GEOPHYSICAL TRANSACTIONS 1992
Vol. 37. No. 2-3. pp. 111-157

GENERALIZATION AND ROBUSTIFICATION OF THE COVARIANCE MATRIX

Béla HAJAGOS* and Ferenc STEINER*

After some unavoidable simplifications in Section 1 typical instances as well as problematic examples are given for the use of the inverse covariance matrix of the classical statistics for weighting. Because of the limited applicability of this matrix, in Section 2 a generalized version of the covariance matrix is introduced; its applicability is mostly shown using *P*-algorithms, i.e., weighted most freqent value calculations. The generalized inverse covariance matrix turned out to be valid also in the case of small samples if we use *P*-algorithms.

In Section 3 the method for determining this generalized covariance matrix is given as the robustification of the classical one; this determination is (from the viewpoint of the computation technique) consistent with the basic algorithms of the most frequent value procedures.

Keywords: correlation coefficient, asymptotic variance, general covariance matrix, robustification, most frequent value, dihesion

1. Introduction

1.1 Some definitions of classical statistics

The probability theory defines the covariance σ_{xy} of the random variables ξ and η well known as

$$\sigma_{xy} = E\left\{ [\xi - E(\xi)] \cdot [\eta - E(\eta)] \right\} \tag{1}$$

(E denotes the expected value). If the common density function is denoted by $f(x,y)$, the covariance is to be calculated as the following integral:

* University of Miskolc, H-3515 Miskoc-Egyetemváros
Manuscript received: 10 January, 1991

$$\sigma_{xy} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-E_x)\,(y-E_y)f(x,y)\;\mathrm{d}x\,\mathrm{d}y, \tag{2}$$

and if n pairs of data (x_i,y_i) are given, a very simple expression corresponds to Eq. (2):

$$\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})\,(y_i-\bar{y})\,; \tag{3}$$

the arithmetic means $\bar{x}$ and $\bar{y}$ are the estimates of the expected values E_x and E_y, respectively.

The degree of stochastic connection between ξ and η is measured by the correlation coefficient ρ_{xy}:

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}\;; \tag{4}$$

σ_x and σ_y denote the corresponding scatters, i.e., the square roots of the variances, consequently the integral formula for ρ_{xy} is

$$\rho_{xy} = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-E_x)\,(y-E_y,f(x,y)\;\mathrm{d}x\mathrm{d}y}{\sqrt{\int_{-\infty}^{\infty}(x-E_x)^2 f(x)\,\mathrm{d}x}\cdot\sqrt{\int_{-\infty}^{\infty}(y-E_y)^2 f(y)\,\mathrm{d}y}} \tag{5}$$

and the estimate of this value is calculated clearly as

$$\rho_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}\,. \tag{6}$$

In the case of J random variables $\xi_1, \xi_2,\ldots,\xi_J$ the notations ρ_{jk} and σ_{jk} represent the correlation coefficient and the covariance of the variables ξ_j and ξ_k. If, similarly, σ_j and σ_k represent the corresponding scatters, the matrix of all covariances, i.e., the so-called covariance matrix $\mathbf{s}$, can be written as

$$\mathbf{S} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\,\sigma_1\,\sigma_2 & \rho_{13}\,\sigma_1\,\sigma_3 & \cdots & \rho_{1J}\,\sigma_1\,\sigma_J \\ \rho_{12}\,\sigma_1\,\sigma_2 & \sigma_2^2 & \rho_{23}\,\sigma_2\,\sigma_3 & \cdots & \rho_{2J}\,\sigma_2\,\sigma_J \\ \cdot & & & & \\ \cdot & & & & \\ \rho_{J1}\,\sigma_1\,\sigma_J & \rho_{J2}\,\sigma_2\,\sigma_J & \rho_{J3}\,\sigma_3\,\sigma_J & \cdots & \sigma_J^2 \end{pmatrix}. \tag{7}$$

This way of presentation of the covariances— used also in the famous article of INMAN [1975]— clearly shows their meaning and suggests that the interpretation of the correlation coefficients, presented in the correlation matrix

$$\rho = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1J} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2J} \\ \cdot & & & & \\ \cdot & & & & \\ \rho_{J1} & \rho_{J2} & \rho_{J3} & \cdots & 1 \end{pmatrix}, \tag{8}$$

is much more a primary quantity than that of the covariances, although the latter can directly be calculated according to Eqs. (2) or (3) and seemingly (see Eq. 4) the correlation coefficient is the secondary (inasmuch as it is derived) quantity.

1. 2 Simplifications

The significance of the correlation matrix of the errors in respect of inversion algorithms lies in the fact that its inverse can be appropriately used for weighting, see e.g. the already cited paper of INMAN, where a well defined geophysical inversion task is thoroughly treated (including the way of linearization, ridge regression, etc.). In contrast to this way of treatment, in the present paper from the point of view of the possible geophysical tasks only the simplest situation is supposed: direct measurements are made for the same quantity, say, for the rock porosity n, but not all data are characterized by the same probable error and/or by the same type of the error distribution. After this simplificaton we can pay full attention to problems of generalization and robustification.

For the clearest presentation of our trains of thought leading to the necessary generalization and robustification of the covariance matrix, however, some further simplifications are unavoidable. In all cases discussed in the present paper the inverse of the covariance matrix shall give

the same weight, for n_a data another weight, for the remaining n_b data ($n_b = n - n_a$)i.e., n_a data will have equal weights and this is true for the remaining ones, too.

Consequently, if we give the weight w ($0 \leq w \leq 1$) for the n_a statistically equally behaving data and the weight $(1-w)$ for the n_b (from the statistical aspect similarly equal) data, the error-curve of the results obtained by statistical algorithm as a function of w will have its minimum place w_{opt} determined by the inverse of the covariance matrix.

In this Introduction the 'statistical algorithm' mentioned above refers to the classical one, i.e., the minimization of the L_2-norm, — which is here simply the calculation of weighted means as according to our supposition direct measurements are made for the unknown geophysical quantity (say for the porosity of the rock at a given depth level).

Our simplifications enable us to write the variance σ_0^2 of the result x_0 in a simple analytical form, on the basis of the well known general expression

$$\sigma_0^2 = \sum_{i=1}^{n} c_i^2 \, \sigma_i^2 + \sum_{\substack{j,k=1 \\ j \neq k}}^{n} c_j \, c_k \, \rho_{jk} \, \sigma_j \, \sigma_k \tag{9}$$

which belongs to the random variable

$$\xi_0 = \sum_{i=1}^{n} c_i \, \xi_i \ . \tag{10}$$

(In Eq. (9), instead of the often used notations VAR (ξ_i) and COV(ξ_j,ξ_k) σ_i^2 and $\rho_{jk} \, \sigma_j \, \sigma_k$, respectively, are written in accordance with the notations already used in Eq. (7).)

If the random variables ξ_i are pair-wise independent, Eq. (9) reduces to

$$\sigma_0^2 = \sum_{i=1}^{n} c_i^2 \, \sigma_i^2 \ , \tag{11}$$

and this is to be written as

$$\sigma_0^2 (w) = \frac{1}{[n_a \, w + n_b \, (1-w)]^2} \left\{ n_a \, w^2 \, \sigma_a^2 + n_b \, (1-w)^2 \, \sigma_b^2 \right\} \tag{12}$$

if (as was agreed earlier),

$$x_0 = \frac{1}{n_a w + n_b (1-w)} \left\{ w \cdot \sum_{i=1}^{n_a} x_i + (1-w) \cdot \sum_{i=n_a+1}^{n} x_i \right\} . \tag{13}$$

In Eq. (12) σ_a and σ_b represent the scatters (i.e., the square roots of the variances) of the n_a and n_b data.

The minimum place $\sigma_0^2(w)$ (see Eq. 12) is

$$w_{opt} = \frac{1/\sigma_a^2}{1/\sigma_a^2 + 1/\sigma_b^2} \tag{14}$$

(which is simply to verify by differentiation). This is in accordance with the following general form of the inverse of the covariance matrix valid for pair-wise independence of the random variables:

$$\mathbf{S}^{-1} = \begin{pmatrix} 1/\sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & 1/\sigma_2^2 & 0 & \dots & 0 \\ 0 & 0 & 1/\sigma_3^2 & \dots & 0 \\ . & . & . & & . \\ . & . & . & & . \\ 0 & 0 & 0 & \dots & 1/\sigma_n^2 \end{pmatrix} \tag{15}$$

The general form of matrix $\mathbf{S}^{-1}$ in the case of different mutual dependences of the random variables cannot be given in a simple way.

We shall discuss, however, only such cases when only n_b random variables are pair-wise dependent for the neighbouring indices characterized by the same correlation coefficient ρ. (Naturally n_b must now be an even number). Otherwise the simplification made in the independent case remains, i.e., there exits only two different variances σ_a^2 and σ_b^2, consequently the covariance matrix will have the following form:

$$\mathbf{S} = \begin{pmatrix} \sigma_a^2 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & & & & & & & & & \cdot \\ 0 & \cdot & 0 & \sigma_a^2 & & & & & & 0 \\ 0 & \cdot & 0 & 0 & \sigma_b^2 & \rho\sigma_b^2 & 0 & \ldots & & 0 \\ 0 & \cdot & 0 & 0 & \rho\sigma_b^2 & \sigma_b^2 & 0 & \ldots & & 0 \\ \cdot & & & & & & & & & \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \sigma_b^2 & \rho\sigma_b^2 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \rho\sigma_b^2 & \sigma_b^2 \end{pmatrix} \begin{matrix} \\ \\ n_a \text{ rows} \\ \\ \\ \\ n_b \text{ rows} \\ \\ \end{matrix} \tag{16}$$

n_a columns $\qquad$ n_b columns

In this special case $\mathbf{S}^{-1}$ also has a simple form: (17)

$$\mathbf{S}^{-1} = \begin{pmatrix} 1/\sigma_a^2 & 0 & 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\ 0 & \ldots & 0 & 1/\sigma_a^2 & & & & & & \cdot \\ 0 & \ldots & 0 & 0 & \dfrac{1}{\sigma_b^2(1-\rho^2)} & -\dfrac{\rho}{\sigma_b^2(1-\rho^2)} & 0 & \ldots & & 0 \\ 0 & \ldots & 0 & 0 & -\dfrac{\rho}{\sigma_b^2(1-\rho^2)} & \dfrac{1}{\sigma_b^2(1-\rho^2)} & 0 & \ldots & & 0 \\ \cdot & & & & & & & & & \cdot \\ 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 & \dfrac{1}{\sigma_b^2(1-\rho^2)} & -\dfrac{\rho}{\sigma_b^2(1-\rho^2)} \\ 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 & -\dfrac{\rho}{\sigma_b^2(1-\rho^2)} & \dfrac{1}{\sigma_b^2(1-\rho^2)} \end{pmatrix}$$

(the product of both matrices in Eqs. (16) and (17) should result in the unity matrix).

Now, the 'proposal' of the matrix $\mathbf{S}^{-1}$ for the weights to be applied is:

$$\begin{array}{ll} \dfrac{1}{\sigma_a^2} & \text{for } n_a \text{ data} \\ \dfrac{1}{\sigma_b^2(1+\rho)} & \text{for } n_b \text{ data} \end{array} \qquad (18)$$

In our simplified case the notations of INMAN [1975] correspond to the following: $M=1$ and $P^0=0$ therefore $\Delta P=x_0$; $A=I$ (the unity matrix); $[\Delta G]_i = x_i$; $N^{-1} = \mathbf{s}^{-1}$. Consequently, the expression to be minimized (see Eq. (8) in the just cited article) simplifies to

$$\Phi = \sum_{i,k=1}^{n} \sigma_{ik}^{-1} (x_i - x_0)(x_k - x_0) \qquad (19)$$

(in our notations) having clearly the minimum place defined by

$$c_i = \frac{\sum_{k=1}^{n} \sigma_{ik}^{-1}}{\sum_{i=1}^{n}\sum_{k=1}^{n} \sigma_{ik}^{-1}} \qquad (20)$$

(compare our Eq. (10); σ_{ik}^{-1} represent an element of the inverse covariance matrix $\mathbf{s}^{-1}$. The same c_i-s are the results of course if we minimize the expression in our Eq. (9).) The weights in Eq. (18) clearly correspond to the general expression in Eq. (20), taking the actual $\mathbf{s}^{-1}$ (see Eq. 17) into consideration.

Expressing the results in Eq. (18) by the earlier introduced variable weight w,

$$w_{opt} = \frac{\dfrac{1}{\sigma_a^2}}{\dfrac{1}{\sigma_a^2} + \dfrac{1}{\sigma_b^2(1+\rho)}}, \qquad (21)$$

(and this clearly reduces to the form given in Eq. (14) if $\rho = 0$). The same result is reached if the general expression in Eq. (9) is concretized to our actual case to get the analogue of Eq. (12):

$$\sigma_0^2(w) = \frac{1}{[n_a w + n_b(1-w)]^2}\left\{n_a w^2 \sigma_a^2 + n_b (1-w)^2 (1+\rho)\sigma_b^2\right\} \qquad (22)$$

and we look for the minimum place of this function by differentiation.

It is instructive to show in figures the variation of the error committed by the statistical algorithm used for the whole range of w ($0 \leq w \leq 1$). The 'probable error' (defined by Bessel) and denoted by q is frequently used by engineers: if F is the probability distribution function and F^{-1} its inverse,

$$q = \frac{1}{2}\left[F^{-1}(3/4) - F^{-1}(1/4)\right] \tag{23}$$

and therefore it is also called 'semi-interquartile range'. The 'semi-intersextile range' Q is similarly defined:

$$Q = \frac{1}{2}\left[F^{-1}(5/6) - F^{-1}(1/6)\right] . \tag{24}$$

It is obvious that it is twice as probable that absolute errors are *less* than Q, than *greater* than Q whilst these probabilities are equal using q as the error characteristic: one half of the errors are expected inside the interquartile interval, the other half outside of it. We shall use throughout the probable error q, which is to be calculated for a Gaussian error distribution well known as

$$q = 0.6745 \cdot \sigma \quad . \tag{25}$$

Sometimes the curve of the semi-intersextile range Q in function of w will be also shown; for the Gaussian case

$$Q = 0.9674 \cdot \sigma \tag{26}$$

holds.

Our demands for accuracy are less rigorous in the case of the errors than in the case of the geophysically important quantities (depths, porosities, etc.) to be determined. If, say, by a geophysical depth-determination the relative error is 4%, this is meant as an error-range between 3.5% and 4.5 %, consequently the relative error of the error is here accepted to be 25 %. Indeed, the relative error of a simple determination of the scatter σ is $100/\sqrt{2n}$ % even in the least problematic case: when the data have Gaussian distribution. For example, if n=8, this error is obviously 25 %. (The just cited formula is a special variant of the expression in Eq. (55), namely for $a \to \infty$, *viz*, this limiting case corresponds to the Gaussian distribution of the errors.) Consequently, the difference between Q and σ of only some per cent for the gaussian distribution (see Eq. 26) can be neglected in nearly all practical cases. By theoretical investigations, however, naturally the exact connection given by Eq. (26) is to be taken into consideration.

1.3 Examples, problems, meditations

It is time to show some examples of classical fashion: firstly for Gauss-distributed data. Although on the grounds of Eqs. (22), (25) and (26) both theoretical curves $q(w)$ and $Q(w)$, respectively, are easy to construct for various situations, we also show Monte Carlo results for w = 0; 0.1; 0.2; ...; 0.9; 1.0. More exactly, for all situations random numbers in question N=1000 times were generated and the just investigated statistical algorithm applied in all investigated cases of this paper (this algorithm is, in the first examples, simply the calculation of the weighted average according to Eq. (13), i.e., an L_2-norm algorithm,) and the empirical q-value (sometimes also the Q-value) of these 1000 results were determined. For information about the statistical fluctuation of these probable errors, the whole procedure was repeated three times, and the values obtained were, on the one hand, separately demonstrated in all Figures, and on the other hand, also the interval of the actual fluctuation was indicated as short perpendicular straight line for all w-values. The self-consistence belonging to this demonstration of our Monte Carlo results throughout the paper proved to be fruitful.

In the first example are only (pair-wise) independent errors, all have Gaussian distribution but n_b=4 of them are characterized by $\sigma_b = \sqrt{2}$ and n_a=4 by σ_a = 1. (As random variables here and later have the meaning of error, the parameter of location will always be equal to zero.)

Fig. 1 shows that Monte Carlo results are in full agreement with the theoretical curve based on Eq. (12), including naturally the fact that w_{opt} is the same both for theoretical and Monte Carlo results. (The theoretical value of the w_{opt} in question will always be indicated on such Figures, sometimes the theoretical q-values for w=0 and 1, will be indicated too.)

In our second example all n=9 data have standard Gaussian distribution but only n_a=3 are independent, ξ_4 and ξ_5, ξ_6 and ξ_7, ξ_8 and ξ_9 are correlated with a correlation coefficient of ρ = 0.6. (This is in full agreement with our simplification in Section 1.2 and therefore it would have been enough to write n_b=6 and ρ = 0.6.) The Monte Carlo results and the theoretical curve $q(w)$ based on Eqs. (22) and (25) are shown in *Fig. 2*; the conclusions are the same as for Fig. 1. (The ρ-scale also shown in Fig. 2 indicates the w_{opt}-values for the corresponding ρ-s on the ground of Eq. (21) which simplifies in this concrete case to $w_{opt} = (1+\rho)/(2+\rho)$.)

It is illusory, however, always to expect such excellent agreement between theoretical and Monte Carlo results, e.g., if n_q data are Gauss-distributed characterized by σ_G = 3 and n_b have the density distribution

$$f_T(x) = (1-p) \cdot f_G(1; x) + p \cdot f_G(\sigma_c; x), \quad (p < 0.5), \tag{27a}$$

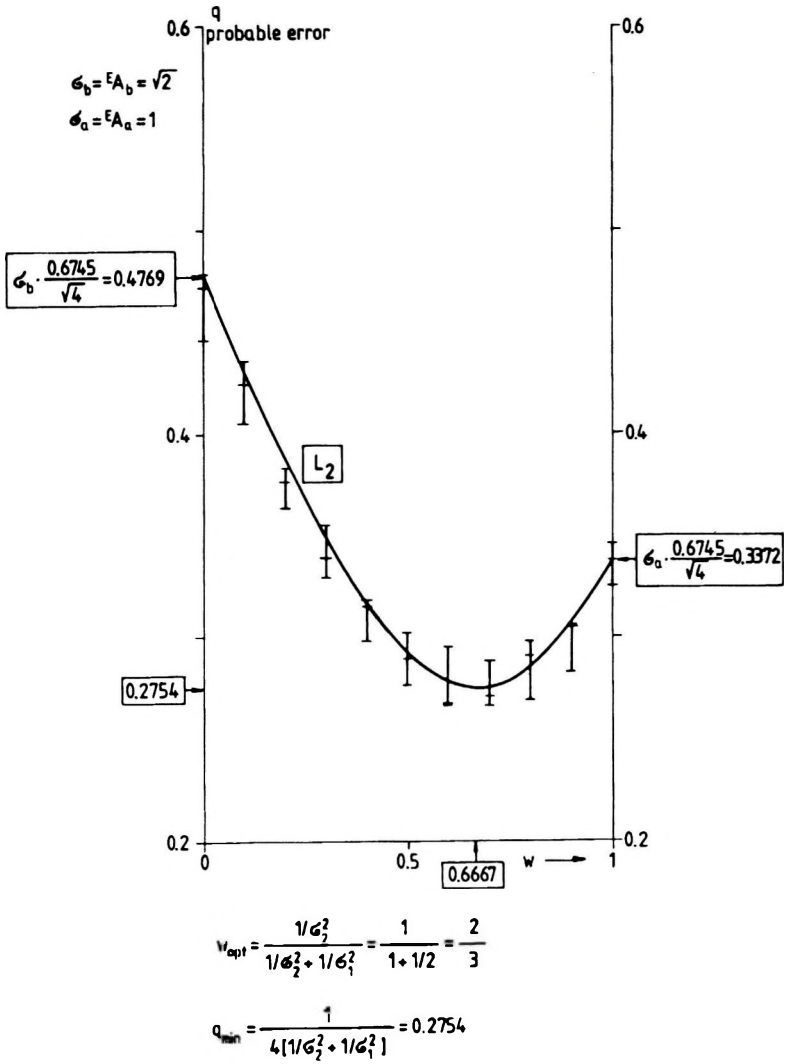

Fig. 1. Error-curve of the results from the conventional L_2-algorithm, using different weights. The data are independent and Gauss-distributed; $n_a=n_b=4$. The best choice corresponds to the covariance matrix of classical statistics

1. ábra. A hagyományos L_2-algoritmussal nyert eredmények hibagörbéje különböző súlyok alkalmazásakor. A primer adatok függetlenek és Gauss-eloszlásúak ; $n_a=n_b=4$. Az a legjobb súlyválasztás, amely a klasszikus statisztika szerinti kovarianciamátrixnak felel meg

Рис. 1. Кривая погрешности результатов, полученных при традициональном альгорифме L_2 при применении разных весов. Первычные данные независымые и имеют Гауссовское распределение, $n_a=n_b=4$. Лучшим выбором веса является тот, который отвечает ковариационной матрице классической статистики

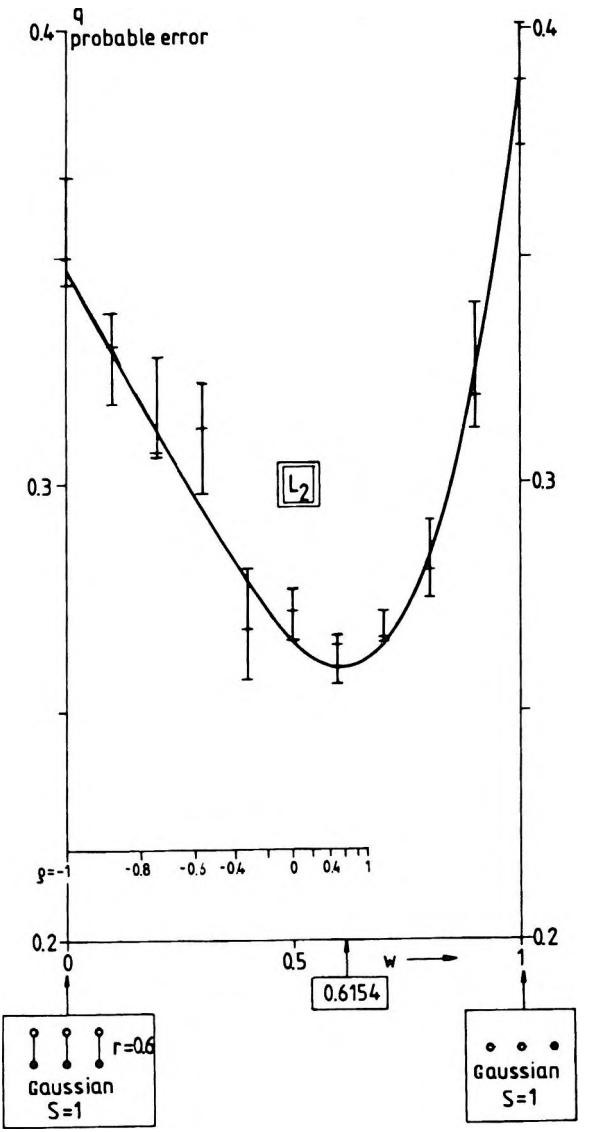

Fig. 2. Error-curve of the results from the conventional L_2-algorithm, using different weights (n_a= 3, n_b= 6; r_{true}=ρ= 0.6) . The best choice corresponds to the covariance matrix of classical statistics

2. ábra. A hagyományos L_2-algoritmussal nyert eredmények hibagörbéje különböző súlyok alkalmazásakor (n_a= 3, n_b= 6; r_{true}=ρ= 0,6). Az a legjobb súlyválasztás, amely a klasszikus statisztika szerinti kovarianciamátrixnak felel meg

Рис. 2. Кривая погрешности результатов, полученных при традициональном алгорифме L_2 при применении разных весов (n_a= 3, n_b= 6; r_{true}=ρ= 0,6). Лучшим выбором веса является тот, который отвечает ковариационной матрице классической статистики

i.e., n_b data are Tukey-distributed. (See Eq. (83) in STEINER [1988]; in Eq. (27a) f_G represents the Gaussian density function

$$f_G(\sigma;x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-x^2/2\sigma^2} \tag{27b}$$

In the monograph just cited the density function of arithmetic means for Tukey-distributed data is also given analytically, see Eq. (85).)

The scatter of the Tukey distribution is obviously

$$\sigma_T = \sqrt{(1-p) + p \cdot \sigma_c^2} \tag{27c}$$

and this means a value of $\sigma_T = 67.09$ if we take p=0.2 and σ_c=150 (this case is separately discussed in STEINER [1988]). As independence is also supposed, Eq. (14) would yield the appropriate w_{opt}-value but we have in this way $w_{opt} \approx 0$ in contradiction to the Monte Carlo results for n_a=n_b=1 shown in *Fig. 3a* where the choice w=0.9 seems the best (both $q(w)$ and $Q(w)$ curves are presented in the figure). Seemingly — at least belonging the minimum place — the Monte Carlo results correspont to the 'theoretical' expectations if n_a=n_b=4 (see also Fig. 3a) but this is not true for the whole range of w: we see in *Fig. 3b* (where other ordinate scaling is used) that the q-value from Monte Carlo calculations for w=1 is only a small fraction of the 'theoretical value' according to Eq. (12) which is to be calculated in this case simply as $0.6745 \cdot \sigma_T / \sqrt{4}$.

The foregoing calls attention to the fact the scatters in the covariance matrix would be misinterpreted as the minimum of the L_2 norm of the differences $(x-x_0)$ for the other distribution: they are to be understood, on the contrary, as asymptotic scatters of the arithmetic means.

The asymptotic scatter is defined for estimates as

$$A = \lim_{n \to \infty} \sqrt{n} \cdot \sigma_{est} \tag{28}$$

where σ_{est} is the empirical scatter of the estimates. As the estimates frequently approximate Gaussian distribution [see e.g. HUBER 1981] if $n \to \infty$ the approximation $q = 0.6745 \cdot A/\sqrt{n}$ or $Q = 0.9674 \cdot A/\sqrt{n}$ can be adequately used (see also Eq. (25) and (26)). If we say only 'variance', it means the asymptotic variance of the arithmetic averages.

The foregoing means that the weigthing given by the inverse covariance matrix is to be understood asymptotically, — but what about small samples? In the extreme case n_a=1 shown in Fig. 3a for the weight w=1 is clearly the semi-interqartile range of the mother distribution and it is authentic and not the value $0.6745 \cdot \sigma_T$; the former is about an order less in the cited case than in the latter. This means that acceptable results by using the inverse covariance matrix for weighting are to be expected for

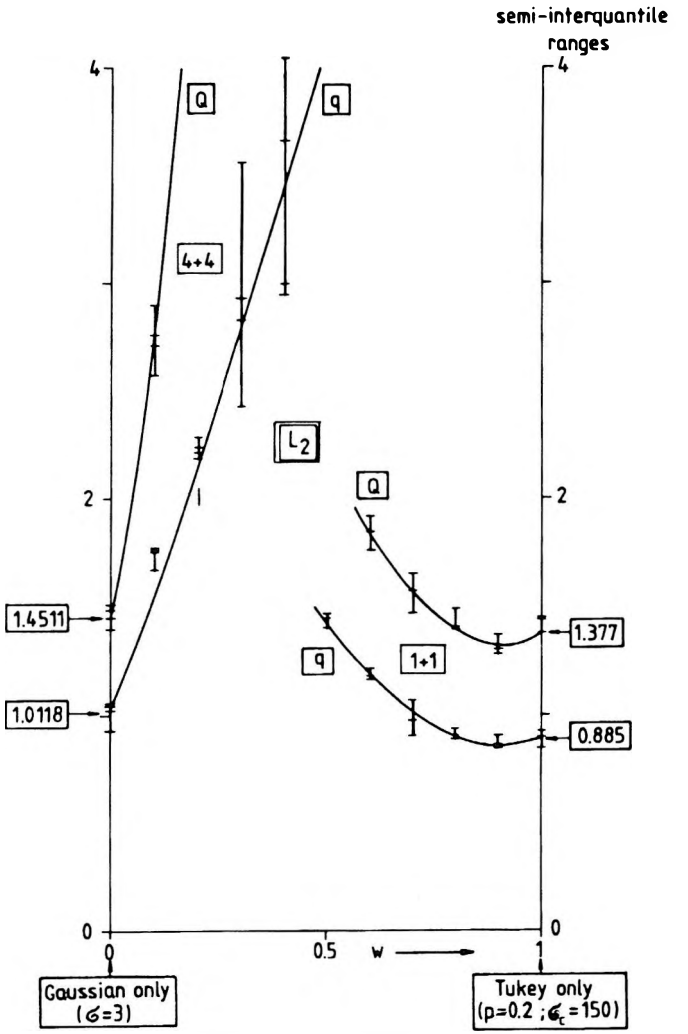

Fig. 3a. Error-curve of the results from the conventional L_2-algorithm, using different weights. The best choice for $n_a=n_b=1$ does not correspond to the classical covariance matrix: w_{opt} can be calculated from the q- and Q-values, respectively, of the mother distributions

3a. ábra. A hagyományos L_2-algoritmussal nyert eredmények hibagörbéje különböző súlyok alkalmazásakor. Az $n_a=n_b=1$ esetben a legjobb súlyválasztás nem felel meg a klasszikus kovarianciamátrixból következőknek: w_{opt} az anyaeloszlások q- ill. Q-értékeiből számítható

Рис. 3а. Кривая погрешности результатов, полученных при традиционный алгорифме L_2 при применении разных весов. В случае $n_a=n_b=1$ лучший выбор веса не отвечает следтствиям классической ковариационной матрици : w_{opt} вычисляется по значениям q и Q матерных распределений

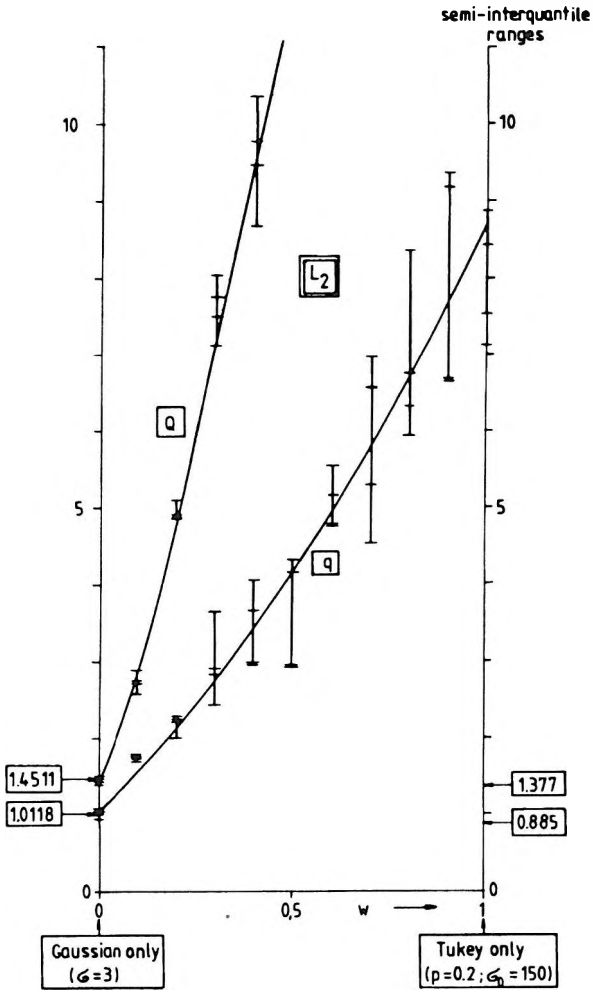

Fig. 3b. Error-curve of the results from the conventional L_2-algorithm, using different weights (n_a=n_b=4). The error-value at w=1 is only a small fraction of the value calculated on the basis of the variance

3b. ábra. A hagyományos L_2-algoritmussal nyert eredmények hibagörbéje különböző súlyok alkalmazásakor (n_a=n_b=4). A w=1-nél csak egy törtrésze az eredmény hibája a szórásból számíthatónak

Рис. 3b. Кривая погрешности результатов, полученных при традиционный алгорифме L_2 при применении разных весов (n_a=n_b=4). При w =1 погрешность результата является лишь дробной частью величины

small samples only if the q- and/or the Q-values of the mother distribution do not differ significantly from $0.6745 \cdot \sigma$ and $0.9674 \cdot \sigma$, respectively. If this is not the case, n must be large for the adequate use of $\mathbf{s}^{-1}$ for weighting.

In the next example the samples are not yet small: $n_a = n_b = 100$ holds. Two types of the $f_m(x)$-supermodel were chosen (see Eq. (143) in STEINER [1988]), namely, for the type parameter values $m=1$ and $m=0.5$, leading on the one hand to the well known Laplace distribution characterized by the density function

$$f_L(x) = \frac{1}{2} e^{-|x|} \tag{29}$$

and, on the other hand, to a very peaky density function

$$f_{ne}(x) = e^{-2\sqrt{|x|}} \tag{30}$$

and therefore it seems to be appropriate to call this latter 'needle distribution'.

Both density functions are shown in *Fig. 4*. The flanks of the needle distribution are heavier than those of the Laplace distribution, therefore also the scatter ($\sigma_{ne} = 2.7386$) is significantly greater than that of the Laplace distribution ($\sigma_L = \sqrt{2}$) yielding, according to Eq. (14), a w_{opt}-value of 0.2105.

The Monte-Carlo results for the weighted means are near to the theoretical ones (see Fig. 5); the just given value of w_{opt} is fully satisfactory but some differences still exist in the neighbourhood of $w=1$ (i.e., $n_a=100$ is not yet enough in this case to speak justifiably about the practically 'total fulfilment' of the asymptotic rule, see the L_2-curve in *Fig. 5*, because of the long tails of the f_{ne} distribution).

In Fig. 5 Monte-Carlo results for weighted medians are also demonstrated (see the L_1-curve); the optimum value of w for this case (=0.8) has really nothing to do with the yet known value $w_{opt}=0.2105$ which applies to the conventional statistical algorithm (i.e., with the minimization of the L_2-norm, which is in our simplified situations nothing more than weighted mean-calculations). — This L_1 curve will be discussed later more thoroughly; we indicate here only the important fact that the *optimum values of the weights strongly depend upon the statistical algorithm used* and therefore appropriate generalization of the covariance matrix is unavoidable. This generalization must also be able to solve the problem that scatters (figuring in the covariance matrix of Eq. (7)) are often, i.e., for very many type-models of probability distributions, infinite, and also the primarily given definition of σ_{xy} (in Eq. (2)) can lose completely its meaning, e.g., if $f(x,y)$ is the density function of a two-varibale Cauchy distribution. As also

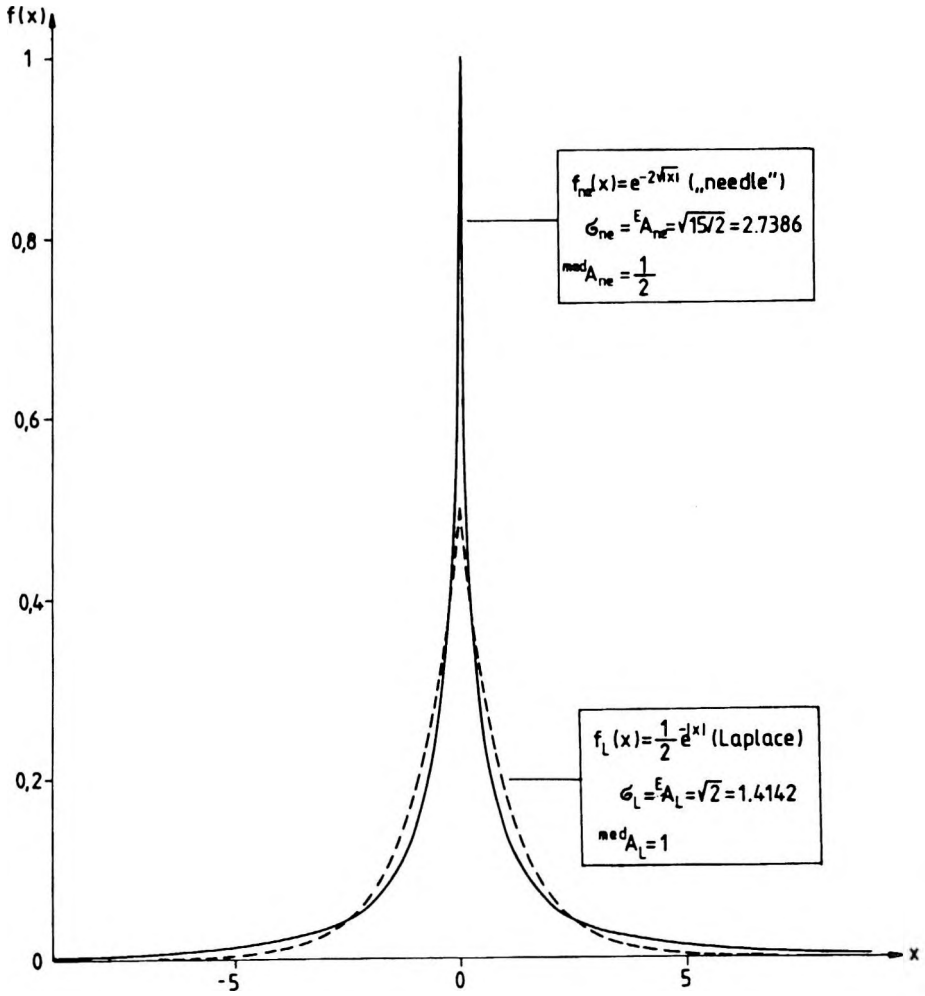

Fig. 4. Probability density function of the Laplace and the so-called 'needle' distribution; the latter is very peaky as well as having much heavier flanks than the Laplace distribution

4. ábra. A Laplace- és az ún. "tű"-eloszlás sűrűségfüggvénye. Az utóbbi egyrészt nagyon hegyes, másrészt viszont jelentősen súlyosabbak a szárnyai, mint a Laplace-eloszlásnak

Рис. 4. Функция плотности распределения Лапласа и иглового распределения. Последное с одной стороны слишком острое, однако его криля имеют больше весов, чем распределение Лапласа

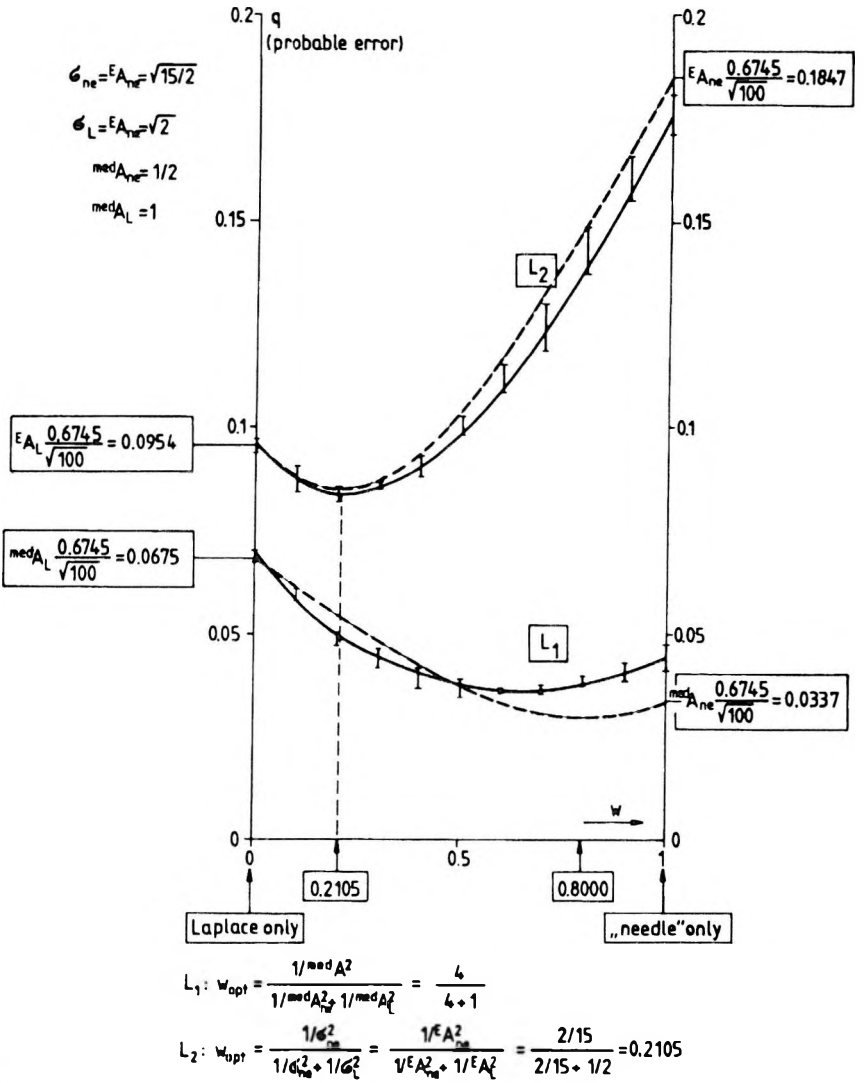

Fig. 5. Error-curves of the results from the L_2-algorithm and the L_1-algorithm, respectively (n_a=n_b=100). The minimum place of the L_1-curve has nothing to do with the conventional covariance matrix but it is easy to interpret on the basis of the generalized one

5. ábra. Az L_2- ill. L_1-algoritmusokkal nyert eredmények hibagörbéi (n_a=n_b=100). Az L_1-görbe minimumhelyének semmi köze sincs a hagyományos kovarianciamátrixhoz, de könnyen magyarázható az általánosított kovarianciamátrix alapján

Рис.5. Кривые погрешности результатов, полученных по алгорифмам L_2 и L_1 (n_a=n_b=100). Минимум кривой L_1 не имеет ничего общего с традиционной ковариационной матрицей, но легко объясняется по обобщеной ковариационной матрице

ρ_{xy} in Eq. (4) has no meaning in these cases, obviously the correlation matrix must also be generalized.

2. Generalizations

2.1 Generalization of the covariance matrix if the errors are independent

Let $f(x)$ be the density function of the actual but unknown probability distribution. If we choose instead of $f(x)$ a well defined density function $g(x)$ as 'substituting distribution', the so-called I-divergence $I_g(f)$ is defined as

$$I_g(f) = \int_{-\infty}^{\infty} \ln \frac{f(x)}{g(x)} f(x)\, dx \tag{31}$$

(see HAJAGOS [1982] or STEINER [1988]). As the I-divergence can be interpreted as a measure of the information loss, that T-value of the location parameter figuring in g can be with reason accepted as the most characteristic value for f, which minimizes the information loss, i.e., for them the relations

$$\frac{\partial I_g(f)}{\partial T} = 0 \tag{32}$$

and

$$\frac{\partial^2 I_g(f)}{\partial T^2} > 0 \tag{33}$$

are simultaneously fulfilled. It is easy to verify (see either of the just cited papers,) that both demands are fulfilled if the following integrals result in zero values:

$$\int_{-\infty}^{\infty} \frac{\frac{\partial g}{\partial T}}{g} f(x)\, dx = 0 \tag{34}$$

and

$$\int_{-\infty}^{\infty} \frac{\frac{\partial^2 g}{\partial T^2}}{g} f(x)\,dx = 0 \quad . \tag{35}$$

These equations define the statistical algorithm: Eq. (34) gives the algorithm for the determination of the location parameter T, Eq. (35) defines that of the parameter of scale (S). For example, if g is the Gaussian density function, Eq. (34) defines the formula for the expectant value E as the location parameter, Eq. (35) defines the formula for the variance σ^2; if g is the Cauchy density function, the Eqs. (34) and (35) define the twofold iteration for determining the most frequent value (as T) and the dihesion ε (as S); etc.

If the following integral denoted by ${}^gI$ is calculated such T and S simultaneously satisfy Eqs. (34) and (35), we get the 'developed information' using the statistical algorithm defined originally by the substituting distribution g:

$${}^gI = \int_{-\infty}^{\infty} \left(\frac{\frac{\partial g}{\partial T}}{g} \right)^2 \cdot f(x)\,dx \quad . \tag{36}$$

(Another name for ${}^gI$ is 'relative information', see HAJAGOS [1982] where the definition of ${}^gI$ is first given.) The maximum value I of ${}^gI$ presents itself in the case of $g=f$; I is the well known Fisher-information. In this case the whole information is exhausted and therefore the straightforward definition of the efficiency e of the statistical algorithm defined by g is

$$e = {}^gI/I \quad ; \tag{37}$$

and it also seems appropriate to use ${}^gI$ as weights. HAJAGOS [1985] has shown that

$${}^gI = \frac{1}{{}^gA^2} \quad , \tag{38}$$

where ${}^gA^2$ is the asymptotic variance of the estimates if g defines (by means of Eqs. (34) and (35)) the statistical algorithm. Substituting the expression of ${}^gI$ given by Eq. (38) in Eq. (37), and taking also into considetarion that $I = 1/A^2_{\min}$, we in fact get the well known formula fot the efficiency

$$e = A^2_{\min}/{}^gA^2 \quad .$$

If Eq. (38) defines the weights for data characterized by different error distributions, the diagonal matrix of these weights can be considered as the inverse ($^{g}A^{-1}$) of the following general covariance matrix:

$$ {}^{g}A = \begin{pmatrix} {}^{g}A_1^2 & 0 & \dots & 0 \\ 0 & {}^{g}A_2^2 & \dots & \cdot \\ \cdot & & & 0 \\ 0 & \dots & 0 & {}^{g}A_n^2 \end{pmatrix} ; \tag{39} $$

$^{g}A_i^2$ means here the asymptotic variance if the estimates concern the random variable ξ_i. From the general covariance matrix ^{g}A for independent random variables obviously follows the classical one if g is Gaussian:

$$ \mathbf{s} = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & & \cdot \\ \cdot & & & 0 \\ 0 & \dots & 0 & \sigma_n^2 \end{pmatrix} \tag{40} $$

(the inverse of $\mathbf{s}$ was given in Eq. (15)), as the expected value E is defined by Eq. (34) if g is Gaussian, and the asymptotic variance of the estimates for E is well known the variance itself (i.e., σ_i^2 in case of ξ_i). — It is convenient to denote also by an arbitrary statistical algorithm the asymptotic scatter ^{g}A by the characteristic itself which is estimated, e.g., if g is the Laplace distribution, the median (med) of ξ is estimated (as it can easily be verified on the ground of Eq. (34)), therefore ^{med}A is written in this case (and $^{E}A \equiv \sigma$ evidently holds). The only exception is the standard version of the most frequent value calculations, i.e., if the theoretical value of M is estimated: simply A is written instead of ^{M}A and thus the covariance matrix has the form, in case of independence if the P-norm is minimized:

$$ A = \begin{pmatrix} A_1^2 & 0 & \dots & 0 \\ 0 & A_2^2 & & \cdot \\ \cdot & & & 0 \\ 0 & \dots & 0 & A_n^2 \end{pmatrix} . \tag{41} $$

Instead of citing many papers we prefer to give in brief here the basic definitions and relations belonging to the P-norm, the standard version of the most frequent value calculations, etc.

The P-norm is defined as

$$P = \varepsilon \cdot \exp\left\{\frac{1}{2}\cdot\int_{-\infty}^{\infty}\ln\left[1+\left(\frac{x-x_0}{2\varepsilon}\right)^2\right] f(x)\,dx\right\} \tag{42}$$

where for ε (i.e., for the so-called dihesion)

$$\varepsilon^2 = \frac{3\int_{-\infty}^{\infty}\frac{(x-x_0)^2}{\left[\varepsilon^2+(x-x_0)^2\right]^2} f(x)\,dx}{\int_{-\infty}^{\infty}\frac{1}{\left[\varepsilon^2+(x-x_0)^2\right]^2} f(x)\,dx} \tag{43}$$

must hold iteratively. The minimum P-value is reached if $x_0=M$ is defined by

$$M = \frac{\int_{-\infty}^{\infty} s(x)\,x f(x)\,dx}{\int_{-\infty}^{\infty} s(x)\,f(x)\,dx} \,. \tag{44}$$

M is called the most frequent value (in standard version); the functions $s(x)$ is defined by

$$s(x) = \frac{4\varepsilon^2}{4\varepsilon^2+(x-M)^2} \,. \tag{45}$$

Eqs. (44) and (43) simultaneously define M and ε (in Eq. (43) naturally $x_0=M$ must be substituted); in the case of a homogeneous sample $x_1, \ldots, x_n$ the estimates for M and ε are calculated (by means of a twofold iteration) on the ground of the sum-counterparts of Eqs. (43) and (44). These equations are easy to get if we substitute

$$f(x) = \sum_{i=1}^{n} x_i\,\delta(x-x_i) \tag{46}$$

into Eqs. (43) and (44); for more details see the summarizing tables at the end of STEINER [1990] and [1991].

The asymptotic variance of the most frequent values is to be calculated according to

$$A^2 = \frac{\int_{-\infty}^{\infty}\frac{(x-M)^2}{[4\varepsilon^2+(x-M)^2]^4} f(x)\,dx}{\left[\int_{-\infty}^{\infty}\frac{4\varepsilon^2-3(x-M)^2}{[4\varepsilon^2+(x-M)^2]^3} f(x)\,dx\right]^2} \tag{47}$$

(see e.g. Eq (134) with $k=2$ in STEINER [1988]).

The most important question in respect of the inverse algorithms is the following: Do the basic statements concerning the optimum weighting

discussed in Section 1 remain after this generalization of the covariance matrix? In particular if our independent data are distributed according to two different ways (we have agreed that this simplification is consequently used in this paper), does the analogue of Eq. (14), i.e.,

$$w_{opt} = \frac{1/{}^{g}A_a^2}{1/{}^{g}A_a^2 + 1/{}^{g}A_b^2} \tag{48}$$

really give the optimum weight? For example, for $L1$-norm algorithms (i.e., for calculation of sample medians) really

$$w_{opt} = \frac{1/{}^{med}A_a^2}{1/{}^{med}A_a^2 + 1/{}^{med}A_b^2} \quad , \tag{49}$$

in the case of the P-norm (i.e., for calculations of the most frequent values) is

$$w_{opt} = \frac{1/A_a^2}{1/A_a^2 + 1/A_b^2} \tag{50}$$

really the best choice?

We have seen the Monte Carlo results from calculating medians in Fig. 5 (see once more the L_1-curve). The asymptotic scatter of the sample medians can be expressed (see e.g. CRAMÉR [1946]) simply as

$${}^{med}A = \frac{1}{2f(med)} \tag{51}$$

for the distributions defined in the Eqs. (29) and (30). The w_{opt}=0.8 value calculated according to Eq. (49) fully corresponds to the Monte Carlo results. In addition, the following analogue of Eq. (12) also seems to be valid for independent cases

$${}^{g}A_0^2(w) = \frac{1}{[n_a w + n_b(1-w)]^2}\left\{n_a w^2 \cdot {}^{g}A_a^2 + n_b(1-w)^2 \cdot {}^{g}A_b^2\right\} \quad , \tag{52}$$

as by substituting ${}^{med}A$ instead of the general ${}^{g}A$ two times in Eq. (52), the theoretical $q(w)$ calculated as 0.6745. ${}^{med}A_0^2(w)$ is near to the Monte Carlo results.

Two remarks: a/ theoretical curves are drawn in the overwhelming majority of our figures with dashed lines; b/ as far as the applied constant 0.6745 is concerned, we refer once more to the fact that estimates very often have (but not always) Gaussian distribution (see HUBER [1981]).

The small departures between the theoretical curve and the Monte Carlo results in the neighbourhood of $w=1$ are similar using the L_1-norm as in the discussed case of the L_2-algorithm because $n_a=100$ is not enough to get complete accordance with the asymptotic rules. The reasons, however, are quite different: the flanks of f_{ne} are very elongated and therefore the sample size must be very large if we wish the averages to behave according to the asymptotic rule, the behaviour of the central zone has hardly any influence; — on the contrary: the calculation of the sample medians is extremely sensitive in the case of symmetrical and unimodal distributions to the data around the maximum place of the density curve if the latter has a peaky maximum but in the case of the extreme pointedness of f_{ne}-curve (see Fig. 4) $n_a=100$ is really not yet enough to detect accurately this feature of the density curve on the basis of the sample (it is completely indifferent in respect of the sample medians if the flanks are heavy or not).

In Fig. 1 we saw in the case of $n_a=n_b=4$ full agreement between Monte Carlo results and the theoretical curve expressing primarily asymptotic behaviour. The first example for the eventual discordance was shown on the right hand side of Fig. 3a: the Monte Carlo results have expressed well the characteristic of the *mother distribution itself,* — but this was *very far* from the theoretical value calculated according to the asymptotic rule, namely, on the ground of the *asymptotic scatter* ($^EA = \sigma$) of the averages. The just discussed departures in Fig. 5 are traceable to similar origin, not only for L_2 but also for L_1, too.

From the viewpoint of geophysical practice it is very important that small samples also behave at least approximately according to the asymptotic rule and this is achieved if some significant characteristic of the mother distribution, say, q or Q, is near to the asymptotic scatter of the estimates obtained by the statistical algorithm in question. As according to Eq.(26), $Q \approx {}^EA = \sigma$ holds for the very classical, namely for the Gaussian case, we should see whether the intersextile range Q is far from or near to the asymptotic scatter of the standard most freqent value calculations (i.e., of the characteristic of the algorithms based on the P-norm, denoted by A) for frequently occuring mother distribution types.

The distribution types of the '$f_a(x)$-supermodel' defined by

$$f_a(x) = \frac{\Gamma\left(\frac{a}{2}\right)}{\sqrt{\pi} \cdot \Gamma\left(\frac{a-1}{2}\right)} \cdot \frac{1}{(1+x^2)^{a/2}} \qquad (a > 1) \qquad (53)$$

proved to be adequate for modelling actual error distributions which occur in geophysical practice. Introducing $p=1/(a-1)$ as parameter of type it can be proven that p_2-p_1 well approximates the type-distance of the corresponding distributions (defined by a_2 and a_1). Consequently it is straightforward to show all important characteristics as a function of p for a great type

interval, mostly from the Gaussian (p=0) to the Cauchy distribution (p=1; for more details about the supermodel $f_a(x)$, e.g. Cramér-Rao bounds, efficiencies, etc., see STEINER [1988]).

In our case in *Fig. 6* the curve of the quotient A/Q was drawn as a function of p; for comparison the σ/Q curve is also shown in the same figure.

Our conclusion on the ground of the A/Q-curve of Fig. 6 is that $A \approx Q$ is satisfactorily fulfilled for a very broad type interval (from the Gaussian type at least to the Cauchy-distribution); consequently, it is justifiable to expect that *asymptotic rules are applicable* (with the demanded accuracy) *also for small samples if algorithms based on the P-norm are used.* As a lower limit four data in a sample can be expected because for only three data the notion 'most frequent value' is hard to interpret adequately (and for two data all algorithms investigated — based on L_1-, L_2- or P-norm — give the same estimate).

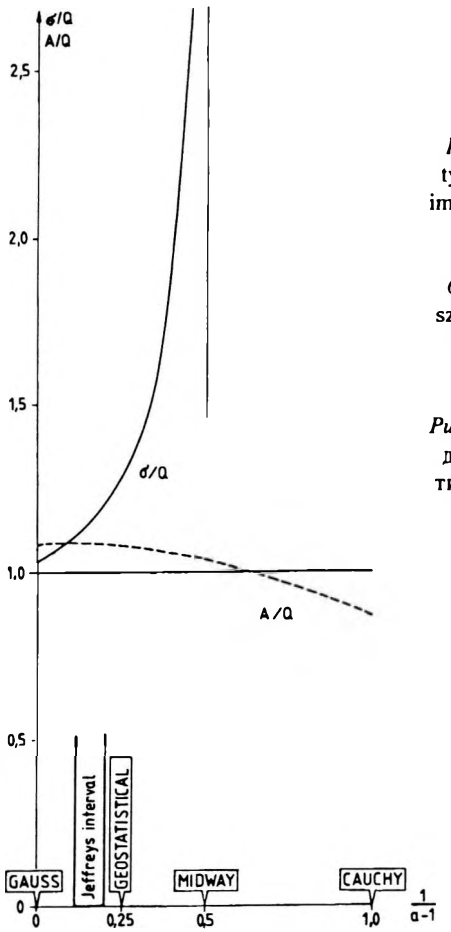

Fig. 6. Relation $A \approx Q$ holds for a large type interval; $\sigma \approx Q$ is valid only for the immediate neighbourhood of the Gaussian

6. ábra. Az $A \approx Q$ közelítő egyenlőség széles típustartományra érvényes; $\sigma \approx Q$ csak a Gauss-eloszlásra és közvetlen környezetére teljesül

Рис. 6. Приближенное равенство $A \approx Q$ действует для широкого диапазона типов; $\sigma \approx Q$ выполняется лишь для гауссовских распределений и их непосредственной близости

The σ/Q-curve of Fig. 6 shows that $\sigma \approx Q$ holds only for a surprisingly small type interval in the very neighbourhood of the Gaussian distribution; the problems shown in the Introduction are therefore to be understood without any difficulty. As

$$\sigma^2 = \begin{cases} \dfrac{1}{a-3} & , \quad \text{if} \quad a>3 \\ \infty & , \quad \text{if} \quad a \leq 3 \end{cases} \tag{54}$$

holds for f_a distributions, σ is infinite for a great variety of probability distribution types. In addition, it can be proved, that the asymptotic scatter of empirical scatters divided by the true value of σ is to be calculated according to the formula

$$A_\sigma/\sigma = \begin{cases} \dfrac{1}{2}\sqrt{3\dfrac{a-3}{a-5}-1} & , \quad \text{if} \quad a>5 \\ \infty & , \quad \text{if} \quad a \leq 5 \end{cases} \tag{55}$$

It means that in the case of the very often occuring $a=5$ (see the density curve of the type-occurence in Fig. 19) the determination of the elements of the covariance matrix becomes problematic (at least from the in point of view of acceptable accuracy).

Let us look at some examples. But first of all a remark which belongs to all the examples shown in this Section: to tell the truth, it would really be necessary to discuss in detail an enormous number of variants but if we were to do this, there would be the danger of losing the clarity of the present paper. The most essential things, however, can be shown, too, if we restrict ourselves to Cauchy-distributed data: the classical covariance matrix is not even defined in this case and therefore these examples stress best of all the importance and necessity of the generalized covariance matrix. In some cases the sample will consist of Cauchy- as well as of Gauss-distributed data. The calculations were made on the ground of the generally applicable P-norm.

Curiously enough also the theoretical $q(w)$-curve of the L_2-results can be simply given if the sample contains only Cauchy-distributed data but n_a of them have a probable error of q_a and n_b are to be characterized by q_b:

$$q_0(w) = \frac{1}{n_a w + n_b (1-w)} \left\{ n_a w \cdot q_a + n_b (1-w) \cdot q_b \right\} . \tag{56}$$

The validity of Eq. (56) can easily be proven on the basis of the Cauchy distribution being a so-called stable one (see Subsection 2.2). The $q_0(w)$ function in Eq. (56) is evidently a monotonous one and therefore no minimum exists for $0<w<1$.

Fig. 7 shows the theoretical $q(w)$ and $Q(w)$-curves, respectively, as 0.7645 and 0.9674 times (see Eqs. (25) and (26)) of the $A_0(w)$-values calculated according to

$$A_0(w) = \frac{1}{n_a w + n_b (1-w)} \cdot \sqrt{n_a w^2 \cdot A_a^2 + n_b (1-w)^2 \cdot A_b^2} \quad , \tag{57}$$

which is clearly a special variant of Eq. (52) for *P*-algorithms. (The numerical values of A obtained by using Eq. (47) are given for some distribution types in STEINER [1990], e.g., $A=1.5$ holds for the standard Cauchy distribution and $A=1.0466$ is valid for the standard Gaussian distribution. If the parameter of scale S differs from unity, A (being valid for the standard case) is to be multiplied by the actual value of S. The Monte Carlo results agree satisfactorily with the theoretical ones in the case of Fig. 7 although only a small sample was investigated ($n_a=n_b=4$). The same is true for the *P*-curve of *Fig. 8*; in this case samples contained data of Cauchy type as well as of Gaussian type. Consequently, the inverse of the generalized covariance matrix informs us adequately about the best weight w_{opt} to be used.

2. 2 *Generalization of the correlation matrix*

The elements of the classical correlation matrix ρ (see Eq. 8) are defined by Eqs. (4) and (5) and calculated on the ground of data pairs according to Eq. (6). These matrix elements, i.e., the correlation coefficients, are used popularly in practice (it is certain that one of the reasons for this is that $-1 \leq \rho \leq 1$ always holds), — although their definition do not make possible a plausible interpretation of this notion for appliers.

An obvious interpretation of ρ, however, can be simply given in the special case if ξ, η and ζ have equally standard Gaussian distribution ξ and ζ are independent and the relation

$$\eta = \rho \cdot \xi + \zeta \cdot \sqrt{1-\rho^2} \tag{58}$$

holds. Namely, it can be proven (see e.g. CRAMÉR [1946]) that in this case Eq. (5) results really in the correlation coefficient ρ having, according to Eq. (58), an immediate meaning: the random variable η 'contains' the random variable ξ to the extent of the proportionality factor ρ.

As Eq. (58) gives an obvious and simple connection of random variables, this relation should be regarded by our generalizing as a primary definition of a value which measures the closeness of the connection between random variables ξ and η. In the general case we shall denote this proportionality factor by r_{true}, — but how can Eq. (58) itself be generalized?

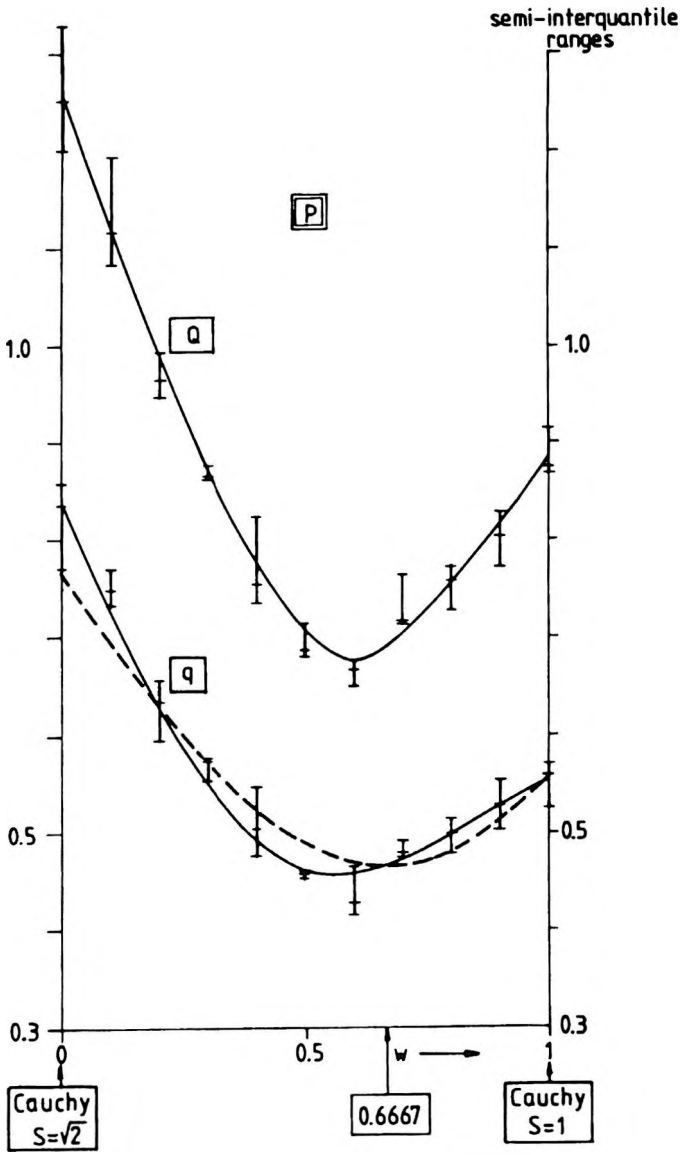

Fig. 7. Error curve of the results from the *P*-algorithm (n_a=n_b=4). The best choice of the weights corresponds to the generalized covariance matrix

7. ábra. P-algoritmussal nyert eredmények hibagörbéje (n_a=n_b=4). A súlyok optimális választása az általánosított kovarianciamátrixnak felel meg

Рис. 7. Кривая погрешностей результатов, полученных по алгорифму *P* (n_a=n_b=4). Оптимальный выбор весов отвечает обобщенной ковариационной матрице

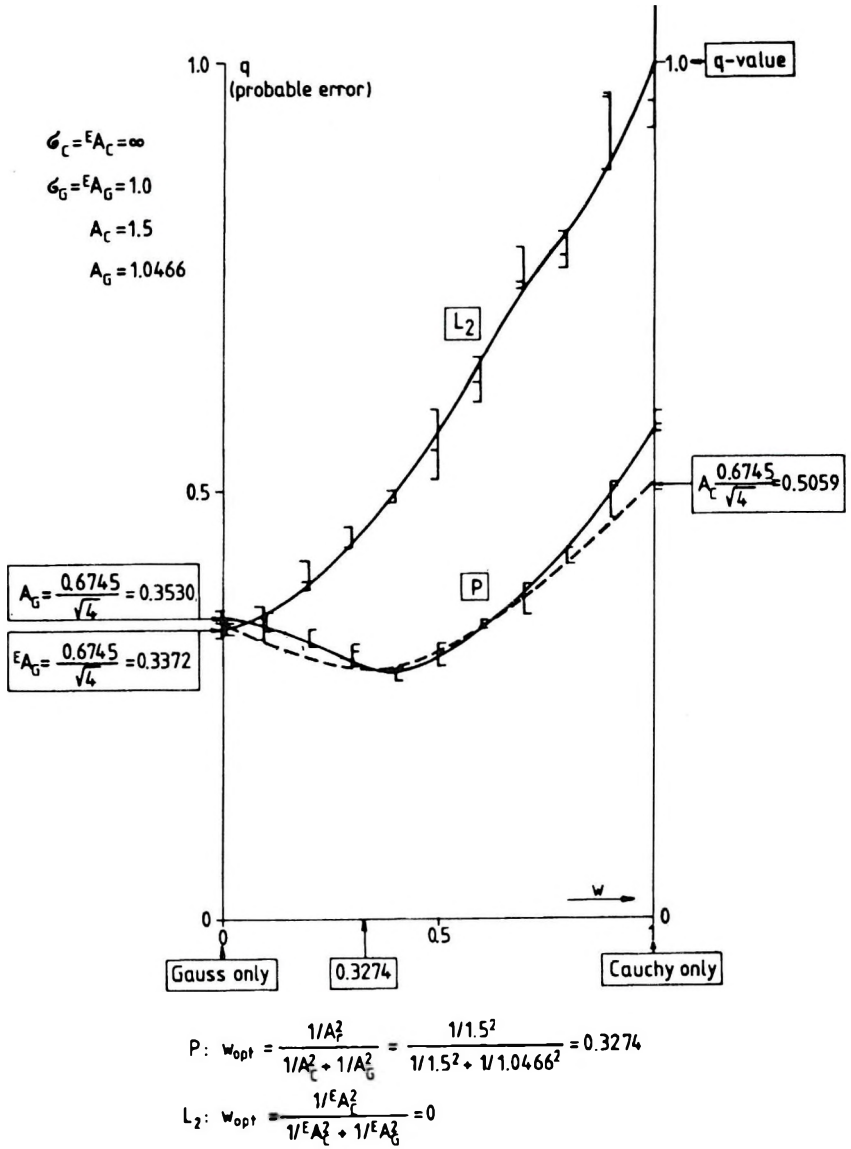

Fig. 8. Error curve of the results from the *P*-algorithm (n_a=4 data are Cauchy-distributed, n_b=4 are Gaussian). The best choice of the weights corresponds to the generalized covariance matrix

8. ábra. P-algoritmussal nyert eredmények hibagörbéje (n_a=4 adat Cauchy-, n_b=4 adat pedig Gauss-eloszlású). A súlyok optimális választása az általánosított kovarianciamátrixnak felel meg

Рис. 8. Кривая погрешностей результатов, полученных по альгорифму *P* (данные n_a=4 имеют распределение Коши, а данные n_b=4 гауссовское распределение). Оптимальный вибор весов отвечает обобщенной ковариационной матрице

In the first step the premise of the Gaussian distribution of the probability distributions should be given up. The supposition of the Gaussian type of all three random variables was convenient in respect of Eq. (58) because of the so-called stability of the Gaussian distributions: the sum of two Gaussian random variables is also Gaussian. There are, however, plenty of stable and symmetrical distributions besides the Gaussian, namely the types of the $f_\alpha(x)$-supermodel defined by

$$f_\alpha(x) = \frac{1}{\pi}\int_0^\infty \exp(-|t|^\alpha/\alpha) \cdot \cos(xt)\,dt \qquad (0 < \alpha \le 2) \qquad (59)$$

(for more about this supermodel see in STEINER [1990]; $f_\alpha(x)$ is the Gaussian density function if $\alpha = 2$, and in the case of $\alpha = 1$ we get the Cauchy distribution). Our generalization is simple: α should have all its possible values, not only the value 2. In this case the following generalized version of Eq. (58) is needed:

$$\eta = r_{true} \cdot \xi + \zeta \cdot (1 - |r_{true}|^\alpha)^{1/\alpha} \qquad (60)$$

(see Eq. (3) in HAJAGOS and STEINER [1989b]) and this corresponds to the following density function:

$$f_\alpha(x,y) = \frac{1}{(1-|r_{true}|^\alpha)^{1/\alpha}} \cdot f_\alpha(x) \cdot f_\alpha\left[\frac{y - r_{true} \cdot x}{(1-|r_{true}|^\alpha)^{1/\alpha}}\right]. \qquad (61)$$

With $\alpha = 2$ this expression clearly gives the well known formula for the two-variable Gaussian distribution.

The f_α and f_a distributions are very similar, see e.g. *Fig*. 9, — but this close connection can also be demonstrated theoretically (based on the investigation of efficiencies, see again the just cited paper. The following empirical formula yields that value of α for a given type parameter a; in such cases f_α is most similar to f_a:

$$\alpha(a) = 2 - 0.92 \cdot \arctan\frac{1.9}{a-1}\,, \qquad (62)$$

see *Fig. 10*). Consequently, the two-variable $f_a(x,y)$ can be written as

$$f_a(x,y) = \frac{1}{(1-|r_{true}|^\alpha)^{1/\alpha}} \cdot f_a(x) \cdot f_a\left[\frac{y - r_{true} \cdot x}{(1-|r_{true}|^\alpha)^{1/\alpha}}\right], \qquad (63)$$

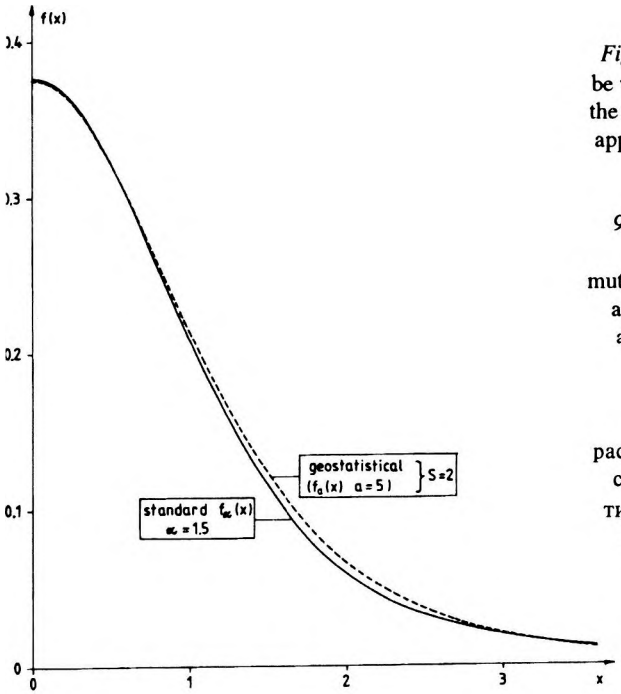

Fig. 9. Stable f_α distributions can be very similar to f_a distributions if the value of the α type parameter is appropriately chosen for the given value of **a**

9. ábra. A stabil f_α-eloszlások nagyfokú hasonlóságot mutathatnak az f_a-eloszlásokkal, ha az α típusparaméter értékét az adott **a**-értéknek megfelelően választjuk

Рис. 9. Стабильные f_α распределения весьма сходные с f_α распределениями, если типовой параметр α выбрать соответственно данному значению **a**

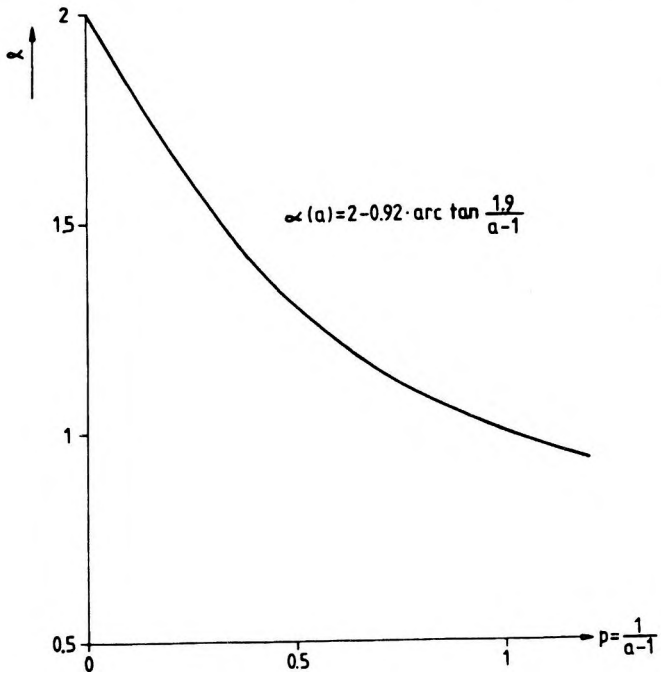

Fig. 10. Curve of the empirical $\alpha(a)$-function

10. ábra Az empirikus $\alpha(a)$-függvény görbéje

Рис. 10. Кривая эмпирической функции $\alpha(a)$

where the α-value is defined by Eq. (62).

After the foregoing the generalized correlation matrix can be written as

$$r_{true} = \begin{pmatrix} 1 & r_{true;1,2} & \cdots & r_{true;1,n} \\ r_{true;2,1} & 1 & \cdots & r_{true;2,n} \\ \cdot & & & \cdot \\ r_{true;n,1} & & \cdots & 1 \end{pmatrix} \tag{64}$$

($r_{true;i,k}$ characterizes the closeness of the statistical connection between ξ_i and ξ_k). Nothing was said, however, till now about the determination of the $r_{true;i,k}$-values; this will be treated from the point of view of practice in Subsection 3.1.

To have some idea about the error committed if we regard the expression for ρ in Eq. (6) as an estimate of r_{true}, see the curves for different α-values in *Fig. 11* (reprinted from HAJAGOS and STEINER [1989b]). For $\alpha = 1.5$, which often occurs ρ is greater by nearly 0.1 for a large r_{true}-interval, and for $\alpha = 1$ the difference can be 0.25, too, — and this value is a

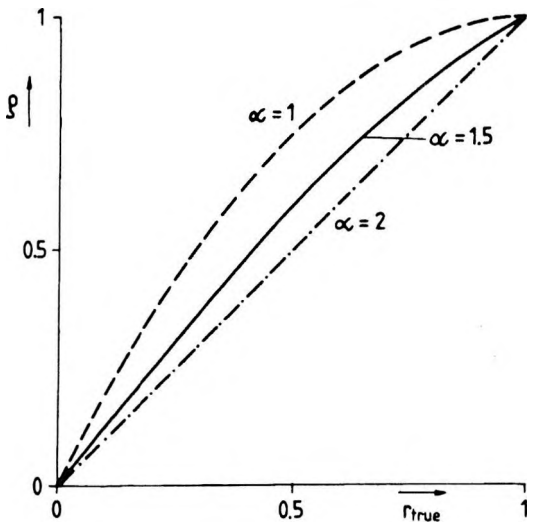

Fig. 11. Value of correlation coefficient ρ defined in classical statistics is systematically larger than r_{true} if the distributions in question are stable distributions characterized by $\alpha<2$ ($\alpha=2$ is the Gaussian case)

11. ábra. A klasszikus statisztika által definiált ρ korrelációs együttható szisztematikusan nagyobb, mint r_{true}, ha $\alpha<2$ ($\alpha=2$ esetén Gauss-eloszlással van dolgunk)

Рис. 11. Коэффициент корреляции ρ, дефинированный по классической статистике систематически выше, чем r_{true}, если $\alpha<2$. (В случае $\alpha=2$ имеем дело с гауссовским распределением)

quarter of the whole $|r_{true}|$-range. No wonder that practicians mostly believe in the existence of any correlation only if ρ is greater (or even significantly greater) than 0.5 (it should not be forgotten that $\alpha=1$ means the Cauchy distribution type — and this type can be utilized for modeling the unavoidable outliers, too, in the case of error distributions originally having not such heavy flanks, see TARANTOLA [1987]).

2. 3 Generalization of the covariance matrix

The elements of the classical covariance matrix are to be written as $\rho_{ik} \cdot \sigma_i \cdot \sigma_k$, see Eq. (7). We have seen, however, that $r_{true;i,k}$ is the proper generalization of ρ_{ik}, on the one hand (see Subsection 2.2), and on the other hand, the asymptotic scatter ${}^{g}A_i$ can in every respect be regarded as the generalization of the scatter σ_i (see Subsection 2.1). The straightforward generalization of the covariance matrix results therefore in

$$ {}^{g}\boldsymbol{A} = \begin{pmatrix} {}^{g}A_1^2 & r_{true;1,2} \cdot {}^{g}A_1 \cdot {}^{g}A_2 & \cdots & r_{true;1,n} \cdot {}^{g}A_1 \cdot {}^{g}A_n \\ r_{true;2,1} \cdot {}^{g}A_2 \cdot {}^{g}A_1 & {}^{g}A_2^2 & & \\ \vdots & & & \vdots \\ r_{true;n,1} \cdot {}^{g}A_n \cdot {}^{g}A_1 & r_{true;n,2} \cdot {}^{g}A_n \cdot {}^{g}A_2 & \cdots & {}^{g}A_n^2 \end{pmatrix}; \quad (65) $$

if all r_{true} are zero we get the already known form of ${}^{g}\boldsymbol{A}$ given in Eq. (39) for the case of independent random variables. If P-norm algorithms are used (i.e., most frequent values are calculated), the covariance matrix **A** clearly has the form

$$ \boldsymbol{A} = \begin{pmatrix} A_1^2 & r_{true;1,2} \cdot A_1 \cdot A_2 & \cdots & r_{true;1,n} \cdot A_1 \cdot A_n \\ r_{true;2,1} \cdot A_2 \cdot A_1 & A_2^2 & & \\ \vdots & & & \vdots \\ r_{true;n,1} \cdot A_n \cdot A_1 & r_{true;n,2} \cdot A_n \cdot A_2 & \cdots & A_n^2 \end{pmatrix} \quad (66) $$

(this is, formally regarded, very nearly the same as Eq. (65): the only difference is that no g-indices figure in Eq. (66)).

The examples (*Figs. 12–18*) for showing the influence of various weighting and especially the a priori weighting according to A^{-1} in the case of *P*-algorithms demonstrate not only Monte Carlo results for small samples but also the corresponding theoretical curve, too. These curves were calculated according to

$$q(w) = \frac{0.6745}{n_a w + n_b (1-w)} \cdot \sqrt{n_a w^2 \cdot A_a^2 + n_b (1-w)^2 \cdot (1+r_{true}) A_b^2} \qquad (67)$$

(this expression is clearly the analogue of Eq. (22) taking also Eq. (25) into consideration). Is was really unavoidable to show these empirical and theoretical $q(w)$-curves (the latter always with dashed line) for some situations to be able to draw important conclusions but the detailed discussion of the shown seven examples were superfluous. The reader will be satisfied in all cases with the agreement between the theoretically obtained (asymptotic) w_{opt}-value calculated on the ground of the corresponding A^{-1} which means in our simplified cases

$$w_{opt} = \frac{\dfrac{1}{A_a^2}}{\dfrac{1}{A_a^2} + \dfrac{1}{A_b^2 \cdot (1+r_{true})}}, \qquad (68)$$

and the optimal w shown by the Monte Carlo results for small samples.

Departures between theoretical and Monte Carlo curves can be really significant if there are only three data (see the neighbourhood of $w=1$ in Figs.13, 15, 16, 17) but minimum places coincide very well also in these cases. If four data are present the departures became insignificant; this was already known, however, on the basis of Figs. 7 and 8.

The error of an inversion is to be calculated also on the ground of the covariance matrix which is to be interpreted asymptotically. Therefore it is important to investigate the departures between the asymptotic and the Monte Carlo values of $q(w_{opt})$, too, but this question is not treated in the present paper in detail. If we remember, however, our train of thought concerning the customarily demanded accuracy of the errors (see the paragraph after Eq. (26)) and we also take into account Figs. 20 and 21 from the next Section, the differences between the asymptotic and the empirical $q(w_{opt})$-values can be regarded in all shown cases to be acceptable; in some cases the agreement is excellent.

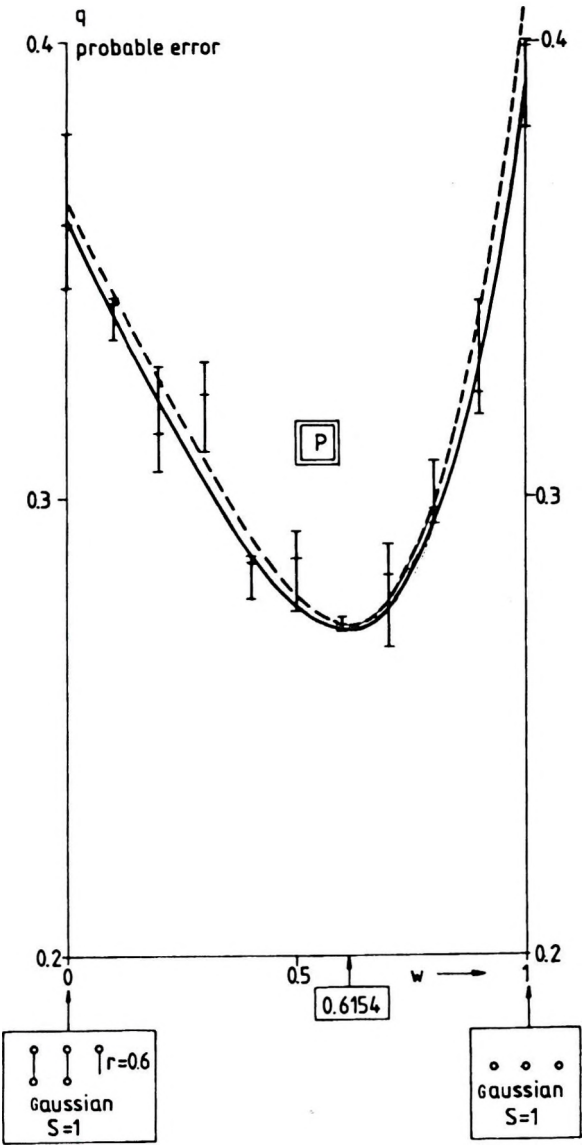

Fig. 12. Error curve of the results from the *P*-algorithm (Gauss-distributed data; n_a=3, n_b=6; r_{true}=0.6). The best choice of the weights corresponds to the generalized covariance matrix

12. ábra. P-algoritmussal nyert eredmények hibagörbéje (az adatok Gauss-eloszlásúak; n_a=3, n_b=6; r_{true}=0,6). A súlyok optimális választása az általánosított kovarianciamátrixnak felel meg

Рис.12. Кривые погрешности результатов, полученных альгорифмом *P* (данные имеют гауссовское распределение, n_a=3, n_b=6; r_{true}=0.6) Оптимальный выбор весов

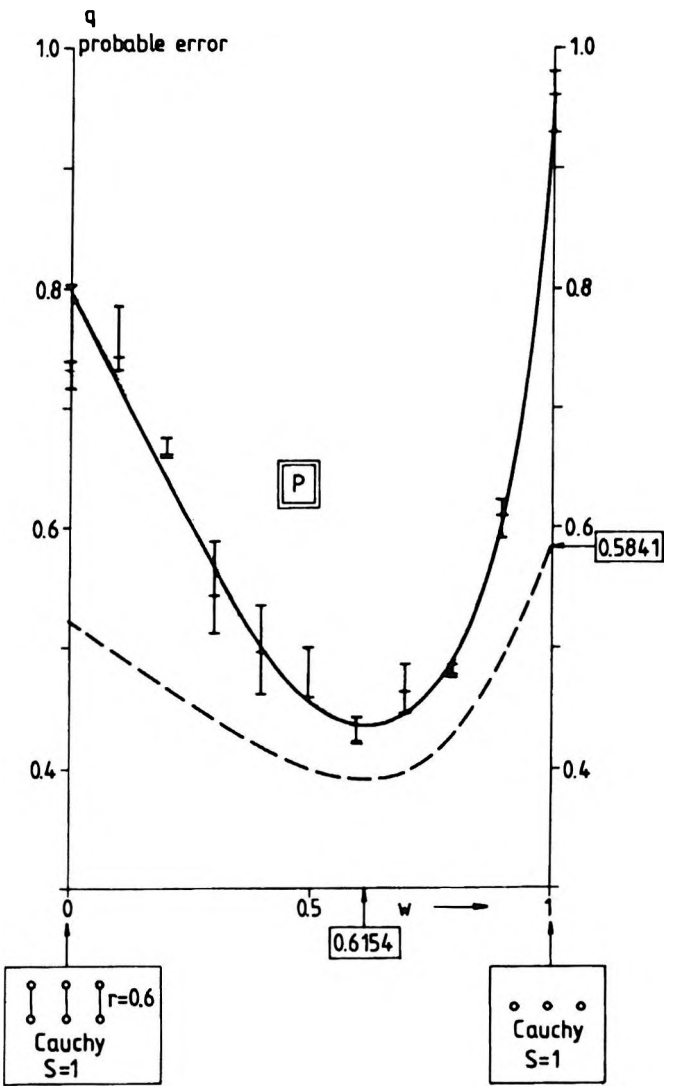

Fig. 13. Error curve of the results from the *P*-algorithm (n_a=3, n_b=6; r_{true}=0.6). The best choice of the weights correspontds to the generalized covariance matrix (the theoretical curve itself differs significantly on both sides)

13. ábra. P-algoritmussal nyert eredmények hibagörbéje (n_a=3, n_b=6; r_{true}=0,6). A súlyok optimális választása az általánosított kovarianciamátrixnak felel meg (noha az elméleti görbe mindkét széle szignifikáns eltérést mutat)

Рис.13. Кривые погрешности результатов, полученных альгорифмом *P* (n_a=3, n_b=6; r_{true}=0.6). Оптимальный выбор весов соответствует обобщенной ковариационной матрице (несмотря на то, что обе края теоретической кривой показывают расхождение)

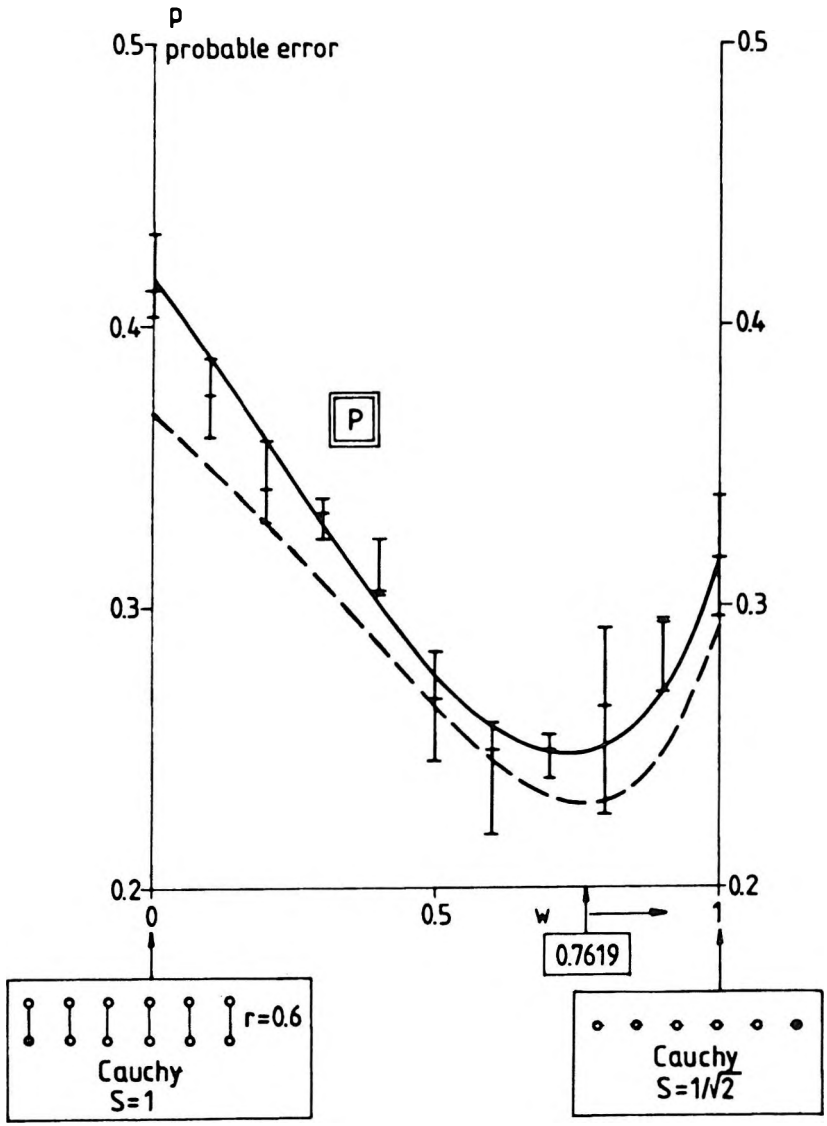

Fig. 14. Error curve of the results from the *P*-algorithm (n_a=6, n_b=12; r_{true}=0.6). The best choice of the weights corresponds to the generalized covariance matrix

14. ábra. *P*-algoritmussal nyert eredmények hibagörbéje (n_a=6, n_b=12; r_{true}=0,6). A súlyok optimális választása az általánosított kovarianciamátrixnak felel meg

Рис. 14. Кривые погрешности результатов, полученных алгорифмом *P* (n_a=6, n_b=12; r_{true}=0.6). Оптимальный выбор весов соответствует обобщенной ковариационной матрице

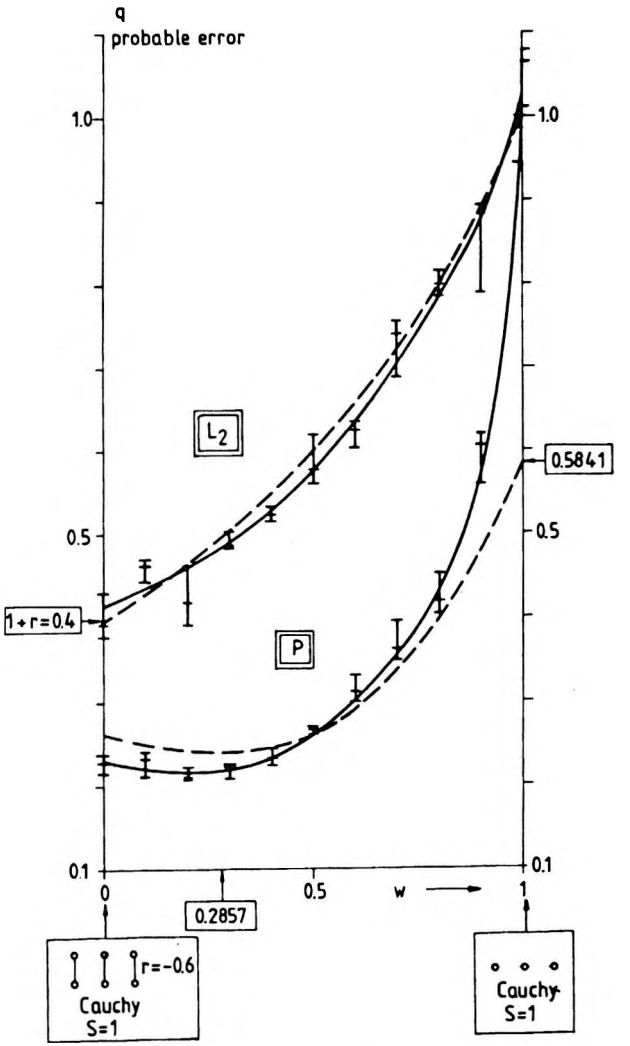

Fig. 15. Error curve of the results from the *P*-algorithm (n_a=3, n_b=6; r_{true}=-0.6). The best choice of the weights corresponds to the generalized covariance matrix. The error curve of the results from the classical L_2-algorithm is also shown

15. ábra. P-algoritmussal nyert eredmények hibagörbéje (n_a=3, n_b=6; r_{true}=-0,6). A súlyok optimális választása az általánosított kovarianciamátrixnak felel meg. A klasszikus statisztikai algoritmussal számított eredmények hibáit az L_2-vel jelölt folytonos görbe mutatja

Рис. 15. Кривые погрешности результатов, полученных альгорифмом *P* (n_a=3, n_b=6; r_{true}= - 0.6). Оптимальный выбор весов соответствует обобщенной ковариационной матрице. Погрешности результатов, полученных альгорифмом классической статистики показаны непрерывной кривой L_2

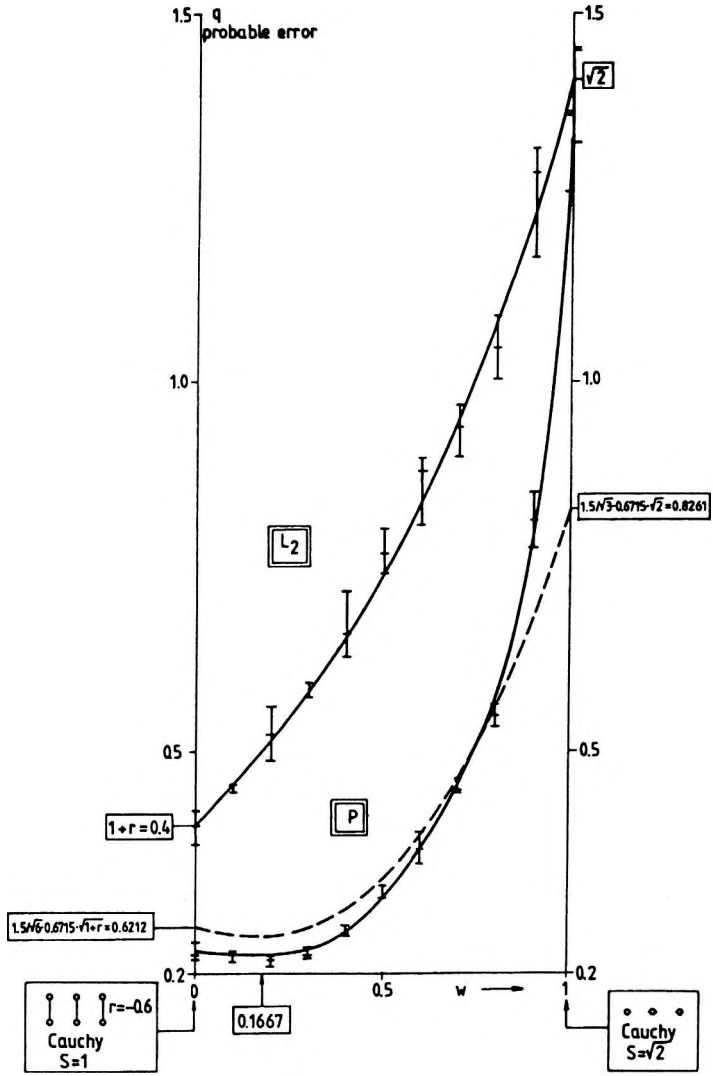

Fig. 16. Error curve of the results from the *P*-algorithm (n_a=3, n_b=6, r_{true}=-0.6; the scale parameters of the mother distributions are: $S_a=\sqrt{2}$ and S_b=1). The best choice of the weights corresponds to the generalized covariance matrix

16. ábra. P-algoritmussal nyert eredmények hibagörbéje (n_a=3, n_b=6, r_{true}=-0,6; az anyageloszlások skálaparaméterei: $S_a=\sqrt{2}$ és S_b=1) . A súlyok optimális választása az általánosított kovarianciamátrixnak felel meg

Рис.16. Кривые погрешности результатов, полученных альгорифмом *P* (n_a=3, n_b=6, r_{true}= - 0.6; параметры шкали распределения $S_a=\sqrt{2}$ and S_b =1). Оптимальный выбор весов соответствует обобщенной ковариационной матрице

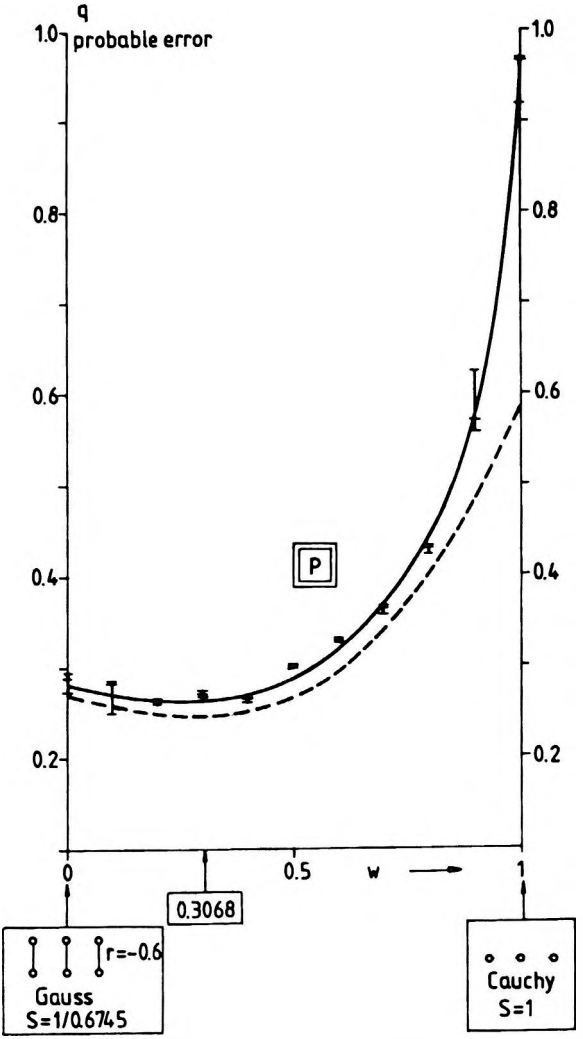

Fig. 17. Error curve of the results from the *P*-algorithm if n_a=3 data are Cauchy-distributed and independent, n_b=6 Gaussian data are pair-wise correlated: r_{true}=-0.6. (Both mother distributions have the same probable error.) The best choice of the weights corresponds to the generalized covariance matrix

17. ábra. P-algoritmussal nyert eredmények hibagörbéje, ha n_a=3 adat Cauchy-eloszlású és független, az n_b=6 db adat Gauss-eloszlású és páronként korrelált: r_{true}=-0,6. (A két anyaeloszlást azonos valószínű hiba jellemzi.) A súlyok optimális választása az általánosított kovarianciamátrixnak felel meg

Рис.17. Кривые погрешности результатов, полученных альгорифмом *P*, если данные n_a =3 независимые и отвечают распределению Коши, а n_b=6 данных имеют гауссовское распределение и попарно коррелированы : r_{true}= - 0.6. (Исходные распределения характеризуются такой же вероятностной погрешностью.) Оптимальный выбор весов соответствует обобщенной ковариационной матрице

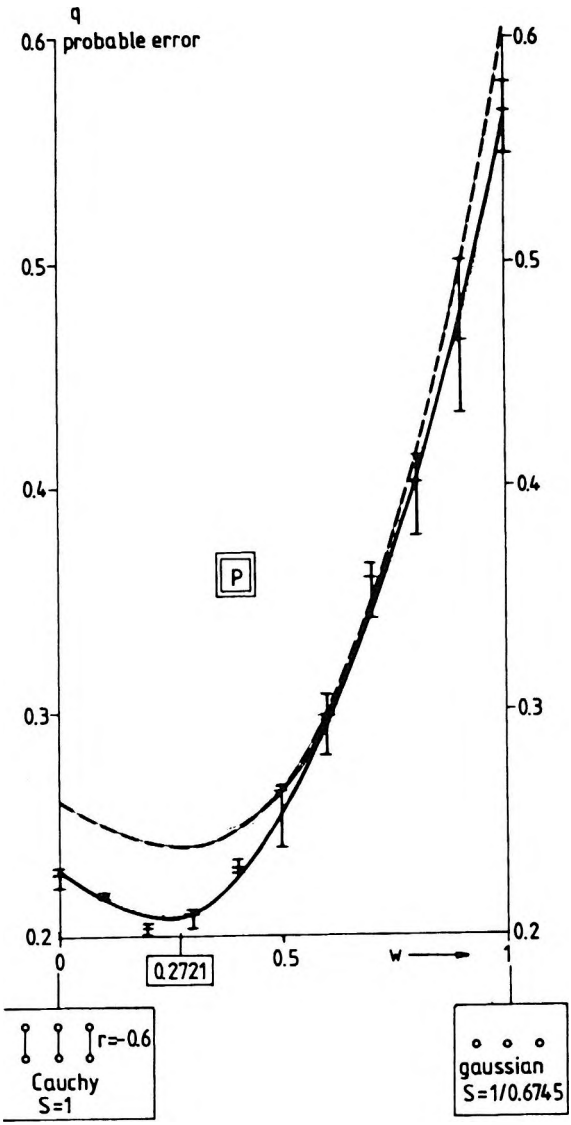

Fig. 18. Error curve of the results from the *P*-algorithm. The mother distribution types are changed, otherwise the samples were generated very similar to the case of Fig. 17. The best choice of the weights corresponds to the generalized covariance matrix

18. ábra. P-algoritmussal nyert eredmények hibagörbéje. A minták képzésénél az anyaeloszlás-típusokat megcseréltük, egyébként a 17. ábra felirata szerint jártunk el. A súlyok optimális választása az általánosított kovarianciamátrixnak felel meg

Рис. 18. Кривые погрешности результатов, полученных альгорифмом *P*. При задавании образцов тип исходных распределений поменяли, прочие условия те же, как на рис. 17. Оптимальный выбор весов соответствует обобщенной ковариационной матрице

3. Robustification

In Section 2 the covariance matrix was always regarded — as is usual — as 'a priori' given. In practice, however, the expression 'a priori' generally means nothing more than that these matrix elements were determined at an earlier stage of the work. In this Section we treat how these determinations are performed.

3.1 Robustification of the correlation matrix

In the literature of robust statistics it is customary to make robustifications of ad hoc type; this can be sanctioned by practice even if there is no satisfactory theoretical background. Let us now make such robustification of the formulae in Eqs. (5) and (6), forgetting for a moment the results of Section 2. Denoting the robustified ρ_{xy} by r_{xy}, our definitions are the following:

$$r_{xy} = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [s(x)\cdot(x-M_x)]\cdot[s(y)\cdot(y-M_y)]\cdot f(x,y)\,dxdy}{\sqrt{\int_{-\infty}^{\infty} s^2(x)\cdot(x-M_x)^2\cdot f(x)\,dx}\cdot\sqrt{\int_{-\infty}^{\infty} s^2(y)\cdot(y-M_y)^2\cdot f(y)\,dy}} \qquad (69)$$

if $f(x,y)$ is known. If data-pairs are given,

$$r_{xy} = \frac{\sum_{i=1}^{n} [s(x_i)\cdot(x_i-M_x)]\cdot[s(y_i)\cdot(y_i-M_y)]}{\sqrt{\sum_{i=1}^{n} s^2(x_i)\cdot(x_i-M_x)^2}\cdot\sqrt{\sum_{i=1}^{n} s^2(y_i)\cdot(y_i-M_y)^2}}\,; \qquad (70)$$

the s-values are calculated according to Eq.(45) with M_x and ε_x in the case of $s(x)$ and with M_y and ε_y in the case of $s(y)$.

The question rightfully arises whether the r_{xy} values, calculated on the ground of Eq. (69), really do not differ significantly from the r_{true}-values defined in Subsection 2.2. As practical cases can be well modelled by f_a-distributions, calculations for checking purposes were made for some values of the type parameter **a**. (By choosing these values, we also took into account the probability density function of the occurrence of different **a**-values in *Fig. 19*; see STEINER [1990]). After this choice, the $f(x,y)$ in

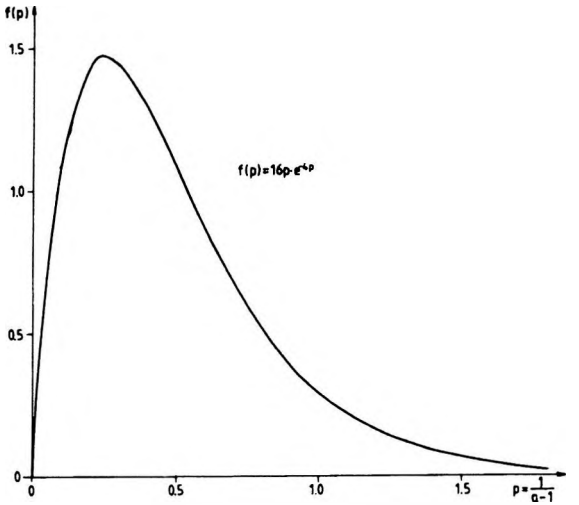

Fig. 19. Density function modelling the occurrence probabilities of different probability distribution type intervals of the $f_a(x)$-supermodel

19. ábra. Sűrűségfüggvény, amely az $f_a(x)$-szupermodell különböző valószínűségeloszlás típus-intervallumainak előfordulási valószínűségét modellezi

Рис. 19. Функция плотности, моделлирующая вероятность наблюдения разных интервалов типа распределения вероятности супермоделли $f_a(x)$

Eq. (69) was taken according to Eq. (63); *Table I.* shows the differences $(r_{xy}-r_{true})$.

The maximum difference is 0.071 (see the row for the Cauchy distribution); if $a \geq 3$ can be supposed in an actual case, all absolute differences are less than 0.05.

These differences (of the character of bias) can be neglected by comparison with the expected statistical fluctuations. To have some idea about the measure of the latter; for n=100, pairs of data (of different probability distribution type) were generated according to Eqs. (60) and (62) in the first step; on the basis of these data-pairs r_{xy}-values were calculated according to Eq. (70) (for comparison the ρ_{xy} values were also determined, see Eq. (6)). This procedure was repeated N=200 times for each type investigated, consequently the interquartile- (full line), the intersextile- (dashed line) and the whole range (thin line) of the data were easy to construct (see *Fig. 20* where also the medians were indicated). The shifts correspond to the already known bias (see Table I) being in general less than the probable error even in the investigated case of n=100. (This statement does not hold for the values of ρ_{xy},— except the very classical but seldom if ever occuring Gaussian distribution.) A further consequence can be made on the basis of Fig. 20: the statistical fluctuation of r_{xy} is nearly

	r_{true} = 0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Gauss (a→∞)	0	-0.008	-0.016	-0.023	-0.029	-0.032	-0.035	-0.033	-0.028	-0.017	0
a=10 } Jeffreys-interval	0	-0.006	-0.010	-0.014	-0.016	-0.017	-0.016	-0.014	-0.009	-0.004	0
a=6 }	0	-0.002	-0.003	-0.003	-0.003	-0.002	0	0.002	0.005	0.007	0
geostatistical (a=5)	0	0	0.001	0.003	0.005	0.007	0.009	0.011	0.013	0.012	0
midway (a=3)	0	0.011	0.023	0.033	0.041	0.047	0.049	0.048	0.043	0.026	0
Cauchy (a=2)	0	0.029	0.050	0.064	0.070	0.071	0.067	0.057	0.041	0.029	0

Table I. The differences $(r_{xy}-r_{true})$ for different values of r_{true} and for various probability distribution types

I. táblázat. Az $(r_{xy}-r_{true})$ különbségek r_{true} különböző értékeire és néhány valószínűségeloszlástípusra

Табл. I. Разницы $(r_{xy}-r_{true})$ для разных значений r_{true} и при изменении типа распределения вероятности

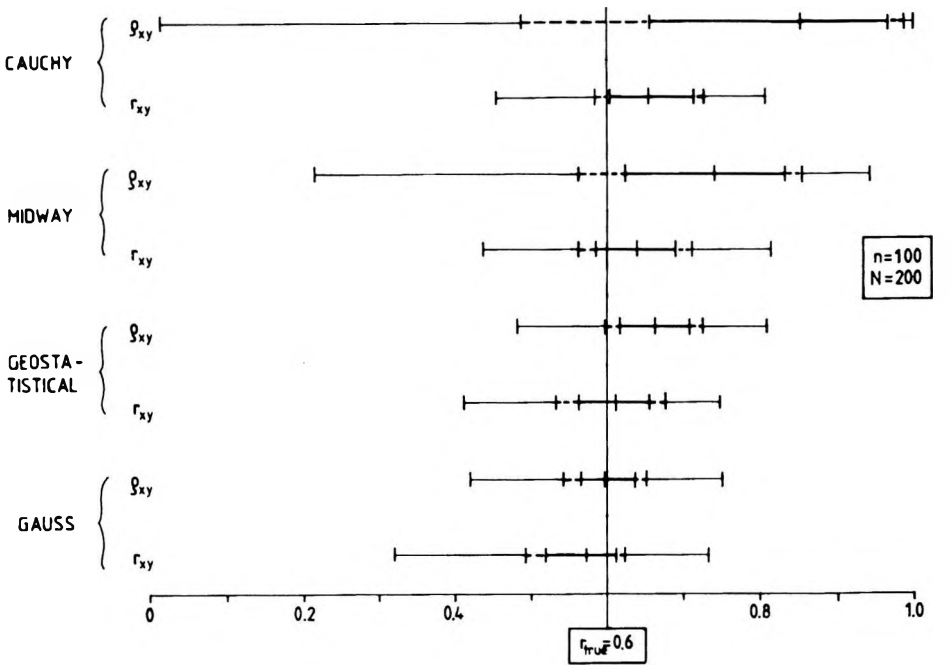

Fig. 20. Statistical fluctuatuions of the r_{xy} - and ρ_{xy}-values in the case of n=100

20. ábra. Az r_{xy}- és ρ_{xy}-értékek statisztikus ingadozása n=100 esetén

Рис. 20. Статистическое колебание значений r_{xy} и ρ_{xy} при n=100

independent of the distribution type; on the contrary, that of ρ_{xy} strongly depends on it and, in addition, we can state that the measure of the statistical fluctuation is also in itself fully unacceptable in the case of the Cauchy distribution (a similar statement characterizes the midway distribution), to say nothing about the great shifts (obtained in full agreement with Fig. 11, see e.g. curve 'α=1').

Concerning statistical fluctuations it is perhaps not superfluous to reprint two figures from CRAMÉR [1946] (see *Fig. 21*): in the frequently occuring cases of n=50 and n=10 the density functions of the ρ_{xy}-values show that unexpectedly large statistical errors occur even in the least problematic case: if the data are Gauss-distributed; this is closely linked with the fact that any kind of correlaton coefficient is characteristic of a *two-varibale* distribution. It can be stated as a conclusion that *the statistical fluctuation of* ρ_{xy} *in the very negihbourhood of the Gaussian as well as that of the* r_{xy}*-values in a broad type interval is significantly greater* at sample sizes used in the overwhelming majority of cases *than the bias of the* r_{xy}*-values*. Consequently, *the robustified correlation matrix defined by*

$$\boldsymbol{r} = \begin{pmatrix} 1 & r_{1,2} & r_{1,3} & \cdots & r_{1,n} \\ r_{2,1} & 1 & r_{2,3} & \cdots & r_{2,n} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ r_{n,1} & & \cdots & \cdots & 1 \end{pmatrix} \tag{71}$$

can be regarded as a satisfactory approximation of the generalized correlation matrix $\boldsymbol{r}_{true}$. As for the computing techniques: the calculation of r_{xy} organically and simply joins with the basic algorithms of the most frequent value procedures (no special programs are needed).

3.2 Robustification of the covariance matrix

The meaning of the word 'robust' involves not only the applicability on a broad type interval but often the resistance, i.e., the insensitivity to outliers, too. In actual fact, the *s*-functions figuring in Eqs. (69) and (70) (i.e., in the formulae of the robust correlation coefficient) guarantees the resistance, too.

The *s*-function, however, is a basic function of the most frequent value calculations, and if we intend to robustify the scatter by means of this function, in a self-consistent way we can only set the aim that the A_i-values should be approximated by the robustified scatters.

A possible robustified form of the scatter is the following:

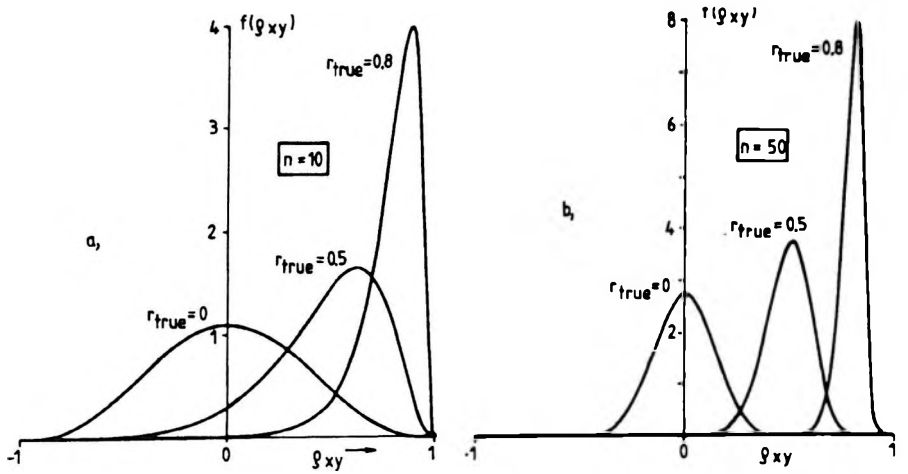

Fig. 21. Density functions of ρ_{xy} (from CRAMÉR [1946]) for n=10 and n=50 if the random variables have Gaussian distribution

21. ábra. A ρ_{xy} sűrűségfüggvényei Gauss-eloszlású valószínűségi változók esetén, ha az adatpárok száma n=10 illetve n=50

Рис. 21. Кривые плотности ρ_{xy} для вероятностных переменных с гауссовским распределением, если количество пар данных n=10, n=50

$$\bar{A} = \sqrt{\frac{9}{5}\int_{-\infty}^{\infty} s(x) \cdot (x-M)^2 f(x)\,dx} \quad ; \tag{72}$$

its estimate is clearly

$$\bar{A} = \sqrt{\frac{9}{5n}\sum_{i=1}^{n} s(x_i) \cdot (x_i-M)^2} \quad . \tag{73}$$

The notation $\bar{A}$ anticipates that $\bar{A} \approx A$ is valid (it is clear that a notation $\bar{\sigma}$ would equally be justifiable). *Table II.* gives the values A and $\bar{A}$ for some probability distribution types and the departures, too (in per cent).

	A	Ā	100.(Ā-A)/A
Gauss	1.0466	1.0369	-0.93 %
a=10	0.3694	0.3666	-0.76 %
a= 6	0.5173	0.5150	-0.44 %
geostatistical	0.5917	0.5904	-0.22 %
midway	0.9236	0.9350	+1.23 %
Cauchy	1.5000	1.5492	+3.28 %

As the fulfilment of $\bar{A} \approx A$ is satisfactory, the robustified covariance matrix $\bar{A}$ (which gives nearly $\bar{A}$, see Eq. (66)) can be written as follows:

$$\bar{A} = \begin{pmatrix} \bar{A}_1^2 & r_{1,2} \cdot \bar{A}_1 \cdot \bar{A}_2 & \dots & r_{1,n} \cdot \bar{A}_1 \cdot \bar{A}_n \\ r_{2,1} \cdot \bar{A}_2 \cdot \bar{A}_1 & \bar{A}_2^2 & & \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ r_{n,1} \cdot \bar{A}_n \cdot \bar{A}_1 & r_{n,2} \cdot \bar{A}_n \cdot \bar{A}_2 & & \bar{A}_n^2 \end{pmatrix} . \tag{74}$$

Thorough investigation of the determination errors of the matrix elements is beyond the scope of the present article (e.g., whether the method given in HAJAGOS and STEINER [1989a] for extraordinary outliers is to be applied or not). Such investigations should preferably be made parallely with the study of practical cases.

The authors are indebted to L. FERENCZY, consultant (Geophysical Exploration Co., Budapest) for having initiated this work, and to their colleagues L. CSERNYÁK, Á. GYULAI and T. ORMOS for fruitful discussions .

REFERENCES

CRAMÉR H. 1946: Mathematical Methods of Statistics. Princeton University Press, Princenton

HAJAGOS B. 1982: Der häufigste Wert, als eine Abschätzung von minimalem Informationsverlust etc. Publications of the Technical University for Heavy Industry Series A Mining **37**, 1-2, pp. 95-114

HAJAGOS B. 1985: Die verallgemeinerten Student-scken t-Verteilungen und die häufigsten Werte. Publications of the Technical University for Heavy Industry Series A Mining **40**, 1-4, pp. 225-238

HAJAGOS B., STEINER F. 1989a: Methods to increase the resistance etc. Acta Geodaetica, Geophysica et Montanistica Acad. Sci. Hung. **24**, 3-4, pp. 289-307

HAJAGOS, B., STEINER F. 1989b: Measure of the linear dependence. Acta Geodaetica, Geophysica et Montanistica Acad. Sci. Hung. **24**, 3-4, pp. 417-439

HUBER P. J. 1981: Robust Statistics. Wiley, New York

INMAN J. R. 1975: Resistivity inversion with ridge regression. Geophysics, **40**, 5, pp. 790-797

STEINER F. 1988: Most frequent value procedures. Geophysical Transactions, **34**, 2-3, pp. 139-260

STEINER F. 1990: Foundations of Geostatistics (in Hungarian). Tankönyvkiadó, Budapest, 363 p.

STEINER F. (Ed.) 1991: The Most Frequent Value. Akadémiai Kiadó, Budapest, 315 p.

TARANTOLA A. 1987: Inverse Problem Theory. Elsevier, Amsterdam

A KOVARIANCIAMÁTRIX ÁLTALANOSÍTÁSA ÉS ROBUSZTIFIKÁLÁSA

HAJAGOS Béla és STEINER Ferenc

A dolgozat első része az egyszerű szemléltethetőség érdekében néhány elkerülhetetlen egyszerűsítést vezet be. Ezután a klasszikus statisztika kovarianciamátrixának inverzével való súlyozás optimális voltát mutatja meg a cikk Gauss-eloszlású hibákra, ugyanakkor példákkal hívja fel a figyelmet arra, hogy ettől eltérő hibaeloszlásoknál, kis mintaelemszám esetén, az optimális súlyozást nem okvetlenül ez a megoldás szolgáltatja.

A második rész a kovarianciamátrix általánosított, az alkalmazott statisztikai algoritmusnak megfelelő alakját definiálja; példákat zömmel *P*-algorimust alkalmazó eljárásokra, azaz súlyozott leggyakoribb érték-számításokra mutat be. Kiderült, hogy az általánosított kovarianciamátrix inverze kis mintaelemszámok esetén is alkalmazható, ha a leggyakoribb értékek szerinti algoritmusokkal dolgozunk.

A harmadik rész ezen általánosított kovarianciamátrixot a klasszikus eset robusztifikálásával javasolja egyszerűség kedvéért meghatározni. Ez a meghatározás ui. számítástechnikai szempontból szorosan illeszkedik a leggyakoribb érték szerinti számítások alapvető algoritmusaihoz.

ОБОБЩЕНИЕ И РОБУСТИФИКАЦИЯ КОВАРИАЦИОННОЙ МАТРИЦЫ

Бела ХАЯГОШ, Ференц ШТЕИНЕР

В первой части статьи вводятся некоторые необходимые для наглядности упрощения. После этого показывается оптимальность взвешивания обратной ковариационной матрицы классической статистики для погрешностей Гауссовского распределения, и обращается внимание на то, что при отличающемся от Гауссовского распределение погрешностей и при небольшом количестве образцов оптимальное взвешивание получается необязательно данным способом.

Во второй части дается определение обобщенного, подходящего для прминенного статистического альгорифма вид ковариационной матрицы. Примеры показываются прежде всего для способов, применяющих альгорифм *P*, то есть показываются расчеты взвешенной наиболее частой величины. Выявилось, что обратнуаю матрицу обобщенной ковариационной можно применять и в случае небольшого количества образцов, если работать с альгорифмами по наиболее частым величинам. В третьей части рекомендуется для простаты определить эти обобщенные ковариационные матрицы при робустификации классической, ввиду того, что такое определение непосредственно связано с основными альгорифмами расчетов по наиболее частым величинам.