GEOPHYSICAL TRANSACTIONS 1997
Vol. 41. No. 1–2. pp. 1–21

New results on the theory of the most frequent value procedures

Ferenc STEINER*

New results have been published in the last few years on the theory of the most frequent value (MFV) procedures. In this paper some of these results are presented in a concise manner to show more clearly the differences between the statistical procedures based upon the L_2, L_1 and P-norms, respectively.

Keywords: most frequent value, robustness, P–norms, statistical analysis, type-distance

1. Different attitudes to the actual probability distribution type of measurement errors result in quite different statistical procedures

To facilitate the presentation of the ne ⁄ results, let us start from the definitions of the likelihood function and from that of the *I*-divergence. There are similarities (or even equalities) and, on the contrary, substantial differences, too. These are summarized in *Table I.*

Table I.

f is assumed to be **unknown** (this corresponds to the overwhelming majority of practical tasks), and therefore it is **substituted** by *g(T,S;x)* of given analytical form. This substitution results in an **information loss** measured by the '*I*-divergence': $$I_g = \int_{-\infty}^{\infty} f(x)\ln\left[f(x)/g(T,S;x)\right]dx.$$	f is assumed to be **a priori known**, including the value of S or T (this latter assumption is implicitly present in the well known book of ANDREWS et al. [1972]). $$\mathcal{L} = \sum_{i=1}^{n} -\ln f(T,S;x_i)$$ is the so-called **likelihood function** which is appropriate for to the basic assumption.

* University of Miskolc, H-3515 Miskolc-Egyetemváros

Manuscript received: 10 October, 1995

Table I. continued

<table>
<tr><td colspan="2">Both I_g and $\mathcal{L}$ should be minimized. If S is known, the accepted value of T fulfils:</td></tr>
<tr><td>$$\frac{\partial I_g}{\partial T} = 0 \qquad (1)$$</td><td>$$\frac{\partial \mathcal{L}}{\partial T} = 0 \qquad (2)$$</td></tr>
<tr><td colspan="2">The integral form of Eq. (2) can be trivially given, and similarly Eq. (1) (which is an integral expression) for the data x_i. Astonishingly enough, both Eqs. (1) and (2) lead to exactly the same formula for T-determination. (In general, iteration is necessary.)
On the contrary, the two approaches belonging to f result in the fact that S-formulae essentially differ from each other:</td></tr>
<tr><td>To be sure that Eq. (1) really results in minimum information loss, the relation $$\frac{\partial^2 I_g}{\partial T^2} > 0 \qquad (3a)$$ must also hold. The fulfilment of this relation, however, is warranted if $$(H \equiv) \int_{-\infty}^{\infty} \frac{\partial^2 g(T,S;x)}{\partial^2 T} \cdot \frac{f(x)}{g(T,S;x)} \mathrm{d}x = 0 \quad (3)$$ holds [see HAJAGOS 1991]. Eq. (3) defines (with an analytically given g-function) the formula for the calculation of S (i.e., of the parameter of scale).</td><td>According to the usual maximum likelihood techniques (see the already cited ANDREWS et al. [1972],) $$\frac{\partial \mathcal{L}}{\partial S} = 0 \qquad (4)$$ is also demanded and with an analytically given f Eq. (4) results in a formula for determining the parameter of scale. Unfortunately, Eq. (4) leads to such S-formulae which are not resistant (i.e., they are outlier-sensitive).</td></tr>
<tr><td colspan="2">A new result can already be formulated:
The formulae derived from Eq. (3) and Eq. (4) never coincide with each other except in the Gaussian case.</td></tr>
<tr><td colspan="2">Notations</td></tr>
<tr><td colspan="2">x_i : directly measured value ($1 \le i \le n$)</td></tr>
<tr><td colspan="2">f, $f(x)$, $f(T,S;x)$: the actual probability density function</td></tr>
<tr><td colspan="2">T: parameter of location (this is often the symmetry point). In nearly all of the tasks the unknown of primary interest is the value of T.</td></tr>
<tr><td colspan="2">S: parameter of scale. We are perhaps not intrinsically interested in its value but the resulting T-value can be heavily influenced by the actual value of S.</td></tr>
</table>

Table I. Two different approaches to the type of the actual probability distributions lead to both theoretically and practically different statistics

I. táblázat. Az előforduló valószínűségeloszlások típusára vonatkozó különböző szemléletű alapfeltevések elméleti és gyakorlati szempontból egyaránt eltérő statisztikai eljárásokhoz vezetnek

This new thesis: 'the formulae derived from Eq. (3) and Eq. (4) never coincide with each other except in the Gaussian case', [see CSERNYÁK 1994] is important also concerning the T-determinations inasmuch as using both methods (namely based on the maximum likelihood principle, on the one hand, and on the other hand, on the *I*-divergence), in just the same manner two formulae are to be fulfilled *simultaneously* (Eqs. (1) and (3) or Eqs. (2) and (4)). Consequently, the formula for the determination of the parameter of scale strongly influences the behaviour of the *T*-determination, e.g. in respect of robustness.

Let us show an example. Choosing the Cauchy type distribution as substitute one (i.e. *g*-function) in the first method, and the same Cauchy-distribution as an a priori known *f* density function in the second one (i.e., in the maximum likelihood method), the actual *T*-formulae derived on the basis of Eqs. (1) and (2) are just the same (as was mentioned earlier in the fourth row of Table I). On the contrary, Eqs. (3) and (4) result in quite other formulae for the parameter of scale, if the Cauchy distribution was chosen as substitute distribution *g*, — but this is in full accordance with the thesis at the end of Table I. If X_i means the residual, i.e., measured value x_i minus computed value (in the simplest case $X_i=x_i-T$ obviously holds), from Eq.(3) it follows that:

$$S^2 = \frac{3\sum_{i=1}^{n} X_i^2 \Big/ \left(S^2 + X_i^2\right)^2}{\sum_{i=1}^{n} 1 \Big/ \left(S^2 + X_i^2\right)^2} \tag{5}$$

and the resulting value is called 'dihesion' (and is denoted by ε). Eq. (4) results in quite another formula (without squares and without the factor '3')

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 \Big/ \left(S^2 + X_i^2\right)}{\sum_{i=1}^{n} 1 \Big/ \left(S^2 + X_i^2\right)} \tag{6}$$

if the Cauchy-distribution was chosen as substitute distribution *g*.

The question arises as to whether the choice of the determination method of *S*, i.e. of the parameter of scale, is really of significant importance in respect of the determination of *T*? Determination of the value of *T* (or of the values $p_1, p_2, \ldots, p_j, \ldots, p_J$ as components of the unknown parameter vector $\overline{p}$ in multidimensional cases) always has priority in our practical tasks. The question of errors or bias of the *S*-determinations is treated in general as a secondary one, or simply the conventional formulae are used for error-determinations. BARTA and HAJÓSY [1985] showed an astonishing (or even comical or possibly even rather tragic) example concerning determinations of the universal gravity constant: the error intervals given by three authors to their measurements have no common point (*Fig. 1*). Possibly Gaussian error distribution was supposed but the real error seemingly was that nothing was known about this supposition... (The whole range of the three error-intervals divided by the absolute value of the gravity constant is 0.07%, i.e., by several orders of magnitude greater than in the case of some other universal constant, e.g. the velocity of light in vacuum, etc.) Figure 1 shows that error-determinations of classical manner can lead to completely false results.

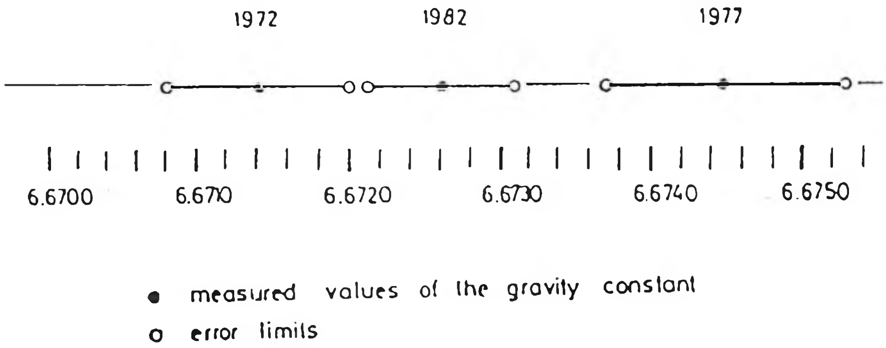

Fig. 1. Error intervals for differing values of the universal gravity constant determined by the measurements of three authors [from BARTA and HAJÓSY 1985]. The error intervals have no common point

1. ábra. Hibaintervallumok az általános tömegvonzási állandónak három szerző által meghatározott értékeihez [BARTA és HAJÓSY 1985 nyomán]. Egyetlen közös pontja sincs a hibaintervallumoknak

2. Essential differences between modern statistics (based e.g. upon the norms P_J, P, P_C, P_{lt}) and the conventional statistical procedures (based on the L_2-norm)

One can perhaps say that some *experimenters* with limited theoretical background work only on the basis of the outdated statistics of the last century. Not at all. Even the Heisenberg relation is formulated for *scatters*, i.e., for minimum values of the old L_2-norm. Incidentally, nowadays primarily norm representations of statistical algorithms are more and more given. The following table (*Table II*) therefore gives the simple expressions of six norms (expressed by the X_i residuals defined earlier) based on HAJAGOS and STEINER [1993a].

Norm	Formula	Eigen-distribution (for this type of error distribution the norm in question works optimally)
L_1	$\frac{1}{n}\sum_{i=1}^{n}\lvert X_i\rvert$	Laplace
L_2	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}X_i^2}$	Gauss
P_J	$\varepsilon\cdot\left\{\prod_{i=1}^{n}\left[1+\left(\frac{X_i}{3\varepsilon}\right)^2\right]\right\}^{\frac{1}{2n}}$	Jeffreys
P	$\varepsilon\cdot\left\{\prod_{i=1}^{n}\left[1+\left(\frac{X_i}{2\varepsilon}\right)^2\right]\right\}^{\frac{1}{2n}}$	geostatistical
P_C	$\varepsilon\cdot\left\{\prod_{i=1}^{n}\left[1+\left(\frac{X_i}{\varepsilon}\right)^2\right]\right\}^{\frac{1}{2n}}$	Cauchy
P_{lt}	$\varepsilon\cdot\left\{\prod_{i=1}^{n}\left[1+\left(\frac{2X_i}{\varepsilon}\right)^2\right]\right\}^{\frac{1}{2n}}$	(very long-tailed error distributions)
For ε (called dihesion) $\sum_{i=1}^{n}\frac{3X_i^2-\varepsilon^2}{\left(\varepsilon^2+X_i^2\right)^2}=0$ must be fulfilled.		

Table II. Formulae of six statistical norms (the corresponding eigen-distributions are also given)

II. táblázat. Hat statisztikai norma formulája és a normákhoz tartozó sajáteloszlások

Returning to the Heisenberg relation, it is formulated in the literature of the Fourier transforms [see e.g. PAPOULIS 1962] in the following way: if γ is the Fourier-transform of φ, and two density functions are defined as

$$f(x)=\varphi^2(x)\Big/\int_{-\infty}^{\infty}\varphi^2(x)\mathrm{d}x \quad \text{and} \quad g(y)=\gamma^2(x)\Big/\int_{-\infty}^{\infty}\gamma^2(x)\mathrm{d}x \tag{7}$$

(g can be called the 'Heisenberg counterpart' of f), then the product of the scatters (standard deviations) cannot be less than 1/2:

$$\sigma(f)\cdot\sigma(g)\geq\frac{1}{2} \tag{8}$$

It is well known [see e.g. HUBER 1981] that σ is at the same time the asymptotic scatter (A) of the algorithm based on the minimization of the L_2-norm and therefore relation (8) can also be written as

$$A(f)\cdot A(g)\geq\frac{1}{2} \qquad \text{for the } L_2\text{-norm.} \tag{8a}$$

Let us show an example, namely the supermodel $f_a(x)$ (see e.g. in STEINER [1991] the first column of the table at the end of the book):

$$f_a(x)=\frac{\Gamma\left(\frac{a}{2}\right)}{\sqrt{\pi}\cdot\Gamma\left(\frac{a-1}{2}\right)}\cdot\frac{1}{\left(1+x^2\right)^{a/2}} \qquad (a>1). \tag{9}$$

(Incidentally, it is easily seen with this analytically simple density function what an enormous difference exists between the formulae given in Eqs. (5) and (6): if the sum expressions are changed to integral ones and $f_a(x)$ figures in these integrals, Eq. (6) yields $S\to\infty$ if $a\to1$; on the contrary, Eq. (5) yields $S\to2.592$, if $a\to1$, see HAJAGOS and STEINER [1993b].)

The Heisenberg counterpart of $f_a(x)$ is [see STEINER 1991, p. 281]:

$$g_a(y)=\frac{2}{\sqrt{\pi}}\frac{\Gamma\left(\frac{a}{2}\right)}{\Gamma^2\left(\frac{a}{4}\right)\cdot\Gamma\left(\frac{a-1}{2}\right)}\left(\frac{|g|}{2}\right)^{\frac{a-2}{2}}\cdot\left[K_{\frac{a-2}{4}}(|g|)\right]^2 \qquad (a>1). \tag{10}$$

The latter formula is a little bit more difficult (the modified Bessel function K figures in it) but the distributions $g_a(y)$ are symmetric and unimodal (like the $f_a(x)$ distributions). The density curve of $g_a(y)$ e.g. to $a=8$ (*Fig. 2*) proves that for modelling actual cases the types of the supermodel $g_a(y)$ could also be appropriately used.

As was mentioned earlier the scatters in the Heisenberg relation (characterizing primarily the mother distributions themselves) have also the

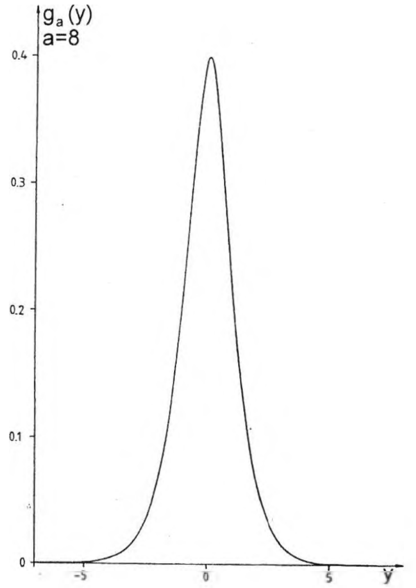

Fig. 2. Probability density function $g_a(y)$ for $a=8$ (see Eq.10); this is the 'Heisenberg counterpart' of the density function $f_a(x)$ for $a=8$

2. ábra. A $g_a(y)$-sűrűségfüggvény $a=8$ esetén (ld. a (10) egyenletet); ez az $a=8$-hoz tartozó $f_a(x)$ valószínűség-sűrűségfüggvény 'Heisenberg-pandantja'

meaning of asymptotic scatters for the case if the L_2-norm is used for determining the parameter of location, e.g. the symmetry point. Consequently, the Heisenberg relation can be written, for the L_2-norm as $A(f_a)\cdot A(g_a)\geq 1/2$ (where equality holds if and only if both f_a and g_a are Gaussian).

And what about the behaviour of the products of the asymptotic scatters of Heisenberg pairs of density functions if the statistical algorithm is defined by minimization of the P_J, P, P_C and P_{lt} norms? Astonishingly enough, they have quite opposite behaviour to that shown by the L_2-norm. The curves of the products $A(f_a)\cdot A(g_a)$ are demonstrated in *Fig. 3: 1/2 is no longer the mininum value for the simultaneously reachable accuracy in the two domains of the Fourier-transformation.* This should also be stressed as a new result [see STEINER and HAJAGOS 1995].

Figure 3 shows the products of asymptotic scatters versus $t=1/(a-1)$ as type-parameter. For comparison in all four cases the same increasing curve concerning the L_2-norm is also shown; if $t=1/(a-1)$ is equal or greater than 0.5 this error-product for L_2 is infinite.

The descending curves of the new statistical norms start for P_J and P at a value which is not significantly greater than 1/2. The curve for P_{lt} starts at about 1, i.e., twice greater value of this mystical 1/2. No wonder, then, that P_{lt} is defined for error distribution with extremely long tails. It should be

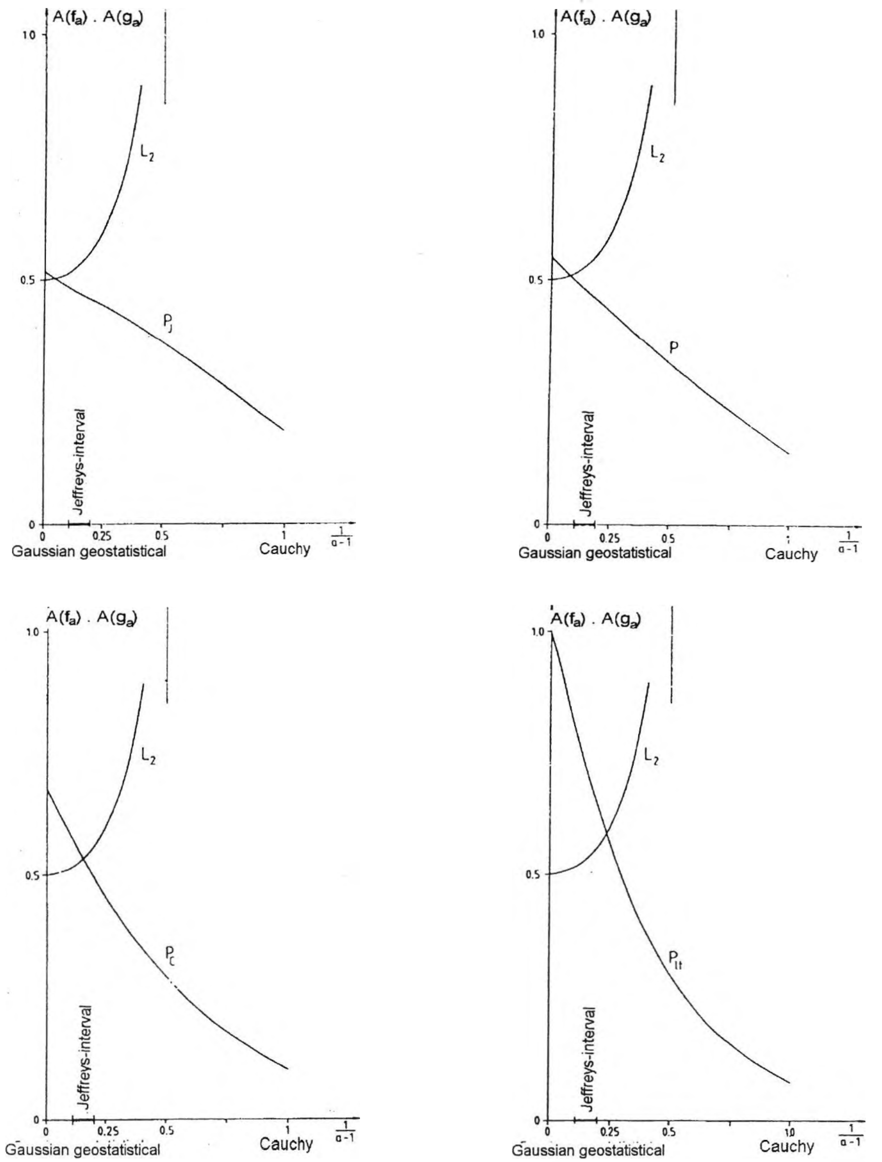

Fig. 3. Descending curves (versus the type parameter $(a\text{-}1)^{-1}$) of the product of asymptotic scatters for 'Heisenberg counterparts' $[A(f_a){\cdot}A(g_a)]$ for the statistical norms P_J, P, P_C and P_{It}. In all four cases the corresponding $A(f_a){\cdot}A(g_a)$ product-curve for L_2 is also demonstrated, for easier comparison [from STEINER and HAJAGOS 1995]

3. ábra. A $t=1/(a\text{-}1)$ típusparaméter növekedésével a 'Heisenberg-pandantok' aszimptotikus szórásainak $A(f_a).A(g_a)$ szorzatára monoton csökkenő görbék adódnak a P_J, P, P_C és P_{It} normák eseteiben. Mind a négy esetben feltüntettük az L_2 normára vonatkozó, meredeken emelkedő $A(f_a).A(g_a)$ szorzatgörbét is mentioned, however, that even in this case the product of error-characteristics for $a=5$ is only by 12% greater than 1/2; in cases of P_J, P and P_C is for $t=1/(a-1)=0.25$ i.e., for the so-called geostatistical distribution *the product is significantly smaller than 1/2.*

It is perhaps useful to tabulate this opposite behaviour (see *Table III*).

L_2 -norm:	$A(f_a)\cdot A(g_a)\geq\frac{1}{2}$,	
The *opposite relations* for the *P*-norms at the same limit above are as follows:		
P_J -norm:	$A(f_a)\cdot A(g_a)<\frac{1}{2}$,	if $a<20$ ($f_a(x)$ for $a=20$ is very near the Gaussian type).
P -norm:	$A(f_a)\cdot A(g_a)<\frac{1}{2}$,	if $a<10$ (this means that the opposite relation is valid already for the whole Jeffreys-interval of types).
P_C -norm:	$A(f_a)\cdot A(g_a)<\frac{1}{2}$,	if $a<6$ (i.e., the opposite relation is already valid in the neighbourhood of the so-called geostatistical distribution).
P_{lt} -norm:	$A(f_a)\cdot A(g_a)<\frac{1}{2}$,	if $a<4$ (this limit is yet in the domain of distributions of finite variance).

Table III. The product of asymptotic scatters belonging to f_a and to its Heisenberg-counterpart g_a behaves in cases of modern norms inversely to the classical case of the L_2-norm

III. táblázat. Az f_a-hoz és a g_a-val jelölt Heisenberg-megfelelőjéhez tartozó aszimptótikus szórások szorzata a modern normák esetében az L_2 klasszikus esetéhez viszonyítva ellentetten viselkedik

This advantageous behaviour is partly due to the fact that in the new norms the parameter of scale is a constant times the dihesion ε: P_J works with $S=3\varepsilon$, P with $S=2\varepsilon$, P_C with the dihesion itself, and P_{lt} with half of it. Let us recall that ε is defined by Eq. (5), i.e., due to Eq. (3) which is based on a train of thought concerning the *I*-divergence.

3. Significance of the right choice of *S* (parameter of scale) in respect of the behaviour of *T*-determinations (the location parameter is denoted by *T*)

Well, we have returned to the significance of the right choice of the parameter of scale *S* in respect of the behaviour of the determination of *T*. In *Fig. 4* two curves of statistical efficiencies are presented for the distribution type-interval $0 \leq t \leq 2$. For $t = 1/(a-1) = 1$ i.e., for the Cauchy type itself, it is evidently indifferent whether Eq. (5) or Eq. (6) is used for the *S*-determination: in both cases maximum efficiency is achieved. With regard to other types, however, significant differences occur between the efficiencies depending upon whether Eq. (5) ('MFV'-curve, from most frequent value), or Eq. (6) (CML-curve, Cauchy maximum likelihood) is used to calculate the parameter of scale.

Figure 4 is ideal for showing what *robustness* really means. This notion does *not* mean outlier-insensitivity: this latter — also very important — behaviour is called *resistance*. The very meaning of robustness is the

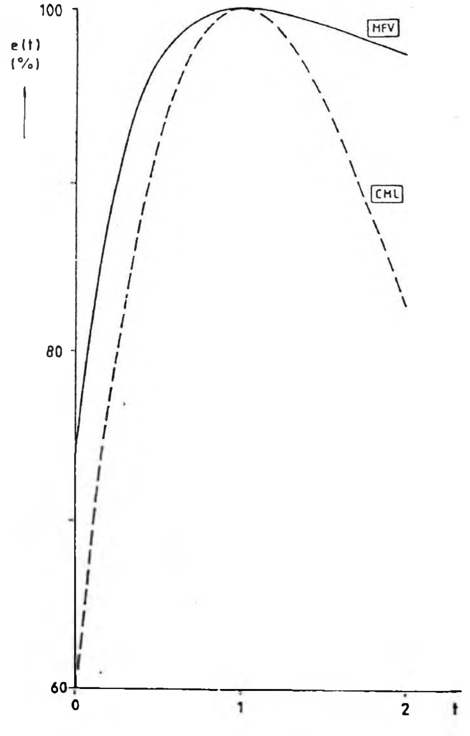

Fig. 4. Significantly different efficiency curves for determining the location parameter [from HAJAGOS and STEINER 1993b], although 'only' the determination method of the parameter of scale is different for the 'MFV' and 'CML' methods

4. ábra. A helyparaméter meghatározására vonatkozó hatásfok-görbék [átvéve HAJAGOS et al. 1993b-ből]; a görbék szignifikáns mértékben különböznek egymástól, noha a helyparamétermeghatározások 'MFV'- és 'CML'-módszerei 'csak' a skálaparaméter számítási módszerében különböznek following: starting from the maximum efficiency place, with increasing type-distance the efficiency slowly decreases. (The correct definition of the type-distance will be given later in Eq. (11); by good fortune it is approximately constant times of the correctly calculated distances in the Gaussian-Cauchy type-interval, — see also Fig. 5 in this article: the proportionality is almost perfect in the broad neighbourhood of the geostatistical type.) It can be easily concluded on the basis of Fig. 4 that the MFV method is much more robust than the CML method though the ψ-function (see HUBER 1981; nowadays also known as 'influence function') is *just the same* in both methods. Consequently, it is far from satisfactory to give the ψ-function only: we have also to decide which formula must be used to determine S on the basis of the data. Surely, it is possible to give an S-value at random — and not on the basis of data — calling this parameter of scale, for example, an 'empirically proven' value or any arbitrary designation; practices of such kind, however, are out of our present scope of interest. (The ψ-function has a well known fundamental role in the theory of Huber as the function for estimating T on the basis of the measured x_i data. The ψ-functions can be constructed from straight lines, too, if no deeper theoretical background is given; a classical example of this is the ψ-function of the Huber-estimate, see HUBER [1964].)

4. Quantitative characterization of the robustnesses of different norms

The two efficiency-curves in Fig. 4 [cited from HAJAGOS and STEINER 1993b] are excellent for the comparison of robustnesses and the conclusion in that special case obviously is that the MFV method is more robust than the CML estimation. This conclusion is, however, only of qualitative nature. We therefore found it necessary to characterize this property by a numerical value. The formula for this value (as a rate of the robustivity denoted by r) is as follows:

$$r = \int_0^\infty e(t) \cdot q(t) \mathrm{d}t \tag{8}$$

where $e(t)$ is the statistical efficiency of the statistical method in question for the probability distribution type characterized by the type parameter t, and $q(t)$

means the probability density of occurrence of type t in the discipline studied [see STEINER and HAJAGOS 1993].

But which $q(t)$ density function can be regarded as adequate for the geosciences? The old statistical literature (not only in the geosciences) often claims that in the overwhelming majority of cases the error-distributions are Gaussian, then called this type 'normal'. (This would mean that $q(t)$ is a Dirac-δ at the point $t=0$.) This dogma was accepted not only because the use of the simplest statistical methods should be justified but also the statistical tests for 'normality' (applying the proposed significance levels) can be misleading. Details about this fact are given in SZŰCS [1993]. For example, the χ^2-test can 'prove' the normality even for data coming from a random variable of a quite different distribution type, if the sample is not sufficiently large.

Only sporadically can reliable information be found about the real type of data although even at the end of the last century NEWCOMB [1886] did not use the L_2-norm for his astronomical data as it was important to exhaust all information contained in the measured values. Briefly speaking, he did not accept the dogma of 'normality'. On the contrary, he concluded that the Gaussian error distribution occurs very rarely.

NEWCOMB is one of the very few scientists who can be regarded as competent in the problem of the really occurring error-types. But also generally: if experts work *only* in the field of geosciences, or *only* in the discipline of mathematical statistics, their opinion about error-types must not be taken into consideration. In this question only such scientists are competent who work deeply enough in *both* disciplines.

As the best example Sir Harold JEFFREYS, the great geophysicist should be cited, who exhausted the information optimally from the seismological data with adequate iteration algorithms, although it was extremely tedious to do this in the thirties of this century: the execution of a single iteration step needed some hours. The book JEFFREYS [1961] written on the theme of probability proves that he was a deep thinking scientist in both disciplines. As for the error-distributions, he concluded that in practice the flanks are not in such a degree short as by the Gaussian, in the best cases $f_a(x)$-like tails can occur for the type-parameter interval $0.1 \le t \le 0.2$, or otherwise written for $6 < a < 10$. This interval is marked in the abscissae of Fig. 3 as 'Jeffreys-interval'.

Rather than listing and discussing all the citations, we opted to accept the following expression as $q(t)$-formula for the geosciences:

$$q(t) = 16t \cdot e^{-4t} \quad . \tag{9}$$

The q density has its maximum at $t=1/4$, i.e., at $a=5$; this corresponds exactly to a statement of DUTTER [1986/87] about the most commonly occurring type of error in geostatistics. Therefore the type belonging to $t=0.25$ is called 'geostatistical'. As is well known we sometimes have to work with Cauchy-distributed data, too. According to Eq. (9) the probability density $q(1)$ of the Cauchy occurrence is really by far not vanishing: it amounts to 20% of the maximum density ($q_{max}=q(0.25)=5.q(1)$). It should be mentioned, too, that the probability densities of the Jeffreys-interval range from $0.77 \cdot q_{max}$ to $0.98 \cdot q_{max}$ and in the very neighbourhood of the Gaussian, namely for $f_a(x)$ $a=40$ the probability density amounts $0.25 \cdot q_{max}$ ($q=0$ holds only for just the Gaussian as modern authors state that this type never occurs as a mother-distribution in practice see MOSTELLER and TUKEY [1977]). Note also that the t type-distance from the Gaussian of the distribution for $a=40$ is only 10% compared with the t type-distance between the geostatistical and the Gaussian types.

Accepting the $q(t)$-expression given in Eq. (9) for the geosciences, for the rate of robustness the values given in per cent in *Table IV* can be calculated according to Eq. (8) [see STEINER and HAJAGOS 1993].

norm	rate of robustness for geosciences
L_2	36%
L_1	80%
P_J	90%
P	96%
P_C	94%
P_{lt}	75%

Table IV. Quantitative characterization of the robustnesses of different norms
IV. táblázat. Különböző normák robusztusságának kvantitatív jellemzése

The norms for the MFV-algorithms show the greatest values (except P_{lt} which prefers the very long tailed distributions and not the types characterizing the geosciences). It should be mentioned that r is not zero even for L_2 although the statement of the modern statistical literature formulates categorically as follows: 'L_2 algorithms are *not robust*'. This is not really astonishing: Table IV (and Eqs. (8) and (9)) gives *new results* that characterize the robustness quantitatively. Consequently, our formulations must be more accurate regarding these new results also in other respects. An example: one decade ago the notations 'robustness' and 'resistance' could be used (with a little opportunism) synonymously, but this is no longer allowed — if confusions are to be excluded. We should avoid (or even forget) the wide-spread practice that *resistance* is proved by one or more example and the consequence is formulated as 'the statistical method is *robust*'. (There are examples showing that decreasing robustness occurs simultaneously with increasing resistance, see HAJAGOS and STEINER [1993b].)

The notion 'type-distance' was mentioned earlier for the quantity $t=1/(a\text{-}1)$ as Eq. (9), i.e., the $q(t)$ function was discussed. This intuitively introduced quantity has also been investigated by us recently, giving an exact definition for type-distances based upon the well known Kolmogorov distance (K) defined for two arbitrary distributions:

$$K = \max_x \left| F(T_F, S_F; x) - H(T_H, S_H; x) \right|. \tag{10}$$

F and H mean probability distribution functions, the T and S values are parameters of location and that of scale of the distributions in question. (Writing 'sup' instead of 'max' the definition is more correct from the point of view of mathematics.)

The Kolmogorov-distance (the K-value) is obviously strongly influenced by the actual values of T_F, S_F, T_H and S_H. If we are interested primarily on the part of the distance which is caused by the difference of the types, we have to eliminate the effects of these four parameters. Consequently, type-distance between F and H (denoted by $D(F,H)$) can be defined (and calculated) in the following way [see HAJAGOS and STEINER 1994]

$$D(F,H) = \min_{T_F, S_F, T_H, S_H} \left\{ \max_x \left| F(T_F, S_F; x) - H(T_H, S_H; x) \right| \right\} \tag{11}$$

(mathematically it is more correct to write 'inf' instead of 'min').

It is generally proven that such 'min-max'definitions fulfil the requirements given for distances (this means, $D(F,F)=0$, $D(F,H)=D(H,F)$ and the triangle-relation is also fulfilled: $D(H_1, H_2) + D(H_2, H_3) \geq D(H_1, H_3)$, see CSERNYÁK [1995]).

The minimization which is to be done actually in our case, is much simpler as is seen in the generally valid Eq. (11). First of all, we compute here only distances of F_a-distributions from the Gaussian; the distribution function of the latter should be denoted for T=0 and S=1 by $G(0,1; x)$ (evidently this is the 'standard version' of the so-called 'normal' distribution tabulated in nearly every book written on the subject of probability and mathematical statistics). If T=0 is also substituted in F_a, the minimization according to T is already made (functions F_a and G both being symmetrical to T and unimodal). As for the S-values: they have during the minimization no separate role, only the ratio of both S-values influences the D-value. Consequently, the type-difference of any F_a from the Gaussian can be calculated simply as

$$D(F_a,G) = \min_{S}\left\{\max_{x}\left|F_a(0,S;x) - G(0,1;x)\right|\right\}. \tag{12}$$

This exact type-difference is still about the Cauchy-type (astonishingly enough) approximately proportional to the intuitively introduced $t=1/(a\text{-}1)$, (see *Fig. 5*). It would therefore be superfluous to make 'more correct' the abscissae of all figures in the MFV-literature which are given in the overwhelming majority of cases for the Gaussian-Cauchy interval. Even the modification of the definition of the robustness for the geosciences (r) in

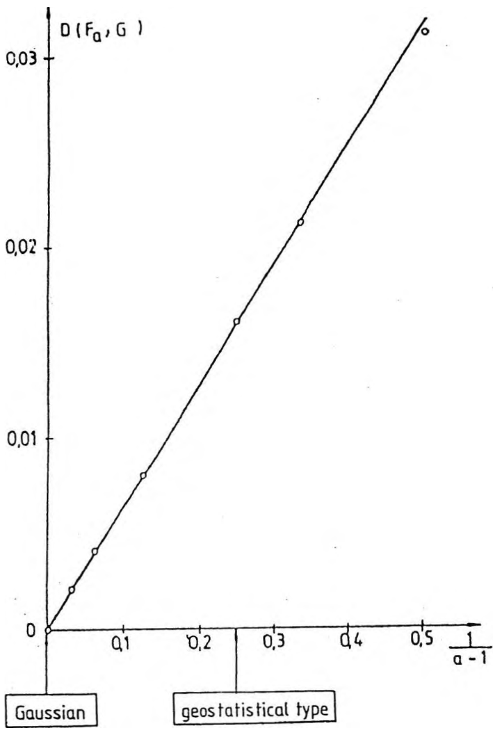

Fig. 5. Type-distance $D(F_a, G)$ (i.e., the distance of type F_a from the Gaussian) versus $t=1/(a\text{-}1)$, showing approximate proportionality between the two quantities in the interval demonstrated [from HAJAGOS and STEINER 1994]

5. ábra. Az F_a típusnak a Gauss-félétől mért $D(F_a, G)$ távolsága az ábrázolt intervallumban közelítő arányosságot mutat a $t=1/(a\text{-}1)$ típusparaméterrel [HAJAGOS et al. 1994 nyomán]

Eq. (8) is not necessary; no significant diminishing is expected if D were to figure in Eq. (8) instead of t.

But what about the much greater D-values, i.e., greater than 0.058? Which value is the Cauchy-Gaussian distance? According to Eq. (8) such types occur very rarely in geophysics and geology practice. We can, however (and, perhaps, we are even obliged) to deal with such questions in a much more general sense, including other geosciences, too, e.g. meteorology, perhaps also astronomy, etc. And finally: exclusion of other disciplines would be unjustifiable if the behaviour of statistical methods in general is to be investigated.

It can be proven [see CSERNYÁK 1995] that $D(F_a, G)$ tends to 0.25 if $a \to 1$ and this value is the maximum possible type-distance among symmetric distributions. This means that the overall behaviour of the statistical methods based on different norms can be easily visualized for the whole supermodel $f_a(x)$ and for $0 \leq D(F_a, G) \leq 0.25$; this is a complete characterization in the sense that characteristic values are shown for all possible type-distances.

In *Fig. 6* the efficiency-curves are shown for the norms listed in Table II versus the type-distance D from the Gaussian from STEINER et al. [1995]. The curve for L_2 quickly drops from 100% to zero. The efficiency curve for the L_1-norm tends to zero if D tends to 0.25; not one of the efficiency-curves of the P_k-norms has such a behaviour. The L_1-curve starts at the value 63.66% ($e=200/\pi$% being valid for the Gaussian distribution) and reaches a maximum over 80% near to the Cauchy-type; for the latter $e=800/\pi^2=81$% holds.

For the whole supermodel $f_a(x)$ (except the close neighbourhood of the Gaussian type) the L_1-curve shows very great advantages over the conventional statistics characterized by the L_2-curve. Fig. 6 shows, however, that the norms P_{Γ}, P-, P_C and P_{lt} have significant advantages even compared with the efficiency curve of the L_1-method. With a slight overstatement this property of these four P_k-norms can be called 'overall robustness'. Note, for example, that the P_C-curve shows for this *as long as theoretically possible type-interval* an e-value never less than 74%.

The 'overall robustness' (OR) for the $f_a(x)$ supermodel, however, can also be characterized quantitatively supposing that all distances $D(F_a, G)$ occur with the same probability density, i.e., q=4 holds for the complete type domain, neglecting special aspects of various disciplines. In this case the definition of the 'overall robustness' is:

$$OR = 4 \cdot \int_0^{0.25} e(D)\mathrm{d}D \qquad (13)$$

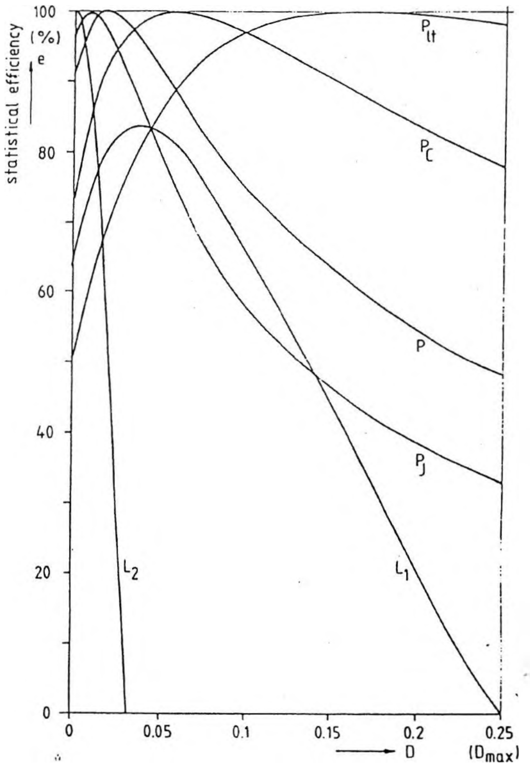

Fig. 6. Efficiency curves for norms listed in Table II versus type-distance $D=D(F_a, G)$ [from STEINER et al. 1995]. The demonstrated D interval is as long as theoretically possible for symmetrical error distribution types

6. ábra. A II. táblázatban felsorolt normák hatásfok-görbéi a Gauss-típustól mért $D=D(F_a, G)$ típustávolság függvényében [STEINER et al. 1995 nyomán]. Az ábrázolt D-intervallum szimmetrikus esetekre vonatkozóan az elmélet szerint a lehető leghosszabb

[see STEINER et al. 1995]; $e \equiv 100\%$ would obviously result in OR = 100%. — For the six norms in Table II, the OR-values are given in *Table V.*

L_p	L_2: 7.8% L_1: 50.1%
norms of the MFV-procedures	P_J: 59.3% P : 71.5% P_C: 90.7% P_{lt}: 92.6%

Table. V. 'Overall robustnesses' for different norms

V. táblázat. Különböző normák általános robusztussága

The conclusions can be formulated as follows:

1. The overall robustness of L_2 can really be regarded as nearly vanishing;
2. The overall robustnesses of the norms resulting in MFV procedures are greater (or even significantly greater) than the overall robustness of the L_1-norm.

Also in respect of the overall robustness of the standard version P has a significantly greater OR-value compared with that of L_1:

$$OR(P)\text{-}OR(L_1) = 21.4\ \%.$$

It is perhaps more interesting that

$$OR(L_1)\text{-}OR(L_2) = 42.3\ \%$$

and

$$OR(P_{lt})\text{-}OR(L_1) = 42.5\ \% \qquad (14)$$

hold and this nearly full coincidence of the differences means not less than (concerning overall robustness) the choice of P_{lt} instead of L_1 has the same advantage that we could have reached using L_1 instead of the conventional L_2-norm.

5. 'Philosophies' in statistics

The rates of robustness (the values of r and OR) are very helpful in choosing the appropriate statistical norm for a given task, but this is far from being satisfactory. It must also be taken into consideration which 'philosophies' are behind the formulae. Finally, let me show a table (*Table VI*) for orientation (from CSERNYÁK et al. [1995]).

All three philosophies can be applied appropriately in different disciplines. For example, in the case of the geosciences (where outliers often occur and modelling can rarely be exact), the statistical method must be sufficiently resistant — and this is warranted in the philosophy of the MFV-procedures. Accepting e.g. the P-norm and its philosophy we have to decide additionally whether or not neglecting more than 50% of the data is acceptable from the point of view of a well defined task to be solved in a framework of a given discipline.

MULTIPARAMETER REGRESSION, GEOPHYSICAL INVERSION	
Philosophy	Theoretically best elaborated and experimentally proved alternative is the minimization of the given norm of the residuals
Classical philosophy: The greatest values of the squared residuals should not be too large, **even if the values characterizing the concentration of the residuals differ significantly from zero**.	L_2-norm
Median-type philosophy: Positive and negative residuals must be in equilibrium with regard to their absolute values, **no group of the data should be neglected.**	L_1-norm
MFV-philosophy: Residuals must concentrate possibly close around zero, even if some data (or occasionally even a significant part of the data) are practically neglected.	P_J-norm, P-norm, P_C-norm, P_{Ir}-norm

Table VI. Different philosophies in statistics and the corresponding norms [CSERNYÁK et al. 1995]

VI. táblázat. Különböző statisztikai filozófiák és az azoknak megfelelő normák [CSERNYÁK et al. 1995]

Acknowledgements

Some of the new results presented were obtained by research work supported by the Hungarian Science Foundation. This whole project was realized in cooperation with researchers of the Institute of Geodesy and Geophysics in Sopron. On behalf of the whole team working in Miskolc on this topic, I should like to express our thanks not only for fruitful consultations and discussions with the researchers of this institute (in the first place with the leader of this project J. SOMOGYI and with J. VERŐ), but also for the

computer-technical support: without the latter some of the new results could not have been presented in the present paper.

REFERENCES

ANDREWS D. F., BICKEL P. J., HAMPEL F. R., HUBER P. J., ROGERS W., TUKEY J. W. 1972: Robust Estimation of Location. Princeton University Press, Princeton. 373 p.

BARTA Gy., HAJÓSY A. 1985: New procedure for determination of the gravity constant. Ann. Univ. Sci. Budapestiensis de Rolando Eötvös Nom. Sectio Geoph. et Meteorologica **1–2**, pp. 5–12

CSERNYÁK L. 1994: A thesis on the difference of the determination methods of the scale parameter (CML≠MFV). Acta Geod. Geoph. Hung. **29,** (1–2), pp. 255–257

CSERNYÁK L. 1995: Distance of probability distribution types. (Study of the distance of the normal and the generalized Student-distributions.) Acta Geod., Geoph. Hung. **30**, pp. 281–284

CSERNYÁK L., HAJAGOS B., STEINER F. 1995: The philosophy of the most frequent value procedures. Acta Geod., Geoph. Hung. **30**, pp. 265–280

DUTTER R. 1986/1987: Mathematische Methoden in der Montangeologie. Vorlesungs-notizen. Leoben 22 p.

HAJAGOS B. 1991: The most frequent value as estimation with minimum information loss. *In* STEINER (ed.) 1991: as Chapter X Eq.(10-11).

HAJAGOS B., STEINER F. 1993a: Determination of fault's position in mines using different norms in inversion. Annales Univ. Sci. Budapestiensis de Rolando Eötvös Nom. Sectio Geophysica et Meteorologica **9**, pp. 101–146

HAJAGOS B., STEINER F. 1993b: Investigations concerning resistance — importance of the choice of the formula determining the scale parameter. Geophysical Transactions **38,** (4), pp. 211–230

HAJAGOS B., STEINER F. 1994: Definition der Entfernung von Wahrscheinlichkeitsverteilungstypen u.s.w. Publ. Univ. of Miskolc, Series D, Natural Sciences **35,** 4, pp. 97–108

HUBER P. J. 1981: Robust Statistics. Wiley, New York, 308 p.

HUBER P. J. 1964: Robust estimation of a location parameter. Ann. Math. Statist. **35,** pp. 73–101

JEFFREYS H. 1961: Theory of Probability. Clarendon Press, Oxford

MOSTELLER F., TUKEY J.W. 1977: Data Analysis and Regression. Addison-Verlag, Reading, Mass

NEWCOMB S. 1886: A generalized theory of the combination of observations so as to obtain the best result. American Journal of Mathematics **8,** pp. 343–366

PAPOULIS A. 1962: The Fourier Integral and its Applications. McGraw-Hill, New York

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

STEINER F., HAJAGOS B. 1993: Practical definition of robustness. Geophysical Transactions **38,** 4, pp. 193–210

STEINER F., HAJAGOS B. 1995: Opposite behaviour of the P-norms compared to that of the L_2-norm with respect to the simultaneously achieved accuracy in the space- and frequency-domain. Acta Geod. Geoph. Hung. **30**, pp. 293–299

STEINER F., HAJAGOS B., PÁSZTOR P. 1995: Curves of statistical efficiency versus type distance of generalized Student-types from the Gaussian. 'Overall robustness' of the P-norms. Acta Geod., Geoph. Hung. **30,** pp. 285–291

SZŰCS P. 1993: Comment to an old dogma: 'The data are normally distributed.' Geophysical Transactions **38,** 4, pp. 231–238

A leggyakoribb érték (MFV) eljárások elméletének új eredményei

Ferenc STEINER

Az MFV-filozófia szerinti elmélet, azaz a leggyakoribb érték-eljárások elvi problémaköre több új eredménnyel gyarapodott az elmúlt néhány évben. A publikációk különböző szakfolyóiratokban jelentek meg, így indokolt, hogy ebben a dolgozatban az eredmények egy részének tömör összefoglalása történjék meg abból a célból, hogy a maximum likelihood-elven, valamint az L_2-, L_1- és P-normákon alapuló statisztikai eljárások lényegi különbségei minél tisztábban álljanak előttünk.