International Institute for Applied Systems Analysis
IIASA
www.iiasa.ac.at

Numerical Techniques for Finding Estimates which Minimize the Upper Bound of the Absolute Deviation

Gaivoronski, A.A.

IIASA Working Paper

January 1987

NUMERICAL TECHNIQUES FOR FINDING ESTIMATES WHICH MINIMIZE THE UPPER BOUND OF THE ABSOLUTE DEVIATION

A. Gaivoronski

January 1987
WP-87-8

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

The paper deals with the numerical techniques for finding the special type of parameter estimates based on the minimization of L_1-norm of error. More specifically, these estimates are derived by minimization of the upper bound of the error, which is evaluated similarly to the upper bounds on the solution of stochastic optimization problem in WP-86-72. The research reported in this paper was performed in the Adaptation and Optimization Project of the System and Decision Sciences Program.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

CONTENTS

NUMERICAL TECHNIQUES FOR FINDING ESTIMATES WHICH MINIMIZE THE UPPER BOUND OF THE ABSOLUTE DEVIATION

A. Gaivoronski

1. INTRODUCTION

We are concerned here with the problem of determining the value of the parameter $a \in R^n$ when the information about it comes in the form of observations

$$\omega_i^j = \varphi^T(z^j)a + \eta_i^j \tag{1.1}$$

where $\varphi(z)$ is a known vector function, z^j are fixed points from finite subset Z of R^m, $Z = \{z^1, \ldots, z^l\}$: and η_i^j are identically distributed for the same j, random variables $\eta_{i_1}^{j_1}$ could depend on $\eta_{i_2}^{j_2}$ for $j_1 \neq j_2$ but should be independent for the same j_1. Some additional information is available in the form of inclusion $a \in A$ where A is a compact subset of R^n. In the simplest case when $n = 1$ and the values of the parameter a are observed directly (1.1) becomes

$$\omega_i = a + \eta_i \tag{1.2}$$

In this latest case the L_1-norm estimate a_s is obtained by minimization with respect to x of the sum

$$F(x, s) = \frac{1}{s} \sum_{i=1}^{s} |\omega_i - x| \tag{1.3}$$

where s is the total number of observations. The properties of this estimate (the sample median) are defined by the properties of the function

$$F(x) = \int |x - \omega| \mathrm{d}H(\omega) \tag{1.4}$$

namely, if $F(a) = \min_{x \in A} F(x)$ then under fairly mild conditions $a_s \rightarrow a$ with probability 1 [3, 16, 23].

If we knew the distribution H of observations ω then the actual value of parameter α can be found from minimization of $F(x)$. This is not the case, however, and we can think of the sampling procedure as means of obtaining information on H. Observations are used to estimate the distribution function H and use this estimate in (1.4) to determine the estimate of α. If we take empirical distribution function of observations and substitute it instead of H in (1.4) we would obtain the function $F(x, s)$ from (1.3) and estimates α_s. The different L_1-estimates would be obtained if different estimates of distribution H are used.

In this paper the numerical techniques for finding the worst-case L_1-estimates are proposed. These estimates minimize the largest possible L_1-error which corresponds to the worst distribution consistent in some sense with observations. The results rely on the techniques developed in [12], which are summarized in section 2. Section 3 is devoted to the algorithm for finding estimates in the simplest case (1.2). More complicated sampling scheme (1.1) is considered in section 3. It should be noted that numerical techniques for finding L_1-estimate was considered in [1–3]. Stochastic optimization techniques and estimation techniques with uncertain noise structure relevant to this problem was discussed in [5–8, 9–15, 17, 21, 24].

2. BOUNDS FOR INTEGRAL FUNCTIONALS

The results from [12] which are relevant to the estimation problem (1.4) are reviewed in this section.

Suppose that we have observations $\omega_1, \dots \omega_s$ with unknown distribution function H. Our aim is to construct the set G_s of distributions which are in some sense consistent with the set of observations.

Let us assume that ω_i belong to some set $\Omega \subset R^m$ with Borel field **B**; probability measure H is defined on this field, thus we have a probability space (Ω, **B**, H). For each fixed s let us consider the sample probability space ($\bar{\Omega}$, $\bar{\mathbf{B}}$, $\bar{\mathbf{P}}$) which is a Cartesian product of s spaces (Ω, **B**, H). The space ($\bar{\Omega}$, $\bar{\mathbf{B}}$, $\bar{\mathbf{P}}$) is the smallest space which contain all (Ω^s, $\mathbf{B}^s$, $\mathbf{P}^s$). In what follows the "convergence with probability 1" will mean the "convergence with probability 1 in the space (Ω^s, $\mathbf{B}^s$, $\mathbf{P}^s$)". With the set of observations $\{\omega_1, \dots, \omega_s\}$ the set of distribution G_s will be associated in the following way.

Let us fix the confidence level $\alpha : 0 < \alpha < 1$. We shall consider events with probability $\mathbf{P}^s$ less than α "improbable" events and discard them. Let us consider arbitrary set $A \subset \mathbf{B}$. Among s observations $\{\omega_1, \ldots, \omega_s\}$ there are i_A observations which belong to set A, $0 \leq i_A \leq s$. The random variable i_A is distributed binominally and its values can be used to estimate $H(A)$ (Mainland [19]). To do this let us consider the following functions

$$\phi(s, k, z) = \sum_{i=k}^{s} \frac{s!}{i!(s-i)!} z^i (1-z)^{s-i}$$

$$\Psi(s, k, z) = \sum_{i=0}^{k} \frac{s!}{i!(s-i)!} z^i (1-z)^{s-i} \tag{2.1}$$

observe that

$$\phi(s, k, z) = \Psi(s, s-k, 1-z)$$

$$\mathbf{P}^s(i_A \geq k) = \phi(s, k, H(A)) \tag{2.2}$$

$$\mathbf{P}^s(i_A \leq k) = \Psi(s, k, H(A))$$

The function $\phi(s, k, z)$ is a monotonically increasing function of z on the interval $[0, 1]$, $\phi(s, k, 0) = 0$, $\phi(s, k, 1) = 1$, $k \neq 0$. Therefore the solution of equation $\phi(s, k, z) = c$ exist for any $0 \leq c \leq 1$. Let us take

$$d(s, k) : \phi(s, k, d(s, k)) = \alpha, \; k \neq 0 \tag{2.3}$$

$$b(s, k) : \Psi(s, k, b(s, k)) = \alpha, \; k \neq s$$

$$d(s, 0) = 0, \; b(s, s) = 1$$

The values $d(s, k)$ and $b(s, k)$ are the lower and upper bounds for the probability $H(A)$ in the following sense.

LEMMA 1. *For any fixed set $A \subset \mathbf{B}$ the bound $d(s, k)$ defined in* (5) *possess the following properties*

1. $\mathbf{P}^s\{d(s, i_A) > H(A)\} \leq \alpha$ *for any measure* H.
2. *If for some function* $c(i)$, $i = 0 : s$, $c(i+1) > c(i)$ *we have* $\mathbf{P}^s\{c(i_A) > H(A)\} \leq \alpha$ *for any* H *then* $c(i) \leq d(s, i)$

This lemma shows that $d(s, i_A)$ is in a certain sense the best lower bound for the probability $H(A)$. The similar result holds for the upper bound $b(s, i_A)$:

LEMMA 1′. *For any fixed set $A \subset B$ $b(s, k)$ defined in (5) possess the following properties:*

1. $\mathbf{P}^s\{b(s, i_A) < H(A)\} \leq \alpha$
2. *If for some function* $c(i)$, $i = 0 : s$, $c(i+1) > c(i)$ *we have* $\mathbf{P}^s\{c(i_A) < H(A)\} \leq \alpha$ *for any* H *then* $c(i) \geq b(s, i)$.

These lemmas are proved in [12].

DEFINITION *The set G_s of the distributions consistent with the set of observations $\{\omega_1, \ldots, \omega_s\}$ for fixed confidence level α is defined as follows:*

$$G_s = \{H : d(s, i_A) \leq H(A) \leq b(s, i_A)\} \tag{2.4}$$

for any measurable A, where $d(s, i_A)$ and $b(s, i_A)$ are defined in (2.3).

Now let us consider the problem of finding upper and lower bounds of functional $\int g(\omega)\mathrm{d}H(\omega)$ on the set G_s. This problem will be used in later sections for defining the special class of L_1-estimates. In this section we are interested in solving the following problem:

minimize (or maximize) with respect to H

$$\int g(\omega)\mathrm{d}H(\omega) \tag{2.5}$$

subject to constraints

$$d(s, i_A) \leq H(A) \leq b(s, i_A), A \in \mathbf{B} \tag{2.6}$$

Let us assume that $g(\omega^0) = \min_{\omega \in \Omega} g(\omega)$ and $g(\omega^{s+1}) = \max_{\omega \in \Omega} g(\omega)$ exist and arrange the set of observations $\{\omega_1, \ldots, \omega_s\}$ in order of increasing values of the function $g(\omega)$:

$$\omega^0, \omega^1, \ldots, \omega^s, \omega^{s+1}$$

Here and elsewhere the original order of observations is indicated by subscript and arrangement in increasing order of the values of g is indicated by superscript. The first element of new arrangement will always be the point with the minimal value of the objective function on the set Ω and the last element (with number $s+1$) will be the point with maximal value. This arrangement depends on the number s of the time interval, but this dependence will not be explicitly indicated for the simplicity of notations.

The solution of the problem (2.5)–(2.6) is given by the following theorem:

THEOREM 1 *Suppose that exist points* ω^0 *and* ω^{s+1} *such that* $g(\omega^0) = \min_{\omega \in \Omega} g(\omega)$, $g(\omega^{s+1}) = \max_{\omega \in \Omega} g(\omega)$. *Then*

1. *The solution of the problem (2.5)–(2.6) exist and among extremal measures always exist discrete one which is concentrated in* $s+1$ *points:*

$$\bar{g}_s = \max_{H \in G_s} \int g(\omega)\,dH(\omega) = \int g(\omega)\,d\bar{H}_s(\omega) = \sum_{i=1}^{s+1} p_s^i g(\omega^i) \tag{2.7}$$

$$\underline{g}_s = \min_{H \in G_s} \int g(\omega)\,dH(\omega) = \int g(\omega)\,d\underline{H}_s(\omega) = \sum_{i=0}^{s} q_s^i g(\omega^i) \tag{2.8}$$

$$\bar{H}^s = \{(\omega^0, p_0^s), \ldots, (\omega^{s+1}, p_{s+1}^s)\}$$

$$\underline{H}_s = \{(\omega^0, q_0^s), \ldots, (\omega^{s+1}, q_{s+1}^s)\}$$

$$p_i^s = d(s, i) - d(s, i-1),\ i = 1:s \tag{2.9}$$

$$q_i^s = d(s, s-i+1) - d(s, s-i),\ i = 1:s$$

$$p_0^s = q_{s+1}^s = 0,\ p_{s+1}^s = q_0^s = b(s, 0)$$

2.

$$\bar{g}_s - \underline{g}_s < \Delta_g \sqrt{\frac{2|\ln \alpha|}{s}}$$

where $\Delta_g = \max_{\omega \in \Omega} g(\omega) - \min_{\omega \in \Omega} g(\omega)$

3.

$$\bar{g}_s \to \int g(\omega)\,dH(\omega)$$

$$\underline{g}_s \to \int g(\omega)\,dH(\omega)$$

with probability 1 as $s \to \infty$.

The proof is contained in [12].

3. THE CASE OF ONE-DIMENSIONAL PARAMETER

Using the results of the previous section we shall obtain estimates of the parameter $a \in R^1$ from observations (1.2). It is assumed that apriori bounds $\underline{a}$ and $\bar{a}$ are known

$$\underline{a} \leq a \leq \bar{a}$$

For the purpose of convergence analysis it is irrelevant how far are the bound $\underline{a}$ and $\bar{a}$ from actual value of a, it is only necessary that $-\infty < \underline{a} < \bar{a} < \infty$. For computational purposes it is preferable of course to have $\underline{a}$ and $\bar{a}$ as close to a as possible. We shall assume for simplicity that $\underline{a} \leq \omega_i \leq \bar{a}$ for all i. The different case can be treated in the same manner, but requires more complicated notation.

Let us take some confidence level α and define the admissible set of distributions G_s from (2.4). It is possible to utilize this information in two different ways. One approach is associated with the case when not only the value of parameter itself is of interest but it is also important to guarantee the smallest possible values of error functional $\int |x - \omega| dH(\omega)$. In this case the estimate is constructed which minimize the worst in the set G_s value of the error functional. The second approach is to define the region to which the actual parameter belongs provided the distribution H can take arbitrary values from admissible set G_s. We shall consider both approaches for one-dimensional case starting with the worst-case estimate.

The worst-case approximation $\bar{F}(x, s)$ to the function $F(x)$ based on the set of distributions G_s consistent with observations is defined as follows:

$$\bar{F}(x, s) = \max_{H \in G_s} \int |x - \omega| d H(\omega) \tag{3.1}$$

The values of this function can be computed using the Theorem 1.

DEFINITION *The worst-case L_1-estimate $\tilde{a}_s$ of parameter a is defined by minimization of the function $\bar{F}(x, s)$ from (3.1):*

$$\bar{F}(\tilde{a}_s, s) = \min_{\underline{a} \leq x \leq \bar{a}} \bar{F}(x, s) \tag{3.2}$$

This estimate depends on the confidence level α. It follows from the part 3 of the theorem 1 that $\bar{F}(x, s) \to \int |x - \omega| dH(\omega)$ with probability 1 for fixed x. The definition of the function $\bar{F}(x, s)$ and bondedness of the regions to which ω and x belong implies that the function $\bar{F}(x, s)$ is convex and uniformly continuous with respect to s. Therefore all limit points of the sequence $\tilde{a}_s$ belong to the set

$$X^* = \{x^* : \int |x^* - \omega| dH(\omega) = \min_{\underline{a} \le x \le \bar{a}} \int |x - \omega| dH(\omega), \underline{a} \le x^* \le \bar{a}\}$$

Therefore $\tilde{a}_s \to a$ if $X^* = \{a\}$

The function $\bar{F}(x, s)$ is convex function and for any fixed x it is possible to compute the values of this function and its subgradients. The convex programming techniques [18, 22] can be used to minimize this function and obtain the estimate $\tilde{a}_s$. However, it is more convenient to develop special algorithm which utilizes the properties of the function $\bar{F}(x, s)$.

Let us start with defining sufficient condition for a point x to minimize the function $\bar{F}(x, s)$. Take arbitrary $x : \underline{a} \le x \le \bar{a}$ and define

$$\underline{\gamma}(x, \omega) = \begin{cases} 1 & \text{if } x > \omega \\ -1 & \text{if } x \le \omega \end{cases}$$

$$\bar{\gamma}(x, \omega) = \begin{cases} 1 & \text{if } x \ge \omega \\ -1 & \text{if } x < \omega \end{cases}$$

Let us arrange observations $\{\omega_1, \ldots, \omega_s\}$ in two orderings. Members of the first ordering will be denoted by $\underline{\omega}^i(x)$ and of the second ordering by $\bar{\omega}^i(x)$, $i = 1:s$. For each i exist j, k such that $\bar{\omega}^i(x) = \omega_j$, $\underline{\omega}^i(x) = \omega_k$ and

$$i > j \Longrightarrow \begin{matrix} |x - \bar{\omega}^i(x)| \ge |x - \bar{\omega}^j(x)| \\ |x - \underline{\omega}^i(x)| \ge |x - \underline{\omega}^j(x)| \end{matrix}$$

$$\begin{matrix} |x - \underline{\omega}^i(x)| = |x - \underline{\omega}^j(x)| \\ x < \underline{\omega}^i(x), x > \underline{\omega}^j(x) \end{matrix} \quad \Longrightarrow \quad i > j$$

$$\begin{matrix} |x - \bar{\omega}^i(x)| = |x - \bar{\omega}^j(x)| \\ x < \bar{\omega}^i(x), x > \bar{\omega}^j(x) \end{matrix} \quad \Longrightarrow \quad i < j$$

In other words both orderings arrange observations in nondecreasing order of the values $|x - \omega|$. They differ only for the observations equidistant from x. Ordering with the members $\underline{\omega}_i(x)$ places first the observations which are to the left of x while ordering with the members $\bar{\omega}_i(x)$ places first the observations which are to the right of x.

Let us define

$$W^i(x) = |x - \bar{\omega}^i(x)|$$

Denote for all $\underline{a} \le x \le \bar{a}$

$$\underline{\omega}^{s+1}(x) = \begin{cases} \bar{a} \text{ if } & \bar{a} - x \ge x - \underline{a} \\ \underline{a} \text{ otherwise} \end{cases}$$

$$\bar{\omega}^{s+1}(x) = \begin{cases} \bar{a} \text{ if } & \bar{a} - x > x - \underline{a} \\ \underline{a} \text{ otherwise} \end{cases}$$

THEOREM 2 *Suppose that $\tilde{a}_s$ is the solution of the problem (3.2) and $\underline{a} < a < \bar{a}$. Then*

$$\underline{\Delta} = \sum_{i=1}^{s+1} p_i^s \underline{\gamma}(\tilde{a}_s, \underline{\omega}^i(\tilde{a}_s)) \le 0 \tag{3.3}$$

and

$$\bar{\Delta} = \sum_{i=1}^{s+1} p_i^s \bar{\gamma}(\tilde{a}_s, \bar{\omega}^i(\tilde{a}_s)) \ge 0 \tag{3.4}$$

where

$$p_i^s = d(s, i) - d(s-1, i), \quad i = 1:s$$

$$p_{s+1}^s = 1 - a(s, s) \quad \text{(see (2.9))}$$

Conversely, if for some $x = \tilde{a}_s$ conditions (3.3) and (3.4) are satisfied then $\tilde{a}_s$ is the solution of the problem (3.2).

The proof of the theorem follows directly from the results of the previous section and from the necessary condition for minima of convex function, namely $0 \in \partial F(x, s)$ where ∂ denotes the subdifferential of the convex function.

It is clear from the theorem that one of the solutions of the problem (3.2) will be among points where the sums (3.3) and (3.4) change sign. This can occur either in points ω_i or where $x - \omega_i = \omega_j - x$ for some i, j or at $\underline{x} = \frac{\bar{a} + \underline{a}}{2}$. This observation leads to the following algorithm:

ALGORITHM 1.

1. Start with selecting arbitrary point x^0 such that $\underline{a} \le x^0 \le \bar{a}$ and exist i, j with $\omega_i \le x^0$, $\omega_j \ge 0$. Arrange initial orderings $\bar{\omega}^i(x^0)$ and $\underline{\omega}^i(x^0)$

2. Suppose that we obtained the point x^k. Then the method proceeds as follows.

2a. Compute $\underline{\Delta}$ and $\bar{\Delta}$ from (3.3), (3.4). Now there could be three possibilities:

if $\underline{\Delta} \le 0$, $\bar{\Delta} \ge 0$ then go to step 2e

if $\underline{\Delta} < 0$, $\bar{\Delta} < 0$ then go to step 2b

if $\underline{\Delta} > 0$, $\bar{\Delta} > 0$ then go to step 2c

2b. Find

$$\varepsilon_1 = \min_i \left\{ \begin{array}{l} W^j(x^k) - W^i(x^k) | \bar{\omega}^i(x^k) < x^k, \\ j = \min_l \{l \,|\, l > i, \bar{\omega}^l(x^k) > x^k\} \end{array} \right\}$$

$$\varepsilon_2 = \min_i \{\omega_i - x^k | \omega_i > x^k\}$$

$$\varepsilon_3 = \frac{\bar{a} + \underline{a}}{2} - x^k, \ \varepsilon_3 > 0$$

If some ε_i does not exist take $\varepsilon_i = \bar{a} - \underline{a}$. Obtain x^{k+1}:

$$x^{k+1} = x^k + \min\{\frac{\varepsilon_1}{2}, \varepsilon_2, \varepsilon_3\}$$

go to step 2d

2c. Find

$$\varepsilon_1 = \min_i \left\{ \begin{array}{l} W^j(x^k) - W^i(x^k) | \underline{\omega}^i(x^k) > x^k, \\ j = \min_l \{l \,|\, l > i, \underline{\omega}^l(x^k) < x^k\} \end{array} \right\}$$

$$\varepsilon_2 = \min_i \{x^k - \omega_i | \omega_i < x^k\}$$

$$\varepsilon_3 = x^k - \frac{\bar{a} + \underline{a}}{2}, \ \varepsilon_3 > 0$$

If some ε_i does not exist, take $\varepsilon_i = \bar{a} - \underline{a}$. Obtain x^{k+1}:

$$x^{k+1} = x^k - \min\left\{\frac{\varepsilon_1}{2}, \varepsilon_2, \varepsilon_3\right\}$$

go to step 2d

2d. Obtain new orderings $\overline{\omega}^i(x^{k+1})$, $\underline{\omega}^i(x^{k+1})$ for the new point x^{k+1} and new $W^i(x^{k+1})$. Go to the step 2a.

2e. The estimate $\tilde{a}_s$ is found: $\tilde{a}_s = x^k$. Terminate the execution.

This method finds the estimate $\tilde{a}_s$ in a finite number of steps.

Now let us consider the problem of constructing the confidence region which contain all the solutions of the problem

$$\min_{\underline{a} \le x \le \overline{a}} \int |x - \omega| dH(\omega)$$

for $H \in G_s$.

Let us consider the ordering with elements ω_i^*, $i = 0 : s + 1$

$$\{\omega_i^*, i = 0 : s + 1\} = \{\omega_i, i = 1 : s\} \cup \{\underline{a}\} \cup \{\overline{a}\} ,$$

$$\omega_{i+1}^* \ge \omega_i^*, i = 0 : s$$

and define two distributions each concentrated in $s + 1$ points

$$H_s^{**} = ((\omega_1^*, p_1^s), \ldots , (\omega_i^* p_i^s), \ldots , (\omega_{s+1}^*, p_{s+1}^s))$$

$$H_s^* = ((\omega_0^*, q_0^s), \ldots , (\omega_i^*, q_i^s), \ldots , (\omega_s^*, q_s^s))$$

where p_i^s and q_i^s are defined in (2.9). It appears that minimum of the functions

$$F^{**}(x, s) = \int |x - \omega| dH_s^{**}(\omega) \tag{3.5}$$

$$F^*(x, s) = \int |x - \omega| dH_s^*(\omega)$$

define right and left end points of the confidence interval $[a^*, a^{**}]$ where

$$a^* = \inf_{\substack{\underline{a} \le y \le \overline{a} \\ H \in G_s}} \{y : \int |y - \omega| dH(\omega) = \min_{\underline{a} \le x \le \overline{a}} \int |x - \omega| dH(\omega)\}$$

$$a^{**} = \sup_{\substack{\underline{a} \le y \le \overline{a} \\ H \in G_s}} \{y : \int |y - \omega| dH(\omega) = \min_{\underline{a} \le x \le \overline{a}} \int |x - \omega| dH(\omega)\}$$

more specifically, the following result holds:

THEOREM 3 *Suppose that*

$$k = \min_{i \geq 0} \{i : b(s, i) \geq 0.5\} \tag{3.6}$$

Then for any x^* *such that*

$$a^* \leq x^* \leq a^{**},\ a^* = \omega_k^*,\ a^{**} = \omega_{s+1-k}^* \tag{3.7}$$

exist $\tilde{H} \in G_s$ *such that*

$$\min_{\underline{a} \leq x \leq \bar{a}} \int |x - \omega| \mathrm{d}\tilde{H}(\omega) = \int |x^* - \omega| \mathrm{d}\tilde{H}(\omega) \tag{3.8}$$

and if for some $\tilde{H} \in G_s$ *and some* x^* *condition (3.8) is satisfied then*

$$a^* \leq x^* \leq a^{**} .$$

PROOF According to the necessary and sufficient conditions the point $x^* : \underline{a} < x^* < \bar{a}$ is the minima of the function $\int |x - \omega| \mathrm{d}H(\omega)$ if and only if

$$\int_{\underline{a} \leq \omega < x} \mathrm{d}H(\omega) \leq \int_{x^* \leq \omega \leq \bar{a}} \mathrm{d}H(\omega)$$

and

$$\int_{\underline{a} \leq \omega \leq x^*} \mathrm{d}H(\omega) \geq \int_{x^* < \omega \leq \bar{a}} \mathrm{d}H(\omega) \tag{3.9}$$

It was assumed for simplicity of notation that $\int_{\underline{a} \leq \omega \leq \bar{a}} \mathrm{d}H(\omega) = 1$ thus (3.8) implies

$$x^* \geq \inf_y \left\{ y : \int_{\underline{a} \leq \omega \leq y} \mathrm{d}H \geq 0.5 \right\}$$

which gives

$$a^* \geq \inf_y \left\{ y : \sup_{H \in G_s} \int_{\underline{a} \leq \omega \leq y} \mathrm{d}H \geq 0.5 \right\}$$

and therefore $a^* \geq \omega_k^*$ where k is defined in (3.6). On the other hand ω_k^* is the minimum of the function $F^*(x, s)$ defined in (3.5). Therefore $a^* = \omega_k^*$. Similarly we obtain that $a^{**} = \omega_{s+1-k}^*$.

From convexity of the set G_s and function $\int|x - \omega|dH(\omega)$ now follows that for any point $x^* : a^* \le x^* \le a^{**}$ exists $\tilde{H} \in G$ such that (3.8) is satisfied. The proof is completed.

The same results can be obtained in the totally similar fashion for the case when the set Ω is bounded, but not coincide with $A = \{x : \underline{a} \le x \le \bar{a}\}$.

4. THE CASE OF VECTOR PARAMETER

Let us consider a more complicated case when the estimated parameter a belongs to R^n. It will be assumed that additional input parameters z are present, $z \in R^m$ and some finite set $Z = \{z^1, \dots, z^l\}$ is selected. The information comes with observations

$$\omega_i^j = \varphi^T(z^j)a + \eta_i^j \tag{4.1}$$

where $\varphi(z)$ is a known vector-function, η_i^j are independent identically distributed for the same j observations errors. Using the same type of argument as in introduction observe that various types of L_1-estimates can be obtained by minimizing the following function

$$F(x) = \sum_{j=1}^{l} \beta_j \int |\varphi^T(z^j)x - \omega^j| d H^j(\omega^j) \tag{4.2}$$

where β_j are the weights assinged to the points z^j. Suppose that s_j is the number of observations performed at the point z^j. Substitution of empirical distributions in (4.2) gives the following functional

$$F(x, s) = \sum_{j=1}^{l} \frac{\beta_j}{s_j} \sum_{i=1}^{s_j} |\varphi^T(z^j)x - \omega_i^j| \tag{4.3}$$

The minimization of this functional is the most common way of obtaining L_1-estimate a^s in this case. The worst-case L_1-estimates will be obtained similar to the simpler case in section 3, namely by minimizing the upper bound of the L_1-error in (4.2):

$$\bar{F}(x, s) = \sum_{j=1}^{l} \beta_j \max_{H^j \in G_{s_j}^j} \int |\varphi^T(z^j)x - \omega^j| d H^j(\omega) \tag{4.4}$$

Here the admissible sets of distributions $G^j_{s_j}$ are defined similar to (2.4) after fixing the confidence level α.

DEFINITION *The worst-case L_1-estimate $\tilde{a}_s$ of parameters a from (4.1) is defined as follows*

$$\bar{F}(\tilde{a}_s, s) = \min_{x \in A} \bar{F}(x, s), \; \tilde{a}_s \in A \tag{4.5}$$

In fact this estimate depends on all s_j, not only on s, but this will be skipped in notations. The values of the function $\bar{F}(s, x)$ can be computed using results of the Theorem 1 and the problem of its minimization can be formulated as a linear programming problem in case the set A is defined by linear constraints. This problem, however, can be of very large scale. Therefore the method based on generalized linear programming [4] will be described here. This method requires the solution of the linear programming problem of comparatively small dimension to be performed at each iteration. In what follows it will be assumed that the set A is defined by linear constraints and it is bounded. The observations ω belong to the set Ω which may or may not coincide with A.

ALGORITHM 2.

1. At the beginning select initial point $x^1 \in A$. For each j make ordering $i(k, j)$:

   $$1 \le k \le s_j + 1, \; |\varphi^T(z^j)x^1 - \omega^j_{i(k,j)}| \le |\varphi^T(z^j)x^1 - \omega^j_{i(k+1,j)}| \, ,$$

   $$i(k+1, j) = s_j, \; |\varphi^T(z^j)x^1 - \omega^j_{s_j}| = \max_{\omega \in \Omega} |\varphi^T(z^j)x^1 - \omega|$$

   Compute

   $$\bar{F}(x^1, s) = \sum_{j=1}^{l} \beta_j \sum_{k=1}^{s_j+1} p_k^{s_j} |\varphi^T(z^j)x^1 - \omega^j_{i(k,j)}|$$

   $$\bar{F}_x(x^1, s) = \sum_{j=1}^{l} \beta_j \, \varphi(z^j) \sum_{k=1}^{s_j+1} p_k^{s_j} \operatorname{sign}(\varphi^T(z^j)x^1 - \omega^j_{i(k,j)})$$

   where $p_k^{s_j}$ is defined according to (2.9)

2. Suppose that the method arrived at point x^r. We have a collection of points $\{x^1, \ldots, x^r\}$ and values $\bar{F}(x^1, s), \ldots, \bar{F}(x^r, s)$, $\bar{F}_x(x^1, s), \ldots, \bar{F}_x(x^r, s)$. At this point the algorithm proceeds as follows:

2a. Solve linear programming problem

$$\min_{\sigma, x} \sigma \qquad (4.6)$$

$$\bar{F}(x^i, s) + \langle \bar{F}_x(x^i, s), x - x^i \rangle \le \sigma \quad i = 1:r$$

$$x \in A$$

and obtain the point x^{r+1} as a solution of this problem and σ^r as its optimal value.

2b. For each j make ordering $i(k, j)$:

$$1 \le k \le s_j + 1, |\varphi^T(z^j)x^{r+1} - \omega^j_{i(k,j)}| \le$$

$$|\varphi^T(z^j)x^{r+1} - \omega^j_{i(k,j)}|$$

$$i(k+1, j) = s_j, |\varphi^T(z^j)x^{r+1} - \omega^j_{s_j}| = \max_{\omega \in \Omega} |\varphi^T(z^j)x^1 - \omega|$$

Compute

$$\bar{F}(x^{r+1}, s) = \sum_{j=1}^{l} \beta_j \sum_{k=1}^{s_j+1} p_k^{s_j} \operatorname{sign}(\varphi^T(z^j)x^r - \omega^j_{i(k,j)})$$

2c. If $\bar{F}(x^{r+1}, s) = \sigma^r$ then assign $\tilde{a}_s = x^{r+1}$ and stop, the estimate has been found. Otherwise go to step 2a.

This technique produces estimate $\tilde{a}_s$ in a finite number of steps because the function $\bar{F}(x, s)$ is piecewise linear.

REFERENCES

1 Arthanari, T.S. and Y. Dodge (1982): Mathematical Programming in Statistics. Wiley, New York.

2 Barrodale, I. and F.D.K. Roberts (1977): Algorithms for restricted least absolute value estimation. Commun. Statist.-Simula. Compta, B6(4), 353–363.

3 Bloomfield, P. and W.L. Steiger (1983). Least Absolute Deviations, Theory, Applications and Algorithms. Birkhäuser, Boston.

4 Dantzig, G.B.: *Linear Programming and Extensions*, Princeton University Press (1980).

5 Dupačová, J.: Minimax Stochastic Programs with Nonconvex Nonseparable Penalty Functions, in: A. Prékopa (ed.): Coloquia Mathematica Societatis János Bolyai 12. Progress in Operations Research, Eger 1974 North-Holland (1976) pp. 303–316.

6 Dupačová, J.: Minimax Stochastic Programs with Nonseparable Penalties, in: K. Iracki, K. Malanowsky, S. Walukiewicz (eds.): Optimization Techniques, Part 1. LN Control and Information Sciences, Springer-Verlag, Berlin, vol. 22 (1980) pp. 157–163.

7 Ermoliev, Yu.: "Method for stochastic programming in randomized strategies," *Kibernetica* 1 (1970).

8 Ermoliev, Yu.: *Methods of Stochastic Programming* (in Russian), Nauka, Moscow (1976).

9 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.

10 Gaivoronski, A.: Linearization methods for optimization of functionals which depend on probability measures. Mathematical Programming Study 28 (1986) 157–181.

11 Gaivoronski, A.: Stochastic optimization techniques for finding optimal submeasures in: V.I. Arkin, A. Shiraev, R. Wets eds. Stochastic optimization. Proceedings of the International Conference, Kiev, 1984. Lecture Notes in Control and Information Sciences, 1986, Springer-Verlag, Berlin, Heidelberg.

12 Gaivoronski, A.: Computing bounds for the solution of the stochastic optimization problem with incomplete information on distribution of random parameters. International Institute for Applied Systems Analysis, Working Paper WP-86-72.

13 Kall, P.: *Stochastic Linear Programming*, Springer Verlag, Berlin, 1976.

14 Kall, P.: Stochastic programs with recourse: an upper bound and the related moment problem. Institut für Operations Research der Universität Zürich. Working Paper, 1986.

15 Kankova, V.: An approximative solution of a stochastic optimization problem. Transaction of the eight Prague conference on information theory, statistical decision functions, random processes v. A, P. Reidel Publishing Company, Boston, 1978.

16 Krishnaiah, P.R. and Sen, P.K. eds: Handbook of Statistics. v. 4, Nonparametric methods, North Holland, 1984.

17 Kurzhanskii, A.B.: *Control and Observation under Conditions of Uncertainty*, Nauka, Moscow, 1977.

18 Lemarechal, C.: "Bundle methods in nonsmooth optimization", in: C. Lemarechal and R. Mifflin, eds., Nonsmooth Optimization, IIASA Proceedings Series, Vol. 3, 79–102. Laxenburg, Austria (1977).

19 Mainland, D. (1948): Statistical methods in medical research, Canadian Journal of Research, 26, (section E), 1–166.

20 Okamoto, M. (1958): Some inequalities relating to the partial sum of binomial probabilities, Annals of the Institute of Statistical Mathematics, Tokyo, 10, 29–35.

21 Prékopa, A. (1973): Contributions to the theory of stochastic programming. *Mathematical Programming* **4**, 202–221.

22 Rockafellar, R.T.: "Monotone operators and the proximal point algorithm," SIAM Journal on Control and Optimization **14**(1976), 877–898.

23 Serfling, R.: Approximation Theorems in Mathematical Statistics, Wiley, 1980.

24 Wets, R. Stochastic programming: solution techniques and approximation schemes, in *Mathematical Programming: The State of the Art*, eds., A. Bachem, M. Grötschel, and B. Korte. Springer Verlag, Berlin, 1983, pp. 566–603.