CHALLENGES IN STOCHASTIC PROGRAMMING [†]

Roger J-B Wets

Department of Mathematics
University of California, Davis

Abstract. Remarkable progress has been made in the development of algorithmic procedures and the availability of software for stochastic programming problems. However, some fundamental questions have remained unexplored. This paper identifies the more challenging open questions in the field of stochastic programming. Some are purely technical in nature, but many also go to the foundations of designing models for decision making under uncertainty.

Keywords: stochastic programming, decisions under uncertainty, chance-constraints, probabilistic constraints, distribution problem, Markowitz portfolio model

Date: May 1, 1994

Today: August 20, 2001

[†] Research supported grants of the National Science Foundation and the U.S.-Israel Binational Science Foundation

Recent work in stochastic programming has mostly been aimed at the design of solution procedures and the development of accompanying software; an overly brief review of the present state-of-the-art is provided in §1. This effort should be continued and expanded, and should remain the central concern of the research in stochastic programming. However, to support the application of stochastic programming in a practical environment, there are a number of fundamental questions that still go begging for appropriate answers. This paper, based on my lecture at the International Conference on Stochastic Programming in Udine (Italy) in 1992, takes stock and goes through a list of the challenges that must be met if one is going to have adequate technical tools to validate the stochastic programming model in the context of decision making under uncertainty and justify the approximations that must be accepted to render the problem solvable by existing or projected computational tools.

1. Algorithmic procedures

As already mentioned, remarkable progress has been made during the last few years in the development of solution procedures for stochastic programming problems. Certainly the expanded capacity and the ever increasing celerity of computers have played a role in these advances. Much of the credit has to go to the development of new techniques, the setting up of data structures, and not least, to the ingenuity of those that have been involved in the writing of the software. In very broad terms, software and algorithmic development in the field is in the following state:

1. There are now a number of *general purpose subroutines*, even commercial level software, for problems of the following type: With

$$t \in (1, \ldots, T) \qquad \xi = (\xi_1, \ldots, \xi_T),$$

$$\xi_t^s = (A_{t1}^s, \ldots, A_{tt}^s, b_t^s, c_t^s), \qquad \text{for } s \in S,$$

$$\text{prob}\,[\,\boldsymbol{\xi} = (\xi_1^s, \xi_2^s, \ldots, \xi_T^s)\,] = p_s,$$

$$\underset{\to t}{\xi} := (\xi_1, \ldots, \xi_t),$$

one is interested in the solution, in particular in the first stage solution $x_1 = (x_1^1, \ldots, x_1^L)$, of the following optimization problem:

$$\begin{aligned}
\min_x \qquad & \sum_{s \in S} p_s \sum_{t=1}^{T} c_t^s x_t^s, \\
\text{subject to } & \sum_{\tau=1}^{t} A_{t\tau}^s x_\tau^s = b_t^s, && t = 1, \ldots, T; \quad s \in S, \\
& x_t^s = x_t^r \ \text{ if } \ \underset{\to t}{\xi}^s = \underset{\to t}{\xi}^r && t = 1, \ldots, T; \quad s, r \in S, \\
& v_t^s \le x_t^s \le u_t^s, && t = 1, \ldots, T; \quad s \in S.
\end{aligned}$$

Such problems are known as *multistage (linear) recourse problems* with discretely distributed random elements. The constraints $x_t^s = x_t^r$ if $\underset{\rightarrow}{\xi}{}_t^s = \underset{\rightarrow}{\xi}{}_t^r$ are the *nonanticipative restrictions* that reflect the fact that decision at time t can only depend on the observation made up to stage t.

Usually, there is no randomness in the first stage data, $\xi_1^s = \xi_1^r$ for all $r, s \in S$, and one must have $x_1^s = x_1^r$ for all $r, s \in S$, i.e., the first stage decision doesn't depend on ξ. The motivation for modeling decision problems as stochastic programs comes mostly from the search for a "robust" first stage decision, i.e., one that will put the decision maker in a rather good position whatever, or almost whatever, be the the outcome of future events.

2. There are a number of algorithmic procedures that can be applied in quite general situations: quasi-gradients methods, stochastic decomposition and certain methods based on the aggregation principle. These procedures can't usually compete, at least not at their present stage of development, with procedures that take advantage of the specific structure of the problem (linearity, for example), but they often provide the only possible approach in certain specific instances.

3. Finally, there are a number of methods tailored to specific classes of applications: like the use of logarithmic penalties for problems with chance-constraints, or extended linear-quadratic techniques for (discrete time) stochastic control problems, and so on.

2. Modeling decision making under uncertainty.

The potential user has a rich choice of techniques, but not many modeling tools are available at the theoretical and computational level (preprocessing, for example). One has to confront a number of complex issues, like how to give to uncertainty a probabilistic description, or how to include the attitude of the decision maker towards risk, or still, how to incorporate the gathering of information in the decision making model. Although sophisticated models can be built for each one of these components of a decision under uncertainty problem, the rules of the game are that it must be done so that a computational approach remains possible! Certainly, approximations will enter the picture in a significant manner, but there are also many other questions that need further investigation.

The components of a bare bones stochastic programming model are

- a decision vector that must satisfy certain constraints: $x_1 \in X_1$;
- a random variable $\boldsymbol{\xi}$ whose value will only be observed after x_1 has been selected;
- an evaluation (cost, possibly) of the decision in terms of the observed outcome.

At a more operational level, this involves

a. a cost function $f_{10}(x_1) + Q(\xi, x_1) =: f(\xi, x)$, with $f_{10}(x_1)$, the immediate costs associated with the choice of x_1, and $Q(\xi, x_1)$ the future costs;

b. the probability distribution P of the random variable $\boldsymbol{\xi}$ about which one may have only partial information, usually because there is a lack of statistical data;

c. and, decision criteria, that can take one of the following forms:

- an appraisal function, $v : I\!\!R \to I\!\!R$, leading to the following formulation,

$$\min_x \quad E\{v\,(f(\xi, x_1))\}, \qquad x_1 \in X_1;$$

- probabilistic constraints, such as

$$\text{prob}\,[\,f(\xi, x_1) \le 0\,] \le \alpha,$$

- multicriteria, tracking objectives, and so on.

In theory, at least, one should be able to rely on utility theory to always reduce the problem to one of maximizing the expectation of a utility function. But, such a utility function is often very difficult to assess, and many decision makers do not necessarily view their objective as one of maximizing utility. One can refer, for example, to the theory developed around the concept of "satisficing." The modeler might be induced to introduce constraints, and when these constraints involve random quantities, it leads to probabilistic or *chance constraints.* The stochastic programming model with *joint* chance constraints then reads:

$$\begin{aligned} \min_x \quad & f_{10}(x) \\ & f_{1i}(x) \le 0, \quad i = 1, \dots, s, \\ & \text{prob}\,[\,f_{1i}(\boldsymbol{\xi}, x) \le 0,\ i = s{+}1, \dots, m\,] \le \alpha. \end{aligned}$$

The inclusion of such constraints can be motivated by

- contractual requirements fixing the reliability level of a given system;
- innate knowledge of the level of unreliability that will be accepted by customers, the public, managers, etc.;
- convenience of formulation that doesn't require the sometimes involved calculations required to determine future (or recourse) costs.

However one should realize that whatever be the reason that led the modeler to the inclusion of chance constraints, they need to be "evaluated." In one way or another this means identifying a model where the chance constraints have been replaced by a (recourse) cost function. The question then becomes one of exploring the relationship between various stochastic programming models. So far, there have been very few examples of investigations of this type, consult [39, 9].

There is one result relating stochastic programs with simple recourse and stochastic programs with (separable) chance constraints.

Let's consider the following two problems:

$$\begin{aligned} \min_x \quad & \langle c, x\rangle \\ \text{such that } & \langle a_i, x\rangle = \beta_i, \quad i = 1, \dots, s, \\ & \text{prob}\,[\langle t_i, x\rangle \ge \xi_i\,] \ge \alpha_i, \quad i = s{+}1, \dots, m, \\ & x \ge 0. \end{aligned} \tag{CC}$$

and

$$\begin{aligned} \min_x \quad & \langle c, x\rangle + E\{Q(\boldsymbol{\xi}, x)\} \\ \text{such that } & \langle a_i, x\rangle = \beta_i, \quad i = 1, \ldots, s, \\ & x \geq 0, \end{aligned} \tag{RP}$$

where

$$Q(\xi, x) = \inf\left\{\langle q^+, y^+\rangle \,\middle|\, \langle t_i, x\rangle + y_i^+ \geq \xi_i,\ y_i^+ \geq 0,\ i = s{+}1, \ldots, m\right\}.$$

Suppose both problems are solvable. Let x^{cc} be an optimal solution of the stochastic program with chance constraints (CC) and assume that the chance constraints are active at x^{cc}. Then x^{cc} is also an optimal solution of (RP) if

$$q_i^+ = \frac{\pi_i^{cc}}{1 - \alpha_i}, \quad i = s{+}1, \ldots, m,$$

where π_i^{cc} is the (optimal) multiplier associated with the chance constraint

$$\operatorname{prob}\left[\langle t_i, x\rangle \geq \xi_i\right] \geq \alpha_i.$$

On the other end, if x^{rp} is an optimal solution of the stochastic programs with simple recourse (RP), then x^{rp} is also an optimal solution of (CC) if

$$\alpha_i = \frac{q_i^+ - \pi_i^{rp}}{q_i^+}, \quad i = s{+}1, \ldots, m,$$

where

$$\pi_i^{rp} = \langle -t_i, v^{rp}\rangle, \quad v^{rp} \in \partial EQ(x^{rp}).$$

The proof follows rather directly from writing down the optimality conditions for these two problems.

This result provides a way to evaluate the "cost-price" to be attached to shortages as a function of the α_i. Although, it only covers a very restricted class of models, it demonstrates that there is a "duality" between chance constraints and "recourse" (or penalty) costs. The reference to duality isn't arbitrary; the standard dual of a stochastic program with recourse involves constraints that must be satisfied in the average (a probabilistic constraint).

Another example in the same vein comes from the popular Markowitz model for portfolio management. An investor has a choice between various financial instruments whose rate of return is uncertain. In theory, again, the investor should maximize expected utility, but the catch is that this utility function isn't usually available. Instead, the Markowitz approach is to "draw" the so-called efficient frontier: for given expected return, one solves a quadratic program that identifies the portfolio minimizing variance. A diagram is produced that plots variance versus expectation. It is then up to the decision maker to choose a point on this efficient frontier. We are dealing here with an approximation to the original problem, and the question is to know how appropriate it is to proceed in this manner. As part of this analysis, one would like to identify the class of utility functions that would generate similar solutions. For the Markowitz model such an analysis has been done to a large extent [34, 42, 20]; consult [43, Introduction to Part III] for a detailed introduction to this question.

The challenge is to *clarify the relationship between stochastic programming models, in particular between models involving reliability considerations and those based on costs, and to develop tools that would validate the replacement of a stochastic programming problem by another problem whose structure might be quite different.* Much of this passes through a better understanding of optimality conditions, duality results, etc., but in addition, one has to develop a much deeper understanding of the interpretation to give to these conditions and a better evaluation of the role played by various modeling options.

3. The distribution problem

As far as a decision maker is concerned, the solution of a stochastic optimization model is a feasible point that yields the best distribution of costs. Thus, in terms of the following formulation:

$$\min\ E\{v(f(x,\boldsymbol{\xi}))\}, \qquad x \in S \subset I\!R^n,$$

where $v : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ is an *appraisal function* (i.e., $v(\theta') \leq v(\theta)$ if $\theta \leq \theta'$) and $f : I\!R^n \times \Xi \to \overline{\mathbb{R}}$ is a random lsc function describing costs, a solution $x^\star \in S$ is such that the distribution $\boldsymbol{\theta}_{x^\star} = f(x^\star, \boldsymbol{\xi})$ is preferred to the distribution $\boldsymbol{\theta}_x = f(x, \boldsymbol{\xi})$ for all $x \in S$. Quite often v is just the negative of a utility function, but we are not going to insist on continuity (a basic property of utility functions), and we do not want to load the interpretation to be given to v with the full axiomatic machinery that comes with utility functions.

It is very unusual, at the outset of the modeling process, to have sufficient information about preferences to be able to construct an appraisal function. In fact, the building of a stochastic model is, or should be, used as a means of discovering the shape of the appraisal function. This means that we should have at our disposal the capability of generating for a given x, the distribution of $f(x, \boldsymbol{\xi})$. If this function f is very simple, one may be able to obtain an explicit expression for this distribution by analytic means. And in particular if f is real-valued, we have at our disposal a wide array of results about convergence, statistical estimates, etc., that can be included in the tool kit. For example, if $f(x, \xi) = |x - \xi|$, then $\boldsymbol{\theta}_x$ has density $p(x - \cdot) + p(x + \cdot)$ where p is the density of $\boldsymbol{\xi}$; Figure 1 illustrates the shape of this density for p a beta density functions and x not too far from the expectation of $\boldsymbol{\xi}$.

But f is only exceptionally real-valued, f is usually of the following type:

$$f(x_1, \xi) = f_{10}(x_1) + \inf_{x_2 \in X_2} \left[\, f_{20}(x_1, x_2, \xi) \,\middle|\, f_{2i}(x_1, x_2, \xi) \leq 0,\ i = 1, \ldots, m_2 \,\right]$$

Fig. 1. Density of cost function: $|x - \xi|$.

where f_{20} might itself be defined through an infimum. This is not a “simple” function! And there is little hope of being able to find a closed form expression for the distribution of $\boldsymbol{\theta}_x = f(x, \boldsymbol{\xi})$. Only numerical/graphical procedures can be expected to produce useful information about the distribution of $\boldsymbol{\theta}_x$. Mathematical analysis however can help provide the background that would validate the numerical approaches and, hopefully, improve their efficiency.

Returning to the expression for f above, with

$$S_2(x_1, \xi) := \bigl\{x_2 \in X_2 \bigm| f_{2i}(x_1, x_2, \xi) \leq 0,\ i = 1, \ldots, m_2\bigr\},$$

one has

$$f(x_1, \xi) = \begin{cases} f_{10}(x_1) + f_{20}(x_1, \bar{x}_2(x_1, \xi), \xi) & \text{whenever } X_2 \cap S_2(x_1, \xi) \neq \emptyset; \\ \infty & \text{otherwise;} \end{cases}$$

where

$$\bar{x}_2(x_1, \xi) \in \operatorname*{argmin}_{x_2 \in X_2} [\, f_{20}(x_1, x_2, \xi) \bigm| f_{2i}(x_1, x_2, \xi) \leq 0,\ i = 1, \ldots, m_2\,],$$

assuming that this latter optimization problem has a solution whenever it is feasible. Generally, the argmin function $(x_1, \xi) \mapsto \bar{x}_2(x_1 \xi)$ is rather complicated and isn't conducive to analytic manipulations.

It thus appears that the only viable approach to finding the distribution of $\boldsymbol{\theta}_x$ is by sampling $\boldsymbol{\xi}$, calculating and recording the value of $f(x, \xi)$ and, possibly, making use of some approximation techniques to obtain either the distribution or the density function of $\boldsymbol{\theta}$, or at least an approximating one. The issues that need to be addressed are the following:

1. Can sampling be validated? In other words, if H^ν is the empirical distribution for $\boldsymbol{\theta}_x$ obtained via ν independent samples of $\boldsymbol{\xi}$, does H^ν converge to the true distribution of $\boldsymbol{\theta}_x$, and at what rate?
2. What can be done to make the sampling procedure as efficient as possible? Can the convergence rate be improved by relying on variance reduction techniques, for example?

3. Could one use curve fitting techniques? More specifically, is it possible to know the general shape of the distribution or density functions associated with $\boldsymbol{\theta}_x$ when $f(x,\xi)$ is a "typical" cost function.

These questions can be studied in the following framework: Given $g : \mathbb{R}^n \times \Xi \to \overline{\mathbb{R}}$, a cost function, and $\boldsymbol{\xi}$ a random vector with values in Ξ and distribution P, the *distribution problem* is to find the distribution of

$$\boldsymbol{\theta} = \inf_{z \in \boldsymbol{R}^n} g(z, \boldsymbol{\xi}).$$

We may as well assume that for all $\xi \in \Xi$, $\theta(\xi) := \inf g(z,\xi)$ is finite. The case when $\boldsymbol{\theta} = \infty$ with positive probability is without interest. (In the stochastic optimization model(s) that motivated our interest in the distribution problem, $\boldsymbol{\theta} = \infty$ with positive probability corresponds to $f(x, \boldsymbol{\xi}) = \infty$ with positive probability. Such a x is not an acceptable solution of the stochastic optimization problem.) Let H be the distribution of $\boldsymbol{\theta}$, and H^ν the distribution obtained via ν independent samples of $\boldsymbol{\xi}$, say

$$\theta^1 = \inf_z g(z, \xi^1),\ \theta^2 = \inf_z g(z, \xi^2),\ \dots\ , \theta^\nu = \inf_z g(z, \xi^\nu).$$

Under the (purely technical) assumption that g is a random lsc function, one has that $\boldsymbol{\theta}$ is a well-defined random variable with values in $\mathbb{R}$. From standard results in mathematical statistics, it then follows that

$$H^\nu \underset{n}{\to} H \qquad \text{a.s.}\ \ ;$$

by $\underset{n}{\to}$, one means *narrow convergence* (equivalently called weak or weak* convergence) of the probability measures H^ν to H. The proof of a.s.-narrow convergence usually relies on the law of large numbers. We are thus dealing with a classical question to which one can apply all the results we know about the convergence of empirical measures [29, 38], in particular in the study of the convergence rate. However, the situation isn't totally classical in that the function $\xi \mapsto \theta(\xi)$ doesn't always have all the desired differentiability, or even continuity, properties. The challenge is *to extend the classical results about the convergence of empirical measures so that one could apply them to the distribution problem.*

Ideally, one would want to draw samples of $\boldsymbol{\xi}$ using a measure on Ξ based on the distribution of $\boldsymbol{\theta}$!. However, this is exactly what we don't know. We could however achieve some gain (*variance reduction*), if we had access to a function that would approximate $\xi \mapsto \inf_{z \in \boldsymbol{R}^n} g(z,\xi)$.

The use of curve fitting techniques to obtain an approximating distribution or density function for $\boldsymbol{\theta}$ after a few samples have been collected would means that there is prior information about the shape of such functions. At first, it may appear that no particular characteristics could be identified for either the distribution function H or the density function h of $\boldsymbol{\theta}$. Indeed, a special case of the distribution problem is

$$\theta = \min \left[\, x \,\middle|\, x = \xi \,\right],$$

in which case $\boldsymbol{\theta}$ has the same distribution as $\boldsymbol{\xi}$, whatever it might be. But this is hardly the type of problem that one would have to deal with in stochastic programming. An appropriate paradigm for the distribution problem might be the following: for all ξ, the function $g(\cdot,\xi)$ is convex, inf-compact (bounded level sets and lower semicontinuous) and bounded below by 0. Good examples could be:

$$g(z,\xi) = \begin{cases} \langle c, z\rangle & \text{if } Az = \xi,\ z \geq 0; \\ \infty & \text{otherwise;} \end{cases}$$

with c a nonnegative vector, or the simple function $g(z,\xi) = |z - \xi|$. This might suggest that the shape of the density function depicted in Figure 1 is “typical”. Figure 2 gives the empirical (based on 200 samples) density of the value of the random linear program:

$$\begin{aligned} \min \quad & x_2 + 1.2x_3 + x_4 \\ \text{such that } \; & 3x_1 + x_3 - 4x_4 = \boldsymbol{\xi}_1 \\ & -x_1 + 2x_2 - 2x_3 + x_4 = \boldsymbol{\xi}_2 \\ & 2x_1 + x_2 - 4x_3 + x_4 = \boldsymbol{\xi}_3 \\ & x_1 \geq 0, \ldots, x_4 \geq 0 \end{aligned}$$

with $\boldsymbol{\xi}_1$ uniformly distributed on $[-1.5, 1]$, $\boldsymbol{\xi}_2$ with a triangular distribution on $[-1, 1]$ and $\boldsymbol{\xi}_3$ normally distributed with mean 1 and standard deviation 1.

Fig. 2. Empirical density of the cost function of a linear program.

The challenge is: *to identify a parametric class of densities and distribution functions associated with the random variable* $\boldsymbol{\theta} = \inf g(\cdot,\boldsymbol{\xi})$ *when* g *is of a certain type.*

In [30], Prekopa obtains a central limit theorem for random linear programs whose coefficients are stochastically perturbed when the unperturbed problem has a basic solution that is “highly stable.” This might suggest that the distribution of $\boldsymbol{\theta}$ is asymptotically normal, but this can only exceptionally be the case. On the other hand, the analysis in [19] fits the present context, in particular refer to Example 4.4 and the discussion related to asymptotic normality.

4. Modeling uncertainty

The basic premise in the building of stochastic programming models for decision making under uncertainty is that all parameters are known in a probabilistic sense, i.e., that their probability distribution is well determined. In the context of the following model:

$$\min_{x \in X} E\{f(x, \boldsymbol{\xi})\} = \int f(x, \xi)\, P(d\xi),$$

it means that the probability distribution P is given. This is seldom the case in practice! Of course, if $\boldsymbol{\xi}$ is a random variable that describes certain natural phenomena (weather patterns, the distribution of elementary particles, e.g.) or the characteristics of a large population (voting patterns, genetic properties, e.g.), then sufficient statistical data might be available so that P can be determined with a very high level of reliability. But more often, not enough information is available and the assumptions under which one can rely on statistical methods to obtain the probability distribution of random phenomena are not satisfied. Statistical methods require that enough data be available so that one can claim that the "asymptotic domain" has been reached. In most stochastic optimization problems $\boldsymbol{\xi}$ is a multi-dimensional random vector. Statistical procedures for estimating such probability distributions require a huge amount of data that is almost never available, especially if $\boldsymbol{\xi}$ is modeling behavioral patterns such as future demands for given products, price systems for raw commodities, etc..

One may wonder if there is any value in even considering stochastic programming models for such problems. Possibly, deterministic models might be equally valuable in helping the decision maker. At this point, however, a reality check might be needed. The basic objective of the modeling of a decision problem as a mathematical program is usually to provide help in reaching a "good" —if not optimal— decision. Assuming that there is uncertainty about the value to be assumed by some parameter, would it be better to base the decision on the solution of a model that assigns a fixed (unique) value to this parameter, or on the solution of a model that would a least allow for the possibility that this parameter could take on two different values? Think of this parameter as representing sales forecasts or flood predictions, for example. In these simple terms, it is obvious that the second model will lead to a better decision. And, if rather than two possibilities, one allows for a number, finite or not, of possible values for this parameter, the solution of the model is bound to provide the basis for an even better decision. In a nutshell, stochastic optimization models are *highly superior* to deterministic models whenever the parameters of the problem are not known with certainty and complete accuracy!

Allowing for the possibility that parameters might take on more than one value raises immediately the question of assigning relative weights, or equivalently frequencies, or still equivalently, a probability distribution, to the possible values of the parameters. At this point, one would have to come to grips with the fact that in all, but a few rare cases, not enough information will be available to calculate a reliable estimate of the probability distribution. In fact, in the worst of cases, no statistical data will be available at all, when a parameter describes a phenomenon that has never been observed (think of the market response to innovative products). In such instances the "experts" will have created the probability distribution. Almost all stochastic optimization models that one has to deal with in practice will have been formulated with a probability distribution for the random parameters that is a quite rough approximation of the true distribution.

The question is then: what level of confidence should one have in the solution generated by the stochastic optimization model? One way to settle this, is to show that the dependence of the solution on the probability measure isn't too strong. More precisely, if

$$x(P) \in \operatorname*{argmin}_{x \in X} \int f(x,\xi)\, P(d\xi)$$

then, for $\varepsilon > 0$ relatively small,

$$x(P) \in \varepsilon\text{-}\operatorname*{argmin}_{x \in X} \int f(x,\xi)\, Q(d\xi)$$

whenever Q isn't too far from P. This is what has been observed in practice, so the question seems to be more one of validating what is already known intuitively. There are also other issues, mostly related to the validating of sampling procedures —which also has the effect of replacing one probability measure by another— that require similar results about the sensitivity of the solution on the probability measure.

The study of the dependence of solutions on parameters leads in one form or another to proving epi-continuity for the (essential) objective function. In turn, this will yield the continuous dependence of the solutions since epi-continuity implies generally the continuity of the argmin mapping. In the context of our earlier formulation of stochastic programming problems, it means showing that

$$P \;\mapsto\; \int f(\cdot,\xi)\, P(d\xi)$$

is *epi-continuous* at an optimal solution $x^\star$:
(a) for all $P^\nu \underset{n}{\to} P$, $x^\nu \to x^\star$: $\liminf_\nu \int f(x^\nu,\xi)\, P^\nu(d\xi) \geq \int f(x^\star,\xi)\, P(d\xi)$;
(b) for all $P^\nu \underset{n}{\to} P$, there exists $x^\nu \to x^\star$: $\limsup_\nu \int f(x^\nu,\xi)\, P^\nu(d\xi) \leq \int f(x^\star,\xi)\, P(d\xi)$.
One then writes,

$$\int f(\cdot,\xi)\, Q(d\xi) \xrightarrow{e} \int f(\cdot,\xi)\, P(d\xi) \quad \text{as} \quad Q \underset{n}{\to} P.$$

If the P^ν are the empirical measures generated from a sample, say $\xi^1, \xi^2, \ldots, \xi^\nu$, of $\boldsymbol{\xi}$, i.e.,

$$P^\nu(A) = \frac{1}{\nu} \sum_{k=1}^{\nu} \mathbb{1}_A(\xi^k) \qquad \forall\, A \text{ measurable} \;\;,$$

the $\mathbf{P}^\nu$ are random measures (they depend on the sample), and one has to establish *a.s. epi-consistency*, i.e.,

$$\int f(\cdot,\xi)\,\mathbf{P}^\nu(d\xi) \xrightarrow{e} \int f(\cdot,\xi)\,P(d\xi) \quad \text{a.s.}\ .$$

Quite a number of results of this type are already known; a complete bibliography would fill at least a couple of pages. For far as epi-continuity, one could refer to [8, 26, 32, 1], and for a.s. epi-consistency, the question is basically solved by a law of large numbers [22, 4].

Although these results address the continuity of the solution of stochastic programs, that's not quite enough. We need "quantitative" statements, e.g., a statement of the type:

$$P \mapsto \varepsilon\text{- argmin} \int f(x,\xi)\,P(d\xi), \qquad \varepsilon > 0,$$

is Lipschitz continuous with "small" Lipschitz constant; Lipschitz continuity for mappings being appropriately defined and an appropriate metric being selected on the space of probability measures. That such results are possible is suggested by the Lipschitz continuity of ε-argmin with respect to the epi-distance for convex optimization problems [3]. Theoretically, it would be sufficient to relate the epi-distance between two stochastic optimization problems, say

$$\min_{x\in X} \int f(x,\xi)\,P(d\xi) \quad \text{and} \quad \min_{x\in X} \int f(x,\xi)\,Q(d\xi),$$

to some distance between the probability measures P and Q. Some results of this nature are already available [33, 36], deriving usually a rough upper bound for the Lipschitz constant. That's not quite satisfactory yet. The challenge is to *come up with Lipschitz constants for the ε-*argmin *mapping that are as sharp as possible and relatively easy to calculate*, i.e., are implementable! A different approach that could achieve the same objective would be to obtain error bounds, example of such efforts, that started with the work of Madansky and Zienba, are [7, 11, 13, 14, 15].

When the approximating measures have been obtained through sampling, "quantitative" statements can only be probabilistic in nature. One would expect that the main probabilistic tools to obtain convergence rates, viz. the central limit theorem, the law of iterated logarithms and results about large deviations, could be adapted to the problem at hand. However, this adaptation isn't straightforward. All the classical results rely on a certain level of smoothness of the functions involved. some of these hurdles have been overcome in a number of situations [10, 16, 21, 27, 28, 35, 37]. In this area, which one could label "asymptotic analysis of stochastic programs", there remain many open questions some of which are technically quite challenging.

However much "asymptotic analysis" might lead to a better understanding of the behavior of the solutions of stochastic programs under increased sampling when the probability distribution of the random variables is known, it falls short of coming to grips with the all too common problem of small samples, i.e., when all the information available consists of a few observed values of the random vector $\boldsymbol{\xi}$. The challenge is *to devise procedures and their justifications that would establish confidence levels for the solution of a stochastic program when the random vector* $\boldsymbol{\xi}$ *is only known through a small sample.* More than purely technical results may be required. There is some statistical literature dealing with the problem of small samples; one can interpret the finding of the solution of a stochastic program with limited information about the probability distribution as one of estimating the optimal solution. But the estimation problems considered in the statistical literature are orders of magnitudes simpler than "estimating" the solution of a stochastic programming problem. It might be necessary to frame the question in a new paradigm in order to arrive at an appropriate interpretation of the question and its solution.

5. Probabilistic structures for multistage problems

A multistage stochastic problem differs from a deterministic dynamic model in that at each stage t the decision maker will be allowed to observe the values assumed by some of the random variables, say ξ_t, and that the decision x_t at stage t *can depend only on the observations recorded so far*:

$$\underset{\rightarrow}{\xi}_t := (\xi_1, \xi_2, \dots, \xi_t), \qquad \underset{\rightarrow}{\xi}_T = \xi$$

Schematically the decision process consists of a sequence of observations followed by decisions:

$$(\xi_1 \to x_1(\underset{\rightarrow}{\xi}_1)) \to (\xi_2 \to x_2(\underset{\rightarrow}{\xi}_2)) \cdots \to (\xi_T \to x_T(\underset{\rightarrow}{\xi}_T)).$$

The probabilistic information available to the decision maker can be represented in the form of a tree, to be called the *scenario tree.* Each path through the tree identifies a particular sequence of realizations $\xi = (\xi_1, \xi_2, \dots, \xi_T)$; the nodes, where the branching occurs, corresponds to "stages". Figure 2 represents a scenario tree when the random vector $\boldsymbol{\xi}$ has a discrete distribution. This is the canonical form in which the information about the random elements of a T-stage stochastic program will be fed to a solution routine, cf. [6] for a description of the SMPS-format.

Fig. 3. Scenario tree, discrete distribution.

There are many issues related to the organization of the data base describing the scenario tree, not the least of which is to provide the user with tools that would allow for compact representations via network-type representations like Markov chains. Work is needed to provide modeling tools that would simplify the task of feeding the data to solver routines. This might be done in GAMS, AMPL or AIMS, for example. The importance of these endeavors in making stochastic programming an effective decision making tool can not be overstated. But technically, these issues are more organizational in nature, in many ways more directly related to data-base manipulations, and will not be discussed here.

On the more mathematical end of the spectrum, the basic challenge is *the replacement of a scenario tree by a much trimmer one!* Of course, the resulting stochastic program will still generate the same, or nearly the same, optimal solution. There are various versions of this question:

1. The scenario tree is trimmed by relying exclusively on the probabilistic structure. For example, certain branches are replaced by their expectations or conditional expectations, certain subtrees are removed, etc..
2. The scenario tree is trimmed by relying on the probabilistic structure but also on the problem structure. For example, information is used about the costs associated with the various branches of the tree.
3. Slim scenario trees are identified that provide upper and lower bounds for the problem at hand and procedures are laid down for improving these bounds.

An issue that needs to be clarified is what must be understood by the "same optimal solution." Clearly, if one tree is replaced by another tree, one cannot expect that the solutions of the stochastic programs:

$$\xi \mapsto x(\xi) = \big(x_1(\underset{\to}{\xi}_1), x_2(\underset{\to}{\xi}_2), \dots, x_T(\underset{\to}{\xi}_T)\big),$$

be identical, since their domain of definition might be quite different. Nor do we expect necessarily that the values of these programs be similar. Since our main concern is with the decisions suggested by the stochastic program it is also in those terms that we need to define the preservation of optimal solutions. This means that we expect the two problems would generate solutions that would be identical, or nearly identical, in the first stage.

Another approach to scenario tree trimming is by sampling. At this point, the challenge is to *develop a consistent sampling procedures to generate scenario trees* that would provide, once the resulting stochastic program is solved, reliable estimates of the optimal solution. One possibility would be to generate the tree sequentially, i.e., sample $\boldsymbol{\xi}_1$, then given ξ_1, sample $\boldsymbol{\xi}_2$, and so on. Another possibility, is to generate samples of $\boldsymbol{\xi}$ and use these samples to build a scenario tree with a given architecture (with a pre-determined number of branchings at each stage).

6. Dynamic stochastic optimization problem

Section 5 already touched on some questions related to discrete time dynamical models, here we want to address more specifically those questions raised by continuous time models, in particular, stochastic control models. Let's use the following formulation of a stochastic optimal control problem as the framework for our discussion:

$$\begin{aligned} \min_{u \in \mathcal{U}} E\Big\{ \int_0^T j(t, \boldsymbol{x}_t, \boldsymbol{u}_t)\, dt \Big\}&, \\ d\boldsymbol{x}_t = g(t, \boldsymbol{x}_t)\, dt + h(t, \boldsymbol{u}_t)\, d\boldsymbol{W}_t&, \\ \boldsymbol{x}_0 = \boldsymbol{\xi}_0&, \end{aligned}$$

where the control $\boldsymbol{u}$ and the state $\boldsymbol{x}$ are stochastic processes defined on $[0, T]$, $\boldsymbol{W}$ a semimartingale, for example, a d-dimensional Brownian motion, and $\boldsymbol{\xi}_0$ is a random vector determining the initial conditions. The functions j, g and h are assumed to be such that the objective's integral functional and the stochastic differential equations describing the dynamics are well defined. The space $\mathcal{U}$ of admissible controls is restricted to those that are nonanticipative with respect to information available at time t, and typically that is to be understood as meaning that the controller has been able to observe the actual state-trajectory from time 0 up to t, i.e.,

$$\forall\, t: \qquad \boldsymbol{u}_t \text{ must be } \mathcal{F}_t\text{-measurable}$$

with $\mathcal{F}_t := \sigma(\boldsymbol{x}_s, 0 \leq s \leq t)$ the sigma-field generated by past observations. More sophisticated formulations will also allow for partial observation of the state, for observations possibly corrupted by (random) noise, and so on.

The main objective pursued in stochastic optimal control is the design of a "feedback control", i.e., a function that will map the available information into an optimal decision. This is also the case for stochastic programming models when recourse decisions are viewed as functions of the past realizations $\underset{\to}{\xi}_{t}$. In terms of the preceding stochastic control problem, this would mean that we would observe $\boldsymbol{W}_t$, and choose u_t as a function of the sigma-field generated by the $\boldsymbol{W}_s$ for $s \leq t$ instead of $\mathcal{F}_t$. This is the basic difference between stochastic programming and stochastic optimal control models. In some important applications, it is possible to infer from the observation of the state, the values taken on by the stochastic process $\boldsymbol{W}_t$, but that is not the case in general. For a more detailed discussion of the relationship between stochastic programming and stochastic control models, one can refer to [41]. The fact that the multistage stochastic programming model always implies a discrete time modeling of the underlying dynamics, whereas stochastic control models allow for both continuous and discrete time modeling of the dynamics isn't all that significant.

It may appear at first that restricting the control processes to those that are measurable with respect to one sigma-field rather than another wouldn't really have that much impact on the structural properties of the problem. The reason it does is that the $\mathcal{F}_t$ depend on $(\boldsymbol{u}_s, s < t)$ and this dependence could be nonlinear. For 1-dimensional unconstrained stochastic optimal control problems, or 1-dimensional problems with no constraints on the state and simple constraints on the controls, such as the one formulated below, it may be possible to use stochastic calculus techniques to find a closed-form expression for the optimal control function. But unfortunately, an approach purely based on analytic tools can only go so far. Usually one has to resort to the calculation of solutions by means of algorithmic procedures. There are a number of possibilities. For stochastic control problems with state space of relatively low dimension (≤ 4), techniques based on Markov chain approximation are developed in [25]. Another approach that replaces the (stochastic) differential equations by difference equations via the discretization of the time axis, brings us in the realm of stochastic programming models. However, modeling a problem so that it be well suited to the potential use of stochastic calculus leads quite often to a model that is ill-suited for the use of stochastic programming techniques. To render this somewhat more concrete, let's consider the following example.

The Black and Scholes model for the option pricing of risky assets leads to the following system of stochastic differential equations:

$$\begin{aligned}\frac{dv_0(t)}{v_0(t)} &= r\,dt \qquad \text{(riskless asset, rate of return: } r\text{)}\\ \frac{dv_i(t)}{v_i(t)} &= \alpha_i\,dt + \langle e^i D, d\boldsymbol{W}(t)\rangle, \quad i = 1, \ldots, N,\end{aligned}$$

where v_i is the price associated with the i-th asset, α_i is the average rate of return (drift), $\boldsymbol{W}$ is a N-dimensional Wiener process, $D^T D$ is positive definite, and e^i the unit vector with a 1 in the i-th component. With this as the underlying price process, one can write down a stochastic differential equation for the evolution of the value of a portfolio. Let X_t be the total wealth at time t with π_{it} the proportion of the total wealth invested in asset i at time t, then

$$dX_t = \langle \alpha - r\mathbb{1}, \pi_t\rangle X_t\,dt + (rX_t - C_t)\,dt + X_t\langle \pi_t, D\,d\boldsymbol{W}_t\rangle, \quad X_0 = x^0,$$

where C_t is the consumption at time t, $\pi_{0t} = 1 - \sum_{i=1}^N \pi_{it}$ is the portion of the wealth invested in the riskless asset, and $\mathbb{1}$ is a vector of 1's. The manager of this portfolio is interested in a nonanticipative investment policy $(\pi_t,\, t \leq \tau)$ and consumption plan $(C_t, t \leq \tau)$, also nonanticipative, that would maximize

$$V_{C,\pi}(x^0) = E\Big\{\int_0^\tau e^{-\beta t} U(C_t)\,dt + \gamma e^{-\beta T}\Big\}$$

where $\tau = \inf\,[\,t \geq 0 \,|\, X_t = 0\,]$ is a stopping time (bankruptcy), γ a bankruptcy payment, β is a discount factor, and U is a strictly increasing concave utility function.

This model, proposed by Merton, was studied extensively in a series of far reaching articles [17,18] that revealed the nature of the solutions. It could serve as the basis of a pratical asset/liability management model, but then additional constraints have to be included (limitation on the portion of the wealth that could be invested in certain particular assets, etc.). When this is the case, it isn't possible to use the characterization of the optimal solution provided by [17], and analytic methods, in particular the stochastic differential calculus, are of limited help. One has to resort to computational schemes to find a solution of this problem. In practical terms, this means a time discretization which leads to a stochastic programming model. One version of such a problem would be: with $t = 0, 1, \ldots, T$, x^0 the initial wealth, and δ the discounting rate,

$$\begin{aligned}
&\max_{c_t, x_t} \quad E\Big\{ \sum_{t=0}^{T-1} \delta^t U(c_t) + \delta^T \gamma(x_T) \Big\} \\
&x_{t+1} = (x_t - c_t)(r + \sum_i \pi_{it}(\boldsymbol{\xi}_{it} - r)),\ t = 0, \ldots, T-1, \\
&c_t \geq 0, \quad \pi_{it} \geq 0, \quad \sum_{i=1}^{N} \pi_{it} \leq 1, \quad x_t \geq 0, \\
&c_t,\ \pi_t,\ x_t \text{ nonanticipative }.
\end{aligned}$$

At each stage, the wealth available for investment is $x_t - c_t$ and the averaged interest is $r + \sum_i \pi_{it}(\boldsymbol{\xi}_{it} - r)$.

The concavity of the objective function has been preserved, but the discretization of the stochastic differential equations describing the evolution of wealth, leads to an equation involving bilinear terms $x_t \pi_{it}$. We end up with a concave function to be maximized on a nonlinear manifold. This is not a convex optimization problem.

We have here a clear example of a problem whose formulation was ideally suited to the techniques of stochastic calculus but which when discretized leads to a (stochastic programming) problem that would be considered quite ungainly. Convexity, which is very important in the design of algorithmic procedures for mathematical programming models, turns out to be somewhat irrelevant when using stochastic calculus, where convexity only gets used to guarantee that the local solution obtained by setting some gradient equal to 0 is also a global solution.

When the information available to the controller is based on the state x_t of the system instead of on information collected about the underlying stochastic process $\boldsymbol{W}_t$, convexity —assuming it was present at the outset— is usually lost. In such cases, one must resort to dynamic programming techniques and these can only be used effectively when it makes sense to work with discrete state and control spaces (of low cardinality) [24, 25].

In summary, efficient solution procedures for (stochastic) optimization problems are mostly going to be based on mathematical programming techniques. But the formulation of stochastic optimization problems as stochastic control problems doesn't lead in general to problems that are well-suited to the application of mathematical programming techniques. The challenge is *to develop "continuous" versions of the stochastic programming (with recourse) models that would allow for the description of the dynamics by differential equations.* One possible approach is to study the limit of multistage stochastic programs as the number of stages tends to ∞. We already know that limit problems must be obtained as epi-limits, but with respect to what topology? And also, the class of problems for which such a passage to the limit is possible will have to be identified. Another approach is to immediately pass to a continuous-time formulation; cf. [5] for one such attempt.

7. The value of information

Let's proceed with the discussion in terms of the following problem:

$$\min_{x\in \boldsymbol{R}^n} \; E^P\{f(x,\boldsymbol{\xi})\} \;:=\; \int f(x,\xi)\,P(d\xi)$$

with optimal solution x^P. The actual cost $f(x^P,\xi)$ to the decision maker is only determined after the value of $\boldsymbol{\xi}$ has been observed, say ξ. Let us assume that more precise information can be collected about the distribution of $\boldsymbol{\xi}$ at some cost. For example, a more detailed statistical analysis of the environment might yield a more precise description of the distribution of $\boldsymbol{\xi}$. Before engaging in such an analysis, we would like to obtain an estimate of the value of such information. More generally, we might have access to more and more refined forecasts for the value to be assumed by $\boldsymbol{\xi}$, for example by means of samples of larger and larger size, and the question is then to choose the level at which the investigation should be conducted.

Let's designate by Q the (new) probability distribution that would be assigned to the random vector $\boldsymbol{\xi}$ after some further information has been collected. Let x^Q be the solution of the stochastic program

$$\min_{x\in \boldsymbol{R}^n} \; E^Q\{f(x,\boldsymbol{\xi})\} \;:=\; \int f(x,\xi)\,Q(d\xi)$$

where P has been replaced by Q. The actual cost $f(x^Q,\xi)$ to the decision maker again can only be determined after the value ξ of $\boldsymbol{\xi}$ has been observed. The value of the additional information is then,

$$\mathrm{val}(Q) \;=\; E^Q\{f(x^P,\boldsymbol{\xi}) - f(x^Q,\boldsymbol{\xi})\}.$$

However at the outset, the new probability measure Q that would be assigned to $\boldsymbol{\xi}$ isn't known with certainty. Typically what is known, is that the new distribution Q will belong to a certain class of probability measures, and that some Q's are more likely than others. This leads to introducing the notion of a *sensor*: a probability measure on the space $\mathcal{M}$ of probability measures on Ξ, the support of the random vector $\boldsymbol{\xi}$. For example, if $\boldsymbol{\xi}$ is real-valued, the support of a sensor S could be the family of gaussian distributions with standard deviation σ (fixed) and mean μ, where $\boldsymbol{\mu}$ is a random variable with distribution P. Such a sensor would provide a model for the following situation: In order to collect more information about the value to be assumed by $\boldsymbol{\xi}$, we shall get a sample from the population. The size of the sample determines σ, but we do not know a priori the sample mean (which should have the same distribution as $\boldsymbol{\xi}$). The value of a sensor S is then

$$\operatorname{val} S := \int \operatorname{val}(Q)\, S(dQ).$$

This model for assessing the value of information was proposed in [2]. There have been alternative proposals for modeling changes in information via nested sigma-fields. It can be shown that such models are special cases of models based on sensors, which furthermore aren't able to fully capture the type of statistical investigation described earlier.

So far the full calculation of the value of sensors has only been experimental and involved only very simple stochastic programming problems. The same approach applied to a more realistic class of problems could very well be computationally prohibitive. The challenge is to *incorporate information gathering in the decision process in a computationally feasible fashion.* To do so it will be necessary to study the mathematical properties of sensors, and in particular make available a (sub)differential calculus for sensors that could be exploited in an algorithmic setting.

8. Partial information.

The decision process of a multistage stochastic program follows the pattern:

$$(\xi_1 \to x_1(\underset{\rightarrow}{\xi}_1)) \to (\xi_2 \to x_2(\underset{\rightarrow}{\xi}_2)) \cdots \to (\xi_T \to x_T(\underset{\rightarrow}{\xi}_T)).$$

Let

$$\mathcal{F}_t = \sigma(\underset{\rightarrow}{\boldsymbol{\xi}}_t) = \text{ sigma-field of events observed up to time } t\ .$$

Since more and more events can be observed as t increases, one always has $\mathcal{F}_t \nearrow$ with t. If we let $x = (x_1, \dots, x_T)$, and (re)define $x_t : \Xi \to I\!R^{n_t}$ with

$$\Xi = \{\xi = (\xi_1, \dots, \xi_t, \dots \xi_T)\}, \qquad \text{support of } \boldsymbol{\xi}$$

the set of solutions must be restricted to

$$\mathcal{N}_{\mathcal{F}} = \{x \,|\, x_t \text{ is } \mathcal{F}_t\text{-measurable }, \quad t = 1, \dots, T\}$$

i.e., to those x's that satisfy the *nonanticipativity* constraints.

Often one doesn't have access to full information about the value assumed by ξ_t, the sigma-field of observable events is $\mathcal{G}_t \subset \mathcal{F}_t$. Again assuming that we have complete memory of all events that have been observed so far, one also has $\mathcal{G}_t \nearrow$. The decisions must be restricted to

$$\mathcal{N}_{\mathcal{G}} = \{x \,|\, x_t \;\mathcal{G}_t\text{-measurable} \;, \quad t = 1, \dots, T\}.$$

Since $\mathcal{G}_t \subset \mathcal{F}_t$, $\mathcal{N}_{\mathcal{G}}$ is a linear subspace of $\mathcal{N}_{\mathcal{F}}$. Passing from $x \in \mathcal{N}_{\mathcal{F}}$ to $x \in \mathcal{N}_{\mathcal{G}}$ doesn't really much change the nature of the original problem, in particular convexity will be preserved.

A well developed duality theory has shown that there is price system that can be associated with the nonanticipativity constraints, similarly one can attach a price system to the constraints generated by the further restricting x to $\mathcal{N}_{\mathcal{G}}$ [31, 5].

One new challenge would be to consider models where only *partial information is available, but one doesn't necessarily have* $\mathcal{G}_t \subset \mathcal{G}_{t+1}$, *i.e. information is not increasing with* t. That is the case when earlier observations are lost, or the information about earlier observations is only available in condensed form. Depending on the process that will generate the sigma-fields $\mathcal{G}_t$, convexity might be lost, in particular if the condensing or loss of information depends on earlier decisions.

9. Stochastic integer programming.

Let's consider the two-stage stochastic integer program:

$$\min_x \langle c, x\rangle + E\{Q(\boldsymbol{\xi}, x)\}, \quad x \in X \subset I\!\!R^n,$$

where

$$Q(\xi, x) = \min\{\langle q, y\rangle \,|\, Wy = h(\xi) - T(\xi)x, \; y \in \mathbb{Z}_+^{n'}\}.$$

The function $x \mapsto Q(\xi, x)$ has the same properties as that of the marginal value of a linear integer program. Although, the overall shape might be that of a convex function, it is discontinuous; a typical 1-dimensional example appears in Figure 3.

Fig. 4. The function $x \mapsto Q(\xi, x)$.

But when P is absolutely continuous, the function EQ, with $EQ(x) = \int Q(\xi, x)\, P(d\xi)$, turns out to be continuous under rather general conditions [40]. Even when P is discrete, the size of the discontinuity jumps will be relatively small (by comparison to the discontinuity jumps of the functions $Q(\xi, \cdot)$). Integrating with respect to P corresponds to taking a convex combinations of the functions $Q(\xi, \cdot)$, and this has both a smoothing and convexifying effect.

In some cases it is possible to proceed via smoothing which replaces discontinuous functions by approximating ones with prescribed differentiability properties, consult [12], for example. But such an approach has definite limitations. A more challenging, but also more promising, approach is *to find a substitute for the recourse problem that wouldn't require the second stage variables to be integer.* More specifically, we want to replace Q by a function Q^a which would also defined as the value of a certain optimization problem (but without integer restrictions on the variables), so that $EQ \approx EQ^a$. For stochastic programs with simple recourse, there have already been some path breaking contributions in this direction [23].

Acknowledgement. I appreciated very much the comments received from Jitka Dupačová (Charles University, Prague) and Alan King (IBM Watson Research Center, Yorktown Heights).

References

[**1**] Zvi Artstein & Roger J-B Wets, "Sensors and information in optimization under stochastic uncertainty," *Mathematics of Operations Research* **18** (1993), 523– 547.

[**2**] Zvi Artstein & Roger J-B Wets, "Stability of stochastic programs with possibly discontinuous objective functions," *SIAM J. on Optimization* **4** (1994), 537–550.

[**3**] Hedy Attouch & Roger J-B Wets, "Epigraphical processes: laws of large numbers for random lsc functions," 1991, to appear.

[**4**] Hedy Attouch & Roger J-B Wets, "Quantitative stability of variational systems: III. ε-approximate solutions," *Mathematical Programming* **61** (1993), 197–214.

[**5**] Kerry Back & Stanley Pliska, "The shadow price of information in continuous time decision problems," *Stochastics* **22** (1987), 151–168.

[**6**] John R. Birge, Michael A. Dempster, Horand I. Gassmann, E.A. Gunn, Alan J. King & Stein Wallace, “A standard input format for multiperiod stochastic linear programs,” *COAL (Mathematical Programming Society, Committee on Algorithms) Newsletter* **17** (1987).

[**7**] John R. Birge & Stein W. Wallace, “A separable piecewise linear upper bound for stochastic linear programs,” *SIAM J. on Control and Optimization* **26** (1988), 725–739.

[**8**] Jitka Dupačová, “Stability and sensitivity-analysis for stochastic programming,” *Annals of Operations Research* **27** (1990), 115–142.

[**9**] Jitka Dupačová, Alexei Gaivoronski, Zdenek Kos & Tomas Szántai, “Stochastic programming in water resources system planning: A case study and a comparison of solutions techniques,” *European J. of Operational Research* **52** (1991), 28–44 .

[**10**] Jitka Dupačová & Roger J-B Wets, “Asymptotic behavior of statistical estimators and of optimal solutions for stochastic optimization problems,” *The Annals of Statistics* **16** (1988), 1517–1549.

[**11**] N.C.P. Edirisinghe & W.T. Ziemba, “Bounds for two-stage stochastic programs with fixed recourse,” *Mathematics of Operations Research* **19** (1994), 292–313.

[**12**] Yuri Ermoliev, Vladimir I. Norkin & Roger J-B Wets, “The minimization of discontinuous functions: mollifier subgradients,” *SIAM J. on Control and Optimization* **32** (1994).

[**13**] Karl Frauendorfer, *Stochastic two-stage programming*, Springer-Verlag Lecture Notes in Economics and Mathematical Systems 392, Berlin, 1992.

[**14**] Alexei Gaivoronski, “Stochastic optimization techniques for finding optimal submeasures,” in *Stochastic Optimization*, V. Arkin, A. Shiraev & R. Wets, eds., Springer Verlag Lecture Notes in Control and Information Sciences 81, Berlin, 1986, 351–363.

[**15**] Peter Kall, “Stochastic programs with recourse: an upper bound and the related moment problem,” *Zeitschrift für Operations Research* **31** (1987), A 119–A 141.

[**16**] Yuri M. Kaniovski, Alan King & Roger J-B Wets, “Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems,” *Annals of Operations Research* (1995), (to appear).

[**17**] Ioannis Karatzas, John P. Lehoczky, Suresh P. Sethi & Steven Shreve, “Explicit solution of a general consumption/investment problem,” *Mathematics of Operations Research* **11** (1986), 261–294.

[**18**] Ioannis Karatzas, John P. Lehoczky & Steve E. Shreve, “Optimal portfolio and consumption decisions for a ”Small Investor” on a finite horizon,” *SIAM J. on Control an Optimization* **25** (1987), 1557–1586.

[**19**] Alan J. King, "Generalized Delta theorems for multivalued mapping and measurable selections," *Mathematics of Operations Research* **14** (1989), 720–736.

[**20**] Alan J. King, "Asymmetric risk measures and tracking models for portfolio optimization under uncertainty," *Annals of Operations Research* **45** (1993), 165–177.

[**21**] Alan J. King & R. Tyrrell Rockafellar, "Asymptotic theory for solutions in statistical estimation and stochastic programming," *Mathematics of Operations Research* **18** (1993), 148–162.

[**22**] Alan J. King & Roger J-B Wets, "Epi-consistency of convex stochastic programs," *Stochastics and Stochastics Reports* **34** (1990), 83–92.

[**23**] Willem K. Klein Haneveld, Leen Stougie & Maarten H. Van der Vlerk, "On the convex hull of the simple integer recourse objective function," Research Memorandum nr. 516, University of Groningen, 1993.

[**24**] P.R. Kumar & Pravin Varaiya, *Stochastic Systems. Estimation, Identification and Adaptive Control*, Prentice-Hall, Englewood Cliffs, 1986.

[**25**] Harold J. Kushner & Paul G. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer-Verlag, New York, 1992.

[**26**] Riho Lepp, "Approximations to stochastic programs with complete recourse," *SIAM J. on Control and Optimization* **28** (1990), 382–394.

[**27**] Georg Ch. Pflug, "Asymptotic dominance and confidence for solutions of stochastic programs," Czechoslovak Journal for Operations Research, 1992.

[**28**] Georg Ch. Pflug, "Asymptotic stochastic programs," manuscript, University of Vienna, 1993.

[**29**] David Pollard, *Convergence of Stochastic Processes*, Springer Verlag, Berlin, 1984.

[**30**] András Prékopa, "On the probability distribution of the optimum of a random linear program," *SIAM J. on Control* **4** (1966), 211–222.

[**31**] R.T. Rockafellar & Roger J-B Wets, "Deterministic and stochastic optimization problems of Bolza type in discrete time," *Stochastics* **10** (1983), 273–312.

[**32**] Werner Römisch & Rüdiger Schultz, "Stability analysis for stochastic programs," *Annals of Operations Research* **30** (1991), 241–266.

[**33**] Werner Römisch & Rudiger Schultz, "Lipschitz stability for stochastic programs with complete recourse," Schwerpunktprogramm DFG, Anwendungsbezogene Optimierung und Steuerung-1992, 1992.

[**34**] Paul A. Samuelson, "The fundamental approximation theorem in portfolio analysis in terms of means, variances and higher moments," *Review of Economic Studies* **37** (1970), 537–542.

[**35**] Alexander Shapiro, "Asymptotic analysis of stochastic programs," *Annals of Operations Research* **30** (1991), 169–186.

[**36**] Alexander Shapiro, "Asymptotic behavior of optimal solutions in stochastic programming," *Mathematics of Operations Research* **18** (1993), 829–845.

[**37**] Alexander Shapiro, "Quantitative stability in stochastic programming," *Mathematical Programming* (1994), Manuscript, University of South Africa, Pretoria.

[**38**] Galen R. Shorack & Jon Wellner, *Empirical Processes and Applications to Statistics*, John Wiley & Sons, New York, 1986.

[**39**] László Somlyódy & Roger J-B Wets, "Stochastic optimization models for lake eutrophication management," *Operations Research* **36** (1988), 660–681.

[**40**] Leen Stougie, "Design and analysis of algorithms for stochastic integer programming," Thesis, Centrum for Wiskunde en Informatica, Amsterdam, 1985.

[**41**] Pravin Varaiya & Roger J-B Wets, "Stochastic dynamic optimization, approaches and computation," in *Mathematical Programming, Recent Developments and Applications*, M. Iri & K. Tanabe, eds., Kluwer Academic Publisher, Dordrecht, 1989, 309–332.

[**42**] William T. Ziemba, "Choosing investment portfolios when the returns have stable distributions," in *Mathematical Programming in Theory and Practice*, P.L. Hammer & G. Zoutendijk, eds., North-Holland, Amsterdam, 1974, 443–482.

[**43**] William T. Ziemba & R.G. Vickson (eds.), *Stochastic Optimization Models in Finance*, Academic Press, New York, 1975.