1

Optimization under Exogenous and Endogenous Uncertainty

Jitka Dupačová
Charles University, Prague, Faculty of Mathematics and Physics, Dept. of Probability and Math. Statistics, Sokolovská 83, CZ-18675 Prague

Abstract

Customary stochastic programs aim at the best feasible decision made before the realization of the random element is observed. The common assumption is that the probability distribution does not depend on decisions — the case of the exogenous uncertainty. This paper focuses on stochastic programming models for which through decisions, a decision-dependent, endogenous randomness is put into effect. Problem structure then becomes important. Examples point out at tractable cases and solution techniques.

Keywords

Stochastic programs, distributions dependent on decisions, exogenous and endogenous uncertainty, contamination
JEL: C61, C44

1 Stochastic programming problems

Customary stochastic programs, see e.g. [3], [5], [9], [14], aim at the best feasible decision x made before the relevant random element $\omega \in \Omega$ is unveiled. The cost of choosing x depends on ω as quantified by a real-valued function $f(x, \omega)$. The problem is

$$\text{minimize } F(x; P) := E_P f(x, \omega) \tag{1}$$

over a closed nonempty subset $\mathcal{X}$ of a finite-dimensional (Euclidean) space. P denotes a known probability distribution of ω on Ω which is x-independent — the case of the exogenous uncertainty. In practice, however, there are important decision problems in which through decisions, a decision-dependent, *endogenous* randomness is put into effect; see e.g. [6], [7], [8], [10].

One may try to *remove the dependence of P on x by formulating a simpler model*: The settlement of revenues of a pension fund is influenced by the attained fund investments profitability whose probability distribution depends on the investment decisions. To an extent, this problem may be circumventioned by fixing the valorization of the accumulated wealth of individual participants to a predetermined guaranteed minimal level and penalizing the deviations; cf. [13]. This, however, is not a general approach.

In this paper we shall deal with stochastic programs of the form

$$\text{minimize } F(x) = \int_\Omega f(x, \omega) P_x(d\omega) \quad \text{on } \mathcal{X} \tag{2}$$

which differ from the standard version (1) in making explicit a possible dependence of probability distributions on decisions. We shall assume for simplicity that the expectations $\int_\Omega f(x, \omega) P_x(d\omega)$ are finite for all $x \in \mathcal{X}$ and an optimal solution exists.

Under specific assumptions, dependence of P on x in (2) can be removed by a suitable *transformation of the decision-dependent probability distribution* P_x, cf. [16] and the Push-in technique explained in [11], [15]:

[1]This work is partly supported by the project "Methods of modern mathematics and their applications" – MSM 00216120839 and by Grant Agency of the Czech Republic (grants 201/05/2340 and 402/05/0115).

Assume that there *exist densities* $p(x,\omega)$ *of probability distributions* P_x *with respect to a common probability measure* Q. Then the objective function in (2) can be rewritten as

$$F(x) = \int f(x,\omega)p(x,\omega)Q(d\omega)$$

Thus we recover the common form $F(x) := \int_\Omega \tilde{f}(x,\omega)Q(d\omega)$ with $\tilde{f}(x,\omega) := f(x,\omega)p(x,\omega)$ and with a decision-independent probability distribution Q. However, it is obtained at the cost of losing convenient properties of the original random objective function $f(x,\omega)$. The properties of the resulting objective function depend on the structure of the problem, namely, on type of dependence of P on x. For example, assume that

$$P_x(B) = Q(B \oplus Hx) \tag{3}$$

for every Borel set $B \subset \Omega$, with Q a probability distribution, $\oplus$ the direct sum and H a given matrix of the matching dimension. Changing variables in $\int f(x,\omega)P_x(d\omega)$ transforms the objective function to $\int f(x,\zeta - Hx)Q(d\zeta)$, whose properties depend on properties of $f(x,u)$ viewed as a function of (x,u) *jointly*.

The acceptance of the decision-dependent model (2) may cause various technical difficulties: For instance, if $f(x,\omega)$ is a convex function of x for each ω, then so is $F(x;P)$, whereas the convexity property of $F(x) := F(x;P_x)$ may be lost. This in turn puts limitations on the choice of numerically tractable optimization techniques even if evaluation of the objective function at any point x is no more complicated then for the classical model (1). Depending on the structure, recursive optimization methods and search techniques, cf. [11], numerical enumeration techniques including branch-and-bound method and disjunctive programming can be used.

The decisions may partly aim at *enhancement of the knowledge of the probability distribution*: For instance, in simple inventory-type stochastic programs the demand observed in the first stage may serve to collect more precise information about the probability distribution of the future demand. However, a demand higher than a certain cut-off point, such as the supply available for the first stage, will not be observed. The wish to obtain as precise information as possible may lead to increasing the order for the first stage. Such decision process can be then formulated by means of *sensors*, cf. [1] [2].

Special attention is needed for multistage problems with a decision-dependent probability distribution. Here not only the first-stage decision, but also the later decisions affect the information about the probability distribution available to the decision maker as they may influence the marginal and conditional probability distributions in subsequent stages; [2] displays examples of this kind. Moreover, decisions may influence the time at which uncertainty gets resolved, i.e., *nonanticipativity conditions may be decision dependent*; see [6], [7].

We shall see that tractability of problems (2) depends essentially on their structure and that there are several favorable problem classes, e.g.,

- The probability distribution is of a known type and the decisions influence only its parameters, see Section 2;
- There is a fixed finite set of probability distributions, see Section 3. The dependence on decisions may often be modeled by Boolean variables and the decisions may be partly related to the choice of a probability distribution from the given set.

Stability of the optimal solution of problem (2) with respect to changes of the probability distribution will be discussed in Section 4.

2 Decision-dependent parameters

2.1 Stochastic PERT problem

Several modeling issues on the subject of the stochastic PERT problem are discussed in [9], [14]. The primary concern is to minimize the expected duration of a project defined as a set of activities which consume time and resources and have to reflect certain temporal precedence relationships. The project can be described by an acyclic directed network with nonnegative arc lengths and with two specific nodes "Start" and "End" of the project.

If the durations of individual activities (lengths of arcs) $g := (g_1, \ldots, g_n)$ are known, the shortest time in which the project can be completed while observing the prescribed preference relations is equal to the length $l^*(g)$ of the critical path, the longest path connecting the Start and End nodes, which is a nonnegative convex function of g. Durations g_j of activities $j = 1, \dots, n$, may be reduced for an additional cost. More specifically, assume that g_j are convex functions of parameters $x_j := (x_{j1}, \dots, x_{jn_j}),\ j = 1, \dots, n$; then the project duration — the composite function $f(x) := l^*(g_j(x_j), j = 1, \dots, n)$ — is convex in x. As to the additional cost $k(x)$, assume that it is convex and separable in individual components $x_{ji},\ i = 1, \dots, n_j,\ j = 1, \dots, n$. The problem is to choose the best parameter values with respect to both the project duration and the additional costs and considering constraints. Thanks to convexity, it is possible to rewrite it in the form with one aggregated objective function, such as $\lambda_1 f(x) + \lambda_2 k(x)$, with parameters $\lambda_1, \lambda_2 > 0$.

Assume now that durations of individual activities are random, $\omega := (\omega_1, \dots, \omega_n)$, and that *parameters*, say $x_j,\ j = 1, \dots, n$, of their probability distributions may be changed for a cost. In [12], this problem is discussed in detail for the class of independent uniform distributions with a fixed spread around changing expected values x_j and for independent triangular distributions determined by the lower/upper bounds $a_j x_j$ / $b_j x_j$ and moduses $m_j x_j,\ j = 1, \dots, n$. These parameters are subject to linear constraints which reflect limits on resources and on activity durations. The applied cost function is $k(x) := \sum_{j=1}^n k_j x_j^{-1}$ and the objective function reflects then two convex criteria: Minimize the expected project completion time when using probability distribution identified by parameters $x_j,\ j = 1, \dots, n$, and minimize the costs $k(x)$ for chosen parameter values $x_j\ \forall j$. In [12], sample-path optimization is presented as an efficient solution method.

2.2 Queuing networks

Let $\omega = (\omega(t), 0 \le t \le T)$ be a stochastic process in continuous time controlled by the parameter $x = (x_1, x_2, \dots, x_r)^\top$, the probability distribution of ω being thus dependent on x, denoted P_x. We should minimize (or maximize) the expectation of a functional $f(x, \omega)$ of that process. For instance, ω may describe performance of a queuing network where customers pass through r service stations, according to definite rules. The inter-arrival times as well as service times are exponentially distributed, the formers with some fixed intensities, the latters with intensities $x_1, x_2, \dots, x_r$. The set $\mathcal{X}$ may be given by budget limitations as $\{x \ge 0 : c(x) \le K\}$, $c(x)$ the cost of running the system under control parameter x. The functional f may be the number of customers whose service was completed during the time interval $[0, T]$. For a fixed x which may be chosen or controlled by the manager we can get the value of $f(x, \omega)$ by simulating histories of all customers who entered the system.

To get the optimal decision, one needs to use a recursive optimization method, such as stochastic approximation procedures; see [11]. Sometimes, a random process can be simulated which leads to P_x in the limit. For example, [10] applies a stochastic quasigradient algorithm in the context of optimal control of a system with a decision dependent transition operator and with an unknown steady-state probability law P_x, which may be found by simulation.

3 Decision-dependent scenario trees

Assume now that there is a finite number of possible probability distributions and that each of them has been approximated by a discrete distribution carried by a finite number of atoms — scenarios. For multistage stochastic programs, each of these discrete probability distributions is used to create a scenario tree which takes into account the related topology of stages and path probabilities are attached to the scenarios; cf. [5]. The resulting scenario trees and the path probabilities are indexed by a finite number of indices d. In principle, one may apply a full enumeration with respect to d, or a version of the branch-and-bound method, cf. [8], or disjunctive programming techniques, cf. [6], [7].

3.1 Probing for information

This simple example illustrating the case of decision-dependent probabilities has been motivated by Chapter 21 in [17]:

A large oil exploration company holds a lease that must be either sold out immediately for a know market price, or after one year for a price which depends on an exogenous factor — uncertainty in future oil prices, or sold after some exploration, e.g. after an experimental drilling. There are three possible outcomes of drilling — discovery of a dry or wet well or a gusher. The crucial issue is to determine their probabilities. In principle, the decision maker may use probabilities based on past experience. However, the cost of drilling is high and it would be useful to eliminate drilling if the well is dry. The suggestion is to precede drilling by another, substantially cheaper exploration method — seismic analysis. Based on its outcome probabilities of the three possible scenarios can be revised.

With the finite and small action space this problem can be modeled using decision trees and no special optimization technique is needed.

3.2 Sizes problem [8]

A production line must meet the demand for a certain number of products ordered according to their attributes (e.g., sizes) $a_1 \prec \ldots \prec a_r$. If the demand for a given product cannot be satisfied, after an additional treatment (e.g., cutting) it is possible to exploit a higher category product for additional substitution costs. Production costs consist of fixed set-up costs for every category that will be actually produced, of random per unit costs of the initial production and of random substitution costs whenever it applies. The additional substitution costs depend on the first-stage decision which determines the initial production levels $x_1, \ldots, x_r$.

The decision about producing category j or not is modeled by a Boolean variable d_j. Hence, there is a collection indexed by vectors $d \in \mathcal{D} = \{0, 1\}^r$ of a finite number of stochastic programs

$$\min_x E_{P_d} f(x, \omega) \text{ subject to } x \in \mathcal{K}(d), \tag{4}$$

where $\mathcal{K}(d)$ denotes coupling constraints on x given d (for example, $d_j = 0$ implies $x_j = 0$, or $\sum_j x_j d_j$ equals the total demand). The random vector ω consists of components related with the production costs on the both initial production and additional treatment levels. Its joint probability distribution P_d depends on the decision d which determines the structure of production.

The producer tries to select the best production scheme, i.e., to decide according to the best stochastic program

$$\min_{d \in \mathcal{D}} \{ \min_{x \in \mathcal{K}(d)} E_{P_d} f(x, \omega)\}. \tag{5}$$

The choice of d (and of x as well) is the first-stage decision, the second-stage decisions concerning the additional treatment enter the random production costs $f(x, \omega)$. Random demand, an exogenous uncertainty, can also be incorporated.

3.3 Project selection

Consider the possibility of investment into I projects of uncertain potential which may be initiated at time instants $t = 1, \ldots, T$. When a project starts, it cannot be interrupted or closed before the considered investment horizon T. There exist several versions of each of these projects, at most one of them can be used. The probability distribution of the uncertain characteristics θ_i of project i is discrete, carried by scenarios θ_i^s, $s \in \mathcal{S}_i$ with probabilities p_i^s; the scenario and its time evolution gets revealed only after the investment is started. The problem is to decide when and to which project (or project version) to invest and at what level of investment subject to various cost, capacity or technological constraints so that the expected net present value of the total investment over the whole horizon is maximal.

Exploration of new oil fields, cf. [6], belongs into this class of problems. It involves decisions about building work platforms at certain places, pipelines networks, production platforms, etc. Oil fields are characterized by their initial capacities γ_i and deliverabilities δ_i. They are the endogenous source of uncertainty because their realizations $\theta_i^s = (\gamma_i^s, \delta_i^s)$ can be observed only after the project i was accepted and started. The time evolution of capacity and deliverability over time is described by a linear reservoir model. This means that all uncertain parameters and the related coefficients are known after their initial values were observed at the starting time, say t_i, of field i exploration. Let the versions $\upsilon \in \mathcal{V}_i$ of project i differ only by different time points $t_i = 0, 1, \ldots T - 1$, at which the exploration of the field i starts.

Decision variables $d_{i\upsilon}$, $i = 1, \ldots, I$, $\upsilon = 0, \ldots, T - 1$ equal 1 if version υ of project i is chosen (i.e., if project i starts at $t_i = \upsilon$) and 0 otherwise. These $0 - 1$ variables have to fulfil conditions

$$\sum_\upsilon d_{i\upsilon} \leq 1 \; \forall i.$$

Operational variables for version υ of project i at time t and for scenario s are continuous, denoted as $y_{i\upsilon}^{ts}$, $s \in \mathcal{S}_i$, $\upsilon \in \mathcal{V}_i$, $i = 1, \ldots, I$. Assume for simplicity that they are subject to a system of linear inequalities

$$q^t \leq \sum_i \sum_{\upsilon \in \mathcal{V}_i} f_{i\upsilon}^{ts} y_{i\upsilon}^{ts} d_{i\upsilon} \leq Q^t, \; t = 0, \ldots, T - 1, \; s \in \mathcal{S}_i, \; i = 1, \ldots, I$$

with coefficients $f_{i\upsilon}^{ts}$ known for $t \geq \upsilon$ and capacity constraints

$$L d_{i\upsilon} \leq y_{i\upsilon}^{ts} \leq c_i^s d_{i\upsilon}, \; t = 0, \ldots, T - 1, \; s \in \mathcal{S}_i, \; i = 1, \ldots, I$$

$$\sum_t y_{iv}^{ts} d_{iv} \le \gamma_i^s,\ s \in \mathcal{S}_i,\ i = 1, \dots, I.$$

Nonanticipativity conditions for each field separately are influenced by the selected version of the i-th project and reflect the simple fact that for project i the decisions d_{iv} and for $d_{iv} = 1$ also y_{iv}^{ts} for $t \le v$ are equal for all scenarios $s \in \mathcal{S}_i$. The reason is that individual scenarios θ_i^s cannot be distinguished before the project i starts and the form of the nonanticipativity conditions is thus influenced by decisions d_{iv}.

With discounted unit costs r_{iv}^{ts} corresponding to version v of project i at period t and for scenario $s \in \mathcal{S}_i$, the objective function is

$$\sum_t \sum_i \{ \sum_{v \in \mathcal{V}_i} \sum_{s \in \mathcal{S}_i} p_i^s r_{iv}^{ts} y_{iv}^{ts} d_{iv} \}.$$

4 Stability of optimal solutions

Consider first the decision-independent case (1). The influence of changes in the probability distribution P can be modeled using the *contamination approach*, see e.g. [4], i.e., using contaminated distributions

$$P_\lambda = (1 - \lambda)P + \lambda \tilde{P}, \tag{6}$$

with $\lambda \in [0, 1]$ and with $\tilde{P}$ another probability distribution under consideration. We suppose that for all considered distributions, stochastic program (1) has an optimal solution.

The objective function $F(x; \lambda) := E_{P_\lambda} f(x, \omega)$ is linear in λ and its derivative with respect to λ equals $E_{\tilde{P}} f(x, \omega) - E_P f(x, \omega)$.

Define the optimal value function

$$\varphi(\lambda) := \min_{x \in \mathcal{X}} F(x; \lambda).$$

If $\hat{x}$ is the *unique* minimizer of (1), then under mild conditions (e.g. [4]) the one-sided derivative exists and

$$\varphi'(0^+) = E_{\tilde{P}} f(\hat{x}, \omega) - E_P f(\hat{x}, \omega), \tag{7}$$

i.e., the local change of the optimal value function caused by a small change of P in the direction $\tilde{P} - P$ is asymptotically the same as that of the objective function at $\hat{x}$. Moreover, $\varphi(\lambda)$ is a concave function of λ on $[0, 1]$, hence the bounds

$$(1 - \lambda)\varphi(0) + \lambda\varphi(1) \le \varphi(\lambda) \le \varphi(0) + \lambda\varphi'(0^+)$$

are valid for all $\lambda \in [0, 1]$.

Consider now the decision-dependent case (2). Let $\hat{x}$ be the true or approximated minimizer of $F(x) = F(x, P_x)$ on $\mathcal{X}$. If the probability distributions $(P_x, x \in \mathcal{X})$ are contaminated by $(\tilde{P}_x,\ x \in \mathcal{X})$ as in (6), then the derivative of $F(\hat{x}, P_{\hat{x}})$ in the direction $\tilde{P}_{\hat{x}} - P_{\hat{x}}$ measures again sensitivity of the objective function at $\hat{x}$ against small changes of $P_{\hat{x}}$ in that direction. However, the assertion (7) about the optimal value function $\varphi(\lambda)$ is no longer true in general. It is true in some special cases, e.g., in the case (3), where

$$P_x(\bullet) = Q(\bullet \oplus Hx),\ \tilde{P}_x(\bullet) = \tilde{Q}(\bullet \oplus Hx),\ x \in \mathcal{X}.$$

Hence, sensitivity analysis for the decision-dependent case (2) would require development of new quantitative stability results.

The directional derivative can be exploited algorithmically: Assume for example that a search technique for solving problem (2) was stopped at a point $\hat{x}$ which is the true or approximate minimizer of $F(x) = F(x, P_x)$ on the set $\mathcal{X}$. In the context of example 3.2 it means that $\hat{x}$ solves a problem akin to (4) with the probability distribution P_d. To analyze the effect of perturbations of the applied vector d and, possibly, to change d to get an improvement, one can exploit directional derivatives $F(\hat{x}, P_{\tilde{d}}) - F(\hat{x}, P_d)$ of the objective function $F(x, P_d)$ at $\hat{x}$ in the direction of $P_{\tilde{d}} - P_d$ for $\tilde{d} \in \mathcal{D}$; cf. [8].

References

[1] Artstein Z (1994) Probing for information in two-stage stochastic programming and the associated consistency. In: Mandl, P, Hušková M (eds.) *Asymptotic Statistics*, Proc. of the 5th Prague Symposium 1993, Physica-Verlag, Heidelberg, pp. 21–33.

[2] Artstein Z (1999) Gains and costs of information in stochastic programming. *Annals of Oper. Res.* 85: 129–152.

[3] Birge JR, Louveaux F (1997) *Introduction to Stochastic Programming*, Springer-Verlag (Series in Operations Research), New York.

[4] Dupačová J (1990) Stability and sensitivity – Analysis for stochastic programming. *Annals of Oper. Res.* 27: 115–142.

[5] Dupačová J, Hurt J, Štěpán J (2002) *Stochastic Modeling in Economics and Finance*, Kluwer Acad. Publ., Dordrecht Boston London.

[6] Goel V, Grossmann IE (2004) A stochastic programming approach to planning of offshore gas field developments under uncertainty in reserves. *Computers & Chemical Engineering* 28: 1409–1429.

[7] Goel V, Grossmann IE (2005) A class of stochastic programs with decision dependent uncertainty, SPEPS 2004–23.

[8] Jønsbraten TW, Wets RJ-B, Woodruff DL (1998) A class of stochastic programs with decision dependent random elements. *Annals of Oper. Res.* 82: 83–106.

[9] Kall P, Wallace SW (1994) *Stochastic Programming*, J Wiley & Sons, Chichester.

[10] Pflug GCh (1990) On-line optimization of simulated Markovian processes. *Math. Oper. Res.* 15: 381–395.

[11] Pflug GCh (1999) *Optimization of Stochastic Models. The Interface between Simulation and Optimization*, 2nd Printing, Kluwer Acad. Publ., Boston Dordrecht London.

[12] Plambeck EL, Fu B-R, Robinson SM, Suri R (1996) Sample-path optimization of convex stochastic performance functions. *Math. Progr.* 75: 137–176.

[13] Polívka J (2002) Liability side of asset–liability management model. In: Šafránková J (ed.) *WDS'02 Part I*, Matfyzpress, Prague, pp. 61–67.

[14] Prékopa A (1995) *Stochastic Programming*, Kluwer Acad. Publ., Dordrecht Boston London and Akadémiai Kiadó, Budapest.

[15] Rubinstein RY, Shapiro A (1993) *Discrete Event Systems. Sensitivity Analysis and Stochastic Optimization by the Score Function Method*, J. Wiley&Sons, Chichester.

[16] Varayia P, Wets RJ-B (1990) Stochastic dynamic optimization approaches and computation. In: Iri M, Tanabe K (eds.) *Mathematical Programming. Recent Developments and Applications*, KTKSci Publ., Tokyo, pp. 309–332.

[17] Wonnacott TH, Wonnacot RJ (1990) *Introductory Statistics for Business and Economics*, 4th edition, J. Wiley&Sons, New York.