NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

DUALITY RELATIONS AND NUMERICAL METHODS FOR OPTIMIZATION PROBLEMS ON THE SPACE OF PROBABILITY MEASURES WITH CONSTRAINTS ON PROBABILITY DENSITIES

Yuri Ermoliev
Alexei Gaivoronski

June 1984
WP-84-46

Working Papers are interim reports on work of the International Institute for Applied Systems Analysis and have received only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute or of its National Member Organizations.

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

PREFACE

In this paper, the authors look at some quite general optimization problems on the space of probabilistic measures. These problems originated in mathematical statistics but have applications in several other areas of mathematical analysis. The authors extend previous work by considering a more general form of the constraints, and develop numerical methods (based on stochastic quasigradient techniques) and some duality relations for problems of this type.

This paper is a contribution to research on stochastic optimization currently underway within the Adaptation and Optimization Project.

ANDRZEJ WIERZBICKI
Chairman
System and Decision Sciences

DUALITY RELATIONS AND NUMERICAL METHODS FOR OPTIMIZATION PROBLEMS ON THE SPACE OF PROBABILITY MEASURES WITH CONSTRAINTS ON PROBABILITY DENSITIES

Yuri Ermoliev and Alexei Gaivoronski

1. INTRODUCTION

This paper is concerned with some quite general optimization problems on the space of probabilistic measures which originated in mathematical statistics but which also have applications in other areas of numerical analysis and optimization.

Assume that we have a set Y which belongs to Euclidean space R^n; let $B(Y)$ denote the Borel field of subsets of Y. Consider two finite positive Borel measures $H^-(y)$ and $H^+(y)$. We shall investigate the following optimization problem:

$$\max_H \Psi^0(H) \tag{1}$$

$$\Psi^i(H) \leq 0 \tag{2}$$

$$\int_A \mathrm{d}H^-(y) \leq \int_A \mathrm{d}H(y) \leq \int_A \mathrm{d}H^+(y) \quad \text{for any} \quad A \in B(Y) \tag{3}$$

$$\int_Y \mathrm{d}H(y) = 1 \quad . \tag{4}$$

Here the $\Psi^i(H)$, $i = \overline{0,m}$, are functions which depend on measure H, with properties which are specified below. If measures H^+ and H^- have densities $H_y^+(y)$ and $H_y^-(y)$, respectively, then constraint (3) becomes

$$H_y^-(y) \leq H_y(y) \leq H_y^+(y) \quad ,$$

where $H_y(y)$ is the (unknown) density of measure H. There are some special cases of this problem, notably when the following three conditions hold simultaneously: (i) functions $\Psi^i(H)$, $i = \overline{0,m}$, are linear with respect to H, i.e.,

$$\Psi^i(H) = \int_Y q^i(y)\mathrm{d}H(y) \quad .$$

(ii) functions $q^i(y)$ form a Tchebycheff system, and (iii) constraint (3) is either nonexistent or assumes the form $C^- \le H_y(y) \le C^+$. In this case the problem can be treated analytically and provides the subject of moment theory (see [1–3] for more information on this topic). Special duality relations, numerical methods for solving (1)–(4) without constraint (3) and various applications to stochastic optimization problems have been described in [4–9].

The purpose of this paper is to develop numerical methods and some duality theory for (1)–(4) with constraints of general form (3). Let us first consider one example from statistics in which constraint (3) plays an important role. The model under consideration is known as *finite population sampling* [10,11] and has much in common with optimal experimental design [12–14]. Suppose that we have a collection S of N objects labeled $i = 1,\ldots,N$. Each object is described by two variables y_i and x_i, where y_i is known for all i and x_i can be estimated through observations z_i using the expression $z_i = x_i + \tau_i$, where τ_i is random noise. It is usually assumed that $x_i = \psi^T(y)\vartheta$, where $\psi(y) = (\psi_1(y),\ldots,\psi_m(y))$ are known functions and $\vartheta = (\vartheta_1,\ldots,\vartheta_m)$ are unknown parameters. The problem is to choose a subset $s \subset S$ containing n objects in such a way as to get the best possible estimate of parameters ϑ given observations z_i, $i \in S$. Measure H^+ can be associated with the initial distribution of points y_i, $i = \overline{1,N}$, and measure H with the subset s to be found. The variance matrix of the best linear estimate of parameters ϑ in the case where all the τ_i are independent and have the same variance becomes (after substitution) proportional to matrix M, where

$$M^{-1} = \int \psi(y)\psi^T(y)\,\mathrm{d}H(y)$$

and the problem reduces to that of minimizing some function of this matrix, for instance, its determinant:

$$\min_H \det(M)$$

$$\int_A \mathrm{d}H(y) \le \int_A \mathrm{d}H^+(y)$$

for any Borel $A \subset Y$.

This problem is exactly of type (1)–(4); constraints (2) may express, for instance, limitations on the variance of the optimal plan.

2. THE LINEAR PROBLEM

We shall begin with duality relations and the characterization of optimal distributions for the following linear problem:

$$\max_{H} \int_{Y} q^0(y)\mathrm{d}H(y) \tag{5}$$

$$\int_{Y} q^i(y)\mathrm{d}H(y) \le 0\,, \quad i = \overline{1,m} \tag{6}$$

$$\int_{A}\mathrm{d}H^-(y) \le \int_{A}\mathrm{d}H(y) \le \int_{A}\mathrm{d}H^+(y) \quad \text{for any Borel } A \subset Y \subset R^n \tag{7}$$

$$\int_{Y}\mathrm{d}H(y) = 1 \ . \tag{8}$$

Let us first consider the case in which there are no constraints (6). Define

$$Z^0(c,f) = \{y : y \in Y, \quad f(y) = c\}$$

$$Z^+(c,f) = \{y : y \in Y, \quad f(y) > c\}$$

$$Z^-(c,f) = \{y : y \in Y, \quad f(y) < c\}$$

for some function $f(y)$ and let

$$c^* = \inf\{c : \int_{Z^+(c,q^0)} \mathrm{d}(H^+(y) - H^-(y)) \le 1 - \int_{Y}\mathrm{d}H^-(y)\} \ .$$

The following lemma gives all possible solutions of problem (5),(7),(8):

Lemma 1. *Suppose that* $H^*(y)$ *is a solution of problem* (5),(7),(8) *and*

1. $H^+(y)$ *and* $H^-(y)$ *are positive Borel measures such that* $\infty > \int_{Y}\mathrm{d}H^+(y) \ge 1$, $\int_{Y}\mathrm{d}H^-(y) \le 1$, *where* Y *is a compact set in* R^n
2. *Function* $q^0(y)$ *is continuous.*

   Then

(i) $\int_{A}\mathrm{d}H^*(y) = \int_{A}\mathrm{d}H^+(y)$ *for any Borel* $A \subset Z^+(c^*,q^0)$

(ii) $\int_{A}\mathrm{d}H^+(y) \ge \int_{A}\mathrm{d}H^*(y) \ge \int_{A}\mathrm{d}H^-(y)$ *for any Borel* $A \subset Z^0(c^*,q^0)$ *and*

$$\int_{Z^0(c^*,q^0)} \mathrm{d}H^*(y) = 1 - \int_{Z^+(c^*,q^0)} \mathrm{d}H^+(y) - \int_{Z^-(c^*,q^0)} \mathrm{d}H^-(y)$$

(iii) $\int_A dH^*(y) = \int_A dH^-(y)$ *for any Borel* $A \subset Z^-(c^*, q^0)$.

Proof. We may assume without loss of generality that $q^0(y) \geq 0$. If this is not the case we may take $\tilde{q}^0(y) = q^0(y) - \min_{y \in Y} q^0(y)$ instead of $q^0(y)$, which will not affect the optimal distribution. Let us first show that a measure H with properties (i)–(iii) exists. From the continuity of the function $q^0(y)$, the sets $Z^+(c, q^0)$ and $Z^-(c, q^0)$ are open with respect to the set Y while the set $Z^0(c, q^0)$ is closed. Thus, for arbitrary Borel set $A \subset Y$ we have $A = A_+ \cup A_0 \cup A_-$, where $A_+ \subset Z^+(c^*, q^0)$, $A_0 \subset Z^0(c^*, q^0)$, $A_- \subset Z^-(c^*, q^0)$, and sets A_+, A_0, A_- are measurable. Therefore any measure on Y is fully defined by its values on subsets of $Z^+(c^*, q^0)$, $Z^0(c^*, q^0)$, $Z^-(c^*, q^0)$. From the definition of $Z^+(c, q^0)$ we have:

$$Z^+(c^*, q^0) = \bigcup_{c > c^*} Z^+(c, q^0)$$

and $Z^+(c_1, q^0) \subset Z^+(c_2, q^0)$ for all $c_1 > c_2$. This gives

$$\lim_{c \downarrow c^*} \int_{Z^+(c, q^0)} dH^+(y) = \int_{Z^+(c^*, q^0)} dH^+(y)$$

and therefore

$$1 - \int_{Z^+(c^*, q^0)} dH^+(y) - \int_{Z^-(c^*, q^0)} dH^-(y) - \int_{Z^0(c^*, q^0)} dH^-(y) \geq 0$$

from the definition of c^*.

Now consider the sequence $c_s < c^*$, $c_{s+1} \geq c_s$, $c_s \to c^*$. We have the following relations:

$$Z^-(c_s, q^0) \subseteq Z^-(c_s, q^0) \cup Z^0(c_s, q^0) \subseteq Z^-(c_{s+1}, q^0)$$

$$\bigcup_s Z^-(c_s, q^0) = Z^-(c^*, q^0) \quad .$$

Considering the finite positive Borel measure $\bar{H} = H^+ - H^-$ we obtain

$$\lim_{s \to \infty} \int_{Z^-(c_s, q^0)} d\bar{H}(y) = \int_{Z^-(c^*, q^0)} d\bar{H}(y)$$

$$\int_{Z^-(c_s, q^0)} d\bar{H}(y) \leq \int_{Z^-(c_s, q^0) \cup Z^0(c_s, q^0)} d\bar{H}(y) \quad .$$

Taking into account the finiteness of the measure $\overline{H}$ we obtain:

$$\int_{Z^0(c^*,q^0)\cup Z^+(c^*,q^0)} d\overline{H}(y) = \int_Y d\overline{H}(y) - \int_{Z^-(c^*,q^0)} d\overline{H}(y) = \lim_{s\to\infty} \{\int_Y d\overline{H}(y) - \int_{Z^-(c_s,q^0)} d\overline{H}(y)\} \geq$$

$$\lim_{s\to\infty} \{\int_Y d\overline{H}(y) - \int_{Z^-(c_s,q^0)\cup Z^0(c_s,q^0)} d\overline{H}(y)\} = \lim_{s\to\infty} \int_{Z^+(c_s,q^0)} d\overline{H}(y) \quad .$$

From the definition of c and the fact that $c_s < c^*$ we have

$$\int_{Z^+(c_s,q^0)} d\overline{H}(y) > 1 - \int_Y dH^-(y)$$

which gives

$$1 - \int_{Z^+(c^*,q^0)} dH^+(y) - \int_{Z^-(c^*,q^0)} dH^-(y) - \int_{Z^0(c^*,q^0)} dH^+(y) \leq 0 \quad .$$

All of this proves that (i)–(iii) do not contradict each other and that there is some positive Borel measure $\tilde{H}$ which satisfies (i)–(iii) and also constraints (7) and (8). Now let H' be an arbitrary positive Borel measure which satisfies constraints (7) and (8). Suppose that for this measure there is some set $A' \subset Z^+(c^*,q^0)$ such that (i) does not hold, i.e.,

$$\int_{A'} dH'(y) < \int_{A'} dH^+(y) \quad .$$

Let us consider a sequence $c_s \downarrow c^*$, $c_s > c^*$, and take $A_s = Z^+(c_s,q^0) \cap A'$. We have $A' = \bigcup A_s$, $A_s \subset A_{s+1}$ and therefore

$$\lim_{s\to\infty} \int_{A_s} dH'(y) = \int_A dH'(y) \ , \ \lim_{s\to\infty} \int_{A_s} dH^+(y) = \int_A dH^+(y) \quad .$$

Thus, there must exist a $c_s > c^*$ and a $\gamma > 0$ such that

$$-\gamma = \int_{A_s} dH'(y) - \int_{A_s} dH^+(y) = \int_{A_s} dH'(y) - \int_{A_s} d\tilde{H}(y) \quad .$$

Note that $q^0(y) > c_s > c^*$ for $y \in A_s$. Using the definition of $\tilde{H}$ and constraint (7) we have:

$$\int_A dH'(y) \leq \int_A d\tilde{H}(y)$$

for arbitrary set $A \subset Z^+(c^*, q^0)$, and

$$\int_A dH'(y) \geq \int_A d\tilde{H}(y)$$

for arbitrary set $A \subset Z^-(c^*, q^0)$. This, together with the fact that $q^0(y)$ is positive, implies:

$$\int_{A_s} q^0(y) d(H'(y) - \tilde{H}(y)) \leq c_s \int_{A_s} d(H'(y) - \tilde{H}(y)) \tag{9}$$

$$\int_{Z^+(c^*, q^0) \setminus A_s} q^0(y) d(H'(y) - \tilde{H}(y)) \leq c^* \int_{Z^+(c^*, q^0) \setminus A_s} d(H'(y) - \tilde{H}(y)) \tag{10}$$

$$\int_{Z^-(c^*, q^0)} q^0(y) d(H'(y) - \tilde{H}(y)) \leq c^* \int_{Z^-(c^*, q^0)} d(H'(y) - \tilde{H}(y)) \quad . \tag{11}$$

We shall now use (9)–(11) to estimate the difference between the values of the objective function for measures $\tilde{H}$ and H':

$$\int_Y q^0(y)\, dH'(y) - \int_Y q^0(y)\, d\tilde{H}(y) =$$

$$\int_{A_s} q^0(y) d(H'(y) - \tilde{H}(y)) + \int_{Z^+(c^*, q^0) \setminus A_s} q^0(y) d(H'(y) - \tilde{H}(y)) +$$

$$\int_{Z^0(c^*, q^0)} q^0(y) d(H'(y) - \tilde{H}(y)) + \int_{Z^-(c^*, q^0)} q^0(y) d(H'(y) - \tilde{H}(y)) \leq$$

$$c_s \int_{A_s} d(H'(y) - \tilde{H}(y)) + c^* \int_{Z^+(c^*, q^0) \setminus A_s} d(H'(y) - \tilde{H}(y)) +$$

$$c^* \int_{Z^0(c^*, q^0)} d(H'(y) - \tilde{H}(y)) + c^* \int_{Z^-(c^*, q^0)} d(H'(y) - \tilde{H}(y)) =$$

$$(c_s - c^*) \int_{A_s} d(H'(y) - \tilde{H}(y)) + c^* \int_Y d(H'(y) - \tilde{H}(y)) = -\gamma(c_s - c^*) < 0 \quad .$$

Thus, H' cannot be the optimal measure, thus proving (i). Parts (ii) and (iii) may be proved in the same way.

For the particular case in which we have only the upper measure H^+, which is atomless, the result of Lemma 1 is close to Theorem 1 from [11].

Example 1. Suppose that measures H^+ and H^- have piecewise-continuous density functions H_y^+ and H_y^-, respectively. In this case it is natural to look for the optimal measure among probabilistic measures with piecewise-continuous probability density functions (p.d.f.s) $H_y(y)$, and to replace constraint (7) by:

$$H_y^-(y) \le H_y(y) \le H_y^+(y) \quad , \quad y \in Y \quad . \tag{7a}$$

The optimal p.d.f. $H_y^*(y)$ under these circumstances is defined as follows:

$$H_y^*(y) = \begin{cases} H_y^+(y) \text{ if } y \in Z^+(c^*, q^0) \\ H_y^-(y) \text{ if } y \in Z^-(c^*, q^0) \\ \left. \begin{array}{l} H_y^-(y) \le H_y^*(y) \le H_y^+(y), \\ \int\limits_{Z^0(c^*,q^0)} H_y^*(y)\mathrm{d}y = 1 - \int\limits_{Z^+(c^*,q^0)} H_y^+(y)\mathrm{d}y - \int\limits_{Z^-(c^*,q^0)} H_y^-(y)\mathrm{d}y \end{array} \right\} \text{if } y \in Z^0(c^*, q^0) \end{cases}$$

Example 2. Suppose that measures H^+ and H^- assign positive weights to a finite number of points, i.e.:

$$H^+ = \{(p_i^+, y_i), \ i = \overline{1,l}, \ p_i^+ \ge 0, \ \sum_{i=1}^{l} p_i^+ \ge 1\}$$

$$H^- = \{(p_i^-, y_i), \ i = \overline{1,l}, \ p_i^- \ge 0, \ \sum_{i=1}^{l} p_i^- \le 1\}$$

$$p_i^- \le p_i^+ \quad , \quad i = \overline{1,l} \quad .$$

Define the sets I^+, I^0, I^- as follows:

$$I^+(c, q^0) = \{i : q^0(y_i) > c \ , \ 1 \le i \le l\}$$

$$I^0(c, q^0) = \{i : q^0(y_i) = c \ , \ 1 \le i \le l\}$$

$$I^-(c, q^0) = \{i : q^0(y_i) < c \ , \ 1 \le i \le l\}$$

and take

$$c^* = \inf \{c : \sum_{I^+(c,q^0)} p_i^+ + \sum_{I^0(c,q^0) \cup I^-(c,q^0)} p_i^- \le 1\} \quad .$$

The optimal probabilistic measure is then defined as follows:

$$H^* = \{(p_i^*, y_i), \ i = \overline{1,l}\} \quad ,$$

where

$$p_i^* = \begin{cases} p_i^+ \text{ if } i \in I^+(c^*, q^0) \\ p_i^- \text{ if } i \in I^-(c^*, q^0) \\ p_i^- \le p_i^* \le p_i^+, \ \sum\limits_{I^0(c^*,q^0)} p_i^* = 1 - \sum\limits_{I^+(c^*,q^0)} p_i^+ - \sum\limits_{I^-(c^*,q^0)} p_i^- \quad \text{if } i \in I^0(c^*, q^0) \end{cases}$$

The result of this lemma may be easily generalized to the case in which the Borel field in (7) is replaced by some other σ-field D. In this case it is necessary for the sets $Z^+(c,q^0)$, $Z^-(c,q^0)$, $Z^0(c,q^0)$ to belong to D. The proof of the lemma remains unchanged.

Another easily treated case arises when constraint (7) is replaced by

$$\int_{A_i} dH^-(y) \le \int_{A_i} dH(y) \le \int_{A_i} dH^+(y)\ , \quad i = \overline{1,N}\ ,$$

where sets A_i are closed, $A_i \cap A_j = \phi$ for $i \ne j$, and

$$\sum_i \int_{A_i} dH^+(y) \ge 1 \quad , \quad \int_B dH^+(y) = 0$$

for all B such that $B \cap A_i = \phi$ for $i = \overline{1,N}$.

We shall now turn to a numerical algorithm for solving (5), (7), (8). It is clear from Lemma 1 that the problem of finding the optimal solution of (5),(7),(8) is essentially that of finding the smallest possible c for which

$$\int_{Z^+(c,q^0)} dH^+(y) + \int_{Z^0(c,q^0)\cup Z^-(c,q^0)} dH^-(y) \le 1\ . \tag{12}$$

In what follows we present a simple algorithm for finding such a c. This algorithm is based on stochastic quasigradient methods developed for stochastic nonsmooth optimization problems [15]. We shall assume that H^+ and H^- have density functions H_y^+ and H_y^-, respectively, so that constraint (7) becomes (7a) and (12) is equivalent to the following problem:

Find the smallest possible c such that

$$W(c) \le 1\ , \tag{13}$$

where

$$W(c) = \int_Y v(c,y)dy$$

$$v(c,y) = \begin{cases} H_y^+(y) & \text{if } q^0(y) > c \\ H_y^-(y) & \text{if } q^0(y) \le c \end{cases}\ .$$

Observe that, under the conditions of Lemma 1, $W(c)$ is a nonincreasing function of c such that $\lim_{c \downarrow c^*} W(c) = W(c^*)$ for arbitrary c^*. Consider the multi-valued function

$$\tilde{W}(c) = \{d : a_c - 1 \le d \le b_c - 1,\ a_c = \varliminf_{c' \to c} W(c'),\ b_c = \varlimsup_{c' \to c} W(c')\} \ .$$

Problem (13) now becomes the problem of finding c such that

$$0 \in \tilde{W}(c) \ . \tag{14}$$

From the properties of the function $W(c)$, there exists a concave function $F(c)$ such that $\tilde{W}(c) = F_c(c)$, where $F_c(c)$ is the set of subdifferentials of function $F(c)$ at the point c:

$$F(c) = \int_{c_-}^{c}\int_{Y}\left[v(t,y) - \frac{1}{\mu(Y)}\right]dy\ dt = \int_{Y}\int_{c_-}^{c}\left[v(t,y) - \frac{1}{\mu(Y)}\right]dt\ dy$$

$$= \mu(Y)E_y\left[\int_{c_-}^{c}\left[v(t,y) - \frac{1}{\mu(Y)}\right]dt\right] \ .$$

Here $\mu(Y)$ is the Lebesque measure of Y, y is distributed uniformly on the set Y, E_y denotes expectation with respect to y, and $c_- = \min_{y \in Y} q^0(y)$. Thus, problem (14) becomes one of maximizing the function $E_y f(c,y)$, where

$$f(c,y) = \int_{c_-}^{c}\left[v(t,y) - \frac{1}{\mu(Y)}\right]dt \ .$$

Stochastic quasigradient methods capable of solving this problem can be implemented with little computational effort. One such method can be stated as follows:

$$c^{s+1} = c^s + \rho_s \xi^s \tag{15}$$

where

$$\xi^s = \begin{cases} H_y^+(y) - \dfrac{1}{\mu(Y)} & \text{if } q^0(y) > c^s \\ H_y^-(y) - \dfrac{1}{\mu(Y)} & \text{if } q^0(y) \le c^s \end{cases} \ ,$$

ρ_s is a stepsize such that

$$\rho_s \ge 0\ ,\ \sum_{s=0}^{\infty} \rho_s = \infty\ ,\ \sum_{s=0}^{\infty} \rho_s^2 < \infty \ ,$$

and y is a random variable uniformly distributed on Y.

Let us now consider the problem (5)–(8), i.e., include constraints of type (6). We shall first prove the duality result which reduces it to a minimax problem with an inner problem of type (5),(7),(8). A similar result for problems without constraints (7) was proved in [7].

We begin by defining the set G of feasible distributions:

$$G = \{H: \int_Y q^i(y)dH(y) = 0,\ i = \overline{1,m_1},\ \int_Y q^i(y)dH(y) \le 0,\ i = \overline{m_1,m},$$

$$\int_A dH^-(y) \le \int_A dH(y) \le \int_A dH^+(y),\ A \subset B(Y);\ \int dH(y) = 1,\ \text{supp}\, H \subset Y\}$$

and the set $\bar{G}$ of all distributions satisfying (7):

$$\bar{G} = \{H: \int_A dH^-(y) \le \int_A dH(y) \le \int_A dH^+(y),\ A \subset B(y)\}\ .$$

The problem then becomes

$$\max_{H \in G} \int_Y q^0(y)dH(y)\ .$$

Consider the following set:

$$Z = \{z : z = (z_0,\ldots,z_m),\ z_i = \int_Y q^i(y)dH(y),\ H \in \bar{G}\}\ . \tag{16}$$

Theorem 1. *Suppose that the conditions of Lemma 1 are satisfied and that the following additional assumptions hold:*

1. *Y is compact and the $q^i(y)$, $i = \overline{0,m}$, are continuous*
2. $0 \in \text{int co}\, Z$.

Under these conditions a solution to problem (5)–(8) exists and

(i) $\max_{H \in G} \int_Y q^0(y) = \min_{u \in U^+} \varphi(u)$, *where* $U^+ = \{u : u_i \ge 0,\ i = \overline{m_1,m}\}$ *and*

$$\varphi(u) = \max_H \int [q^0(y) - \sum_{j=1}^m u_j q^j(y)]dH(y)\ , \tag{17}$$

subject to the constraint

$$\int_A dH^-(y) \le \int_A dH(y) \le \int_A dH^+(y) \tag{18}$$

for any Borel A.

(ii) *For any solution H^* of problem* (5)–(8) *there exists a $u^* \in U^+$ such that H^* is a solution of problem* (17)–(18) *for $u = u^*$ and $\varphi(u^*) = \max_{u \in U^+} \varphi(u)$.*

Proof. In what follows we shall consider the reduction of the usual R^n topology to set Y for sets from R^n. In particular, we shall use the term "open set" as an abbreviation for "open set with respect to Y", and so on. Consider first the set Z defined in (16). This set is convex because for any $z', z'' \in Z$ and λ: $0 \le \lambda \le 1$ we have

$$\lambda z_i' + (1-\lambda)z_i'' = \lambda \int_Y q^i(y) dH_1(y) + (1-\lambda) \int_Y q^i(y) dH_2(y)$$

$$= \int_Y q^i(y) d\,[\lambda H_1(y) + (1-\lambda) H_2(y)]$$

for any $H_1, H_2 \in G$, and $\bar{G}$ is convex by definition. We shall now prove that Z is closed. Consider an arbitrary convergent sequence $z^s : z^s \in Z,\ z^s \to z^*$. To prove that Z is closed we have to show that $z^* \in Z$. A probabilistic measure H^s is associated with each point z^s such that $z_i^s = \int q^i(y) dH^s(y)$, $i = \overline{0,m}$, and

$$\int_A dH^-(y) \le \int_A dH^s(y) \le \int_A dH^+(y) \quad .$$

Set Y is compact and therefore, according to the Prohorov theorem [16], there must exist a subsequence $H^{s_k}(y)$ and a measure H^* such that

$$\lim_{k \to \infty} \int f(y) dH^{s_k}(y) = \int f(y) dH^*(y)$$

for every continuous bounded $f(y)$. Now take an arbitrary closed set $A \subset Y$ and consider the open set $A_\varepsilon = \{y : p(y, A) < \varepsilon\}$, where $p(y, A)$ is the Hausdorf distance between y and A:

$$p(y, A) = \inf_{z \in A} \| y - z \| \quad .$$

We have

$$\lim_{\varepsilon \to 0} \int_{A_\varepsilon} dH^+(y) = \int_A dH^+(y)$$

$$\lim_{\varepsilon \to 0} \int_{A_\varepsilon} dH^*(y) = \int_A dH^*(y)$$

due to the fact that A_ε decreases to A as $\varepsilon \downarrow 0$. Now consider the following function (see [16]):

$$\psi_\varepsilon(y) = \tau\left[\frac{1}{\varepsilon} p(y, A)\right] ,$$

where

$$\tau(t) = \begin{cases} 1 & \text{if } t \le 0 \\ 1-t & \text{if } 0 \le t \le 1 \\ 0 & \text{if } 1 \le t \end{cases} .$$

This function is continuous and bounded for $\varepsilon > 0$, and therefore we have

$$\lim_{k \to \infty} \int \psi_\varepsilon(y) \mathrm{d}H^{S_k}(y) = \int \psi_\varepsilon(y) \mathrm{d}H^*(y) .$$

On the other hand,

$$\int \psi_\varepsilon(y) \mathrm{d}H^{S_k}(y) \le \int_{A_\varepsilon} \mathrm{d}H^{S_k}(y) \le \int_{A_\varepsilon} \mathrm{d}H^+(y)$$

and

$$\int \psi_\varepsilon(y) \mathrm{d}H^*(y) \ge \int_A \mathrm{d}H^*(y) ,$$

leading to

$$\int_A \mathrm{d}H^*(y) \le \int_{A_\varepsilon} \mathrm{d}H^+(y)$$

for arbitrary ε. Thus we finally obtain

$$\int_A \mathrm{d}H^*(y) \le \int_A \mathrm{d}H^+(y)$$

for any closed A.

This expression holds for all open sets because for any open set A there exists a sequence of closed sets A_s such that $A_{s+1} \supset A_s$ and

$$\bigcup_s A_s = A .$$

This fact, together with the regularity of measures H^+ and H^*, implies that

$$\int_A \mathrm{d}H^*(y) \le \int_A \mathrm{d}H^+(y)$$

for any $A \subset B(Y)$. Therefore $H^* \in \bar{G}$, and since

$$z_i^* = \lim_{k\to\infty} z_i^{s_k} = \lim_{k\to\infty} \int_Y q^i(y)\mathrm{d}H^{s_k}(y)$$

$$= \int_Y q^i(y)\mathrm{d}H^*(y) \quad ,$$

we must have $z^* \in Z$. This confirms that set Z is closed.

We have now proved that Z is a convex compact set in R^{m+1} and therefore that the optimal value of the problem

$$\sup_{H\in G} \int_Y q^0(y)\mathrm{d}H(y)$$

is equal to the optimal value of the following finite-dimensional problem:

$$\max_{z\in Z} z_0 \tag{19}$$

$$z_i = 0\ , \quad i = \overline{1,m_1} \tag{20}$$

$$z_i \le 0\ , \quad i = \overline{m_1,m} \quad . \tag{21}$$

From assumption 2 of the theorem we deduce that the optimal value of (19)–(21) must be equal to the optimal value of the following minimax problem:

$$\min_{u\in U^+} \max_{z\in Z} [z_0 - \sum_{j=1}^{m} u_j z_j] \tag{22}$$

and thus for any solution z^* of (19)–(21) there must exist a $u^* \in U^+$ such that

$$\bar{\varphi}(u^*) = \min_{u\in U^+} \bar{\varphi}(u)\ , \quad \bar{\varphi}(u) = \max_{z\in Z} [z_0 - \sum_{j=1}^{m} u_j z_j]$$

and

$$z_0^* - \sum_{j=1}^{m} u_j^* z_j^* = \max_{z\in Z} [z_0 - \sum_{j=1}^{m} u_j^* z_j] \quad . \tag{23}$$

We may now deduce that a solution to (5)–(8) exists because a solution to (19)–(21) exists and for each $z \in Z$ there is an $H \in \overline{G}$ such that

$$\int_Y q^i(y)\,dH(y) = z_i\,, \quad i = \overline{0,m} \quad .$$

From (22) we obtain

$$\max_{H \in G} \int_Y q^0(y)\,dH(y) = \min_{u \in U^+} \overline{\varphi}(u) = \min_{u \in U^+} \max_{z \in Z} [z_0 - \sum_{j=1}^{m} u_j z_j]$$

$$= \min_{u \in U^+} \max_{H \in \overline{G}} \int_Y [q^0(y) - \sum_{j=1}^{m} u_j q^j(y)]\,dH(y) = \min_{u \in U^+} \varphi(u)$$

and the first part of the theorem is proved.

Now let H^* be an arbitrary solution of (5)–(8) and

$$z_i^* = \int q^i(y)\,dH^*(y)\,, \quad i = \overline{0,m} \quad .$$

From (23), there exists a $u^* \in U^+$ such that

$$\int_Y [q^0(y) - \sum_{j=1}^{m} u_j^* q^j(y)]\,dH^*(y) = z_0^* - \sum_{j=1}^{m} u_j^* z_j^* =$$

$$\max_{z \in Z} [z_0 - \sum_{j=1}^{m} u_j^* z_j] = \max_{H \in \overline{G}} \int_Y [q^0(y) - \sum_{j=1}^{m} u_j^* q^j(y)]\,dH(y)$$

and

$$\varphi(u^*) = \min_{u \in U^+} \varphi(u) \quad .$$

This completes the proof.

We have now reduced the original problem (5)–(8) to that of minimizing the convex function $\varphi(u)$. According to Lemma 1, the optimal solution of (5)–(8) is then described by the optimal value u^* and a constant c^*, which is the smallest possible value of c such that the following inequality holds:

$$\int_{Z^+(c,u^*)} dH^+(y) + \int_{Z^-(c,u^*) \cup Z^0(c,u^*)} dH^-(y) \le 1 \quad , \tag{24}$$

where

$$Z^+(c,u) = \{y : y \in Y,\ r(u,y) > c\}$$

$$Z^-(c,u) = \{y : y \in Y,\ r(u,y) < c\}$$

$$Z^0(c,u) = \{y : y \in Y,\ r(u,y) = c\}$$

$$r(u,y) = q^0(y) - \sum_{j=1} u_j q^j(y) \quad .$$

The following distributions H^* are then possible candidates for the optimal solution:

$$H^* = \begin{cases} H^+(y) & \forall y \in Z^+(c,r) \\ H^-(y) & \forall y \in Z^-(c,r) \\ \int_A \mathrm{d}H^+(y) \geq \int_A \mathrm{d}H^*(y) \geq \int_A \mathrm{d}H^-(y) & \forall A : A \in B(Y),\ A \subset Z^0(c^*,u^*) \end{cases} \tag{25}$$

In this case, however, not all of the distributions defined by (24)–(25) are optimal. In order to ensure optimality it is necessary to introduce a uniqueness condition which specifies that the point

$$z^* = \{z : z = (z_0,\ldots,z_m)\ ,\ \ z_i = \int_Y q^i(y)\mathrm{d}H^*(y)\}$$

be the same for all H^* defined by (24)–(25). This will be the case if, for instance,

$$\int_{Z^0(c^*,u^*)} \mathrm{d}H^-(y) = 0\ ,\qquad \int_{Z^0(c^*,u^*)} \mathrm{d}H^+(y) = 0 \quad .$$

It is very difficult to obtain u^* by minimization of $\varphi(u)$ using convex optimization methods. This is because it is necessary to solve problem (17)–(18) (or, equivalently, (24)) in order to get the value of $\varphi(u)$ for particular u, which is in itself a computational challenge. However, it is possible to implement stochastic optimization methods for solving this problem using an approach similar to (15). We shall suppose once again that the measures have densities, so that constraint (7) may be replaced by (7a).

The problem of solving (5)–(7) now becomes the following:

Find u^* such that

$$u^* = \arg\min_{u \in U^+} \int_Y r(u,y) H_y(y,c(u))\mathrm{d}y \quad , \tag{26}$$

where

$$H_y(y,c) = \begin{cases} H_y^+(y) & \text{if } r(u,y) > c \\ H_y^-(y) & \text{if } r(u,y) < c \\ H_y^+(y) \ge H_y(y) \ge H_y^-(y) & \text{if } r(u,y) = c \end{cases}$$

and $c(u)$ is a solution of the following problem:

$$\max \; E_y \int_{c_v}^{c} \left[v(t,u,y) - \frac{1}{\mu(Y)} \right] dt \quad , \tag{27}$$

where

$$v(t,u,y) = \begin{cases} H_y^+(y) \text{ if } r(u,y) > t \\ H_y^-(y) \text{ if } r(u,y) \le t \end{cases} .$$

Here $\mu(Y)$ is the Lebesque measure of Y, c_v is a large negative number and y is distributed uniformly over Y. Problems (26)–(27) can be solved simultaneously using stochastic quasigradient techniques. The method in this particular case is:

$$c^{s+1} = c^s + \rho_s \xi^s \tag{28}$$

$$u^{s+1} = \pi_U^+(u^s - \delta_s \eta^s) \quad , \tag{29}$$

where π_U^+ is a projection operator on U^+,

$$\xi^s = \begin{cases} H_y^+(y^s) - \dfrac{1}{\mu(Y)} & \text{if } r(u^s,y^s) > c^s \\ H_y^-(y^s) - \dfrac{1}{\mu(Y)} & \text{if } r(u^s,y^s) \le c^s \end{cases}$$

$$\eta_i^s = q^i(y^s)\, v(c^s,u^s,y^s)$$

and y^s is an observation of the random variable, which is uniformly distributed on Y.

The convergence of this algorithm can be studied using the theory of nonstationary processes in stochastic optimization [17]. Method (29) may be considered as a means of tracking the changing maximum of function (27). Under quite broad assumptions, convergence requires that $\delta_s / \rho_s \to 0$. In this case we have

$$\varphi(u^s) - \max_{H \in G} \int [q^0(y) - \sum_{j=1}^{m} u_j q^j(y)] dH(y) \to 0$$

and algorithm (29) becomes a stochastic quasigradient method for solving $\min_{u \in U^+} \varphi(u)$.

Applying this theory to the problem at hand, we find that method (28)–(29) will solve problem (26)–(27) if, in addition to the conditions specified in Theorem 1, we have the following constraints on the stepsizes:

$$\rho_s \geq 0 \,, \quad \delta_s \geq 0 \,, \quad \sum_{s=0}^{\infty} \delta_s = \infty \,, \quad \sum_{s=0}^{\infty} \rho_s^2 < \infty \,, \quad \delta_s / \rho_s \to 0 \quad . \tag{30}$$

3. THE GENERAL NONLINEAR PROBLEM

We now have all the tools necessary to investigate the general nonlinear problem (1)–(4). The approach is the same as in [8]. We shall assume that functions $\Psi^i(H)$ are directionally differentiable:

$$\Psi^i(H_1 + \alpha(H_2 - H_1)) = \Psi^i(H_1) + \alpha \int_Y q^i(y, H_1) d(H_2(y) - H_1(y)) + \tau^i(\alpha, H_1, H_2) \tag{31}$$

for $i = \overline{0,m}$ and $H_1, H_2 \in \overline{G}$, where

$$\tau^i(\alpha, H_1, H_2) / \alpha \to 0 \text{ as } \alpha \to 0 \quad .$$

In what follows we assume that functions $q^i(y, H)$ are such that expression (31) is meaningful. The following lemma gives conditions which are necessary and in the convex case also sufficient for distribution H to be a solution of problem (1)–(4).

Lemma 2. *Suppose that $\Psi(H^*) \geq \Psi(H)$ for some $H^* \in G$ and all $H \in G$, and that the following conditions are satisfied:*

1. *Functions $q^i(y, H)$, $i = \overline{0,m}$, are bounded on Y for all $H \in G$*
2. $|\Psi^0(H_1) - \Psi^0(\alpha H_2 + (1 - \alpha) H_1)| \leq L\alpha$, $0 \leq \alpha \leq 1$ *and* $L < \infty$
3. *Functions $\Psi^i(H)$, $i = \overline{1,m}$, are convex, i.e.,*

$$\Psi^i(\alpha H_1 + (1 - \alpha) H_2) \leq \alpha \Psi^i(H_1) + (1 - \alpha) \Psi^i(H_2)$$

4. *There exists an $\tilde{H} \in G$ such that $\Psi^i(\tilde{H}) < -\sigma < 0$ for $i = \overline{1,m}$. Then*

$$\sup_{H \in G^*} \int_Y q^0(y,H^*)\mathrm{d}H = \int_Y q^0(y,H^*)\mathrm{d}H^* \quad , \tag{32}$$

where

$$G^* = \{H : \int_Y q^i(y,H^*)\mathrm{d}H(y) \le \int_Y q^i(y,H^*)\mathrm{d}H^*(y),$$

$$H \in \overline{G},\ \int_Y \mathrm{d}H(y) = 1,\ i \in I^0,\ I^0 = \{i : \Psi^i(H^*) = 0\}\} \quad .$$

If, additionally, $\Psi^0(H)$ is concave and the distribution H^* satisfies (32), *then H^* is the solution of problem* (1)–(4).

The proof of this lemma is similar to that of Lemma 1 from [8] and is therefore omitted.

Combining the results of Lemma 2 and Theorem 1, we obtain Theorem 2.

Theorem 2. *Suppose that $\Psi^0(H^*) \ge \Psi^0(H)$ for all $H \in G$, that the conditions of Lemma 2 are satisfied and that the following assumptions hold:*

1. *Set $Y \subset R^n$ is compact*
2. *$H^+(y)$ and $H^-(y)$ are positive Borel measures such that*

$$\infty > \int_Y \mathrm{d}H^+(y) \ge 1\ ,\quad \int_Y \mathrm{d}H^-(y) \le 1$$

3. *Functions $q^i(y,H^*)$, $i = \overline{0,m}$, are continuous on Y.*

Then

(i) *We have*

$$\int_Y q^0(y,H^*)\mathrm{d}H^*(y) = \min_{u \in U^+} \tilde{\Psi}(u) \quad ,$$

where

$$U^+ = \{u : u_i \ge 0\} \quad ,$$

$$\tilde{\Psi}(u) = \max_H \int [q^0(y,H^*) - \sum_{j \in I^0} u_j q^j(y,H^*)]\mathrm{d}H(y) + \sum_{j \in I^0} u_j d_j \quad ,$$

$$d_j = \int_Y q^j(y,H^*)\mathrm{d}H^*(y) \quad ,$$

subject to constraints

$$\int_A \mathrm{d}H^-(y) \le \int_A \mathrm{d}H(y) \le \int_A \mathrm{d}H^+(y)$$

for any Borel $A \subset Y$.

(ii) *There exist a* $u^* \in U^+$ *and a* c^* *such that*

$$\tilde{\Psi}(u^*) = \min_{u \in U^+} \tilde{\Psi}(u)$$

$$\int_A \mathrm{d}H^*(y) = \int_A \mathrm{d}H^+(y) \quad \text{for all Borel } A \subset Z^+(c^*, u^*)$$

$$\int_A \mathrm{d}H^*(y) = \int_A \mathrm{d}H^-(y) \quad \text{for all Borel } A \subset Z^-(c^*, u^*)$$

$$\int_A \mathrm{d}H^-(y) \le \int_A \mathrm{d}H(y) \le \int_A \mathrm{d}H^+(y) \quad \text{for all Borel } A \subset Z^0(c^*, u^*)$$

$$\int_{Z^0(c^*,u^*)} \mathrm{d}H^*(y) = 1 - \int_{Z^+(c^*,u^*)} \mathrm{d}H^+(y) - \int_{Z^-(c^*,u^*)} \mathrm{d}H^-(y) \quad ,$$

where

$$Z^+(c,u) = \{y : y \in Y,\ \tilde{r}(u,y) > c\}$$

$$Z^-(c,u) = \{y : y \in Y,\ \tilde{r}(u,y) < c\}$$

$$Z^0(c,u) = \{y : y \in Y,\ \tilde{r}(u,y) = c\}$$

and

$$\tilde{r}(u,y) = q^0(y,H^*) - \sum_{j \in I^0} u_j q^j(y,H^*) \quad .$$

Numerical methods for solving nonlinear problems with constraints of type (3) will be the subject of a subsequent paper.

REFERENCES

1. M. Krein and A. Nudelman. The Markov Moment Problem and Extremal Problems. Translation of *Mathematical Monographs* 50. American Mathematical Society, Providence, RI, 1977.
2. S. Karlin and W.J. Studden. *Tchebycheff Systems: With Applications in Analysis and Statistics*. Wiley Interscience, New York, 1966.
3. J.H.B. Kemperman. On a class of moment problems. In the *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* 2 (1972) 101–126.
4. J. Zackova. On minimax solutions of stochastic linear programming problems. *Cas. Pest. Mat.* 91 (1966) 423–430.
5. Yu. Ermoliev. Method for stochastic programming in randomized strategies. *Kibernetica* 1 (1970).
6. J. Dupacova. Minimax stochastic programs with nonseparable penalties. In the Proceedings of the 9th IFIP Conference, Warsaw, 1979. Part 1. *Lecture Notes in Control and Information Sciences* 22. Springer-Verlag, Berlin, 1980, pp. 122–136.
7. Yu. Ermoliev, A. Gaivoronski and C. Nedeva. *Stochastic Optimization Problems with Incomplete Information on Distribution Functions*. Working Paper WP-83-113, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1983.
8. A. Gaivoronski. *Optimization of Functionals which Depend on Distribution Functions: 1. Nonlinear Functional and Linear Constraints*. Working Paper WP-83-114, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1983.
9. J. Birge and R. Wets. *Designing Approximation Schemes for Stochastic Optimization Problems, in Particular for Stochastic Problems with Recourse*. Working Paper WP-83-111, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1983.
10. H.P. Wynn. Optimum designs for finite population sampling. In S.S. Gupta and D.S. Moore (Eds.), *Statistical Decision Theory and Related Topics, II*. Academic Press, New York, 1977.
11. H.P. Wynn. Optimum submeasures with application to finite population sampling. Private communication.
12. J. Kiefer and J. Wolfowitz. Optimum designs in regression problems. *Annals of Mathematical Statistics* 30 (1959) 271–294.
13. V. Fedorov. *Theory of Optimal Experiments*. Academic Press, New York, 1972.
14. P. Whittle. Some general points in the theory of optimal experimental design. *Journal of the Royal Statistical Society, Series B* 35 (1973) 123–130.
15. Yu. Ermoliev. *Methods of Stochastic Programming* (in Russian). Nauka, Moscow, 1976.
16. P. Billingsley. *Convergence of Probability Measures*. John Wiley, New York, 1968.
17. Yu. Ermoliev and A. Gaivoronski. *Simultaneous Nonstationary Optimization, Estimation and Approximation Procedures*. Collaborative Paper CP-82-16, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1982.