RM-73-7

A FEW METHODOLOGICAL REMARKS ON OPTIMIZATION RANDOM COST FUNCTIONS

Yuri A. Rozanov

December 1973

A Few Methodological Remarks on Optimization Random Cost Functions

Yuri A. Rozanov

Let $f(\alpha,x)$ be a functional of a variable $x \varepsilon X$, where α is some "unobservable" random parameter with a probability distribution P. Suppose we have to choose some point $x^o \varepsilon X$, and we like to optimize this procedure in some sense of minimization of $f(\alpha,x)$, $x \varepsilon X$, with unknown parameter α.

For example, $f(\alpha,x)$ may be a cost function of some economic model concerning future time, say

$$f(\alpha,x) = \sum_1^n \alpha_j x_j \quad , \qquad x = (x_1,\ldots,x_n) \varepsilon X \quad , \tag{1}$$

where X is a given convex set in n-dimensional vector space formed with inequalities

$$\sum_1^n \alpha_j x_j \geq b_i \quad , \qquad i = 1,\ldots,m \tag{2}$$

(including $x_j \geq 0$; $j = 1,\ldots,n$), and $\alpha = (\alpha_1,\ldots,\alpha_n)$ is a vector of "cost coefficients," which are expected to take values with some probability distribution $P(\cdot|\delta)$ under conditions of some given data δ.

Sometimes one uses a criterion based on minimization of mean value $Ef(\alpha,x)$, $x \varepsilon X$, and considers x^o as the optimal point if

$$Ef(\alpha,x^o) = \min_{x\varepsilon X} Ef(\alpha,x) \quad . \tag{3}$$

This criterion looks quite reasonable if one is going to deal with a big number N of similar models, and the total cost function can be approximately described (according to central limit theorem) as

$$\sum_{k=1}^{N} f(\alpha_k,x) \approx [Ef(\alpha,x)]\cdot N + \Theta\sqrt{N} \quad ,$$

where Θ is a random (normal) variable with mean zero and variance $\sigma^2(x) = Df(\alpha,x)$. But if you have to put in a big investment only once, then mean value criterion may not work well; moreover, the minimum point x^o of mean value function $Ef(\alpha,x)$, $x\varepsilon X$, can be the maximum point of the cost function $f(\alpha,x)$, $x\varepsilon X$, with a great probability.

In order to make this obvious remark clearer, let us mention a model of a non-symmetric coin game with two outcomes: $\alpha = \alpha_1,\alpha_2$, which takes place with corresponding probabilites $p_1,p_2 = 1 - p_1$, and cost function is $f(\alpha,x)$ with $x = x_1,x_2$. One has to pay $f_{ij} = f(\alpha_i,x_j)$ under the outcome α_i if he chooses in advance the strategy x_j $(i,j = 1,2)$. Suppose $f_{ij} = C$ $(i\neq j)$, where C is the all gambler capital (so he will lose this capital C under the strategy x_j if it be the outcome α_i, $i\neq j$), and $f_{ii} = -M_iC$ (he will increase the initial capital C in M_i times). The mean value function is

$$Ef(\alpha,x) = \begin{cases} C(-M_1p_1 + p_2) & \text{if } x = x_1 \\ C(p_1-M_2p_2) & \text{if } x = x_2 \end{cases} .$$

Suppose the outcome α_1 takes place with a great probability p_1 (say $p_1 = 0.999$) and M_2 is so big that

$$p_1-M_2p_2 < -M_1p_1 + p_2 \quad .$$

Using mean value criterion, we obtain $x^o = x_2$ as the optimal point, but obviously this is a very foolish strategy, except in the case when one should very much like to lose his capital (because it will be with the great probability 0.999). Another similar example: suppose the cost function is

$$f(\alpha,x) = \begin{cases} \alpha_{10} + \alpha_{11}x & \text{with probability } p_1 \\ \alpha_{20} + \alpha_{21}x & \text{with probability } p_2 = 1 - p_1 \end{cases}$$

(say $p_1 = 0.999$, $p_2 = 0.001$) where $0 \leq x \leq 1$ and the cost coefficients α_{11},α_{21} are such that $\alpha_{11} > 0$; $\alpha_{11}p_1 + \alpha_{21}p_2 < 0$.

Using mean value criterion, we have to choose $x^o = 1$, though with the great probability p_1 ($p_1 = 0.999$) it will be the maximum point (see Fig. 1) of the actual cost function $f(\alpha,x)$, $0 \leq x \leq 1$.

Concerning the mean value type criterion, we wish to say some other things. It is very easy to realize that one may prefer a random variable $\eta_1 = f(\alpha,x_1)$ in comparison to

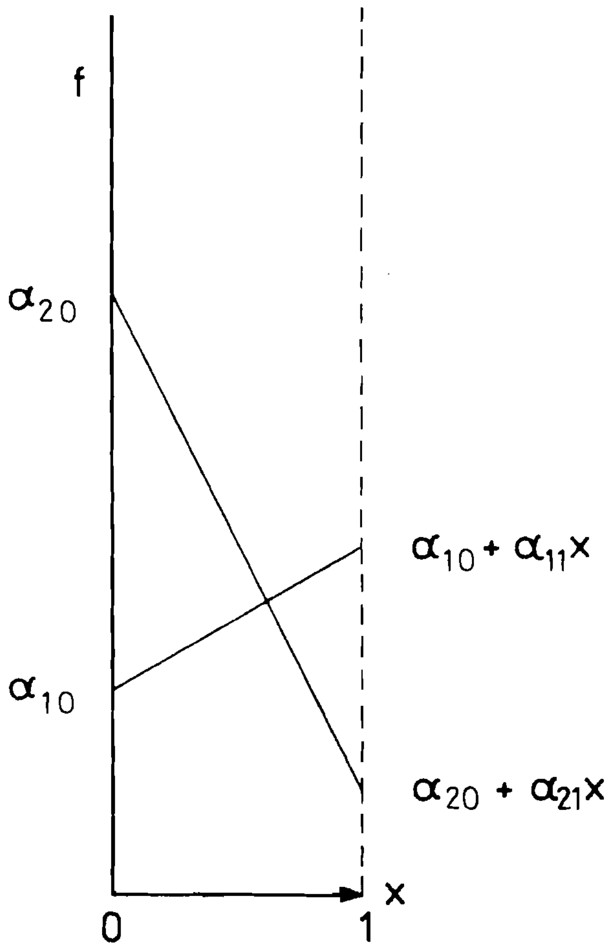

FIGURE 1

another random variable $\eta_2 = f(\alpha, x_2)$ if for some crucial point y

$$F_1(y) = P\{\eta_1 \leq y\} \geq \{\eta_2 \leq y\} = F_2(y) \quad .$$

Of course, there may be a few, in some sense, crucial points $y = y_1, \ldots, y_n$. Suppose it is possible to estimate "an importance" of these points with the corresponding values $u(y)$, $y = y_1, \ldots, y_n$ in such a way that one prefers η_1 (as compared to η_2) if

$$\sum_k F_1(y_k)\ u(y_k) \geq \sum_k F_2(y_k)\ u(y_k) \quad .$$

The preference relation can be rewritten in the form

$$\int F_1(y)\ dU\ (y) \geq \int F_2(y)\ dU\ (y) \quad ,$$

where

$$U(y) = \sum_{y_k \leq y} u(y_k) \ , \qquad -\infty < y < \infty \quad .$$

Because for any distribution function $F(y)\,(F(-\infty) = 0,\ F(\infty) = 1)$ we have

$$\int F(y)\ dU\ (y) = - \int U(y)\ dF(y) + U(\infty) \quad ,$$

the preference criterion can be represented in the form

$$EU(\xi_1) \leq EU(\xi_2) \quad , \tag{4}$$

where $E(\cdot)$ is the corresponding mean value.

One can consider (4) for arbitrary distribution type function $U(y)$, $-\infty < y < \infty$ as the general <u>mean value criterion</u>. Obviously, if the corresponding density $u(y)$, $-\infty < y < \infty$ is <u>positive</u>, then $U(y)$, $-\infty < y < \infty$ is a monotone increasing function. Besides, if for any $y_1 \leq y_2$ on some interval we consider y_1 as "more important" in comparison with y_2, more precisely if

$$u(y_1) \geq u(y_2) \quad , \qquad y_1 \leq y_2 \quad ,$$

i.e. the density $u(y)$, $x \varepsilon I$ is a monotone decreasing function on the interval I, then the preference function $U(y)$, $y \varepsilon I$, is <u>convex</u> (see Fig. 2).

We are going to suggest below a few other types of criteria of optimization for random cost functions.

1. Let $f(\alpha,x)$, $x \varepsilon X$ be a cost function which depends on a random parameter α. Suppose for some <u>acceptable</u> cost value C we can neglect a probability that the actual cost will exceed C. Suppose that minimal (random) cost

$$C(\alpha) = \min_{x \varepsilon X} f(\alpha,x)$$

has a probability distribution with a rather <u>small range</u> and corresponding minimum point $\xi \varepsilon X$:

$$f(\alpha,\xi) = \min_{x \varepsilon X} f(\alpha,x)$$

has a <u>discrete distribution</u> (maybe with a very big dispersion).

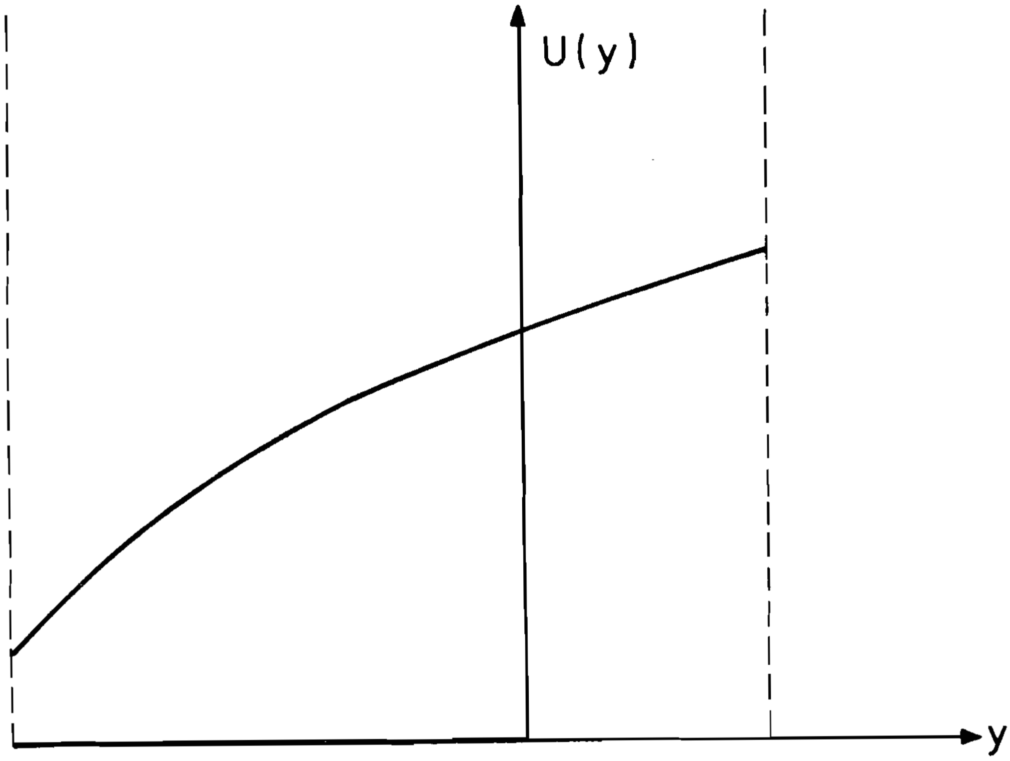

FIGURE 2

It seems quite reasonable to take a risk to choose such point $x^o \varepsilon X$ for which

$$P\{f(\alpha,x^o) = C(\alpha)\} = \max_{x\varepsilon X} P\{f(\alpha,x) = C(\alpha)\} \quad . \tag{5}$$

Note that if the probability in the relation (5) equals to 1, in other words, there is a point $x^o \varepsilon X$ for which

$$f(\alpha,x^o) = \min_{x\varepsilon X} f(\alpha,x) \text{ with probability } 1 \quad ,$$

then our criterion gives the usual minimum of cost function.

Let us consider the linear cost function

$$f(\alpha,x) = \sum_1^n \alpha_j x_j$$

of $x = (x_1,\ldots,x_n)\varepsilon X$, where $\alpha = (\alpha_1,\ldots,\alpha_n)$ is the random vector with a given probability distribution P, and X is a simplex in n-dimensional vector space of the type (2):

$$\sum_1^n a_{ij}x_j \geq b_i \quad ; \qquad i = 1,\ldots,m \quad .$$

Denote $x^1,\ldots,x^N$ extreme points of simplex X. As well known, a minimum point $\xi\varepsilon X$ (ξ depends on α) can be chosen among $x^1,\ldots,x^N$, so $x^o = x^1,\ldots,x^N$ is the optimal point in the sense of the criterion (5) if

$$P\{\xi = x^o\} = \max_{1\leq k\leq N} P\{\xi = x^k\} \quad . \tag{6}$$

Thus, the problem is to find all probabilities*

$$p_k = P\{\xi = x^k\} \quad ; \qquad k = 1,\dots,N$$

and to choose the optimal x^o as the point among x^k; $k = 1,\dots,N$, with the greatest probability p_k; $k = 1,\dots,N$.

We have $p_k = P(Y^k)$ where Y^k is the set of all vectors $y = (y_1,\dots,y_n)$ for which the corresponding linear function

$$f(y,x) = \sum_1^n y_j x_j \quad , \; x \varepsilon X$$

has x^k as the minimum point:

$$f(y,x^k) = \min_{x \varepsilon X} f(y,x) \quad .$$

In order to make our elementary consideration more clear, let us shift x^k to the origin point $x = 0$. Obviously, the extreme point $x^k = 0$ gives a minimum of $f(y,x)$, $x \varepsilon X$, iff

$$\sum_1^n y_j x_j \geq 0 \text{ for all } x \varepsilon X \quad ,$$

(in other words, iff the vector $y = (y_1,\dots,y_n)$ belongs to so-called <u>polar cone</u>).

Let us take <u>all</u> hyperplains

$$\sum_1^n a_{ij} x_j = b_i \quad , \; i \varepsilon I_k \tag{7}$$

---

* Note the events $\{\xi = x^k\}$; $k = 1,\dots,N$ generally are not disjoined and $\sum_1^N p_k$ not necessary equals to 1.

--see (2)--containing the extrème point x^k. (In the case $x^k = 0$ we have $b_i = 0$, $i \varepsilon I_k$.) Let us introduce a cone

$$X^k = \bigcap_{i \varepsilon I_k} \{x: \sum_1^n a_{ij}x_j \geq 0\} \quad .$$

The corresponding polar cone is exactly the set Y_o^k of all vectors $y = (y_1,\dots,y_n)$ such that $\sum_1^n y_j x_j \geq 0$, $x \varepsilon X^k$ (see Fig. 3). This polar cone Y_o^k is formed by all linear combinations

$$y = \sum_{i \varepsilon I_k} \lambda_i a_i \quad ; \qquad \lambda_i \geq 0 \tag{8}$$

of the vectors $a_i = (a_{i1},\dots,a_{in})$, $i \varepsilon I_k$ because a dual polar cone for the set of all vectors (8) coincides with X^k: obviously,

$$\sum y_j x_j = \sum_{i \varepsilon I_k} \lambda_i (\sum a_{ij} x_j) \geq 0$$

for all $\lambda_i \geq 0$, iff $x \varepsilon X^k$. (See, for example, duality theorem in [1].) Thus, $\underline{Y^k = x^k + Y_o^k}$ is the set of all vectors

$$y = x^k + \sum_{i \varepsilon I_k} \lambda_i a_i \quad , \qquad \lambda_i \geq 0 \quad , \tag{9}$$

where $a_i = (a_{i1},\dots,a_{in})$ are all vectors such that for $x = x^k$ at the relations (2) we have strict equalities, and the optimal point can be found among x^k, $k = 1,\dots,N$ as a point with maximum probability

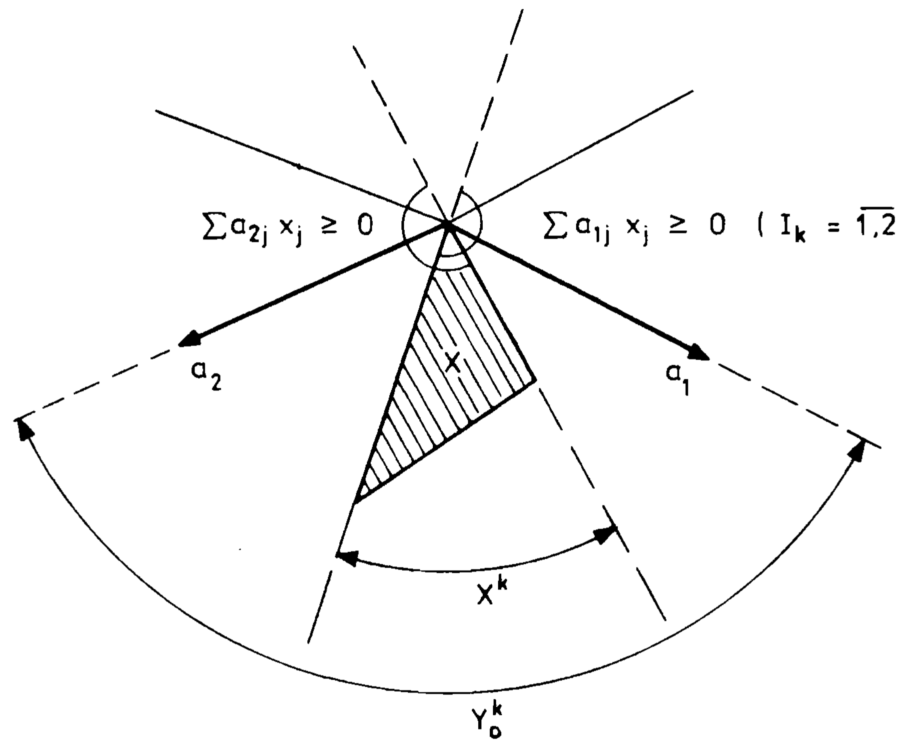

FIGURE 3

$$P(Y_k) = P\{\alpha \varepsilon Y_k\} \quad ; \quad k = 1,\dots,N \quad . \tag{10}$$

2. Suppose, as above, there is the acceptable cost, which can be exceeded only with a corresponding small probability, but the situation is different in the sense that the range of the minimum cost distribution is considerably big. (For example, the minimum point $\xi = x^1, x^2$ can be distributed with almost equal probabilities $P_1 > P_2$, but corresponding cost values are such that $f(\alpha, x^1) >> f(\alpha, x^2)$, so there is no reason to choose the point x^1 with the greatest probability P_1 as optimum.)

Suppose that one is going to risk in order to make the cost value less than some level C_o. (Probability $P\{C(\alpha) \leq C_o\}$ has to be considerably big.) Then one can choose optimal point $x^o \varepsilon X$ in the sense that

$$P\{f(\alpha, x_n^o) \leq C_o\} = \max_{x \varepsilon X} P\{f(\alpha, x) \leq C_o\} \quad . \tag{11}$$

This criterion is of mean value type (4) concerning a new cost function $EU(f(\alpha,x))$, $x \varepsilon X$ where

$$U(y) = \begin{cases} 1 & \text{if } y \leq C_o \\ 0 & \text{if } y > C_o \end{cases} \quad ,$$

namely,

$$EU(f(\alpha, x^o)) = \min_{x \varepsilon X} EU(f(\alpha, X)) \quad . \tag{12}$$

(Note it is impossilbe to restrict "y" in order to deal with the convex function $U(y)$, $y \varepsilon I$.)

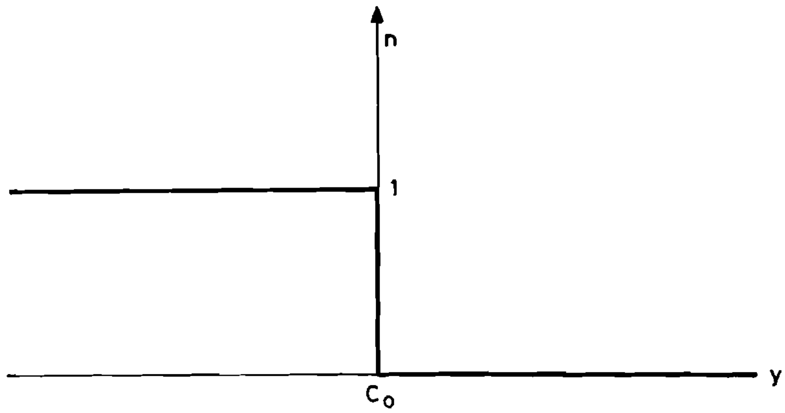

FIGURE 4

3. Suppose, now, there is a good deal of risk to pay a big amount if we use "extreme strategy" x^0 of types (5) or (11), because with considerably big probability, cost value $f(\alpha,x^0)$ may be too much. Suppose one should like to prevent a danger of dealing with the "almost worst" outcome α, and the problem is to find optimal strategy against "very clever random enemy." In this situation, the following criterion seems quite reasonable (similar to the minimax principal of game theory).

Namely, suppose one agrees (roughly speaking) to risk only with a small probability $\varepsilon \geq 0$. Let $C(x)$ be the "ε-quantil" for the random variable $f(\alpha,x)$:

$$C(x) = \min C \mid P\{f(\alpha,x) \leq C\} \leq 1-\varepsilon \quad . \tag{13}$$

One can choose the point $x^o \varepsilon X$, which is optimal in the sense that

$$C(x^o) = \min_{x\varepsilon X} C(x) \quad . \qquad (14)$$

In the case of $\varepsilon = 0$, our criterion of optimality coincides with well known minimax principal of the game theory, which was mentioned above, because if $\varepsilon = 0$, then

$$C(x) = \sup_{\alpha} f(\alpha,x) \quad .$$

(We mean so-called essential sup $f(\alpha,x)$ concerning the probability distribution P of the random variable α.)

For the linear cost function (1) with the coefficients $\alpha = (\alpha_1,\ldots,\alpha_n)$ which are weakly dependent, one can expect the random variable $f(\alpha,x) = \sum_1^n \alpha_j x_j$ is normally distributed (due to the central limit theorem) with a mean value

$$(c,x) = \sum_1^n c_j x_j$$

and variance

$$\|\sigma^{\frac{1}{2}} x\|^2 = \sum_1^n \sum_1^n \sigma_{ij} x_i x_j$$

$(c_i = E\alpha_i;\ \sigma_{ij} = E(\alpha_i - c_i)(\alpha_j - c_j);\ i,j = 1,\ldots,n)$.

If it holds true, then

$$C(x) = \sum_1^n c_i x_i + y_\varepsilon\Big(\sum_1^n \sum_1^n \sigma_{ij} x_i x_j\Big)^{\frac{1}{2}},\ x\varepsilon X \quad ,$$

where y_ε denotes ε-quantil for the standard normal distribution:

$$\frac{1}{\sqrt{2\pi}} \int_{y_\varepsilon}^{\infty} e^{-y^2/2}\, dy = \varepsilon \quad .$$

This function

$$C(x) = (c,x) + y_\varepsilon \, \| \sigma^{\frac{1}{2}} x \| \quad , \ x \varepsilon X$$

(where $\sigma^{\frac{1}{2}}$ means the square root of the positive matrix $\{\sigma_{ij}\}$) for $y_\varepsilon > 0$ is concave because

$$\| \sigma^{\frac{1}{2}} \frac{x_1 + x_2}{2} \| \leq \frac{1}{2}\left(\| \sigma^{\frac{1}{2}} x_1 \| + \| \sigma^{\frac{1}{2}} x_2 \| \right)$$

and the minimum point x^0 can be found with well known concave programming methods. (See, for example, [1].)

References

[1] Karlin, S. Mathematical Methods and Theory in Games, Programming, and Economics, Vol. 1. Reading, Mass. Addison-Wesley, 1959.