Working Paper

STOCHASTIC PROGRAMMING WITH INCOMPLETE INFORMATION

Jitka Dupačová

February 1986
WP-86-08

International Institute for Applied Systems Analysis
A-2361 Laxenburg, Austria

STOCHASTIC PROGRAMMING WITH INCOMPLETE INFORMATION

Jitka Dupačová

February 1986
WP-86-08

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

FOREWORD

One of the activities of the Adaptation and Optimization Project of the System and Decision Sciences Program is to develop mathematical methods and approaches for treating models of systems characterized by limited information about parameter distribution.

This paper presents three approaches which reflect different assumptions about the incomplete knowledge of the distribution and which can be applied to model building as well as to sensitivity analysis, approximation and robustness studies in stochastic programming problems. The suggested methods build a bridge between the purely deterministic approaches of nonlinear programming stability and the tools of mathematical statistics.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

CONTENTS

STOCHASTIC PROGRAMMING WITH INCOMPLETE INFORMATION

Jitka Dupačová

Abstract

The possibility of successful applications of stochastic programming decision models has been limited by the assumed complete knowledge of the distribution F of the random parameters as well as by the limited scope of the existing numerical procedures.

We shall introduce selected methods which can be used to deal with the incomplete knowledge of the distribution F, to study robustness of the optimal solution and the optimal value of the objective function relative to small changes of the underlying distribution and to get error bounds in approximation schemes.

The research was mostly carried out at the Department of Statistics, Charles University, Prague and it was stimulated by a close collaboration of the author with the ADO project of SDS. The present version of the paper was written at IIASA Laxenburg.

1. Introduction

Quite a large class of stochastic programming decision problems can be transformed to the following mathematical programming problem

$$\begin{aligned} &\text{maximize } g_0\ (x\ ;F) \\ &\text{subject to } g_i(x\ ;F) \geq 0\ ,\ 1 \leq i \leq m\ , \\ &\qquad g_i(x\ ;F) = 0\ ,\ m+1 \leq i \leq m+p\ , \\ &\qquad x \in \mathbf{X} \end{aligned} \tag{1.1}$$

where $\mathbf{X} \subset R^n$ is a given nonempty set. The functions g_i , $0 \leq i \leq m+p$, do not depend on random parameters directly but by means of their distribution F only.

An example of (1.1) is when a nonlinear program

$$\begin{aligned} &\text{maximize } h_0\ (x\ ;\ \omega) \\ &\text{subject to } h_k\ (x\ ;\ \omega) \geq 0\ ,1 \leq k \leq l\ ,\ h_k\ (x\ ;\ \omega) = 0\ ,l+1 \leq k \leq s\ , \\ &\qquad x \in \mathbf{X}_0 \end{aligned} \tag{1.2}$$

contains random parameters ω in $h_k\ (x\ ;\ \omega)$, $0 \leq k \leq s$, and the decision $x \in \mathbf{X}_0$ has to be chosen before the values of these parameters are observed.

Among others, two well known decision models of stochastic programming can be evidently written in form (1.1):

Stochastic program with recourse

$$\begin{aligned} &\text{maximize } E_F\ \{h_0\ (x\ ;\ \omega) - \varphi(x\ ;\ \omega)\} \\ &\text{subject to } x \in \mathbf{X} \subset \mathbf{X}_0 \end{aligned} \tag{1.3}$$

where the penalty function $\varphi(x\ ;\ \omega)$ evaluates the loss corresponding to the case that the chosen $x \in \mathbf{X}$ does not fulfill the constraints $h_k\ (x\ ;\ \omega) \geq 0$, $1 \leq k \leq l$, $h_k(x;\omega) = 0, l+1 \leq k \leq s$, for the observed values of the random parameters. The set $\mathbf{X} \subset \mathbf{X}_0$ is defined by induced constraints which guarantee that φ is well defined.

Stochastic program with probabilistic constraints

$$\begin{aligned} &\text{maximize } E_F\ \{h_0\ (x\ ,\ \omega)\} \\ &\text{subject to } P_F\ \{h_k\ (x\ ;\ \omega) \geq 0\ ,\ k \in I_i\} \geq \alpha_i\ ,\ 1 \leq i \leq m\ , \\ &\qquad x \in \mathbf{X} \subset \mathbf{X}_0 \end{aligned} \tag{1.4}$$

where $I_i \subset \{1, \ldots, l\}, \alpha_i \in \langle 0,1 \rangle$, $1 \leq i \leq m$, are given in advance.

For both mentioned basic types of decision models, numerous remarkable theoretical results were achieved and numerical approaches suggested. However, the numerical solution is rather complicated in general, mainly due to the fact that repeated evaluation of function values and gradients is needed which is rather time consuming and demands special simulation and/or approximation techniques. The question of error bounds is evidently both of practical and theoretical interest.

The optimal solution $x(F)$ and the optimal value of the objective function in (1.1) depend on the chosen type of model and on the distribution F which is usually assumed to be completely known and independent of the chosen decision x. However, the distribution F is hardly known completely in real situations. The numerical results obtained should thus be at least complemented by an additional information about sensitivity of the optimal solution with respect to eventual changes of the distribution F. In the robust case, a small change in the distribution F should cause only a small change in the optimal solution.

A first idea could be to study stability of the optimal solution of program (1.1) with respect to the underlying distribution F directly. However, the space of probability measures provided with a metric corresponding to the weak topology is not a linear one, so that the general results of parametric programming are not applicable directly.

In this paper three approaches will be presented. They reflect different assumptions on the (incomplete) knowledge of the distribution F. As we shall see, they may be used to perform sensitivity analysis and postoptimality studies, to get error bounds and to solve problems of stochastic programming under an explicitly given assumption of incomplete knowledge of the distribution F.

(i) Assuming that the considered distribution is known to belong to a parametric family of distributions, say $F \in \{F_y , y \in Y\}$, we can rewrite program (1.1) making the dependence on the parameter vector y explicit:

$$\begin{aligned} &\text{maximize } g_0(x\ ;\ y) \\ &\text{subject to } g_i(x\ ;\ y) \geq 0\ ,\ 1 \leq i \leq m\ , \\ &\qquad g_i(x\ ;\ y) = 0\ ,\ m+1 \leq i \leq m+p\ , \\ &\qquad x \in \mathbf{X} \end{aligned} \tag{1.5}$$

where $g_i(x\ ;\ y)$, $0 \leq i \leq m+p$, are used instead of $g_i(x\ ;\ F_y)$, $0 \leq i \leq m+p$, respectively. The stability of the optimal solution of program (1.5) with respect to the parameter vector $y \in Y$ can be studied to a certain extent through the methods of parametric programming and through the methods developed for non-linear programming stability studies (see e.g. Armacost and Fiacco (1974), Garstka (1974)).

Having in mind the statistical background of the parameter values which are typically statistical estimates of the true parameter values, the results of parametric programming have been complemented by statistical approaches (see Dupačová (1983), (1984) for problem (1.3), Dupačová (1986a) for problem (1.4)). The results are summarized in Section 3.

(ii) The local behaviour of the optimal solution $x(F)$ with respect to small changes of the underlying distribution F can be studied via t-contamination F by a suitably chosen distribution G, i.e., instead of F, distributions of the form

$$F_t = (1-t)F + tG\ ,\ 0 \leq t \leq 1$$

are considered (see Dupačová (1983), (1985a) for problem (1.3), Dupačová (1986a) for problem (1.4)). The original stability problem thus reduces to that linearly perturbed by a scalar parameter t. This approach gives a basis for performing sensitivity analysis of the optimal solution $x(F)$ and for post-optimality studies. (See Section 4.)

(iii) In typical cases of incomplete knowledge of the distribution, F is known to belong to a specified set $\mathbf{F}$ of distributions. One approach is via minimax. We shall discuss in Section 5 the case when the constraints in (1.1) do not depend on F and are incorporated into $\mathbf{X}$. For convex compact set $\mathbf{F}$, the ***minimax solution*** $x(F^*)$

is the optimal solution of the problem (1.1) corresponding to the least favourable distribution $F^{*} \in \mathbf{F}$ and, similarly, the ***maximax solution*** $x(F^{**})$ corresponds to the most favourable distribution $x(F^{**}) \in \mathbf{F}$. Even without compactness of $\mathbf{F}$ we may get minimax and maximax bounds

$$\max_{x \in \mathbf{X}} \inf_{F \in \mathbf{F}} g_0(x ; F)$$

and

$$\max_{x \in \mathbf{X}} \sup_{F \in \mathbf{F}} g_0(x ; F)$$

which provide an interval estimate for the optimal value $\max_{x \in \mathbf{X}} g_0(x ; F)$ for any $F \in \mathbf{F}$. This fact can be used to draw conclusions about the dependence of the optimal solution on changes of F within the given set $\mathbf{F}$ and to get error bounds in numerical methods.

In specific cases (reliability, worst case analysis) the minimax solution itself is of great interest. In addition, it is possible to combine complete and incomplete knowledge of the distribution of specific random parameters of the given problem (Dupačová (1985b)). Even in the minimax approach, however, the solution depends on the choice of the set of distributions $\mathbf{F}$ and it is necessary to choose such a set which fits to the presented problem as well as possible, using all the available information. For getting error bounds (as a part of an iterative algorithm) one cannot probably increase the level of information too much.

As the set $\mathbf{F}$ is often defined by prescribing values of certain moments of the distributions $F \in \mathbf{F}$, the results of the moment problem can be used to get computable minimax/maximax solutions and bounds (see Dupačová (1977) ,(1978)). When the prescribed values η of moments are not known precisely enough, namely, when they are estimated on the basis of observed data, the problem of stability of the minimax solution comes to the fore and to solve it, methods mentioned sub (i) can be applied.

2. Examples

To get some motivation, let us consider first a few examples.

Example 2.1. The cattle-feed problem (van de Panne and Popp (1963)). The problem is to find the amounts x_j of input j which lead to the minimum cost of the final mixture in which restraints on the nutrition contents are satisfied. In the formulation, the protein content weight percentages per ton, a_j, for each of four considered inputs are assumed to be normally distributed random variables with means μ_j and variances σ_j^2, $1 \leq j \leq 4$. Besides of deterministic linear constraints, one probabilistic constraint

$$P\{\sum_{j=1}^{4} a_j x_j \geq p\} \geq 1-\alpha \tag{2.1}$$

is constructed.

Under normality assumption, (2.1) can be written in the following way

$$\sum_{j=1}^{4} \mu_j x_j + \Phi^{-1}(\alpha)(\sum_{j=1}^{4} \sigma_j^2 x_j^2)^{1/2} \geq p$$

where $\Phi^{-1}(\alpha)$ denotes the α − quantil of the $N(0,1)$ distribution. The parameters μ_j, σ_j^2, $1 \leq j \leq 4$, are estimated by sampling and in applications, the estimates are used instead of the true parameter values. In Armacost and Fiacco (1974) the problem of stability of the optimal solution with respect to parameter values was solved, namely, derivatives of the optimal solution with respect to the parameter values were obtained.

Having in mind the statistical background of the considered parameters we shall aim to complement the deterministic stability results by statistical ones.

Example 2.2. A simple stochastic model of water reservoir design. The problem is to minimize the required capacity c of the reservoir subject to the following constraints:

Freeboard constraint

$$P\{s_i \le c - v_i\} \ge \alpha_1 , \quad 1 \le i \le n ,$$

Minimum storage constraint

$$P\{s_i \ge m_i\} \ge \alpha_2 , \quad 1 \le i \le n ,$$

Minimum release constraint

$$P\{x_i \ge y_i\} \ge \alpha_3 , \quad 1 \le i \le n ,$$

where, in the particular time interval, s_i is the storage, v_i is the flood control freeboard storage, m_i is the minimum storage, x_i is the total release and y_i is the prescribed minimum release.

Using linear decision rule, the variables x_i , s_i are expressed via monthly inflows r_i , r_{i-1} whose marginal distributions F_i are supposed to be known. Usually, log-normal distribution is used and its parameters are estimated on the basis of relatively long time series of the observed monthly inflows. However, in particular months, specific deviations from the assumed distribution may appear: in spring, the distribution may be relatively close to the normal one. Under these circumstances, we can accept the hypothesis that the true marginal distributions are mixtures of given log-normal and normal ones. We are interested to describe changes of the original optimal decision due to the influence of the alternative distribution.

Even in this simple example, three different types of variables typical for stochastic models of water resources systems can be distinguished at first sight:

- constant coefficients and parameters, such as system reliabilities, flood control freeboard storage, minimum storage or rule curve and penalty coefficients in the corresponding recourse model
- random variables with a known distribution (i.e., with a well estimated distribution), e.g. the monthly inflows

- random variables with an incomplete knowledge of distribution, such as the future demands (Dupačová (1985b)).

A deeper insight into the modelled real life problem, however, leads to the conclusion that the parameters are far from being known precisely, that the distribution has been estimated from time series of data which are observed with a relatively high measurement error or that the type of the distribution follows from the past experience and the parameters of the distribution are estimated on the basis of random input data. On the other hand, the final decision should not be too sensitive to the changes of the parameters and distributions, it should be robust enough.

Example 2.3. The STABIL model (Prékopa et al. (1980)) was applied to the fourth Five-Year Plan of the electrical energy sector of Hungary. Besides numerous deterministic linear constraints, one joint probabilistic constraint

$$P_F\{\sum_{j=1}^{n} a_{ij}\, x_j \geq \omega_i\ , \quad 1 \leq i \leq 4\} \geq p$$

was used; the four right-hand sides ω_i, $1 \leq i \leq 4$, were regarded stochastic and the joint distribution of these random variables was supposed to be normal. Due to the lack of reliable data, some of the correlations could not be given precisely enough. That is why two alternative correlation matrices were considered in Prékopa et al. (1980) and the numerical results were compared.

Alternatively, instead of given normal distributions $N(\mu\ ,\ \Sigma_1)$ or $N(\mu\ ,\ \Sigma_2)$ their mixture

$$(1-t)\ N(\mu\ ,\ \Sigma_1) + t\ N(\mu\ ,\ \Sigma_2) \tag{2.2}$$

can be considered which helps to study the changes of the optimal solution in principle for $0 \leq t \leq 1$; (2.2) corresponds to the gross error or contamination model.

Example 2.4. Project planning. The problem is to fix the completion time T of the given project. The reduction of the completion time is profitable at the rate of

$c \geq 0$ and the eventual delay in the completion is penalized with $q \geq c$ per time unit. The project is represented by a network whose arcs correspond to the planned activities. Assume that there is one sink and one source only and that the activities are numbered by indices $1 \leq i \leq n$. Whereas the structure of the project (the network) is supposed to be given, the completion times, say ω_i, of the activities are random variables and so is the total completion time τ. According to our formulation, the decision T has to be made before the realizations of ω_i's are known and one has to solve the stochastic program

$$\min_T \{cT + q\, E_F\, [\tau(\omega) - T]^+\} \tag{2.3}$$

where F denotes the joint distribution of the n-dimensional random vector ω and the explicit form of the $\tau(\omega)$ can be derived.

In practice, the distribution F is hardly known completely. Using the PERT-method, one usually solves the problem (2.3) under assumption that the random completion times ω_i are independently distributed with a Beta-distribution over a given interval. The parameters p , q of the Beta distribution are usually fixed on the basis of the available information about some characteristics of the distribution, such as the mean value, mode and variance.

Example 2.5. (Seppälä (1975)). In his stochastic multi-facility problem Seppälä considers the case of stochastically dependent weight coefficients. In order to eliminate the estimation of the correlation coefficients, he introduces a parameter to the model which weights the totally correlated case and the uncorrelated one.

3. Nonlinear programming stability results and estimated parameters

As our starting point, consider the following deterministic nonlinear program depending on a vector parameter y:

Let $Y \subset R^q$ be an open set, $h : R^n \times Y \to R^{m+p+1}$ be given continuously differentiable functions. For a fixed $y \in Y$, the problem is to

$$\text{maximize } h_0(x\,;\,y)$$
$$\text{subject to } h_i(x\,;\,y) \geq 0\,,\; 1 \leq i \leq m\,, \qquad M(y)$$
$$h_i(x\,;\,y) = 0\,,\; m+1 \leq i \leq m+p\,.$$

The corresponding Lagrange function has the form

$$L(x\,,u\,,v\,;y) = h_0(x\,;\,y) + \sum_{i=1}^{m} u_i\, h_i\,(x\,;\,y) + \sum_{i=1}^{p} v_i\, h_{m+i}\,(x\,;\,y)$$

and by $w(y) = [x(y)\,,u(y)\,,v(y)] \in R^n \times R^m_+ \times R^p$, the Kuhn-Tucker point of $M(y)$ will be denoted. The knowledge of the Kuhn-Tucker conditions of the first and second order as well as the knowledge of the linear independence condition and the strict complementarity conditions (Fiacco (1976), Robinson (1980)) will be assumed throughout the text.

Theorem 3.1. *Let $y^0 \in Y$ and let $w(y^0)$ be the Kuhn-Tucker point of $M(y^0)$ for which the Kuhn-Tucker conditions of the first and second order, the linear independence condition and the strict complementarity conditions hold true. Let on a neighbourhood of $[x(y^0)\,;\,y^0]$, h_i, $0 \leq i \leq m+p$, be twice continuously differentiable with respect to x and continuous derivatives*

$$\frac{\partial^2 h_i\,(x\,;\,y)}{\partial y_k\, \partial x_j} \text{ exists for all } \quad 1 \leq k \leq q\,, \quad 1 \leq j \leq n\,, \quad 0 \leq i \leq m+p\,.$$

Then the following statements hold true:

(a) *For $y \in 0_\varepsilon(y^0)$, there exists a unique once continuously differentiable function $w(y) = [x(y)\,,u(y)\,,v(y)]$ satisfying the Kuhn-Tucker conditions of the first and second order, the linear independence condition and the strict complementarity conditions for $M(y)$.*

(b) *Let $I(y) \subset \{1,\ldots,m\}$ contain the indices of the active inequality constraints*

$$h_i(x(y)\,;\,y) = 0\,,\; i \in I(y)\,,$$

and denote by

$$w_I(y) = [x(y)\,,u_i(y)\,,i \in I(y)\,,v(y)]\,,$$

$\nabla_x h_I (x ; y) = [\nabla_x h_I (x ; y), i \in I(y), \nabla_x h_i (x ; y), m+1 \le i \le m+p]$,
$\nabla_y h_I (x ; y) = [\nabla_y h_I (x ; y), i \in I(y), \nabla_y h_i (x ; y), m+1 \le i \le m+p]$.

Let further

$$D(y) = \begin{bmatrix} \nabla^2_{xx} L (w ; y) & \nabla_x h_I (x ; y) \\ [\nabla_x h_I (x ; y)]^T & 0 \end{bmatrix}_{[w(y) ; y]}, \tag{3.1}$$

$$B(y) = \left[\nabla^2_{yx} L (w ; y), \nabla_y h_I (x ; y)\right]^T_{[w(y) ; y]}. \tag{3.2}$$

Then for $y \in 0_\varepsilon(y^o)$,

$$\frac{\partial w_I(y)}{\partial y} = -D^{-1}(y) \cdot B(y) \tag{3.3}$$

and the remaining components of $\frac{\partial w(y)}{\partial y}$ *equal to* 0.

The statements of Theorem 3.1 are a modification of results by Fiacco (1976), and Robinson (1974). Due to the assumptions, the implicit function theorem can be applied to the system of equations which correspond to the active constraints in the Kuhn-Tucker conditions of the first order. Namely, the strict complementarity conditions play an important role reducing locally the program $M(y^o)$ to a classical maximization problem with equality constraints.

The assumptions can be weakened using results by Robinson (1980): Without assuming the strict complementarity conditions in $M(y)$, let us denote

$$I^+(y) = \{i \in I(y) : u_i(y) > 0\}$$
$$I^0(y) = \{i \in I(y) : u_i(y) = 0\}$$

and formulate the ***strong second order sufficient condition:***

For each $\kappa \ne 0$ with

$$\kappa^T \nabla_x h_i (x ; y) = 0, \quad i \in I^+(y)$$
$$\kappa^T \nabla_x h_i (x ; y) = 0, \quad m+1 \le i \le m+p ,$$

the inequality $\kappa^T \nabla^2_{xx} L(w(y);y)\kappa < 0$ holds true.

Except for the differentiability of the Kuhn-Tucker points $w(y)$, the first assertion of Theorem 3.1 can be parallelly reformulated. The differentiability property was studied, e.g., by Jittorntrum (1984). It is possible to get directional derivatives of $w(y)$ in any direction under the strong second order sufficient condition without assuming strict complementarity conditions. We shall use this result later in connection with the contamination method (see Section 4). The most general result on differentiability is due to Robinson (1984); for its application see the forthcoming paper Dupačová (1986b).

As we shall see later, the parameter vector y may correspond to the parameters of the underlying distribution F (see Theorem 3.2), to the parameter of contamination (see Section 4) and, eventually, to the probability levels α_i , $1 \le i \le m$, in (1.4) or to other parameters used to build a specific decision model of stochastic programming.

Assume now that the parameter vector y in $M(y)$ is connected with statistical assumptions about the distribution F of random coefficients in a stochastic programming decision model. It comes typically when F is known to belong to a parametric family of distributions $\{F_y \, , \, y \in Y\}$, so that y is the parameter vector identifying the distribution.

For the stochastic program with recourse (1.3) it means that for a fixed distribution F_y, $M(y)$ is the program

$$\text{maximize } g_0(x \, ; \, y) := E_{F_y} \{h_0 \, (x \, ; \, \omega) - \varphi(x \, ; \, \omega)\}$$

on a set $\mathbf{X}$ which does not depend on F_y, e.g.,

$$\mathbf{X} = \{x \in R^n : g_i(x) \ge 0 \, , \, 1 \le i \le m \, , \, g_i(x) = 0 \, , \, m+1 \le i \le m+p \, \},$$

for the stochastic program with probabilistic constraints (1.4), $M(y)$ is the program

$$\text{maximize } g_0 \, (x \, ; \, y) := E_{F_y}\{h_0(x \, ; \, \omega)\}$$
$$\text{subject to } g_j(x \, ; \, y) := P_{F_y}\{h_\kappa(x \, ; \, \omega) \ge 0 \, , \, \kappa \in I_i\} - \alpha_i \ge 0 \, , \, 1 \le i \le m \, ,$$
$$x \in \mathbf{X}_0$$

In general our aim is to solve program (1.5) for the true parameter vector, say $\eta \in Y$. However, our decision can only be based on the knowledge of an estimate, say y^N, of η. As a result, the substitute program $M(y^N)$ is solved instead of $M(\eta)$. Under the asymptotic normality assumption on the distribution of the estimate y^N in $M(y^N)$, the deterministic stability results of Theorem 3.1 can be complemented by statistical ones.

Theorem 3.2. *Let y^N be an asymptotically normally distributed estimate of the true parameter vector η that is based on the sample of size N:*

$$\sqrt{N}\,(y^N - \eta) \sim N(0\,,\,\Sigma)$$

with a known variance matrix Σ. Let the assumptions of Theorem 3.1 be fulfilled for $M(\eta)$. Then the optimal solution $x(y^N)$ of $M(y^N)$ is asymptotically normal

$$\sqrt{N}\,(x(y^N) - x(\eta)) \sim N(0\,,\,V) \tag{3.4}$$

with the variance matrix

$$V = \left[\frac{\partial x(\eta)}{\partial y}\right] \Sigma \left[\frac{\partial x(\eta)}{\partial y}\right]^T$$

where $\left[\frac{\partial x(\eta)}{\partial y}\right]$ is the (n,q) submatrix of (3.3).

Proof. Under assumptions of Theorem 3.1, $x(y)$ is a continuously differentiable (vector) function on a neighbourhood of $x(\eta)$. Using the normality assumption and the δ-method [Rao, 1973, p.388], we get the result immediately.

Remark 3.3. All elements of $\left[\frac{\partial x}{\partial y}\right]$ are continuous on a neighbourhood of η, so that the asymptotic distribution (3.4) can be substituted by

$$N\left[0,\ \left[\frac{\partial x(y^N)}{\partial y}\right] \Sigma \left[\frac{\partial x(y^N)}{\partial y}\right]^T\right];$$

see Rao (1973, p. 388).

Example 3.4. The application of Theorem 3.2 to Example 2.1 is straightforward. Let y be the vector consisting of asymptotically normal estimates s_j, $1 \le j \le 4$, of the true variances σ_j, $1 \le j \le 4$. According to Theorem 3.1, the derivatives $\frac{\partial x}{\partial y}$ exist and their values were obtained by Armacost and Fiacco (1974). We have thus asymptotic normality of the optimal solution. To get the variance matrix of the resulting distribution, the variance matrix Σ (diagonal in our case) should be known besides of $\frac{\partial x}{\partial y}$.

Special cases 3.5. In some special cases, it is possible to get explicit formulas for the derivatives $\frac{\partial x}{\partial y}$ and thus for the variance matrix V of the asymptotic distribution (3.4). We shall introduce the results applied to the ***simple recourse problem*** (see Dupačová (1984)):

$$\text{maximize } g_0(x\,;\,y) := c^T x - E_{F_y}\left[\sum_{i=1}^{m} q_i \left[\sum_{j=1}^{n} a_{ij}\, x_j - \omega_i\right]^+\right]$$

on the set

$$\mathbf{X} = \{x \in R^n : Px = p\ ,\ x \ge 0\}\ ,$$

where P is a given $(r\ ,\ n)$ matrix of rank r, c and p are fixed vectors, $q_i > 0$, $1 \le i \le m$, are given and $A = (a_{ij})$ is of the full column rank.

To get regularity we assume that $\mathbf{X}$ is nonempty, bounded with nondegenerated vertices. Further we assume asymptotic normality of the estimates y^N of the true parameter vector η. The differentiability properties of $g_0(x\,;\,\eta)$ in a neighbourhood of $[x(\eta)\ ,\ \eta]$ are implied by assuming that the marginal densities f_i are continuous and positive in neighbourhoods of the points $\left[\sum_{j=1}^{n} a_{ij}\, x_j(\eta)\,;\,\eta\right]$, $1 \le i \le m$, respectively.

Two types of parametric families will be considered:

3.5.1. y_i, $1 \le i \le m$ are ***location parameters***. Then we have for the nonzero components $x_j\ (\eta)$, $j\ \in J$ of the optimal solution $x(\eta)$

$$\frac{\partial x_J(\eta)}{\partial y} = -[I - C^{-1} P_J^T (P_J C^{-1} P_J^T)^{-1} P_J] CB ,$$

where

$$P_J = (p_{ij})_{\substack{1 \le i \le r \\ j \in J}} , C = -A^T QA , B = A^T Q$$

with

$$Q = diag \{q_i f_i \left[\sum_{j=1}^{n} a_{ij} x_j (\eta) ; \eta_i \right], 1 \le i \le m \} . \tag{3.5}$$

3.5.2. y_i, $1 \le i \le m$ are *scale parameters*, $y_i > 0\ \forall_i$. Then

$$\frac{\partial x_J(\eta)}{\partial y} = -[I - C^{-1} P_J^T (P_J C^{-1} P_J^T)^{-1} P_J] CB ,$$

where

$$C = -A^T QA , B = A^T Q\ diag \left[\frac{1}{\eta} \sum_{j=1}^{n} a_{ij} x_j (\eta) , 1 \le i \le m \right]$$

and Q is given by (3.5).

4. Contaminated distributions

Throughout this section, the functions g_i , $0 \le i \le m + p$ in (1.1) will be assumed to depend *linearly* on the distribution F. This assumption is evidently satisfied for the stochastic programs with recourse as well as for those with probabilistic constraints, and in all cases when g_i are expectations of suitable functions derived from h_i. Furthermore, we shall assume that $\mathbf{X} = R^n$, it means only that the original deterministic constraints and the induced ones have been incorporated into the explicit constraints in (1.1) (with $g_i(x\ ; F)$ independent of F, of course).

The local behaviour of the optimal solution $x(F)$ of the program (1.1) with respect to small changes of the distribution F can be studied via t-contamination of the distribution F by a suitably chosen distribution G, i.e., instead of F, distribution of the form

$$F_t = (1-t)F + tG \, , \quad 0 \le t \le 1 \tag{4.1}$$

will be considered. In (4.1), F_t is called ***distribution F t-contaminated by distribution G***. Due to our assumption, the original stability problem thus reduces to that linearly perturbed by a scalar parameter $t \in < 0 \, , 1 >$:

$$\begin{aligned} &\text{maximize } (1-t)\, g_0\,(x\,;F) + t g_0(x\,;G) \\ &\text{subject to } (1-t)\, g_i\,(x\,;F) + t g_i(x\,;G) \ge 0, \quad 1 \le i \le m \, , \\ &\qquad\qquad (1-t)\, g_i\,(x\,;F) + t g_i(x\,;G) = 0, \quad m+1 \le i \le m+p \, . \end{aligned} \tag{4.2}$$

In principle, it is possible to get the trajectory of the optimal solutions $x(F_t)$, $0 \le t \le 1$; for an appropriate method see e.g. Gfrerer et al. (1983).

We shall aim to obtain the Gâteaux differential $dx(F\,;G-F)$ of the optimal solution of (1.1) in the direction of $G-F$. To get the explicit results, one has to check the differentiability and regularity assumptions of Theorem 3.1 and to compute matrices B(0), D(0) corresponding to the contamination parameter $t = 0$.

The knowledge of the Gâteaux differential of $x(F)$ at F in the direction of $G-F$ is useful not only for the first order approximation of the optimal solutions corresponding to distributions belonging to a neighbourhood of F but also for deeper statistical conclusions on robustness, namely, in connection with statistical properties of the estimate $x(F_\nu)$ of $x(F)$, which is based on the empirical distribution F_ν. For fixed constraints in (4.2) and for the special choices $G = \delta_w$ (degenerated distribution concentrated at one point w), the Gâteaux differential $dx(F\,;\delta_w - F)$ corresponds to the influence curve $\Omega_F(w)$ widely used in asymptotic statistics. Different characteristics of $\Omega_F(w)$ suggested by Hampel (1974) measure the effect of contamination of the data by gross errors, the local effect of rounding or grouping of the observations, etc. For an example see Dupačová (1985a).

Theorem 4.1. *For the program*

$$\begin{aligned} &\text{maximize } g_0\,(x\,;F) \\ &\text{subject to } g_i\,(x\,;F) \ge 0, \quad 1 \le i \le m \, , \\ &\qquad\qquad g_i\,(x\,;F) = 0, \quad m+1 \le i \le m+p \end{aligned} \tag{4.3}$$

assume:

(i) $g_i(\cdot\,;F) : R^n \to R^1$ *are twice continuously differentiable*, $0 \le i \le m+p$,

(ii) *the Kuhn-Tucker conditions of the first and second order, the linear independence condition and the strict complementarity conditions are fulfilled for* $w(F) = [x(F)\,,\,u(F)\,,\,v(F)] \in R^n \times R^m_+ \times R^p$,

(iii) *there is a neighbourhood* $\vartheta(x(F)) \subset R^n$ *on which* $g_i(\cdot\,;G)\,,\,0 \le i \le m+p$ *are twice continuously differentiable.*

Then:

(a) *There is a neighbourhood* $\vartheta(w(F)) \subset R^n \times R^m_+ \times R^p$, *a real number* $t_0 > 0$ *and a continuous function* $w : \langle 0\,,\,t_0) \to \vartheta(w(F))$, $w(0) = w(F)$ *such that for any* $t \in \langle 0, t_0)$, $w(t) = [x(t)\,,\,u(t)\,,\,v(t)]$ *is the Kuhn-Tucker point of (4.2) for which the second order sufficient condition, the linear independence condition and the strict complementarity conditions are fulfilled.*

(b) *The Gâteaux differential* $dw(F\,;G-F)$ *of the Kuhn-Tucker point* $w(F)$ *of (4.3) in the direction of* $G-F$ *is given by*

$$dw_I(F\,;G-F) = -\begin{bmatrix} \nabla^2_{xx} L\,(w\,(F)\,;F) & \nabla_x\, g_I\,(x\,(F)\,;F) \\ [\,\nabla_x\, g_I\,(x\,(F)\,;F)\,]^T & 0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \nabla_x L\,(w\,(F)\,;G) \\ g_I\,(x\,(F)\,;G) \end{bmatrix} \quad (4.4)$$

The remaining components of $dw(F\,;\,G-F)$, *which correspond to the non-active constraints in (4.3), equal to 0.*

Proof is a straightforward application of Theorem 3.1. We took the liberty of adopting the notation to our case; namely

$$\nabla_x L\,(w\,(F)\,;G) = \nabla_x\, g_0\,(x\,(F)\,;G) + \sum_{i=1}^{m} u_i\,(F)\, g_i\,(x(F);G) + \sum_{i=m+1}^{p} v_i\,(F)\, g_i\,(x\,(F);G)\,,$$
$$w_I(F) = [x\,(F)\,,\,u_i\,(F)\,,\,i \in I(F),\, v\,(F)]$$

with $I(F) \subset \{1,\,...\,,m\}$ containing the indices of the active inequality constraints $g_i\,(x\,(F)\,;\,F) = 0$, $g_I(x\,(F)\,;\,G)$ and $g_I(x\,(F)\,;\,F)$ are vectors consisting of components $g_i\,(x\,(F)\,;\,G)$ and $g_i\,(x\,(F)\,;\,F)$ for $i \in I(F)\,,\,m+1 \le i \le m+p$, respectively,

$\nabla_x g_I (x (F) ; F)$ is a matrix consisting of columns $\nabla_x g_i (x (F) ; F)$ for $i \in I(F)$ and $m+1 \le i \le m+p$.

Due to the fact that (4.2) is linearly perturbed, we have

$$\nabla^2_{xt} L (w ; t) = \nabla_x L (w ; G) - \nabla_t L (w ; F) ,$$
$$\nabla_t g_I (x ; t) = g_I (x ; G) - g_I (x ; F) ,$$

so that

$$\nabla^2_{xt} L (w (F) ; t) = \nabla_x L (w (F) ; G)$$
$$\nabla_t g_I (x (F) ; t) = g_I (x (F) ; G) .$$

Remark 4.2. For fixed constrains in (4.2), i.e., for $g_i(x ; F)$ independent of F, we have evidently $g_i (x (F) ; G) = g_i (x (F) ; F) = 0$ for $i \in I(F)$ or $m+1 \le i \le m+p$ in (4.4). In the case of stochastic program with recourse

$$\text{maximize } g_0(x ; F)$$

on a set **X** described by fixed constraints $g_i(x) \ge 0$, $1 \le i \le m$, $g_i(x) = 0$, $m+1 \le i \le m+p$, we have thus

$$dw_I(F ; G-F) = -\begin{bmatrix} \nabla^2_{xx} L (w (F) ; F) & \nabla_x g_I (x (F)) \\ [\nabla_x g_I (x (F))]^T & 0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \nabla_x L (w (F) ; G) \\ 0 \end{bmatrix} \quad (4.5)$$

Theorem 4.3. *Let assumptions of Theorem 4.1 hold true.*

(a) *Let the matrix*

$$L = \nabla^2_{xx} L (w (F) ; F)$$

be nonsingular. Then the Gâteaux differential of the isolated local maximizer $x(F)$ of (4.3) in the direction of $G-F$ is given by

$$dx (F;G-F) = -C^{-1}\nabla_x L (w (F);G)$$

where

$$C^{-1} = [I - L^{-1}P(P^T L^{-1}P)^{-1}P^T]L^{-1}$$
$$P = \nabla_x g_I(x (F);F)$$

and I is the n-dimensional unit matrix.

(b) *Let the matrix* $P = \nabla_x g_I(x(F);F)$ *be of rank* n. *Then the Gâteaux differential of the isolated local maximizer* $x(F)$ *of (4.3) in the direction of* $G-F$ *is given by*

$$dx(F\,;\,G-F) = -(P^T)^{-1}\, g_I\,(x(F)\,;\,G)\,.$$

Proof follows from (4.4) by well known formulas for inversion of the matrix

$$D(0) = \begin{bmatrix} L & P \\ P^T & 0 \end{bmatrix}$$

which is nonsingular and which contains the nonsingular square submatrix L in the case of (a) or the nonsingular square submatrix P in case (b).

The assumptions of strict complementarity play an essential role in the proof of Theorem 4.1. They guarantee that the interval $< 0,\, t_0)$ on which $w(t)$ is the Kuhn-Tucker point of (4.2) is nonempty. Alternatively, the strict complementarity conditions can be replaced by the strong second order sufficient condition which was stated in Section 3. Thanks to the fact, that we have a scalar parameter only and that we are in fact interested in the right-hand derivatives of the optimal solution with respect to the parameter at the given point $t = 0$, the result of Jittorntrum (1984) applied to our problem gives the desired assertion on the Gâteaux differential.

Denote $I^+(F) = \{i \in I(F) : u_i(F) > 0\}$, $I^0(F) = \{i \in I(F) : u_i(F) = 0\}$. Under strict complementarity conditions, $I^0(F) = \phi$.

Theorem 4.4. *Let in assumptions of Theorem 4.1, strict complementarity conditions be replaced by the strong second order sufficient condition. Then:*

(a) *There is a neighbourhood* $\vartheta(w(F)) \subset R^n \times R^m_+ \times R^p$, *a real number* $t_0 > 0$ *and a continuous function* $w : < 0,t_0) \to \vartheta(w(F))$, $w(O) = w(F)$ *such that for any* $t \in < 0\,,\, t_0)$, $w(t) = [x(t)\,,\, u(t)\,,\, v(t)]$ *is the Kuhn-Tucker point of (4.2) for which the strong second order sufficient condition and the linear independence condition are fulfilled.*

(b) *There is a set R of indices such that*

$$I^+(F) \subset R \subset I^+(F) \cup I^0(F) = I(F) \,,$$

for which the nonzero components of the Gâteaux differential $dw(F\,;\,G-F)$ *are given by*

$$dw_R(F\,;\,G-F) = -\begin{bmatrix} \nabla^2_{xx} L\ (w\ (F)\,;F) & \nabla_x\, g_R\ (x\ (F)\,;F) \\ [\nabla_x\, g_R\ (x\ (F)\,;F)]^T & 0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \nabla_x L\ (w\ (F)\,;\,G) \\ g_R\ (x\ (F)\,;\,G) \end{bmatrix}.$$

Special cases 4.5. For specific decision models of stochastic programming, we can get correspondingly the specific form of the assumptions as well as the explicit formulas for the Gâteaux differentials. The assumptions can be subdivided into three categories:

(A) The basic model assumptions, including the absolute continuity of the distribution F.

(B) The general assumptions such as the existence of the Kuhn-Tucker point for which strict complementarity conditions are fulfilled.

(C) The assumptions of differentiability, the linear independence condition and the 2nd order sufficient condition, which can be fitted to the considered model.

In the following survey of results, we shall list mostly the form of the assumptions of the last category and we shall give explicit formulas for the reduced vectors of the Gâteaux differentials containing the nonzero components only. The full statements can be found in Dupačová (1983), (1985a), (1986a).

4.5.1. Simple recourse problem (Dupačová (1983))

$$\max_{x \geq 0} E_F \left\{ c^T x - \sum_{i=1}^{m} q_i \left(\sum_{j=1}^{n} a_{ij} x_j - \omega_i \right)^+ \right\} \qquad (4.6)$$

with $q_i > 0$, $1 \leq i \leq m$, $A = (a_{ij})$ and c given and F such that $E_F \omega$ exists.

Assumptions:

(i) Denote $J = \{j : x_j(F) > 0\}$; the matrix $A_J = (a_{ij})_{1 \le i \le m, \, j \in J}$ has full column rank.

(ii) The marginal densities f_i , $1 \le i \le m$, are continuous and positive at the points $X_i(F) = \sum_j a_{ij} \, x_j(F)$, $1 \le i \le m$, respectively.

(iii) G is an m-dimensional distribution whose marginal distribution functions G_i have continuous derivatives on neighbourhoods of the points $X_i(F)$, $1 \le i \le m$, respectively.

Gâteaux differential

$$dx_J\,(F\,;\,G-F) = (A_J^T KA_J)^{-1}\,(C_J - A_J^T k)\;,$$

where

$$c_J = (c_j)_{j \in J}\,,\; k = (k_i)\;,\; 1 \le i \le m\;,$$

with

$$k_i = q_i\; G_i\;(X_i(F))\;,\quad 1 \le i \le m\;,$$

and

$$K = diag\,\{q_i\; f_i(X_i(F))\;,\; 1 \le i \le m\;\}\;.$$

4.5.2. *Individual probabilistic constraints* (Dupačová (1986a))

$$\text{maximize } c(x) \qquad (4.7)$$
$$\text{subject to } P_F\{\, \sum_{j=1}^{n} a_{ij}\, x_j \ge \omega_i \,\} \ge \alpha_i\;,\quad 1 \le i \le m\;\}$$

with $\alpha_i \in (0\,,\,1)$, , $1 \le i \le m$ and $A = (a_{ij})$ given; a^i denotes the i-th row of A.

Assumptions:

(i) $c : R^n \to R^1$ is twice continuously differentiable.

(ii) The rank of $A_I = (a_{ij}\,)_{i \in I, \, 1 \le j \le n}$ equals to card $I(F)$ = card I.

(iii) The marginal densities f_i , $1 \le i \le m$, are continuously differentiable on neighbourhoods of $X_i(F) = \sum_j a_{ij}\, x_j(F)$, $1 \le i \le m$, respectively, and $f_i(X_i(F)) > 0$, $i \in I(F)$.

(iv) G is an m-dimensional distribution whose marginal distribution functions G_i are twice continuously differentiable on neighbourhoods of the points $X_i(F)$, $1 \le i \le m$, respectively.

(v) For all $l \in R^n$, $l \ne 0$, for which $A_I l = 0$, inequality $l^T \nabla^2_{xx} c(x(F)) l < 0$ holds true and the matrix

$$L = \nabla^2_{xx} L(w(F) ; F) = \nabla^2_{xx} c(x(F)) + \sum_{i \in I(F)} u_i(F) f_i'(X_i(F)) a^{iT} a^i$$

is nonsingular.

Gâteaux differential

$$dx(F ; G-F) = -L^{-1} A_I^T (A_I^T L^{-1} A_I^T)^{-1} f_I^{-1} (G_I - \alpha_I)$$

where

$$f_I = diag\ \{f_i\ (X_i(F)) ,\ i \in I(F)\ \} \tag{4.8}$$

$$G_I = [G_i(X_i(F))]_{i \in I(F)} \text{ and } \alpha_I = [\alpha_i]_{i \in I(F)} \ . \tag{4.9}$$

4.5.3. For the case of individual probabilistic constraints and a *linear* objective function $c(x) = c^T x$, substantially weaker assumptions can be used to get a result comparable with that of Theorem 4.3; see Dupačová (1986a).

Assumptions:

(i) $c(x) = c^T x$

(ii) The optimal solution $x(F)$ is unique and nondegenerated.

(iii) The marginal densities f_i , $1 \le i \le m$ are continuous and positive at the points $X_i(F)$, $1 \le i \le m$, respectively.

(iv) G is an m-dimensional distribution whose marginal distribution functions G_i have continuous derivatives on neighbourhoods of the points $X_i(F)$, $1 \le i \le m$, respectively.

Gâteaux differential

$$dx(F ; G-F) = -A_I^{-1} f_I^{-1} [G_I - \alpha_I] \tag{4.10}$$

where f_I , G_I , α_I are given by (4.8) and (4.9).

Comment. In the last case, the assumptions on the distributions are comparable with those for the simple recourse problem, which is quite natural. Contrary to the case of nonlinear objective function, $x(F)$ is the optimal solution of the linear program

$$\text{maximize } c^T x \text{ subject to } Ax \geq b \tag{4.11}$$

where $b_i = F_i^{-1}(\alpha_i)$. Similarly, $x(t)$ is the optimal solution of the linear program

$$\text{maximize } c^T x \text{ subject to } Ax \geq b_t ,$$

where $b_{it} = F_{it}^{-1}(\alpha_i)$, $1 \leq i \leq m$, are the quantities of the contaminated marginal distribution function

$$F_{it}(\kappa) = (1-t) F_i(\kappa) + tG_i(\kappa) .$$

Let us approximate $F_{it}^{-1}(\alpha_i)$ linearly (see e.g. Serfling (1980)):

$$F_{it}^{-1}(\alpha_i) = F_i^{-1}(\alpha_i) + t \frac{\alpha_i - G_i(F_i^{-1}(\alpha_i))}{f_i(F_i^{-1}(\alpha_i))} , \quad 1 \leq i \leq m$$

and approximate $x(t)$ by the optimal solution $\hat{x}(t)$ of the following linear parametric program:

$$\text{maximize } c^T x$$

$$\text{subject to } \sum_{j=1}^{n} a_{ij} x_j \geq F_i^{-1}(\alpha_i) + t \frac{\alpha_i - G_i(F_i^{-1}(\alpha_i))}{f_i(F_i^{-1}(\alpha_i))} , \; 1 \leq i \leq m .$$

Let $B = A_I^T$ be the optimal basis of the linear program dual to (4.11); then $x(F) = A_I^{-1} b_I$ and $X_i(F) = F_i^{-1}(\alpha_i)$, $i \in I$. According to our assumptions, $x(F)$ is unique and nondegenerated, so that B is optimal for t belonging to a neighbourhood of zero and

$$\hat{x}(t) = x(F) + t \; A_I^{-1} \left[\frac{\alpha_i - G_i(F_i^{-1}(\alpha_i))}{f_i(F_i^{-1}(\alpha_i))} \right]_{i \in I} \tag{4.12}$$

$$= x(F) + t \; dx(F ; G-F)$$

using the result (4.10).

4.5.4. One joint probabilistic constraint (Dupačová (1986a))

$$\text{maximize } c(x) \qquad (4.13)$$
$$\text{subject to } P_F\{Ax \geq \omega\} \geq \alpha$$
$$\text{with } \alpha \in (0,1),\ A = (a_{ij}) \text{ given.}$$

Assumptions:

(i) $c : R^n \to R^1$ is twice continuously differentiable.

(ii) There is a Kuhn-Tucker point $w(F) = [x(F)\,;\,u(F)]$ for (4.13) such that $u(F) > 0$ and the second-order sufficient condition is fulfilled.

(iii) In a neighbourhood of $X(F) := Ax(F)$, the distribution functions F and G are twice continuously differentiable and

$$A^T \nabla_X F(X(F)) \neq 0 .$$

Gâteaux differential

$dw(F\,;\,G-F) =$

$$= \begin{bmatrix} -L^{-1} A^T \{u(F) \left[\nabla_X G(X(F)) - \frac{l(G)}{l(F)} \nabla_X F(X(F))\right] + \frac{1}{l(F)} \nabla_X F(X(F)) \left[G(X(F)) - \alpha\right]\} \\ u(F) \left[1 - \frac{l(G)}{l(F)}\right] + \frac{1}{l(F)} \left[G(X(F)) - \alpha\right] \end{bmatrix}$$

where

$$L = \nabla^2_{xx} L\,(w\,(F)\,;\,F) = \nabla^2_{xx}\, c\,(x\,(F)) + u(F)\ A^T\, \nabla^2_{XX} F\,(X\,(F))A\ ,$$

$$l(G) = \nabla_X F\,(X\,(F))^T\ A\,L^{-1} A^T \nabla_X\, G\,(X\,(F))\ ,$$

$$l(F) = \nabla_X F\,(X\,(F))^T\ A\,L^{-1} A^T \nabla_X F\,(X\,(F))\ .$$

Comment. Having solved the original problem (4.13), we know $x(F)$ and we have to compute $u(F)$, L^{-1} and to evaluate $G(Ax(F)) - \alpha$, $\nabla_X G\,(Ax\,(F))$, $\nabla_X F\,(Ax\,(F))$ to get the Gâteaux differential. For a given $x(F)$, $u(F)$, F and G, it depends on the difference between the values of the distribution functions $F(Ax(F)), G(Ax(F))$ and on the relative differences of their gradients which are measured by $\frac{l(G)}{l(F)}$ and

$\nabla_X G(Ax(F)) - \frac{l(G)}{l(F)} \nabla_X F(Ax(F))$.

For the gradient of F we have

$$\nabla_X F(X) = f.\ \tilde{F}^{(1)}(X)$$

where $f = diag\{f_i(X_i),\ 1 \le i \le m\}$ and $\tilde{F}^{(1)}(X)$ is the m-vector of the conditional distribution functions $F(X^{(i)} \mid X_i)$, $1 \le i \le m$; here $X^{(i)}$ denotes the $(m-1)$-dimensional subvector of X in which the i-th component, X_i, was deleted. Similar formulas hold true for $\nabla_X G(X)$.

In Example 2.3, the two considered distributions F and G are multinormal ones and differ by their correlation matrices only. In this case, $g_i = f_i$, $1 \le i \le m$, the conditional distributions $F(X^{(i)} \mid X_i)$, $G(X^{(i)} \mid X_i)$, $1 \le i \le m$, are normal and

$$\nabla_X G(X) = f\ \tilde{G}^{(1)}(X)$$

with

$$\tilde{G}^{(1)}(X) = \left[G(X^{(i)} \mid X_i)\right]_{1 \le i \le m} .$$

These circumstances make the numerical evaluation of the Gâteaux differential realistic.

5. The minimax approach

Assume now that the set **X** of admissible solutions is defined by fixed constraints only and that the objective function $g_0(x\ ; F)$ is linear with respect to F (for a generalization to the nonlinear case see Gaivoronski (1985)). In this case, we can set

$$g_0(x\ ; F) = E_F f_0(x\ ; \omega),$$

where $f_0(x\ , \omega)$ may e.g. correspond to the difference $h_0(x\ ; \omega) - \varphi(x\ ; \omega)$ in the general stochastic program with recourse (1.3). Let **F** be a given set of distributions to which F is known to belong. (The case of the complete knowledge of the distribution corresponds to $\mathbf{F} = \{F\}$.) Consider the two-person zero-sum game

$$H = (\mathbf{X}, \mathbf{F}, g_0) \tag{5.1}$$

where $\mathbf{X}$ is the set of strategies of the first player, $\mathbf{F}$ is the set of strategies of the second player and g_0 is the pay-off function. Any optimal pure strategy of the first player in the game (5.1) will be called the ***minimax solution*** of stochastic program

$$\max_{x \in \mathbf{X}} g_0(x\,;F) \qquad \text{for } F \in \mathbf{F}\ . \tag{5.2}$$

Under quite general assumptions on $\mathbf{F}$, $\mathbf{X}$ and f_0, a minimax solution exists and

$$\sup_{x \in \mathbf{X}} \min_{F \in \mathbf{F}} g_0(x\,;F) = \min_{F \in \mathbf{F}} \sup_{x \in \mathbf{X}} g_0(x\,;F)$$

(See e.g. Žáčková (1966), Theodorescu (1969).)

To find a minimax solution means in general to solve an optimization problem

$$\text{maximize } \inf_{F \in \mathbf{F}} g_0(x\,;F) \text{ on the set } \mathbf{X}\,. \tag{5.3}$$

If the set $\mathbf{F}$ of distributions is defined, inter alia, by prescribed values of certain moments of the distributions $F \in \mathbf{F}$, it is possible to use general results of the moment problem to get

$$\inf_{F \in \mathbf{F}} g_0(x\,;F)$$

in a form suitable for further computations (see Dupačová (1977)). We shall outline the results of the moment problem briefly and we shall indicate their application in stochastic programming.

Let $\kappa := (\kappa_1, \ldots, \kappa_k) : \Omega \to R^k$, $\kappa_0 : \Omega \to R^1$ be Borel measurable mappings. Denote $\kappa(\Omega)$ the image of the set Ω under the mapping κ, by $Y := \operatorname{conv} \kappa(\Omega)$ the convex hull of $\kappa(\Omega)$ and assume that $\operatorname{int} Y \neq 0$. For $y \in \operatorname{int} Y$ denote by $\mathbf{F}_y$ the set of distributions of a random vector ω on $(\Omega, \mathbf{B})$ such that $\kappa_1, \ldots, \kappa_k$, κ_0 are integrable with respect to all elements $F \in \mathbf{F}$ and

$$E_F\, \kappa_i(\omega) = y_i\ ,\ 1 \le i \le k, \text{ for all } F \in \mathbf{F}_y\ . \tag{5.4}$$

The moment problem is to find

$$U(y) := \sup_{F \in \mathbf{F}_y} E_F \,\kappa_0(\omega) \tag{5.5a}$$

or

$$L(y) := \inf_{F \in \mathbf{F}_y} E_F \,\kappa_0(\omega) \;. \tag{5.5b}$$

Under the above assumptions,

$$L(y) = \sup_{d \in D} \{d_0 + \sum_{j=1}^{k} d_j \, y_j\} \tag{5.6}$$

where

$$D := \{d \in R^{k+1} : d_0 + \sum_{j=1}^{k} d_j \, \kappa_j(\omega) \le \kappa_0(\omega) \;\forall \omega \in \Omega\} \,.$$

In many important cases, e.g., for Ω compact, $\kappa_1, \ldots, \kappa_k$ continuous, κ_0 lower semi-continuous, the infimum (5.5b) and the supremum (5.6) are achieved. In this case, there exists a distribution $F^* \in \mathbf{F}$ and a vector $d^* \in D$ such that

$$L(y) = E_{F^*} \,\kappa_0(\omega) = d^*_0 + \sum_{j=1}^{k} d^*_j \, y_j \tag{5.7}$$

and for the given $y \in \text{int}\, Y$, problem (5.5b) reduces to

$$L(y) = \max_{d \in D} \{d_0 + \sum_{j=1}^{k} d_j \, y_j\} \;.$$

Evidently, as a function of the parameter y, $L(y)$ is convex. Parallel results can be given for the upper bound $U(y)$. It is important from the point of view of computation that F^* in (5.7) is a *discrete* distribution. The corresponding probability measure must evidently be concentrated in the points $\omega \in \Omega$, for which $d_0{}^* + \sum_{j=1}^{k} d_j{}^* \,\kappa_j(\omega) = \kappa_0(\omega)$. Denote

$$B(d^*) = \{u = \kappa(\omega) : d_0{}^* + \sum_j d_j{}^* \,\kappa_j(\omega) = \kappa_0(\omega) \,,\; \omega \in \Omega\} \,;$$

Then for almost all $y \in \text{int}\, Y$, there is a unique $d^* \in D$ such that $y \in \text{conv}\, B(d^*)$, i.e.

$$y = \sum_{i=1}^{l} p_i \,\kappa(\omega^i) \text{ with } \kappa(\omega^i) \in B(d^*) \,, 1 \le i \le l \;, \tag{5.8}$$

and

$$\sum_{i=1}^{l} p_i = 1 \, , \, p_i \geq 0 \, , \, 1 \leq i \leq l \, .$$

Corresponding to this representation,

$$L(y) = \sum_{i=1}^{l} p_i \, \kappa_0(\omega^i) = d_0^* + \sum_{j=1}^{k} d_j^* \, y_j \, .$$

For these and other related results see e.g. Kemperman (1968), the case of the inequality constraints on the moments was studied by Kemperman (1972), different approaches to the case of the noncompact Ω can be found e.g. in Richter (1957), Kemperman (1972), Cipra (1985), Gassman and Ziemba (1985).

When applying the above results to the minimax problem, it is quite natural to put $\mathbf{F} = \mathbf{F}_y$ and

$$\kappa_0(\omega) = f_0(x \; ; \; \omega) \, .$$

The dependence of f_0 on the decision variables x together with the final goal - to solve the "outer" maximization problem

$$\text{maximize} \inf_{F \in \mathbf{F}} E_F \, f_0(x \; ; \; \omega) \text{ on the set } \mathbf{X} -$$

are the reason why the direct application is possible only in special cases.

This will be the case if the set of the considered discrete distributions possessing the properties (5.8) would be relatively small and independent on x or if it would be possible to reduce the corresponding moment problem to finite number of one-dimensional moment problems. As an example of the first mentioned possibility, we have

Theorem 5.1. *Let $\Omega \subset R^k$ be a convex polyhedron with extreme points $\omega^1, \ldots, \omega^N$ and let $y \in \operatorname{int} \Omega$.*

Let $f_0 : \mathbf{X} \times \Omega \to R^1$ be a concave function of ω for any fixed $x \in \mathbf{X}$. Denote by $\mathbf{F}_y$ the set of distributions F for which

$$P_F(\omega \in \Omega) = 1, \; E_F \, \omega = y \quad \forall F \in \mathbf{F}_y \, .$$

Then

$$\min_{F \in \mathbf{F}_y} E_F f_0(x \,;\, \omega) = L(y)$$

where

$$L(y) := \max \sum_{\nu=1}^{N} p_\nu f_0(x \,;\, \omega^\nu)$$

$$\text{subject to} \sum_{\nu=1}^{N} p_\nu \omega_i^\nu = y_i \,,\; 1 \le i \le k \;,$$

$$\sum_{\nu=1}^{N} p_\nu = 1,\; p_\nu \ge 0,\; 1 \le \nu \le N \,.$$

The proof follows form the more general result with a piecewise concave function $f_0(x \,;\, \cdot)$ which was studied in Dupačová (1976); see also Dupačová (1980a).

The "worst" (i.e. the extreme) distribution $F^* \in \mathbf{F}_y$ is a discrete distribution concentrated on at most $k+1$ extreme points of Ω. The set $\mathbf{F}^* \subset \mathbf{F}_y$ of all distributions concentrated on at most $k+1$ extreme points of Ω is well specified and does not depend on x.

For f_0 separable in ω, i.e., for

$$f_0(x \,;\, \omega) = \sum_{i=1}^{k} f_{0i}(x \,;\, \omega_i)$$

where for arbitrary fixed x, $f_{0i}(x \,;\, \omega_i)$, $1 \le i \le k$, is a concave function of one-dimensional variable ω_i, and for $\Omega = \mathop{X}_{i=1}^{k} < \underline{\omega}_i \; \overline{\omega}_i >$, a version of the Edmundson -Madansky bound follows easily from Theorem 5.1 (see Dupačová (1977)):

$$\min_{F \in \mathbf{F}_y} E_F f_0(x \,;\, \omega) = \sum_{j=1}^{k} \lambda_j f_{0j}(x \,;\, \underline{\omega}_j) + \sum_{j=1}^{k} (1-\lambda_j) f_{0j}(x \,;\, \overline{\omega}_j) \tag{5.9}$$

with

$$\lambda_j := \frac{\overline{\omega}_j - y_j}{\overline{\omega}_j - \underline{\omega}_j} \,,\; 1 \le i \le k \,.$$

In addition, by Jensen's inequality

$$\max_{F \in \mathbf{F}_y} E_F f_0(x \,;\, \omega) = f_0(x \,;\, E_F\omega) = f_0(x \,;\, y) \,. \tag{5.10}$$

Repeated application of (5.9) and (5.10) to sets of conditional distributions with respect to subintervals of $\overset{k}{\underset{i=1}{X}} < \underline{\omega}_i , \overline{\omega}_i >$ possessing fixed conditional mean values leads to closer bounds (see Ben-Tal and Hochman (1972)). This idea has been successfully applied to algorithmic solution of the complete recourse problem by Kall and Stoyan (1982), see also Wets (1983), Birge and Wets (1983). From the just explained theoretical background it follows in addition, that the ***bounds are the best*** with respect to the considered set $\mathbf{F}_y$. It means that without an additional information about the distributions, these bounds cannot be improved.

The above results on the moment problem hold out the possibility of constructing bounds of more general type using all available information on the set **F**, formulated either in terms of higher order moment conditions or qualitative conditions like unimodality. The knowledge of these bounds enables one to draw conclusions on ***robustness*** of the optimal value of (5.2) with respect to the distributions belonging to **F** . On the other hand, the applicability of the more complicated bounds in approximation schemes for solving stochastic programs is limited by their numerical complexity.

From the point of view of real-life applications the case of distributions with given mean values and second order moments is quite typical. As an example of the related result, we have

Theorem 5.2. *Let* $\Omega \subset R^k$,

$$\mathbf{F} = \{F : E_F\, \omega_i = y_i \text{ , } \mathrm{var}_F\, \omega_i = \sigma_i^2 \text{ , } 1 \le i \le k \} \tag{5.11}$$

and

$$f_0(x \,;\, \omega) = \min_{1 \le j \le J} \{\varphi^j(x) + \omega^T f^j\}$$

where $f^j \in R^k$ *and* $\varphi^j : \mathbf{X} \to R^1$, $1 \le j \le J$, *are given and such that* $f_0(x \,;\, \omega)$ *is bounded from above on* $\mathbf{X} \times \Omega$.

Then $\min_{F \in \mathbf{F}_y} E_F f_0(x ; \omega)$ *equals to the maximal value of the linear program*

$$\mu_0 + \sum_{i=1}^{k} \mu_i y_i + \sum_{j=1}^{k} \mu_{ii} (\sigma_i^2 + y_i^2)$$

subject to conditions

$$\mu_{ii} > 0 , 1 \le i \le k ,$$

$$\frac{1}{4} \sum_{j=1}^{k} \frac{1}{\mu_{ii}} (f_i^j - \mu_i)^2 \ge \mu_0 - \varphi^j(x) , 1 \le j \le J .$$

For the details see Dupačová (1980b). The resulting program is a convex one, its form is suitable for stability analysis with respect to parameters y , σ^2 and with respect to the decision variables x.

Special case 5.3. Using the general approach for the *simple recourse problem*

$$\text{maximize} \min_{F \in \mathbf{F}} E_F \{c^T x - \sum_{i=1}^{m} q_i (\sum_{j=1}^{n} a_{ij} x_j - \omega_i)^+\} \text{ on a set } \mathbf{X}$$

with $\mathbf{F}$ given by (5.11), we evidently can write

$$\min_{F \in \mathbf{F}} E_F \{c^T x - \sum_{i=1}^{m} q_i (\sum_{j=1}^{n} a_{ij} x_j - \omega_i)^+\} =$$
$$c^T x - \sum_{i=1}^{m} q_i \max_{F \in \mathbf{F}_i} (\sum_{j=1}^{n} a_{ij} x_j - \omega_i)^+$$

where $\mathbf{F}_i$ is the set of marginal distributions F_i corresponding to $F \in \mathbf{F}$ and as a result, we get an explicit formula

$$\min_{F \in \mathbf{F}} E_F \{c^T x - \sum_{i=1}^{m} q_i (\sum_{j=1}^{n} a_{ij} x_j - \omega_i)^+\} =$$

$$c^T x - \sum_{i=1}^{m} \frac{q_i}{2} \{ \sqrt{\sigma_i^2 + (y_i - \sum_{j=1}^{n} a_{ij} x_j)^2} + \sum_{j=1}^{n} a_{ij} x_j - y_i \} .$$

For the *proof* see Dupačová (1977) , (1980a), Jagannathan (1977).

The complexity of the problem substantially increases if the information about covariances is considered. A similar situation is well known both in stochastic linear programs with recourse (e.g., the case of simple versus complete recourse) and in probabilistic programming (e.g., the case of individual versus joint probabilistic constraints).

A method of solving problem (5.3) for the set $\mathbf{F}$ described by general moment conditions was suggested in Ermoliev, Gaivoronski and Nedeva (1985). Using the fact that the extreme distribution $F \in \mathbf{F}$ can be found among the discrete distributions together with the corresponding duality relations helps to reformulate the problem

$$\max_{x \in \mathbf{X}} \min_{F \in \mathbf{F}} E_F f_0(x ; \omega) ,$$

with f_0 concave in x,

$$\mathbf{F} = \{F : P_F(\Omega) = 1 , E_F \kappa_i(\omega) \leq y_i , 1 \leq i \leq k \}$$

and $\mathbf{X}$, Ω convex, compact, into the following one

$$\max_{x \in \mathbf{X}, \mu \in R_+^k} \min_{\omega \in \Omega} \left\{ f_0(x ; \omega) - \sum_{i=1}^{k} \mu_i [\kappa_i(\omega) - y_i] \right\}$$

which is solvable by means of stochastic quasigradient methods or other methods suitable for nondifferentiable optimization.

In many real-life situations, one has at their disposal besides knowledge of some moments, a qualitatively different information about the distribution such as its unimodality. We shall see, that it is often possible to remove this additional condition by a suitable transformation and to reduce the problem to the original moment problem.

Definition 5.4. A random variable ω is ***unimodal*** if there is a number y_0 such that the distribution function of ω is convex on $(-\infty , y_0 >$ and concave on $< y_0 , + \infty)$. Any number y_0 with the given property is called the ***mode*** of the distribution.

Denote by $\mathbf{F}_y[y_0]$ the set of unimodal distributions on R^1 with the given mode y_0 such that for all $F \in \mathbf{F}_y[y_0]$, the moment conditions (5.4) are fulfilled. Define a transformation T on Borel measurable functions Φ which are integrable over any closed sub-interval of $(-\infty , y_0)$ and $(y_0 + \infty)$ as follows:

$$\Phi^*(u) := (T\Phi)(u) = \frac{1}{u - y_0} \int_{y_0}^{u} \Phi(v)\, dv \quad \text{for } u \neq y_0 \tag{5.12}$$

$$= \Phi(u) \qquad\qquad \text{for } u = y_0 \, .$$

Denote further by $\hat{\mathbf{F}}$ the set of all distributions such that

$$E_F\, T(|\kappa_j|) < \infty \, , 0 \le j \le k \; , \; \forall F \in \hat{\mathbf{F}} \, .$$

Then

$$\inf_{F \in \mathbf{F}_y[y_0]} E_F\, \kappa_0(\omega) = \inf \{E_F\, \kappa_0^*(\omega) : F \in \hat{\mathbf{F}} \, , E_F\, \kappa_j^*(\omega) = y_j \; , 1 \le j \le k\} \tag{5.13}$$

The *transformed* moment problem (5.13) has the form of the classical moment problem (5.4), (5.5b), the number of moment conditions remains unchanged whereas the qualitative condition of unimodality does not appear any more. All extremal points of the set $\mathbf{F}_y\, [y_0]$ are mixtures of uniform distributions over $(u \, , y_0)$ or $(y_0 \, , u)$ and of the degenerated distribution concentrated at y_0. (For these and related results see Cipra (1978).)

For application to the minimax problem (5.3), the following result (Dupačová (1977)) is important:

Theorem 5.5. *Let Φ be a real valued concave (convex) function on R^1 that is integrable over any closed subinterval of $(-\infty \, , y_0)$ and $(y_0 \, , +\infty)$. Then $\Phi^* = T\Phi$ defined by (5.12) is concave (convex).*

Special case 5.6. Let

$$f_0(x \; ; \omega) = \sum_{i=1}^{k} f_{0i}\, (x \; ; \omega_i) \, ,$$

where for an arbitrary fixed $x \in \mathbf{X}$, $f_{0i}(x \; ; \cdot)$, $1 \le i \le k$ is a concave function of the one-dimensional variable ω_i. Let ω_i be independent random variables whose distributions F_i are unimodal with a given mode y_{i0}, given range $<\underline{\omega}_i \, , \overline{\omega}_i>$ and mean value y_i , $1 \le i \le k$. Solving the k-dimensional moment problem reduces to solving one-dimensional one of the following type:

$$\inf_{F \in \mathbf{F}_y[y_0]} E_F f_0(x\,;\,\omega)$$

where $\omega \in \langle \underline{\omega}, \overline{\omega} \rangle$, the distribution F is unimodal with a given mode y_0 and fulfills the moment condition $E_F w = y$ and $f_0(x\,;\,\cdot)$ is concave in ω for any fixed $x \in \mathbf{X}$. Using (5.12) for $\kappa_1(\omega) = \omega$ we get

$$\kappa_1^*(u) = \frac{1}{u - y_0} \int_u^{y_0} \omega \, d\omega = \frac{u + y_0}{2} \text{ for } u \neq y_0$$

$$= u = \frac{u + y_0}{2} \quad \text{for } u = y_0 \ .$$

The moment condition in the transformed moment problem (5.13) reads

$$E_F\, \kappa_1^*(\omega) = E_F \left\{ \frac{\omega + y_0}{2} \right\} = y$$

so that

$$E_F\, \omega = 2y - y_0 \ .$$

Using now the concavity property of $f_0^*(x\,;\,u)$ together with (5.9), (5.12) and (5.13), we get for $y_0 \in \operatorname{int} \langle \underline{\omega}, \overline{\omega} \rangle$

$$\inf_{F \in \mathbf{F}_y[y_0]} E_F f_0(x\,;\,\omega) = \lambda f_0^*(x\,;\,\underline{\omega}) + (1 - \lambda) f_0^*(x\,;\,\overline{\omega}) =$$

$$= \lambda \frac{1}{y_0 - \underline{\omega}} \int_{\underline{\omega}}^{y_0} f_0(x\,;\,\omega)\, d\omega + (1 - \lambda) \cdot \frac{1}{\overline{\omega} - y_0} \int_{y_0}^{\overline{\omega}} f_0(x\,;\,\omega)\, d\omega \ ,$$

where

$$\lambda = \frac{\overline{\omega} - 2y + y_0}{\overline{\omega} - \underline{\omega}} \ .$$

The "worst" distribution $F^* \in \mathbf{F}_y[y^0]$ possesses the density

$$p(\omega) = \frac{\overline{\omega} - 2y + y_0}{(\overline{\omega} - \underline{\omega})(y_0 - \underline{\omega})} \quad \text{for } \omega \in \langle \underline{\omega}, y_0)$$

$$= \frac{2y - y_0 - \underline{\omega}}{(\overline{\omega} - \underline{\omega})(\overline{\omega} - y_0)} \quad \text{for } \omega \in (y_0, \overline{\omega} \rangle$$

or in other words, it is the mixture of two uniform distributions over $\langle \underline{\omega}, y_0)$ and

$(y_0, \bar{\omega} >$ with the weights λ and $1 - \lambda$, respectively.

In the case of an unknown mode, its position can be found in an optimal way (i.e., giving the "worst" distribution again). As a result we have (see Dupačová (1977)):

Theorem 5.7. *Let $\mathbf{F}_y^0$ be a set of unimodal distributions (with an unknown mode) with a fixed support $<\underline{\omega}, \bar{\omega}>$ and a given mean value y. Let κ_0 be concave on $<\underline{\omega}, \bar{\omega}>$. Then the minimum of $E_F\, \kappa_0(\omega)$ for $F \in \mathbf{F}_y^0$ is attained*

(i) *for the uniform distribution F^* over $<\underline{\omega}, \bar{\omega}>$, if $\frac{1}{2}(\bar{\omega} + \underline{\omega}) = y$,*

(ii) *for the distribution $F_1 \in \mathbf{F}_y^0$, which is a mixture of F^* and of the degenerated distribution concentrated at $\underline{\omega}$ if $\frac{1}{2}(\bar{\omega} + \underline{\omega}) > y$,*

(iii) *for the distribution $F_2 \in \mathbf{F}_y^0$, which is a mixture of F^* and of the degenerated distribution concentrated at $\bar{\omega}$ if $\frac{1}{2}(\bar{\omega} + \underline{\omega}) < y$.*

These results are directly applicable in stochastic programming, see e.g. Klein Haneveld (1984) for their application to project planning under incomplete knowledge of distribution: Instead of stochastic program with recourse described briefly in Example 2.4 he solves the problem

$$\text{minimize} \sup_{F \in \mathbf{F}} \{cT + q\, E_F\, [\tau(\omega) - T]^+\}$$

where $\mathbf{F}$ is set of distributions whose marginal distributions are unimodal with a given support and a given mean value.

To summarize - the results on the minimax approach can be used

(i) to approach the specific problems of stochastic programming with an incomplete information about the distribution of the random parameters,

(ii) for computing minimax (maximax) bounds on the optimal value of the objective function; these bounds are of the type

$$L_{\mathbf{F}} = \max_{x \in \mathbf{X}} \inf_{F \in \mathbf{F}} E_F f_0(x ; \omega)$$

and

$$U_{\mathbf{F}} = \max_{x \in \mathbf{X}} \sup_{F \in \mathbf{F}} E_F f_0(x ; \omega)$$

and for any $F \in \mathbf{F}$, we have inequalities

$$L_{\mathbf{F}} \leq \max_{x \in \mathbf{X}} E_F f_0(x , \omega) \leq U_{\mathbf{F}} .$$

The knowledge of the bounds $L_{\mathbf{F}}$ and $U_{\mathbf{F}}$, provides an information about robustness of the optimal value of the objective function $E_F f_0(x ; \omega)$ with respect to distributions $F \in \mathbf{F}$. With an increasing information about $\mathbf{F}$ we get narrower bounds. A special type of the bounds has been successfully used in approximation schemes for numerical solution of the stochastic programs (with the complete knowledge of the distribution).

(iii) It seems reasonable to use the worst distribution $F^* \in \mathbf{F}$ in place of the distribution G in the contamination method (see Section 4) at least in cases when there is no evidence for using any other distribution. The resulting objective function

$$(1 - t) E_F f_0(x ; \omega) + t E_{F^*} f_0(x ; \omega)$$

corresponds to the Hodges-Lehman decision rule (see e.g. Schneeweiss (1967)) and the optimal solutions can be related to the "restricted Bayes strategies" of Nadeau and Theodorescu (1980).

In all the cases mentioned, the advantage of the minimax approach is relatively easy computability of the resulting program

$$\max_{x \in \mathbf{X}} E_{F^*} f_0(x ; \omega)$$

at least in the cases when the "worst" distribution F^* or $E_{F^*} f_0(x ; \omega)$ can be determined explicitly.

References

Armacost, R.L., and A.V. Fiacco (1974): Computational experience in sensitivity analysis for nonlinear programming. *Mathematical Programming* **6**, 301–326.

Ben-Tal, A., and E. Hochman (1972): More bounds on the expectation of a convex function of a random variable, *J. Appl. Prob.* **9**, 803–812.

Birge, J. and R.J-B. Wets (1983): Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse. IIASA, Laxenburg, Austria. WP-83-111.

Cipra, T. (1978): A class of unimodal distributions and its transformations. *Čas. pěst. mat.* **203**, 17–26.

Cipra, T. (1985): Moment problem with given convariance structure in stochastic programming, *Ekonomicko-matematický obzor* **21**, 66–77.

Dempster, M. (1980) (ed.): Stochastic programming. Academic Press, London.

Dupačová, J. (1976): Minimax stochastic programs with nonconvex nonseparable penalty functions. In: A. Prékopa (ed.): Coll. Math. Soc. J. Bólyai 12. Progress in Operations Research, Eger 1974. J. Bólyai Math. Soc. a North-Holland, 303–316.

Dupačová, J. (1977): Minimaxová úloha stochastického lineárního programování a momentový problém. *Ekonomicko-matematický obzor* **13**, 279–307.

Dupačová, J. (1978): Minimax approach to stochastic programming and the moment problem. Selected results. ZAMM **58**, T466–467.

Dupačová, J. (1980): Minimax stochastic programs with nonseparable penalties. In: K. Iracki, K. Matanowski, S. Walukiewicz (eds.): Optimization techniques. (Proc. IXth IFIP Conference, Warszawa 1979) Part I. Lecture notes in control and information sciences 22, Springer, Berlin, 157–163.

Dupačová, J. (1979): Experience in stochastic programming models. In: A. Prékopa (ed.): Survey of Math. Programming. Proc. IXth. Int. Math. Progr. Symp., Budapest 1976. Akadémiai Kiaidó, Budapest, 99–105.

Dupačová, J. (1980): On minimax decision rule in stochastic programming. In: A. Prékopa (ed.): Math. Methods of Oper. Research 1. Studies on Math. Programming, Mátrafüred 1975. Akadémiai Kiaidó, Budapest, 38–48.

Dupačová, J. (1983): Stability in stochastic programming with recourse. *Acta Univ. Carol., Math. et Phys.* **24**, 23–34.

Dupačová, J. (1984): Stability in stochastic programming with recourse – estimated parameters. *Mathematical Programming* **28**, 23–34.

Dupačová, J. (1985a): Stability in stochastic programming with recourse – contaminated distributions. *Mathematical Programming Studies* – to appear.

Dupačová, J. (1985b): Minimax approach to stochastic programming and an illustrative application. *Stochastics* – to appear.

Dupačová, J. (1986a): Stability in stochastic programming – probabilistic constraints. In: Arkin, V.I., A. Shiriaev, R. Wets (eds.): Proceedings of International Conference on Stochastic Optimization, Kiev. Lecture Notes in Control and Information Sciences, Springer-Verlag, Berlin 1986 – to appear.

Dupačová, J. (1986b): On some connections between parametric and stochastic programming. Proc. of Int. Conference "Parametric optimization and related topics", Academic Verlag, Berlin – to appear.

Ermoliev, Yu., A. Gaivoronski and C. Nedeva (1985): Stochastic optimization problems with incomplete information on distribution function, *SIAM J. Control and Optimization* **23**, 696–716.

Fiacco, A.V. (1976): Sensitivity analysis for nonlinear programming using penalty methods. *Mathematical Programming* **10**, 287-311.

Gaivoronski, A. (1985): Linearization methods for optimization of functionals which depend on probability measures, *Math. Progr. Studies* – to appear.

Garstka, S.J. (1974): Distribution functions in stochastic programs with recourse: a parametric analysis. *Mathematical Programming* **6**, 339–351.

Gfrerer, H., J. Guddat and Hj. Wacker (1983): A globally convergent algorithm based on imbedding and parametric optimization. *Computing* **30**, 255–252.

Gassmann, H. and W.T. Ziemba (1985): A tight upper bound for the expectation of a convex function of a multivariate random variable. *Math. Progr. Studies* – to appear.

Hampel, F.R. (1974): The influence curve and its role in robust estimation. *Journal of the American Statistical Association* **69**, 383–397.

Jagannathan, R. (1977): Minimax procedure for a class of linear programs under uncertainty. *Oper. Res.* **25**, 173–177.

Jittorntrum, K. (1984): Solution point differentiability without strict complementarity in nonlinear programming. *Mathematical Programming Study* **21**, 127–138.

Kall, P. (1979): Computational methods for solving two-stage stochastic linear programming problems. *Z. angew. Math. Phys.* **30**, 261–271.

Kall, P. and D. Stoyan (1982): Solving stochastic programming problems with recourse including error bounds. *Math. Operationsforsch. Statist., Ser. Optimization* **13**, 431–447.

Kemperman, J.H.B. (1968): The general moment problem: A geometric approach. *Ann. Math. Statist.* **39**, 93–122.

Kemperman, J.H.B. (1972): On a class of moment problems. In: Proc. of the VIth Berkeley Symposium on Math. Stat and Probability, II. Univ. of California Press, 101–126.

Klein Haneveld, W.K. (1984): Robustness against dependence in PERT: An application of duality and distributions with known marginals. Research report 84-02-OR, Uni. Groningen.

Madansky, A. (1959): Bounds on the expectation of a convex function of a multivariate random variable. *Ann. Math. Statist.* **30**, 743–746.

Madansky, A. (1960): Inequalities for stochastic linear programming problems, *Manag. Sci.* **6**, 197–204.

Nadeau, R. and R. Theodorescu (1980): Restricted Bayes strategies for programs with simple recourse. *Operations Research* **28**, 777–784.

Panne, C. van de and W. Popp (1963): Minimum-cost cattle feed under probabilistic protein constraints. *Management Science* **9**, 405–430.

Prékopa, A., I. Deák, S. Ganczer and K. Patyi (1980): The STABIL stochastic programming model and its experimental application to the electrical energy sector of the Hungarian economy. In: M. Dempster (ed.): Stochastic Programming. Academic Press, London, 369–385.

Rao, C.R. (1973): Linear statistical inference and its applications. Wiley, New York.

Richter, H. (1957): Parameterfreie Abschätzung und Realisierung von Erwartungswerten. *Bl. Dtsch. Ges. Versicherungsmath.* **3**, 147–162.

Robinson, S.M. (1974): Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear programming algorithms. *Mathematical Programming* **7**, 1–16.

Robinson, S.M. (1980): Strongly regular generalized equations. *Mathematics of Operations Research* **5**, 43–62.

Robinson, S.M. (1984): Local structure of feasible sets in nonlinear programming, Part III: Stability and sensitivity. Comp. Sci. Technical Report #552, University of Wisconsin – Madison.

Schneeweiss, H. (1967): Entscheidungskriterien bei Risiko. Springer, Berlin.

Serfling, R.J. (1980): Approximation theorems of mathematical statistics. Wiley, New York.

Seppälä, Y. (1975): On a stochastic multi-facility location problem. *AIIE Trans.* **7**, 56–62.

Theodorescu, R. (1969): Minimax solutions of random convex programs. *Atti Acad. Naz. Lincei*, Ser. **8**, 46, 689–692.

Wets, R.J-B. (1983): Stochastic programming: Solution techniques and approximation schemes. In: A. Bachem, M. Grötschel, B. Korte (eds.): Mathematical Programming – The state of the art, Bonn 1982. Springer, Berlin, 566–603.

Žáčková, J. (1966): On minimax solutions of stochastic linear programming problems. *Čas. pěst. mat.* **91**, 423–430.