic program (9) we have the following corollary.

Corollary 1. *Let* $\omega^1, \ldots, \omega^N$ *be the atoms of* P *and* $\pi_j > 0\, \forall j$, $\sum_{j=1}^N \pi_j = 1$ *their probabilities, let* $\mathcal{X}$ *be a nonempty bounded convex polyhedron and* $f_0(\boldsymbol{x}, \omega)$ *a piece-wise linear convex function of* $\boldsymbol{x}$ *on* $\mathcal{X}$*. Let* P^ν *be empirical distributions based on finite random samples of sizes* ν *from* P*. Assume in addition that there is a unique optimal solution* $\boldsymbol{x}^*(P)$ *of the true problem (1) and* $\boldsymbol{x}^\nu \in \mathcal{X}^*(P^\nu)$*. Then* μ*-a.s.*

$$\boldsymbol{x}^\nu \equiv \boldsymbol{x}^*(P)$$

for ν *large enough.*

This means that *the empirical problem provides* μ*-a.s. the exact optimal solution of the true problem for* ν *large enough.* This result can be extended to the case of multiple true optimal solutions, cf. [36], where it is also proved that for $\nu \to \infty$ the probability of the event $\boldsymbol{x}^\nu \equiv \boldsymbol{x}^*(P)$ approaches 1 exponentially fast. Nevertheless, to determine the sufficient sample size remains a crucial problem.

For empirical stochastic programs (9) it is easy to prove that the obtained optimal values $\varphi(P^\nu)$ have a one-directional bias in the sense that

$$E\varphi(P^\nu) \leq \varphi(P); \tag{10}$$

see, e.g. [26, 27] and this result extends to multistage stochastic programs, cf. [4]. An empirical point estimate of $E\varphi(P^\nu)$ follows by the Law of Large Numbers and an asymptotic $100(1-\alpha)\%$ confidence interval, say $[L^\nu_{1-\alpha}, U^\nu_{1-\alpha}]$, for $E\varphi(P^\nu)$ can be obtained from the Central Limit Theorem. For an arbitrary $\boldsymbol{x} \in \mathcal{X}$, $F(\boldsymbol{x}, P)$ is an upper bound for the true optimal value and the empirical expectation $F(\boldsymbol{x}, P^\nu)$ is an unbiased consistent point estimator of $F(\boldsymbol{x}, P)$. Again, one may construct an interval estimate or an asymptotic confidence interval covering this upper bound with a given probability. Exploitation of these properties of the optimal value $\varphi(P^\nu)$ of the empirical problem and of the upper bounds turned out to be very useful for numerical solution and valuation of results of empirical two-stage stochastic linear programs.

Theoretically, these consistency results apply also to *multistage* stochastic programs. However, the assumption of an infinitely increasing sample size means that at every node of the scenario tree, the number of branches grows to infinity, hence, the number of descending nodes grows to infinity, too, and the sample based problems become very quickly intractable. In addition, random sampling from a continuous probability distribution P provides almost surely different scenarios so that the empirical problem will not reflect the assumed tree structure. It means that the sample-based problem approximates the true stochastic program (7) with *relaxed* nonanticipativity constraints (except for the constraint on the scenario independent first-stage solutions). The theoretical optimal value of this relaxed program, say, $\varphi^R(P)$ is thus approximated by the biased estimator $\varphi^R(P^\nu)$ and inequalities

$$E\varphi^R(P^\nu) \leq \varphi^R(P) \leq \varphi(P)$$

may be used to construct an asymptotic *lower bound* for $\varphi(P)$ – the lower bound of the confidence interval for $E\varphi^R(P^\nu)$. However, $\varphi^R(P^\nu)$ is not a consistent estimate of the true value $\varphi(P)$. One can again think of an *upper bound* for $\varphi(P)$ obtained as the value of the objective function (3) at an arbitrary feasible first-stage solution; this means to evaluate or to estimate the expected value $E_{P_1}\varphi_1(\boldsymbol{x}_1, \omega_1)$. For $T > 2$, function $\varphi_1(\boldsymbol{x}_1, \omega_1)$ in (3) is the optimal value of a $T-1$-stage stochastic program, its empirical expectation is biased below and (3) *with the empirical expectation need not provide an upper bound for* $\varphi(P)$; see [35] for a detailed discussion. Hence, one should be cautious when using stopping rules based on objective function values for *empirical multistage* stochastic programs.

These are the reasons for using *conditional sampling schemes* which take into account the assumed dynamics of the problem, the given horizon and stages:

– Under interstage independence and for a given tree structure one may construct the tree by simple random sampling from marginal distributions of components ω_t of ω.
– The next step is Markov dependence, modeled for instance by

$$\omega_{t+1} = \boldsymbol{G}_t \omega_t + \varepsilon_t \tag{11}$$

with nonrandom transition matrices $\boldsymbol{G}_t$ and interstage independent random $\varepsilon_t \, \forall t$.
– The final choice is sampling from the *conditional* probability distributions $P_t(\omega^{t-1,\bullet})$ of ω_t given the history $\omega^{t-1,\bullet}$, see e.g. [3, 4].

Under mild additional assumptions, consistency of $\varphi(P^\nu)$ can be proved again. For example [35] applies the Law of Large Numbers for the case of a bounded set $\mathcal{X}^*(P)$ of optimal solutions and under further assumptions which guarantee uniform convergence of the empirical objective functions whereas [4] assumes strictly convex objective functions and exploits the discretization scheme of [28].

The related *asymptotic distributions and rates of convergence* require that the corresponding consistency results hold true. Under suitable assumptions (e.g., Theorem 3.3 of [34]) it is possible to extend the asymptotic normality results valid for the empirical optimal value $\varphi(P^\nu)$ to the multistage case, see also [4]. Their extension to rates of convergence for optimal first-stage solutions of multistage stochastic programs has not yet been satisfactorily explored.

3.2 Qualitative and Quantitative Stability Results

For various reasons, empirical estimates of the probability distribution P are not always available and, moreover, they need not provide the best approximation technique: They focus solely on the probability distribution, which is not the only ingredient of the stochastic programming models, they do not take into account any expert knowledge or foresight and for technical reasons, they cannot be based on very large samples. Moreover, the goal is to get a sensible approximation of the optimal solution and of the optimal value, not an approximation of the probability distribution.

We shall look now into stability analysis of (1) with respect to the parameter P under the *additional assumption that the set $\mathcal{X}$ in (1) is convex and that for all $\omega \in \Omega$, $f_0(\boldsymbol{x}, \omega)$ is a convex lower semicontinuous function of $\boldsymbol{x}$ on $\mathcal{X}$*. This implies that the objective function in (1),

$$F(\boldsymbol{x}, P) = E_P f_0(\boldsymbol{x}, \omega) = \int_\Omega f_0(\boldsymbol{x}, \omega) P(d\omega) \tag{12}$$

is convex in $\boldsymbol{x}$ on R^n for any probability measure P (on $(\Omega, \mathcal{B})$) such that the expectation (12) is finite.

For multistage stochastic linear programs with complete recourse a qualitative stability result can be obtained if the set $\mathcal{X}$ is convex polyhedral, the set $\mathcal{X}^*(P)$ of optimal solutions of the true problem is bounded and certain growth assumptions are fulfilled; see Corollary 3.3 in [18]. Namely, the persistence property $\mathcal{X}^*(\hat{P}) \neq \emptyset$ holds true on a neighborhood of P, the mapping $\mathcal{X}^*$ is weakly upper semicontinuous in P and the optimal value satisfies the Lipschitz condition

$$|\varphi(P) - \varphi(\hat{P})| \leq L_\varphi d(P, \hat{P}) \tag{13}$$

for all $\hat{P}$ belonging to a neighborhood of P.

Under suitable continuity assumptions as to the random objectives and constraints, a quantitative stability result akin to (13) can be proved also for *nonlinear* multistage stochastic programs with interstage Markov dependence when a special discretization scheme is applied to approximate the true absolutely continuous probability distribution P carried by a compact support; cf. [23].

The success and applicability of the *quantitative* stability results depend essentially on an appropriate choice of the distance d used to measure the perturbations in the model input. The structure of the convex program (1) suggests to consider a probability distance of the form

$$d_{\mathcal{F}}(P,\hat{P}) := \sup\{|\int_{\Omega} f(\omega)(P(d\omega) - \hat{P}(d\omega))| : f \in \mathcal{F}\} \tag{14}$$

where $\mathcal{F}$ is the class of all measurable functions from Ω to R^1 that may appear as integrands in (12). The probability distance $d_{\mathcal{F}}(P,\hat{P})$ is finite whenever P and $\hat{P}$ belong to the set

$$\mathcal{P}_{\mathcal{F}}(\Omega) := \{Q : \sup_{f\in\mathcal{F}} |\int_{\Omega} f(\omega)Q(d\omega)| < \infty\}$$

of probability distributions (on $(\Omega,\mathcal{B})$) satisfying a uniform moment condition with respect to $\mathcal{F}$. Now, (1) is regarded as a convex parametric program with parameter P belonging to the space $(\mathcal{P}_{\mathcal{F}}(\Omega), d_{\mathcal{F}})$.

An important example is the Fortet-Mourier metric ζ_p of order p defined by

$$\zeta_p(P,\hat{P}) := \sup_{f\in\mathcal{F}_p(\Omega)} |\int_{\Omega} f(\omega)(P(d\omega) - \hat{P}(d\omega))|$$

on the set of probability distributions which satisfy the moment condition $\int_{\Omega} \| \omega \|^p Q(d\omega) < \infty$ and for the set of integrands

$$\mathcal{F}_p(\Omega) := \left\{ f : |f(\omega) - f(\tilde{\omega})| \le \max\left\{1, \|\omega\|^{p-1}, \|\tilde{\omega}\|^{p-1}\right\} \|\omega - \tilde{\omega}\| \forall \omega, \tilde{\omega} \in \Omega \right\}.$$

Whereas ζ_2 turns out to be suitable for two-stage stochastic programs with complete recourse, for T-stage stochastic programs with fixed recourse it is necessary to use the ζ_T metric. An exception are multistage stochastic linear programs with only right-hand sides random; these behave stable with respect to the ζ_1 metric; cf. [32].

The following stability result is a consequence of a more general perturbation theorem in [31].

Proposition 2. *Assume in addition that $\mathcal{X}^*(P)$ is nonempty and bounded, $P \in \mathcal{P}_{\mathcal{F}}(\Omega)$ and the function $F(\boldsymbol{x}, P)$ is locally Lipschitzian on $\mathcal{X}$. Then the optimal solution set mapping $\mathcal{X}^*$ is (Berge) upper semicontinuous at P and there exist constants $L > 0$, $\delta > 0$ such that $\mathcal{X}^*(Q)$ is nonempty and $|\varphi(P) - \varphi(Q)| \le L d_{\mathcal{F}}(P,Q)$ whenever $Q \in \mathcal{P}_{\mathcal{F}}(\Omega)$ and $d_{\mathcal{F}}(P,Q) < \delta$.*

For discrete probability distributions, these stability results form a basis for scenario reduction techniques, cf. [13] for the two-stage problems. In the case of interstage independence, these scenario reduction techniques may be exploited for multistage stochastic programs in a straightforward way (reduction applies separately to the marginal probability distributions) and an extension under Markov dependence (11) is possible as well. Further generalizations are under way.

Also the *contamination method* may be included under quantitative stability approaches. In the convex case considered in this Section, results of [20] apply to stability and sensitivity of the optimal value with respect to contamination of the probability distribution P by another probability distribution Q also for multistage stochastic programs; cf. [8]. Without any specific requirements on the structure of the problem (1) or on the properties of the probability distributions P, Q, contamination allows to test the robustness of the optimal value with respect to an additional, richer branching for a given topology of stages or to inclusion of an additional branching point.

3.3 Moment Bounds and Worst Case Analysis

In the context of (1) with $F(\boldsymbol{x}, P) := E_P f_0(\boldsymbol{x}, \omega)$, one can try to construct *minmin and minmax bounds*

$$\min_{\boldsymbol{x} \in \mathcal{X}} \inf_{P \in \mathcal{P}} F(\boldsymbol{x}, P) \leq \varphi(P) \leq \min_{\boldsymbol{x} \in \mathcal{X}} \sup_{P \in \mathcal{P}} F(\boldsymbol{x}, P) \ \forall P \in \mathcal{P} \tag{15}$$

on the optimal value of the true program to get information about robustness of the optimal value within the considered family of probability distributions.

The objective functions $E_P f_0(\boldsymbol{x}, \omega)$ of the inner minimization and maximization problems

$$\inf_{P \in \mathcal{P}} E_P f_0(\boldsymbol{x}, \omega) \text{ and } \sup_{P \in \mathcal{P}} E_P f_0(\boldsymbol{x}, \omega)$$

are linear in P, which means that for a convex, compact set $\mathcal{P}$, both the infimum and supremum are attained and the optimal *best case* and *worst case* probability distributions P^*, $P^{**} \in \mathcal{P}$ are extremal points of $\mathcal{P}$.

In the framework of the moment problem, these extremal points are well described for $\mathcal{P}$ defined by a given support and by known values of certain generalized moments: For admissible moment values, the extremal distributions are discrete ones, concentrated in a modest number of points; hence, the bounds (15) follow by solution of a *scenario-based program*. However, extremal distributions *independent* of the decisions $\boldsymbol{x}$ appear only exceptionally, under special assumptions (e.g., convexity, concavity or saddle property) about the integrand $f_0(\boldsymbol{x}, \bullet)$ and about the families of distribution functions, e.g., for those with given support and expectations; see [5] for details and references.

A related, though less ambitious problem is to get bounds on the performance of an optimal solution $\boldsymbol{x}(P)$ obtained for a probability distribution $P \in \mathcal{P}$ using the corresponding worst case and best case probability distributions from $\mathcal{P}$. This leads to bounds which are then exploited in various computational schemes for two-stage stochastic programs, as initiated in [22]. Also here a tractable procedure for the (repeated) evaluation of bounds requires certain convexity properties of the function $f_0(\boldsymbol{x},\omega)$ with respect to ω and a special type of family $\mathcal{P}$.

For multistage stochastic programs, convexity or saddle property of $f_0(\boldsymbol{x},\bullet)$ depends not only on the structure of the stochastic program in question (recall the assumption of the fixed recourse needed for two-stage programs) but also upon special additional assumptions about the probability distribution of the stochastic data process: It can be generalized to fixed recourse and interstage independence, see e.g. Section 11.1 of [2] or [7]. In presence of interstage dependences, even with randomness entering just the right-hand sides, convexity of $f_0(\boldsymbol{x},\bullet)$ holds true only under special assumptions about their probability distributions.

Example. To illustrate the problem, consider the following three stage stochastic linear program with a staircase structure, relatively complete fixed recourse and random right-hand sides, written according to the scheme (3)–(4).

Minimize

$$\boldsymbol{c}_1^\top \boldsymbol{x}_1 + E_{P_1}\left\{\varphi_1(\boldsymbol{x}_1,\omega_1)\right\} \tag{16}$$

subject to

$$\boldsymbol{A}_1\boldsymbol{x}_1 = \boldsymbol{b}_1$$

$$\boldsymbol{l}_1 \le \ \boldsymbol{x}_1 \le \ \boldsymbol{u}_1,$$

with function φ_1 defined by

$$\varphi_1(\boldsymbol{x}_1,\omega_1) = \min_{\boldsymbol{x}_2}\left[\boldsymbol{c}_2^\top \boldsymbol{x}_2 + E_{P_2(\omega_1)}\varphi_2(\boldsymbol{x}_2,\omega_2)\right] \tag{17}$$

subject to

$$\boldsymbol{B}_2\boldsymbol{x}_1 + \boldsymbol{A}_2\boldsymbol{x}_2 = \boldsymbol{b}_2(\omega_1)$$

$$\boldsymbol{l}_2 \le \boldsymbol{x}_2 \le \boldsymbol{u}_2$$

and

$$\varphi_2(\boldsymbol{x}_2,\omega_2) = \inf_{\boldsymbol{x}_3} \boldsymbol{c}_3^\top \boldsymbol{x}_3 \tag{18}$$

subject to

$$\boldsymbol{B}_3\boldsymbol{x}_2 + \boldsymbol{A}_3\boldsymbol{x}_3 = \boldsymbol{b}_3(\omega_2)$$

$$\boldsymbol{l}_3 \le \boldsymbol{x}_3 \le \boldsymbol{u}_3$$

We assume that the right-hand sides are *linear* in ω_1 and in ω_2. We want to construct bounds for the optimal value $\varphi(P)$ using just the first order moment information about $\omega = (\omega_1, \omega_2)$ whose marginal probability distributions P_1, P_2 do not depend on $\boldsymbol{x}$, have known fixed supports and expectations. For simplicity, we list the assumptions and delineate the problems for the case of one-dimensional ω_1, ω_2.

Assumptions:

- Marginal probability distributions P_1, P_2 have known fixed supports $[\alpha_1, \beta_1]$, $[\alpha_2, \beta_2]$ and expectations $\overline{\omega}_1$, $\overline{\omega}_2$.
- For each fixed realization $\tilde{\omega}_1$ of ω_1, the conditional probability distribution $P_2(\tilde{\omega}_1)$ is carried by interval $[\alpha(\tilde{\omega}_1), \beta(\tilde{\omega}_1)]$ and the conditional expectation $\overline{\omega}_2(\tilde{\omega}_1) := E_{P_2(\tilde{\omega}_1)}\omega_2$ is its interior point.

Lower bound. Using the evident convexity of the function $\varphi_2(\boldsymbol{x}_2, \omega_2)$ with respect to ω_2 and the conditional Jensen inequality, one gets a lower bound, say, $\varphi_1^*(\boldsymbol{x}_1, \omega_1)$ for $\varphi_1(\boldsymbol{x}_1, \omega_1)$ replacing $E_{P_2(\omega_1)}\varphi_2(\boldsymbol{x}_2, \omega_2)$ in (17) by $\varphi_2(\boldsymbol{x}_2, \overline{\omega}_2(\omega_1))$. One more application of the same idea provides the lower bound as the optimal value of the deterministic expected value program. Hence, it turns out that *dependence of random right-hand sides is no obstacle for generalization of the Jensen lower bound for the optimal value of a multistage stochastic linear program with random right-hand sides.*

Upper bound. Concerning the upper bound, the situation is different; see [7]. The upper bound for $E_{P_2(\tilde{\omega}_1)}\varphi_2(\boldsymbol{x}_2, \omega_2)$ follows via the Edmundson-Madansky inequality, hence,

$$\varphi_1(\boldsymbol{x}_1, \tilde{\omega}_1) \le \varphi_1^{**}(\boldsymbol{x}_1, \tilde{\omega}_1)$$

where

$$\varphi_1^{**}(\boldsymbol{x}_1, \omega_1) = \min_{\boldsymbol{x}_2} \left[\boldsymbol{c}_2^\top \boldsymbol{x}_2 + \lambda_2(\tilde{\omega}_1)\boldsymbol{c}_3^\top \boldsymbol{x}_{31} + (1 - \lambda_2(\tilde{\omega}_1))\boldsymbol{c}_3^\top \boldsymbol{x}_{32}\right] \tag{19}$$

subject to

$$\begin{aligned} \boldsymbol{B}_2\, \boldsymbol{x}_1 + \boldsymbol{A}_2\, \boldsymbol{x}_2 &= \tilde{\omega}_1, \\ \boldsymbol{B}_3\, \boldsymbol{x}_2 + \boldsymbol{A}_3\, \boldsymbol{x}_{31} &= \alpha(\tilde{\omega}_1), \\ \boldsymbol{B}_3\, \boldsymbol{x}_2 + \boldsymbol{A}_3\, \boldsymbol{x}_{32} &= \beta(\tilde{\omega}_1), \\ \boldsymbol{l}_2 \le \boldsymbol{x}_2 \le \boldsymbol{u}_2,\ \boldsymbol{l}_3 &\le \boldsymbol{x}_{3i} \le \boldsymbol{u}_3,\ i = 1, 2 \end{aligned}$$

and with

$$\lambda(\tilde{\omega}_1) = \frac{\beta(\tilde{\omega}_1) - \overline{\omega}_2(\tilde{\omega}_1)}{\beta(\tilde{\omega}_1) - \alpha(\tilde{\omega}_1)}. \tag{20}$$

To proceed further to get the upper bound for $E_{P_1}\varphi^{**}(\boldsymbol{x}_1, \tilde{\omega}_1)$, convexity of $\varphi^{**}(\boldsymbol{x}_1, \tilde{\omega}_1)$ with respect to $\tilde{\omega}_1$ is essential. Denote $U_2(\boldsymbol{x}_2, \tilde{\omega}_1)$ the objective function in (19). To get convexity of the optimal value of (19) in ω_1, for the sake of subsequent use of the Edmundson-Madansky upper bound on its expectation, one needs $U_2(\boldsymbol{x}_2, \tilde{\omega}_1)$ *jointly convex* in $\boldsymbol{x}_2, \tilde{\omega}_1$. To this purpose, it is not enough to assume *linearity* of α and β in $\tilde{\omega}_1$ (recall the form of λ in (20)). A sensible *additional* assumption concerning the class of conditional probability distributions $P_2(\tilde{\omega}_1)$ is the Markov property (11), which in our simple example reduces to $\omega_2 = g\omega_1 + \varepsilon$ where ω_1 and ε are independent and $g \in R$ is fixed. Under this assumption, $\lambda(\tilde{\omega}_1)$ is independent of $\tilde{\omega}_1$, hence, $\varphi_1^{**}(\boldsymbol{x}_1, \tilde{\omega}_1)$ is convex in ω_1 and one more application of the Edmundson-Madansky inequality provides the upper bound for the optimal value $\varphi(P)$ for all probability distributions of the considered properties. In our simple example, the upper bound is equal to the optimal value of the approximating stochastic program based on four scenarios, namely, $[\alpha_1, \alpha_2(\alpha_1)], [\alpha_1, \beta_2(\alpha_1)], [\beta_1, \alpha_2(\beta_1)], [\beta_1, \beta_2(\beta_1)]$ with probabilities $\lambda_1\lambda_2,\ \lambda_1(1-\lambda_2), (1-\lambda_1)\lambda_2, (1-\lambda_1)(1-\lambda_2)$.

Of course, interstage independence is a special case of the Markov property (11) and the Markov property is fulfilled, e.g. for *multidimensional normal distribution* of $\omega = (\omega_1, \omega_2)$; however, this probability distribution is not carried by a compact support. This means that the important convexity (or saddle) property needed for construction of upper bounds for multistage stochastic programs via Edmundson-Madansky inequality is rarely fulfilled and *the applied upper bounds are not exact*. Some properties of these upper bounds are discussed in [33].

Let us summarize: In general, the upperbounding techniques based on the first order moment information carry over to multistage stochastic linear programs with relatively complete fixed recourse and with random right-hand sides, linear in random parameters ω, only in special cases, e.g., when *one of the following conditions holds true:*

1. The right-hand sides are interstage independent;
2. For all stages, the right-hand sides can be expressed in the form of a sum of interstage independent random vectors related to preceding stages and to the given stage, compare with (11);
3. For all stages, the conditional distributions of random parameters ω_t are carried by simplices whose extreme points are linear in past values $\omega_1, \ldots, \omega_{t-1}$ whereas the barycentric coordinates of the conditional mean values are fixed, independent of this history.

Example – continuation. Let us detail Case 3. Generalization to T-stage problem means to assume a fixed position of the conditional mean values $\overline{\omega}_t(\omega^{t-1,\bullet})$ (described by fixed values $\lambda_t \in (0,1)$) within intervals $[\alpha_t(\omega^{t-1,\bullet}), \beta_t(\omega^{t-1,\bullet})]$ whose endpoints are linear in $\omega_1, \ldots, \omega_{t-1}$. This type of assumptions can be used to model an increasing uncertainty by a growing range of the variables around some trend described by the conditional mean values. The upperbounding scenarios are sequences

$$\rho_1, \rho_2(\rho_1), \ldots, \rho_t(\rho_1, \rho_2(\rho_1), \ldots), \rho_{T-1}(\rho_1, \rho_2(\rho_1), \ldots)$$

with α_t or β_t substituted for ρ_t; compare with [17].

An extension to random vectors ω_t whose distributions are carried by simplices is also possible. Assumption of fixed values of λ_t independent of past observations translates to fixed barycentric coordinates of the conditional mean values $\overline{\omega}_t(\omega^{t-1,\bullet})$. The general bounding technique based on barycentric scenarios, see [19], follows, inter alia, from the *assumed convexity* or saddle property of the objective functions for all stages, e.g., convexity of the function $\varphi(\boldsymbol{x}_1, \omega_1)$ defined by (17). The same assumption is needed also for the multistage extension of the upperbounding technique in [17]. Our discussions imply that the convexity assumptions refer, besides the interstage independence of random right-hand sides, to a rather special form of interstage dependent right-hand sides so that the conditional distributions fulfil the special moment properties of Case 3 discussed above or possess a *Markov property*, e.g. (11). In such case, the random elements ω_t in stage t may be represented as a *sum of interstage independent random summands* related only to individual stages $1, \ldots, t$; see [9] for an application.

Acknowledgements. This work was partly supported by research project *Mathematical Methods in Stochastics* – MSM 113200008 and by the Grant Agency of the Czech Republic (grants 201/02/0621 and 402/02/1015). Helpful comments of Alex Shapiro are gratefully acknowledged.

References

1. Bertocchi, M., Dupačová, J., Moriggia, V. (2001): Horizon and stages in applications of stochastic programming in finance, submitted to *Ann. Oper. Res.*
2. Birge, J. R., Louveaux, F. (1997): Introduction to stochastic programming. Springer, New York
3. Chen, Z., Consigli, G., Dempster, M. A. H., Hicks-Pedrón, N. (1997): Towards sequential sampling algorithms for dynamic portfolio management. In: Zopounidis, C. (ed) New operational tools for the management of financial risks. Kluwer, Dordrecht, pp. 197–211
4. Dempster, M. A. H. (1998): Sequential importance sampling algorithms for dynamic stochastic programming. WP 32/98, Judge Inst. of Management Studies, University of Cambridge
5. Dupačová, J. (1987): The minimax approach to stochastic programming and an illustrative application. Stochastics, **20**, 73–88
6. Dupačová, J. (1995): Multistage stochastic programs: The state-of-the-art and selected bibliography. Kybernetika, **31**, 151–174
7. Dupačová, J. (1995): Bounds for stochastic programs in particular for recourse problems. WP-95-085, IIASA Laxenburg
8. Dupačová, J. (1995): Postoptimality for multistage stochastic linear programs. Ann. Oper. Res., **56**, 65–78

9. Dupačová, J. (1996): Uncertainty about input data in portfolio management. In: Bertocchi, M. et al. (eds) Modelling techniques for financial markets and bank management. Physica Verlag, Heidelberg, pp. 17–33
10. Dupačová, J. (2001): Output analysis for approximated stochastic programs. In: Uryasev, S., Pardalos, P. M. (eds) Stochastic optimization: Algorithms and applications. Kluwer, Dordrecht, pp. 1–29
11. Dupačová, J. (2002): Applications of stochastic programming: Achievements and questions. European J. Oper. Res., **140**, 281–290
12. Dupačová, J., Consigli, G., Wallace, S.W. (2000): Scenarios for multistage stochastic programs. Ann. Oper. Res., **100**, 25–53
13. Dupačová, J., Gröwe-Kuska, N., Römisch, W. (2003): Scenario reduction in stochastic programming: An approach using probability metrics. Math. Progr. A, **95**, 493–511
14. Dupačová, J., Hurt, J., Štěpán, J. (2002) Stochastic modeling in economics and finance, Part II. Kluwer, Dordrecht
15. Dupačová, J., Sladký, K. (2002) Comparison of multistage stochastic programs with recourse and stochastic dynamic programs with discrete time. ZAMM, **82**, 753–765
16. Dupačová, J., Wets, R. J.-B. (1988): Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. Ann. Statist., **16**, 1517–1549
17. Edirisinghe, N. C. P., Ziemba, W. T. (1992): Tight bounds for stochastic convex programs. Oper. Res., **40**, 660–677
18. Fiedler, O., Römisch, W. (1995): Stability in multistage stochastic programming. Ann. Oper. Res., **56**, 79–93
19. Frauendorfer, K. (1996): Barycentric scenario trees in convex multistage stochastic programming. Math. Progr., **75**, 277–293
20. Gol'shtein, E. G. (1970): Vypukloje programmirovanije. Elementy teoriji. Nauka, Moscow. [Theory of convex programming, Translations of Mathematical Monographs, **36**, American Mathematical Society, Providence RI, 1972].
21. Høyland, K., Wallace, S. W. (2001): Generating scenario trees for multistage decision problems. Manag. Sci., **47**, 295–307
22. Kall, P., Stoyan, D. (1982): Solving stochastic programming problems with recourse including error bounds. Math. Operationsforschung Statist. Ser. Optimization, **13**, 431–447.
23. Kaňková, V. (2000): Multistage stochastic programming; Stability, approximation and Markov dependence. In: Inderfurth, K. et al. (eds) Operations research proceedings 1999, Springer, Berlin, pp. 136–141
24. King, A. J., Wets, R. J.-B. (1990): Epi-consistency of convex stochastic programs. Stochastics and Stochastics Reports, **34**, 83–92
25. Langen, H-J. (1981): Convergence of dynamic programming models. Mathematics of Oper. Res., **6**, 493–512
26. Mak Wai-Kei, Morton, D. P., Wood, R. K. (1999): Monte Carlo bounding techniques for determining solution quality in stochastic programs. OR Letters, **24**, 47–56
27. Norkin, V. I., Pflug, G. Ch., Ruszcynski, A. (1998): A branch and bound method for stochastic global optimization. Math. Progr., **83**, 425–450
28. Olsen, P. (1976): Discretization of multistage stochastic programming problems. Math. Programming Study, **6**, 111–124

29. Olsen, P. (1976): Multistage stochastic programming with recourse: The equivalent deterministic problem. SIAM J. Control Optim., **14**, 495–517
30. Pflug, G. Ch. (2001): Scenario tree generation for multiperiod financial optimization by optimal discretization. Math. Progr. B, **89**, 251–271
31. Rachev, S. T., Römisch, W. (2002): Quantitative stability of stochastic programs: The approach via probability metrics. Mathematics of Operations Research **27**, 792–818
32. Römisch, W., Wets, R. J.-B. (2001): Stability of ϵ-solutions to multistage stochastic programs. Manuscript, Inst. of Mathematics, Humboldt University, Berlin
33. Schürle, M. (1998): Zinsmodelle in der stochastischen Optimierung mit Anwendungen im Asset- & Liability-Management. Bank und finanzwirtschaftliche Forschungen **279**, P. Haupt Verlag, Bern
34. Shapiro, A. (1991): Asymptotic analysis of stochastic programs. Ann. Oper. Res., **30**, 169–186
35. Shapiro, A. (2002): Statistical inference of multistage stochastic programming problems. Preprint, Georgia Inst. of Technology, Atlanta
36. Shapiro, A., Homem-de-Mello, T. (2000): On rate of convergence of Monte Carlo approximations of stochastic programs. SIAM J. Optimization, **11**, 70–86
37. Wets, R. (1966): Programming under uncertainty: The solution set. SIAM J. Appl. Math., **14**, 1143–1151
38. Wets, R. (1972): Stochastic programs with recourse: A basic theorem for multistage problems. Z. Wahrscheinlichkeitstheorie verw. Geb., **21**, 201–206