Statistical applications of optimization methods and mathematical programming

Daniela-Ioana MANEA
The Bucharest University of Economic Studies, Romania
daniela.manea@csie.ase.ro

Emilia ȚIȚAN
The Bucharest University of Economic Studies, Romania
emilia.titan@csie.ase.ro

Radu R. ȘERBAN
Spiru Haret University, Bucharest
radu_ser@yahoo.com

Mihaela MIHAI
The Bucharest University of Economic Studies, Romania
emilia.titan@csie.ase.ro

Abstract. Optimization techniques perform an important role in different domains of statistic. Examples of parameter estimation of different distributions, correlation analysis (parametric and nonparametric), regression analysis, optimal allocation of resources in partial research, exploration of response surfaces, design of experiments, efficiency tests, reliability theory, survival analysis are the most known methods of statistical analysis in which we find optimization techniques. The paper contains a synthetic presentation of the main statistical methods using classical optimization techniques, numerical optimization methods, linear and nonlinear programming, variational calculus techniques. Also, an example of applying the "simplex" algorithm in making a decision to invest an amount on the stock exchange, using a prediction model.

Keywords: optimization techniques, statistics, mathematical programming.

Introduction
Given the complexity of contemporary socio-economic relationships, management decision-making based solely on the experience and intuition of the management team cannot meet the imperatives of efficient management of economic processes. In this sense, numerous studies, researches, appeared in the last decades, in the field of optimal management of economic processes, have demonstrated and required, with necessity, the use of economic-mathematical models.

An important place, in this respect, is the economic-mathematical models of optimization, which although it is an essential chapter of the operational research, at the same time, is an extremely useful tool in making optimal economic decisions.

In many scientific fields, in the economy, both at micro level, as macroeconomic in industry, in any decision-making process, optimization techniques are commonly used.
Optimization issues are a subject of study in many disciplines such as economics, statistics, operational research, game theory, artificial intelligence, engineering (chemical, electrical etc.). Thus, in the Aspects of multidimensional optimization, Mateia & Cechin-Crista (2012) analyzes the use of multidimensional optimization techniques in artificial intelligence and genetic algorithms.

Optimization techniques have developed a lot in recent years and due to the possibility of solving complex applications using specialized software packages.

Because statistical methods have a close connection with optimization techniques, some researchers consider statistics to be a subdomain of optimization. There are many applications of optimization methods in the main branches of statistics and therefore the study of optimization techniques is an area of interest for statisticians.

JS Rustagi (1971) highlights the variety of use of optimization techniques in statistics in the volume Optimizing Methods in Statistics.

Although the wide-ranging use of statistical optimization techniques can be identified, we will, in the following, exemplify the frequently used statistical methods in which we find these techniques.

The examples will be structured into two categories, those using classical optimization methods and those that use numerical optimization methods. Also, will be describe linear and non-linear programming methods, variational computational techniques that are related to dynamic programming and Pontryagin’s Principle.

**Classic optimization techniques used in statistics**

We will continue to analyze some statistical applications that are based on classic optimization methods.

A first applicability can be considered as estimating the parameters of a statistical population based on partial research.

Either $X_1, X_2,..., X_n$ selection data from a population with normal distribution with mean $\mu$ and $\sigma^2$ variance.

Estimating the mean $\mu$ and $\sigma^2$ variance by the maximum verosimility method requires maximizing the $L(\mu, \sigma^2)$ function, called maximum verosimility function, where

$$L(X_1,...,X_n; \mu, \sigma^2) = \prod_{i=1}^{n} f(X_i; \mu, \sigma^2) = \left( \frac{1}{\sigma \sqrt{2\pi}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2}$$

with

$$f(X_i; \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2\sigma^2} \left( \frac{x_i - \mu}{\sigma} \right)^2}, \ x \in R, \ \mu \in R, \ \sigma > 0.$$ 

In this problem, the solution can be obtained by solving the system:

$$\begin{align*}
\frac{\partial \log L(x_1,...,x_n; \mu, \sigma)}{\partial \mu} &= 0 \\
\frac{\partial \log L(x_1,...,x_n; \mu, \sigma)}{\partial \sigma} &= 0
\end{align*}$$

If, in addition, as is often the case, a non-negativity restriction on the mean is required ($\mu > 0$) then it must be taken into account in the optimization process to obtain its estimate $\mu$. 

Similarly, optimization methods are used and for estimating vector averages $\tilde{\mu}$ and covariance matrix $\Sigma$ in a $p$-varied normal population.

Let the selection data $X_1, X_2, ..., X_n$. The logarithm of the maximum verosimility equation is:

$$L(\tilde{\mu}, \Sigma) = -\log|\Sigma| - (\Sigma^{-1}V)' , \text{ cu } V = \frac{1}{n} \left[ \sum (X_i - \bar{X})(X_j - \bar{X}) \right]'$$

Another example is the boundaries of the series of correlation coefficients determined in the time series analysis. The correlation coefficient of order $s$ is:

$$r_s = \frac{n}{n-s} \cdot \frac{\sum_{i=1}^{n} x_{i+s} \sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} x_i^2}$$

It is considered the upper limit of $r_1$ for $s=1$. This is a problem of maximizing it $r_1$, equivalent to:

$$\max \sum_{i=1}^{n} x_i x_{i+s},$$

with the restriction $\sum_{i=1}^{n} x_i^2 = \text{constant}$. The solution of the above problem can be obtained using the Lagrange method.

Contrary to the established opinion that correlation coefficient values belong to the interval $[-1, +1]$, the series of correlation coefficients (Bertsekas, 1999), (Bickel & Doksum, 1977) has limits greater than one.

In the case of using regression analysis with restrictions, the general multiple regression model is:

$$Y = X\beta + \varepsilon$$

where $Y$ is a vector of $n \times 1$ size, $X$ is a known matrix of $n \times p$ size, $\beta$ is a vector of the size parameters $p \times 1$ and $\varepsilon$ is the random vector of $n \times 1$ size, with mean $0$ and covariance $\sigma^2 I$.

The estimators of the $\beta$ vector, obtained by the least squares method, are determined by solving the optimization problem:

$$\min_{\beta} (Y - X\beta)'(Y - X\beta).$$

If some restrictions are applied to the $\beta$ vector (e.g. $\beta \geq \beta_0$), we have a limitation optimization problem.

The optimal allocation of resources in partial research is another example of using optimization methods. In designing and carrying out partial research, issues related to the optimal allocation of resources (financial, time, selection, etc.) have to be solved. Cochran (1977) extensively approaches sampling problems using optimization methods.

We consider, as an example, the case of the cluster survey when we are interested in determining the optimal size of a minimal variation cluster for a given cost.

Suppose $M$ represents the total number of statistical units to be distributed between the $N$ clusters of size $M_0$ each. Let $S^2_w$ the within-cluster variance and $S^2_b$ the between-cluster variance. Let $n$ be the volume of the sample (number of clusters in the sample).
Let \( C_b \) be the cost of cluster research, irrespective of its size, and \( C_w \) the cost associated with the research of each element of the cluster, regardless of the size of the cluster. For a fixed \( C \) cost, we have:

\[
C = nC_b + nM_0C_w
\]

and it is desired to minimize the overall variance of the overall mean \( \bar{y} \),

\[
V(\bar{y}) = \left(1 - \frac{n}{N}\right) \frac{S_w^2 + M_0S_b^2}{nM_0}.
\]

The optimal solution proves to be:

\[
M_{op} = \frac{C_b}{C_w} \frac{S_w^2}{S_b^2}.
\]

**Numerical optimization methods used in statistics**

A first applicability of these methods is found in the problem of estimating Gamma distribution parameters.

\[
f(x; \alpha, \beta) = \begin{cases} 
  x^{\alpha-1} e^{-\frac{x}{\beta}} / \Gamma(\alpha) \beta^\alpha, & x > 0 \\
  0, & x \leq 0
\end{cases}
\]

Estimates of maximum verosimility are obtained, in this case, by equalized with zero the partial derivatives of the logarithm of the maximum verosimility function \((\log L)\) in relation to the parameters \( \alpha \) and \( \beta \), where:

\[
\log L = -\sum \frac{x_i}{\beta} + (\alpha - 1)\sum \log x_i - n\log \Gamma(\alpha) - n\alpha \log \beta.
\]

The equations are:

\[
\frac{\partial \log L}{\partial \alpha} = \bar{x} - \alpha \beta = 0,
\]

\[
\frac{\partial \log L}{\partial \beta} = \sum \log x_i - n \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n \log \beta = 0.
\]

These equations are solved using numerical methods, tables of the digamma function \( \Gamma'(\alpha)/\Gamma(\alpha) \) being available (Pearson, 1970), making the solution easy.

Resolving models from reliability theory and survival analysis is often done using numerical optimization methods.

In the model of failure, the Weibull distribution is used, the probability density being given by the function:

\[
f(t) = \begin{cases} 
  \frac{\beta}{\delta} \left(\frac{t-\mu}{\delta}\right)^{\beta-1} e^{-\frac{(t-\mu)}{\delta}}, & t \geq \mu \\
  0, & t < \mu
\end{cases}
\]

where \( \beta, \delta > 0 \) and \( \mu \geq 0 \) represent the three parameters of the Weibull distribution.

Let's assume that the experiment is make over the period \((0,t_0)\) and the moments of failure for the \( r_0 \) produced from the \( r \) tested are given by \( t_1, t_2, \ldots \). Then the verosimility function for the sample is given by:
Estimates of maximum verosimility for $\mu$, $\beta$, and $\delta$ can only be obtained using numerical methods.

In case of survival analysis we consider that the moment of death of a person is a random variable that follows an exponential distribution of the $\lambda$ parameter.

Let us assume that the volume of sample undergoing study at time $t$ is $n$ persons and we know the number of $r$ deaths observed at moments $t_1, t_2, ..., t_r$. The maximum verosimility function for estimating the $\lambda$ parameter is in this case:

$$L = \lambda^r e^{-\lambda \sum_{i=1}^{r} t_i} (1 - e^{-\lambda t_i})^{n-r}$$

or through logarithm

$$\log L = r \log \lambda - \lambda \sum_{i=1}^{r} t_i + (n-r) \log(1 - e^{-\lambda t_i}).$$

Therefore, the maximal equivalence equation for $\lambda$ is:

$$\frac{r}{\lambda} - \sum_{i=1}^{r} t_i + \frac{ae^{-\lambda t_i}}{1 - e^{-\lambda t_i}} = 0.$$  \hspace{1cm} (19)

For determining the $\lambda$ solution, the corresponding tables are also used in this case.

Another case of using numerical methods is the response surface design. The following relationship is considered between the $\mu$, the response mean of a $y$ variable and the matrix of the independent variables $X$, respectively the $\theta$ vector of the unknown parameters:

$$\mu = f(X, \theta),$$ \hspace{1cm} (20)

where $X = (X_1, X_2, ..., X_p)$ represents a $p$-dimensional "design" variable, and $\theta = (\theta_1, \theta_2, ..., \theta_r)$ represents the $r$-dimensional vector of the parameters.

Since the function $f$ is generally unknown, it is approximated by a polynomial function of a certain degree, and then estimates for $\theta$ are obtained by an estimation method.

Let $\hat{Y}(X)$ be the estimated response to the factorial variables $X$. It is necessary to specify that the independent variables $X_1, X_2, ..., X_p$ are not random variables. The problem in response surface design is finding that $X$ variable that minimizes

$$J = \int (\hat{Y}(X) - \mu(X))^2 dX$$

over the class of all variables of vector $X$.

**Applications of mathematical programming in statistical analysis models**

Mathematical optimization techniques with inequality restrictions lead, in most cases, to mathematical programming problems. There is a rich literature (Arthanari & Dodge, 1981), (Vaduva, 1968) on the application of mathematical programming in statistics.

A prime example of this is linear regression. The common model of linear regression is as follows:

$$Y = X\beta + \varepsilon,$$  \hspace{1cm} (22)

where $Y$ is $n$-dimensional vector, $X$ is a known matrix of dimensions $n \times p$, $\beta$ is the $p$-dimensional vector of unknown parameters and $\varepsilon$ is the $n$-dimensional residue vector.
Suppose we have, in addition, the restriction, \( C\beta \geq 0 \) where \( C \) is a known matrix of size \( g \times p \).

The estimation of the vector \( \beta \) is usually done using the least squares method that consists in determining the minimum expression:

\[
(Y - X\beta)'(Y - X\beta).
\]  
(23)

The above problem is a square programming problem and there are algorithms well known to solve such problems. More authors have significant contributions in this field, such as Judge and Takayama (1976), Maliţa and Dragomirescu (1972), respectively Şerban and Dumitrescu (1998).

The use of mathematical programming in statistics is also found in the case of the sampling problem. One of the most common problems of surveys is the estimation of \( y \) from a total collectivity, on a finite population given by

\[
y = \sum_{i=1}^{T} N_i y_i^*,
\]  
(24)

where \( y_1^*, y_2^*, \ldots, y_T^* \) are possible individual values of population elements, and \( N_i \) represents the number of statistical units that take \( y_i^* \) values, \( i = 1, \ldots, T \). Let \( n_i \) the number of statistical units in the sample with \( y_i^* \) values, so that, in the simple random sampling, \( n \)-size sampling, we have \( n = \sum n_i \), estimator of maximum verosimility of \( y \) is obtained by maximizing the verosimility function:

\[
L = \prod_{i=1}^{T} \frac{\left( N_i \right)^{n_i}}{n!} \left( \frac{N}{n} \right),
\]  
(25)

where \( n = \sum n_i \) is the sample size.

In this case, optimization is reduced to solving a programming problem with integer numbers, studied by Rao (1984).

Another example is the design of experiments, an important class of statistical applications being that of factorial experiments. If we have a large number of factors, all the possible combinations cannot be used and some of these factors are used to solve the problem. This factorial fractionation is the subject of an optimal cost problem. The problem of finding the optimal cost of factorial fractionation leads to solving an optimization problem. In this sense, the results of Neuhardt and Mount-Campbell (1982) are known.

The use of mathematical programming in statistics is also found in the problem of minimizing the sum of absolute value of the errors of a regression model of the form:

\[
y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}, \quad i = 1, 2, \ldots, r, \quad j = 1, 2, \ldots, r, \quad k = 1, 2, \ldots, n,
\]  
(26)

with \( \sum \alpha_i = \sum \beta_i = 0 \). Here, \( y_{ijk} \) is the \( k \) observation for the \( i \) level of the first factor and for the \( j \) level of the second factor.

The smallest estimated absolute values of the \( \mu, \alpha_i, \beta_j \) parameters are obtained by minimizing the expression:
\[ \sum_{i} \sum_{j} \sum_{k} |y_{ijk} - \mu - \alpha_i - \beta_j|. \]  

(27)

This problem is equivalent to the following linear optimization problem:

\[ \min \sum_{i} \sum_{j} \sum_{k} (d_{ijk}^+ + d_{ijk}^-), \]  

(28)

or otherwise

\[ \mu + \alpha_i + \beta_j + d_{ijk}^+ - d_{ijk}^- = y_{ijk}, \text{cu } d_{ijk}^+ \geq 0 \text{ si } d_{ijk}^- \geq 0. \]  

(29)

A significant number of applications of statistical programming methods used to estimate the lowest absolute value are described in the literature by Gentle (2006).

The estimation of Markov chains is another well-known example of the application of mathematical programming in statistics. Consider the probability of the \( p_{ij}(t) \) transition probability of the \( x_i, t = 1,2,...,T \), \( i, j = 1,2,...,r \) Markov chain.

As is known, \( p_{ij}(t) \) probabilities are given by:

\[ p_{ij}(t) = P\{x_i = s_j | x_{i-1} = s_i\}, \]  

(30)

where \( s_j, i = 1,2,...,r \) is a chain number. We make it clear that

\[ \sum_{i} \sum_{i} p_{ij}(t) = 1, \quad 0 \leq p_{ij} \leq 1. \]  

(31)

Suppose the chain is observed for \( N(t) \) independent attempts, \( t = 1,2,...,T \). Let \( w_j(t) \) the proportion of events in category \( j \). Function of maximum verosimility for the sample can be written as follows:

\[ L = \prod_{t=1}^{T} \frac{N(t)!}{\prod_{m} (N(t)w_m(t))! N(t) - \sum_{k} (N(t)w_k(t))} \frac{N(t)!}{\prod_{j} (w_j(t-1)p_{ij}(t))} \frac{N(t) - \sum_{k} (N(t)w_k(t))}{1 - \sum_{i} \sum_{j} (w_j(t-1)p_{ij}(t))} \]  

(32)

The issue of maximizing the verosimility function written above is a nonlinear programming problem. This problem, as well as others, has been studied by Lee, Judge, and Zellner (1970).

**The simplex algorithm for multidimensional optimization**

The algorithm is due to Nelder and Mead (1965) and is one of the most commonly used algorithms for minimizing an assumed \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) function.

Before describing the method, we recall that a simplex in \( \mathbb{R}^n \) is a set \( \{x_1, x_2, ..., x_n, x_{n+1}\} \) of \( (n + 1) \) points in this space.

Suppose the points of the simplex are numbered such that:

\[ f(x_1) \leq f(x_2) \leq \cdots \leq f(x_n) \leq f(x_{n+1}) \]  

and let \( d^0 = x_0 - x_{n+1} \) be the centroid of the first \( n \) points of the simplex.

Three types of new points are calculated as follows:

- an \( x \) point in the direction \( d^0 = x_0 - x_{n+1} \); \( d^0 = x_0 - x_{n+1} \) (33)

point that is obtained by a reflection of the point \( x_{n+1} \) (the point where \( f \) takes the highest value) with respect to the centroid \( c_{x_0} \).

The scalar \( \alpha > 0 \) is the reflection factor (or reflection size).

If the value of function \( f \) in the new point \( x_r \), verifies the inequalities:
\[
f(x_r) \leq f(x_c) < f(x_e)
\]
then the point \( x_{n+1} \) is replaced by the point \( x_r \), better, in the sense of the minimum function \( f \), that is \( x_{n+1} := x_r \).

If \( f(x_r) < f(x_e) \) then the \( d^1 = x_r - x_0 \) direction is a decrease direction of the \( f \) function.

• on this direction a point is sought
\[
x_v = x_0 + \gamma d^1
\]
point that is obtained by an expansion of the simplex in the \( d^1 \) direction of decrease.

The scalar \( \gamma \geq 1 \) is the expansion factor.

• If \( f(x_v) < f(x_c) \), then \( x_{n+1} \) point is replaced with \( x_v \), \( x_{n+1} := x_v \);

• If \( f(x_v) \geq f(x_c) \), then \( x_{n+1} := x_c \).

If \( f(x_r) \geq f(x_{n+1}) \) (which means that the reflected point does not improve the value of the \( f \) function), then in the \( d^2 = x_{n+1} - x_0 \ (d^2 = -d^0) \), direction the point is calculated
\[
x_i := x_0 + \beta d^2
\]
called a contraction point. The scalar \( 0 < \beta < 1 \) is the contraction factor.

• If \( f(x_v) < f(x_{n+1}) \), then \( x_{n+1} \) point is replaced with \( x_{n+1} \) cu \( x_v \) \( x_{n+1} := x_v \);

• If \( f(x_v) \geq f(x_{n+1}) \), then neither the contraction in the \( d^2 \) direction was successful and the whole simplex is contracted around the point \( x_i \) where the value of the function \( f \) is the smallest \( x_i = \frac{1}{\delta}(x_i + x_e) \), \( i = 1,2,...,n+1 \), \( 36 \)

where \( \delta \geq 2 \).

Different stopping criteria are known. Different stopping criteria are known. Nelder and Mead (1965) propose the following stop criterion:

\[
\overline{f} = \frac{1}{n+1} \sum_{i=1}^{n+1} f(x_i) \quad \text{and} \quad s = \left[ \frac{1}{n+1} \sum_{i=1}^{n+1} (f(x_i) - \overline{f})^2 \right]^{\frac{1}{2}}
\]

• if \( s \leq \varepsilon \) (accuracy required), the algorithm also stops \( x^* := x_i \);

• if \( s > \varepsilon \), continue with the new simplex obtained.

**Variational analysis methods used in statistics**

Variational analysis methods refer to the optimization of functionalities in a given class, such as maximizing or minimizing integral functions with certain restrictions. There are many applications in statistics that depend to a large extent on variational analysis techniques. Below are some examples illustrating the use of modern and classical variational analysis techniques in statistics.

In this sense, a first example is the statistics that use variables measured on the ordinal scale (order statistics). Let us assume that \( x_1 \leq x_2 \leq ... \leq x_n \) represents ordered selection data, having a continuous distribution function \( F(x) \). The expected value with the highest rank, i.e. \( x_n \), is given by:
\[
L(F) = \int xd(F^n(x)).
\]
An important problem in using these statistics is to find the lower and upper limits of the $L(F)$ function when the average and variance of the random variable $X$ are known. Similarly, the expected amplitude of the sample $(x_n - x_i)$ can be found. This means:

$$\min(\max) \int xd\left[1 - F^n(x) - (1 - F(x))^n\right],$$

with some restrictions.

Problems of this type appear, as a rule, in the nonparametric statistical inference and have been extensively analyzed by Rustagi (1971).

Mann-Witney-Wilcoxon statistics are a second example of using variational analysis methods. Suppose we are interested in finding the limits of the Mann-Witney-Wilcoxon statistics variance for various applications such as identifying confidence intervals for $p = P(X < Y)$. The minimized integer (maximized) is reduced to:

$$L(F) = \int (F(x) - kx)^2 dx,$$

with the restriction

$$\int F(x)dx = 1 - p.$$  

Efficiency tests are another application of methods in this category. We consider a random sample of a population that has the continuous $F(x)$ distribution function. Suppose we are interested in testing hypotheses.

$$\begin{align*}
H_0 & : F(x) = G(x) \\
H_1 & : G(x) = F(x - \theta)
\end{align*}$$

with $\theta$ identifier parameter. The relative asymptotic efficiency of the Wilcoxon test against the $t$ test (to be used if $F$ and $G$ are normally distributed) is given by:

$$L(f) = \int f^2(x)dx,$$

where $f(x)$ is the probability density of variable $X$.

An issue of interest in non-parametric inference is finding the limits for $L(f)$ under certain conditions such as:

$$\int f(x)dx = 1,$$

and moreover, that the average of the random variable is zero, i.e.

$$\int xf(x)dx = 0.$$  

A comprehensive approach to this problem has been conducted by Lehmann (1986). Another example is that of regression in the projected experiments. The linear regression model is considered:

$$y = \alpha + \beta x + \epsilon,$$

where $\alpha$ and $\beta$ are unknown parameters, $x$ is independent variable and $\epsilon$ is the error with the mean 0 and the variance $\sigma^2$. In the regression of the projected experiments, the researcher is interested in allocating $n$ observations to $x_1, x_2, \ldots, x_n$ in order to optimize a certain parameter estimation function. A frequently used criterion is maximizing the determinant of the covariance matrix of estimated factors parameters. For example, for a $n$ sample size, the $\left(\hat{\alpha} \quad \hat{\beta}\right)$ covariance is $M\sigma^2$, where:
Assuming that $\sigma^2$ is known, the optimal regression problem is to find $x_1,x_2,...,x_n$ values that maximize the determinant of the matrix $M$. More specifically, it is required to solve the nonlinear optimization problem:

$$\max[M(x_1,...,x_n)] = \left( \frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2} \cdot \frac{1}{\sum (x_i - \bar{x})^2} - \frac{\sum x_i}{n \sum (x_i - \bar{x})^2} \cdot \frac{\sum x_i}{\sum (x_i - \bar{x})^2} \right)$$

(49)

There are other criteria best used to solve this problem, described mainly by Federov (1996).

The robustness of the estimators is another example of using variational analysis methods in statistics. Let be a random variable $X$, the $f(x-\theta)$ probability density function and the $F(x-\theta)$ distribution function. A $T_n$ statistic determined from selection data $X_1,X_2,...,X_n$, starting from $f(x-\theta)$ represents an $M$-estimate for the $\theta$ parameter if it maximizes the expression:

$$\sum_{i=1}^n \rho(x_i - T_n).$$

(50)

The $T_n$ statistic is given by the equation:

$$\sum_{i=1}^n \psi(x_i - T_n) = 0, \text{ cu } \psi = \rho'.$$

(51)

If $\rho(x) = -x^2$ we obtain the least squares estimators and if $\rho(x) = \frac{f'(x)}{f(x)}$ we obtain the maximum veracity estimators. Bickel and Doksum (1977) and Collins and Portnoy (1981) describe several robustness problems with estimators, with variational calculation methods.

**Example of application of linear programming methods**

Using the simplex algorithm, the example presented analyses the application of linear programming methods for the decision to invest in a stock on a stock based on a forecasting model.

A prediction model is a representation of the relationship between a set of independent variables (in this case, continuous variables) and one or more dependent variables known as *response variables*. Therefore, the prediction is a *response estimate* based on a set of values of the independent variables entered into a model.

There are situations when the response variable is known and the purpose is to identify a set of values for the independent variables, for which the predictive model provides the desired response value. This is a reverse technique known as *response optimization*. Situations of this type occur frequently in industrial production, in the case of product development, situations in which the response variable is a product quality variable, and the
measurement variables (the independent variables) are the conditions in which the product develops.

The specialized software package Statistics offers the possibility to solve such a problem through Response Optimization. Thus, using the simplex algorithm, a search is performed in the space of the independent variables. For each set of selected independent values, the prediction model is evaluated and compared to the desired response. This process is repeated until a set of independent values is obtained for which the model generates the desired response value.

The independent variables taken into account in the decision to invest a certain amount on the Bucharest Stock Exchange are the results of the transactions of the issuers of the Basic Category (a sample of 500 issuers). The independent variables used in the forecasting model are:

- **Number of traded shares** (in millions of shares);
- **Price variation** (%);
- **Number of transactions** (total for 52 weeks - thousands of transactions);
- **PBV** (price / book value ratio);
- **ROE** (return on equity);
- **Number of issued shares** (million shares);
- **Debts longer than one year** (million u.m.);
- **PER** (ratio between current market price and net profit per share);
- **Average number of employees** (persons);
- **Net result** (million u.m.);
- **Average price** (u.m.);
- **Condition** (1 = marketable, 0 = suspended) - dummy variable;
- the dependent variable is the **Average value of a transaction** (hundreds of u.m.).

The Response Optimization module will automatically generate variable values for those missing values based on the numeric characteristics of the variables (mean, dispersion, minimum and maximum). It is recommended that the value of the desired response is approximately equal to the average of the dependent variable in the dataset. Descriptive statistics indicators (Table 1) for independent variables are used to establish the initial simplex solution, step size as well as the minimum or maximum values of the simplex algorithm variables (Appendix 1).

The initial settings of the simplex algorithm (maximum number of iterations, algorithm stop criterion, target value of the response variable, values of the initial solution of the simplex algorithm, and the counting step for each variable) are also presented in Appendix 1.

### Table 1. Descriptive statistics indicators

<table>
<thead>
<tr>
<th>Variable names</th>
<th>Variables (BSE investment)</th>
<th>Type</th>
<th>Number of categories</th>
<th>Mean</th>
<th>Variance</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>the average value of a transaction</td>
<td>dependent (continuous)</td>
<td></td>
<td></td>
<td>22.5328</td>
<td>84.59</td>
<td>5.0000</td>
<td>5.0000</td>
</tr>
<tr>
<td>number of traded shares</td>
<td>independent (continuous)</td>
<td></td>
<td></td>
<td>3.6135</td>
<td>73.99</td>
<td>0.0063</td>
<td>88.9762</td>
</tr>
</tbody>
</table>
Variable names | Variables (BSE investment) | Type | Number of categories | Mean | Variance | Min | Max |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>price variation</td>
<td>independent (continuous)</td>
<td>11.3636</td>
<td>534.94</td>
<td>0.0000</td>
<td>100.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>number of transactions</td>
<td>independent (continuous)</td>
<td>11.1368</td>
<td>47.06</td>
<td>0.4600</td>
<td>27.7400</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PBV</td>
<td>independent (continuous)</td>
<td>0.5547</td>
<td>0.01</td>
<td>0.0386</td>
<td>0.8710</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROE</td>
<td>independent (continuous)</td>
<td>6.2846</td>
<td>0.49</td>
<td>3.5610</td>
<td>8.7800</td>
<td></td>
<td></td>
</tr>
<tr>
<td>number of shares issued</td>
<td>independent (continuous)</td>
<td>68.5749</td>
<td>792.36</td>
<td>2.9000</td>
<td>100.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>debts over one year</td>
<td>independent (continuous)</td>
<td>3.7950</td>
<td>4.43</td>
<td>1.1296</td>
<td>12.1265</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PER</td>
<td>independent (continuous)</td>
<td>9.5494</td>
<td>75.82</td>
<td>1.0000</td>
<td>24.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>average number of employees</td>
<td>independent (continuous)</td>
<td>408.2372</td>
<td>28404.76</td>
<td>87.0000</td>
<td>711.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>net result</td>
<td>independent (continuous)</td>
<td>18.4555</td>
<td>4.69</td>
<td>12.6000</td>
<td>22.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>medium price</td>
<td>independent (continuous)</td>
<td>12.6531</td>
<td>50.99</td>
<td>1.7300</td>
<td>37.9700</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>independent (categorical)</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors' own research

We consider three scenarios of analysis, scenarios generated by the investment level namely 15, 20 and 40 (u.m.).

![Figure 1. Iterations of the simplex algorithm for an investment of 15 u.m.](image)

Source: Authors' own research.

The results obtained for the three values of the response variable are shown in Appendix 2 and Figures 1, 2 and 3 respectively graphically represent the iterations of the...
algorithm for the three analysis scenarios. It is noted that if the invested amount is 20 u.m. (value close to the average response variable) the number of iterations is relatively low (around 150) to obtain the optimum solution. If the target value of the response variable deviates significantly, in addition to or minus, from the average of the variable, the number of iterations increases significantly, 350 iterations for an investment of 15 u.m., respectively 1000 iterations for a twice the investment the average of the variable.

**Figure 2. Iterations of the simplex algorithm for an investment of 20 u.m.**
Source: Authors’ own research

**Figure 3. Iterations of the simplex algorithm for an investment of 40 u.m.**
Source: Authors’ own research

**Conclusions**

Linear programming techniques allow finding the optimal solution of a problem, considering into account all the limitations and constraints. In the construction of a linear programming problem are particularly important: defining variables, obtaining statistical data, formulating the objective function and defining restrictions. In this paper a presentation was made of the
most known and used statistical methods in which we find optimization methods and linear programming techniques.

The importance and utility of linear programming techniques are found in the efficiency of any process, the fields of activity in which these optimization techniques are and can be applied being very diverse. Thus, in addition to the classical uses in design and manufacturing, in economics and finance, linear programming is a powerful tool in areas such as:

- **Agriculture**: farmers can maximize their income by identifying, for a certain crop, what is the quantity to be produced under limited resources, how to use it effectively;
- **Nutrition**: mathematical modeling offers the possibility of planning the dietary needs for the purpose of prevention, under the conditions of a healthy diet, with a minimum degree of processing and low costs. The objective function may be to minimize the total cost of the food basket, and the restrictions may be represented by certain dietary guides, recommendations for nutrients, cultural identity or a combination thereof.
- **Transport**: the efficiency of the transport systems is based on the solutions offered by the linear programming methods for cost and time efficiency, on the one hand, respectively the optimization of the profit according to the prices and the demand of the customers.
- **Energy**: linear programming offers the possibility to optimize the design of modern electricity systems that include, besides traditional electricity systems and renewable systems, such as wind or solar photovoltaic energy. Linear programming techniques are an indispensable tool for the energy industry, contributing Nelder, J. A., & Mead, R. (1965). A Simplex Method for Function Minimization. The Computer Journal, 7(4), 308-313. Retrieved from https://doi.org/10.1093/comjnl/7.4.308


ting to the optimization of transmission, storage and distribution costs, to the optimization of the electrical charging requirements.

**References**


### Appendix 1

<table>
<thead>
<tr>
<th>Optimization settings</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined models</td>
<td>no</td>
</tr>
<tr>
<td>Optimization type</td>
<td>user defined</td>
</tr>
<tr>
<td>Target</td>
<td>15.00000000</td>
</tr>
<tr>
<td>Maximum number of iterations</td>
<td>1000</td>
</tr>
<tr>
<td>Stopping criteria</td>
<td>0.000100</td>
</tr>
<tr>
<td>Starting value (number of traded shares)</td>
<td>3.613524</td>
</tr>
<tr>
<td>Step size (number of traded shares)</td>
<td>0.860155</td>
</tr>
<tr>
<td>Minimum (number of traded shares)</td>
<td>0.006320</td>
</tr>
<tr>
<td>Maximum (number of traded shares)</td>
<td>88.976200</td>
</tr>
<tr>
<td>Starting value (price variation)</td>
<td>11.363636</td>
</tr>
<tr>
<td>Step size (price variation)</td>
<td>2.332245</td>
</tr>
<tr>
<td>Minimum (price variation)</td>
<td>0.000000</td>
</tr>
<tr>
<td>Maximum (price variation)</td>
<td>100.000000</td>
</tr>
<tr>
<td>Starting value (number of transactions)</td>
<td>11.136779</td>
</tr>
<tr>
<td>Step size (number of transactions)</td>
<td>0.686035</td>
</tr>
<tr>
<td>Minimum (number of transactions)</td>
<td>0.460000</td>
</tr>
<tr>
<td>Maximum (number of transactions)</td>
<td>27.740000</td>
</tr>
<tr>
<td>Starting value (PVB)</td>
<td>0.554695</td>
</tr>
<tr>
<td>Step size (PVB)</td>
<td>0.011588</td>
</tr>
<tr>
<td>Minimum (PVB)</td>
<td>0.385000</td>
</tr>
<tr>
<td>Maximum (PVB)</td>
<td>0.871000</td>
</tr>
<tr>
<td>Starting value (ROE)</td>
<td>6.284634</td>
</tr>
<tr>
<td>Step size (ROE)</td>
<td>0.070262</td>
</tr>
<tr>
<td>Minimum (ROE)</td>
<td>3.561000</td>
</tr>
<tr>
<td>Maximum (ROE)</td>
<td>8.780000</td>
</tr>
<tr>
<td>Starting value (number of issued shares)</td>
<td>68.574901</td>
</tr>
<tr>
<td>Step size (number of issued shares)</td>
<td>2.814886</td>
</tr>
<tr>
<td>Minimum (number of issued shares)</td>
<td>2.900000</td>
</tr>
<tr>
<td>Maximum (number of issued shares)</td>
<td>100.000000</td>
</tr>
<tr>
<td>Starting value (debts over one year)</td>
<td>3.795043</td>
</tr>
<tr>
<td>Step size (debts over one year)</td>
<td>0.210571</td>
</tr>
<tr>
<td>Minimum (debts over one year)</td>
<td>1.129600</td>
</tr>
<tr>
<td>Maximum (debts over one year)</td>
<td>12.126500</td>
</tr>
<tr>
<td>Starting value (PER)</td>
<td>9.549407</td>
</tr>
<tr>
<td>Step size (PER)</td>
<td>0.870726</td>
</tr>
<tr>
<td>Minimum (PER)</td>
<td>1.000000</td>
</tr>
<tr>
<td>Maximum (PER)</td>
<td>24.000000</td>
</tr>
<tr>
<td>Starting value (average number of employees)</td>
<td>408.237154</td>
</tr>
<tr>
<td>Step size (average number of employees)</td>
<td>16.853712</td>
</tr>
<tr>
<td>Minimum (average number of employees)</td>
<td>187.000000</td>
</tr>
<tr>
<td>Maximum (average number of employees)</td>
<td>711.000000</td>
</tr>
<tr>
<td>Starting value (net result)</td>
<td>18.455534</td>
</tr>
<tr>
<td>Step size (net result)</td>
<td>0.216495</td>
</tr>
<tr>
<td>Minimum (net result)</td>
<td>12.600000</td>
</tr>
<tr>
<td>Maximum (net result)</td>
<td>22.000000</td>
</tr>
</tbody>
</table>
### Appendix 2

<table>
<thead>
<tr>
<th>Model name</th>
<th>number of traded shares</th>
<th>price variation</th>
<th>number of transactions</th>
<th>PBV</th>
<th>ROE</th>
<th>number of shares issued</th>
<th>debts over one year</th>
<th>PER</th>
<th>average number of employees</th>
<th>net result</th>
<th>medium price</th>
<th>condition</th>
<th>the average value of a transaction</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Model name</th>
<th>number of traded shares</th>
<th>price variation</th>
<th>number of transactions</th>
<th>PBV</th>
<th>ROE</th>
<th>number of shares issued</th>
<th>debts over one year</th>
<th>PER</th>
<th>average number of employees</th>
<th>net result</th>
<th>medium price</th>
<th>condition</th>
<th>the average value of a transaction</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Model name</th>
<th>number of traded shares</th>
<th>price variation</th>
<th>number of transactions</th>
<th>PBV</th>
<th>ROE</th>
<th>number of shares issued</th>
<th>debts over one year</th>
<th>PER</th>
<th>average number of employees</th>
<th>net result</th>
<th>medium price</th>
<th>condition</th>
<th>the average value of a transaction</th>
</tr>
</thead>
</table>