GEOPHYSICAL TRANSACTIONS

CONTENTS

TARTALOMJEGYZÉK

СОДЕРЖАНИЕ

MOST FREQUENT VALUE PROCEDURES
(A SHORT MONOGRAPH)

Ferenc STEINER*

This monograph deals with modern statistical algorithms derived from the conception of the most frequent value (the latter meaning both its original and its generalized form). Modern methods are necessary as standard statistical procedures based on the least squares principle are not effective enough for the error distributions commonly occurring in the earth sciences; the occurrence of the sterile Gaussian distribution in these disciplines is the exception rather than the rule. The theoretical background is also given as a means of helping in the applications for special purposes, i.e. to further the development of various new algorithms for geophysical interpretation on the grounds of the conception of the most frequent value.

Keywords: robust statistics, efficiency, outlier, resistance, error distribution, model distribution, super-model, most frequent value, dihesion, I-divergence

Introduction

Algorithms for interpreting measured data sets often use certain fitting technique in an explicit or implicit manner; these methods of fitting or adjustment can be deduced, by means of the known notion of 'deviation', from some determination method of the location where the densest values lie.

Let the (ordered) sample be the following: 7, 9, 10, 11, 13, 40 *(Fig. 1)*. Let us try to answer without any preconception the question: which value can be regarded as the most characteristic for the gathering? In other words: if we were
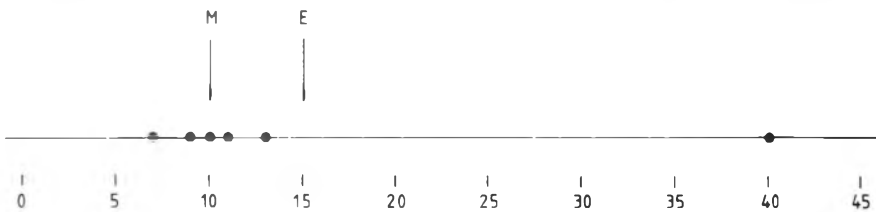
Fig. 1. Arithmetic mean (E) of the sample of six elements does not indicate the most likely value, denoted by M

1. ábra. Az ábrán bemutatott hatelemű minta számtani középértéke (E) nem a leginkább várható értékintervallum közepét jelöli ki. Az utóbbi sokkal inkább elfogadható M-re

Рис. 1. Среднеарифметическое по примеру из шести элементов (E) не попадает в центр наиболее вероятного интервала значений. Последний значительно более приемлем для M.

* Department of Geophysics, Technical University for Heavy Industry, Miskolc-Egyetemváros, H-3515, Hungary

to continue the sampling, which is that value, where, or in what neighbourhood can the next value generally be expected? If we really do not have any preconception, we have nothing to say against the value 10 (marked by *M* on the figure). Visually we are inclined to accept this value, but we need an *algorithm for computation*.

If, on the basis of a sample, we wish to determine the most characteristic value by computation, we customarily calculate the arithmetic mean (or simply 'mean'), without any additional consideration. This value, however, can give a result which flatly contradicts the rational approach: in our example we get the value 15 in this way (see the arrow marked by *E*); it can hardly be stated that the next value is expected with maximum probability in the neighbourhood of the arithmetic mean.

It is clear that the value of the arithmetic mean is strongly influenced by the sample element 40 (it is indifferent if this value is a so-called outlier, or if the occurrence of such values has a small but finite probability in respect of the distribution itself, too). If we are able to determine by an exact algorithm the location of the densest lying points (in our case *M*, which can really be called the most frequent value), this method should fulfil the demand to be insensitive to points lying far away. We are forced now to anticipate notions which will be analysed later in this monograph.

The disadvantageous behaviour of the mean and the proper behaviour of the most frequent value are closely connected with the fact that the arithmetic mean as an estimate of location *is not robust;* on the contrary, the most frequent value is a *robust* estimate for the parameter of location, e.g. for the symmetry point of the distribution (if the distribution is symmetrical). Since robustness is a very important notion, we are obliged to say more about it even in the introduction.

If one wishes to utilize mathematical results for practical purposes, we face the following structure of the mathematical theorems: 'if..., then...'. Given the fulfilment of the premises on the 'if'-side, we can immediately use the statement of the theorem.

It would be impossible that the infinite number of real situations would all be 'covered' by mathematical premises. It is therefore advisable if small changes in the premises cause small changes of the consequences: and this is the very content of the notion 'robustness'. For a given robust procedure it is sufficient if the premises are only approximately fulfilled. From the practical point of view this is of immense importance as the domain of applicability becomes unusually broad. Precisely because of this it is not astonishing that the number of mathematical articles dealing with robustness increased by more than an order of magnitude in the seventies (*Fig. 2.;* the percentual values were calculated on the ground of numbers obtained from the data bank of the Dialog Information Services, Inc.).

Until the introduction of the notion 'robustness' only a few decades ago, engineers had implicitly supposed that mathematical statements do not change significantly if the premises vary only insignificantly, i.e. they supposed that

robustness was fulfilled. TUKEY, in 1960, was the first to show that robustness may not be fulfilled to a catastrophic extent for theorems and characteristics bound by arithmetic means.
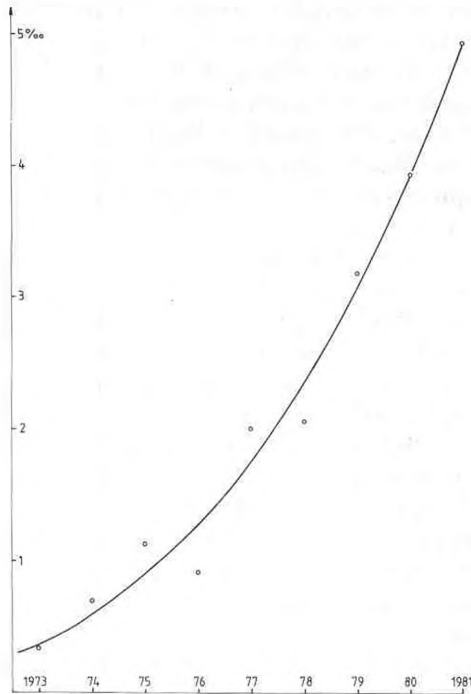
Fig. 2. Rapid increase in the number of mathematical articles treating problems of robustness over a period of about ten years

2. ábra. Alig egy évtized alatt következett be gyors emelkedés a robusztussággal foglalkozó matematikai dolgozatok számában

Рис. 2. Быстрый рост количества математических статей по мощности, происшедний за десятилетие.

We shall see that the computing demand of robust methods is significantly greater than that of the least squares technique (the simplest case of the latter is the calculation of arithmetic means). Two opposite demands are therefore required to be fulfilled: we are obliged to exhaust the maximum information from our data sets, but the costs of the much greater computer time seems to contradict it in the sense that the question arises as to whether the cost/information ratio is really less by using robust methods or not. Fortunately the operation costs continuously decrease in such a drastic manner that in the near future standard methods, i.e. methods based on the least squares principle, will not be able to be applied economically, except in very special cases. Efficiency is of great importance precisely in the earth sciences because data are expensive to obtain; the application of a statistical method, say of 50% efficiency, means nothing other than we have thrown out half of our data.

If it is possible we shall use or apply robust methods. The present monograph shows that robust procedures on the grounds of the most frequent value conception are rather simple and therefore do not need too much computer time. The style of the monograph is similar to that of the great part of robust articles: we must flatly contradict some statements which are still accepted nowadays (and are to be read in handbooks, too). The mental attitude of the reader should always be (in both directions) open but critical and without any preconceptions. Let us recall a remark of Huxley: 'all truths begin as heterodoxy and end as orthodoxy'. Really, some sentences can separately be enhanced also from this monograph in such a manner that orthodox statisticians will indeed find it to be heretical.

The first systematic monograph about robust statistics in general [HUBER 1981] appeared a few years ago but was written for mathematicians and was without concrete algorithms suggested for those applying them. The author of the present monograph is convinced that interpretation methods in the earth sciences (and also elsewhere) must increase effectivity as soon as possible by using robust procedures, and that is the very reason why the method of the most frequent values is the central theme of this work: the advantages of this conception are equally of theoretical, of practical and of a computertechnical nature. The applications of this principle (instead of the least squares one) for the manifold cases of practice may not be the task of only one team consisting of only a few members; here, however, we give all information and foundations for developing special algorithms to solve, economically, interpretation problems of different kinds.

The mathematical background of the reader may be very different: the spectrum begins with those who 'believe' even nowadays in the predominant occurrence of the Gaussian distribution, and reaches as far as to those experts who already know or even apply robust statistical procedures.—Therefore the following three tables may be of interest.

Table I shows among other things that experts working both in practice and in the field of theoretical statistics have never believed in the general occurrence of Gaussian distribution. However, because of the poor possibilities of computing technique, it was for a long time worth while to accept this hypothesis and to develop practical algorithms on the grounds of the least squares principle. *Table II* shows connections, similarities and differences among different statistical procedures and hypotheses, defining the place of the conception of the most frequent value calculations. If the reader has some experience and theoretical knowledge in statistics, the author hopes that *Table II* is fully understandable (although drastic but unavoidable simplifications were made). Should this not be the case the preliminary details contained in *Table II* might serve as sign-posts so the reader does not lose his way in the particulars of this monograph. *Table III* lists the demands of practice which must in any case be fulfilled by every applied statistical procedure. The enhancement of these demands is important inasmuch as a great number of 'home-made', 'own' procedures can easily be defined (defining them either by the substituting distribution g or directly by

Some important dates and periods in the development of statistics

First appearance of the *maximum likelihood principle* in the eighteenth century [D. BERNUOLLI]. This principle was also utilized later by Gauss but the manifold applicability became obvious only one and a half centuries later

Practical application of the *least squares principle* began at the turn of the eighteenth and nineteenth centuries [LEGENDRE, GAUSS]. During one and a half centuries many scientists utilized this principle in various disciplines for solving special interpretation problems

FISHER [1912] revived the use of the maximum likelihood. CRAMÉR and DUGUÉ derived important theorems concerning this principle, e.g. by means of this principle the estimation algorithm of maximum efficiency is given for very general distribution types

Minimization of the loss of information (i.e. the minimization of the *I*-divergence) leads to the best modelling of reality; the practical algorithms partly agree with those algorithms which follow from the maximum likelihood principle [KULLBACK, PEREZ, RÉNYI]

1778   1700   1800   1900   1912   1960   2000

1812   1886   1938   1977

What kinds of probability distributions do occur and do not occur in reality? Only those experts are competent who deal both with practical problems and with theoretical statistics

The line of thought of Gauss (he himself also declared it an inverse one): if the calculation of arithmetic means is the best estimation procedure, then the errors distribute according to the Gaussian density function $f_G(x)$ (Gaussian distribution, so-called 'normal distribution')

NEWCOMB remarked on the Gaussian distribution: 'As a matter of fact, however, the cases are quite exceptional in which the errors are found to really follow the law'

JEFFREYS: carefully, under unchanged conditions measured values are in general at the very most *Student-distributed with 5-9 degrees of freedom* (and have not a Gaussian distribution)

MOSTELLER and TUKEY: 'Some misinterpret the word 'normal' to mean 'the ordinarily occurring' but, so far as we know, distributions that exactly fit this formula never occur in practice...'

GAUSSIAN   GAUSSIAN   GAUSSIAN   THE APPROPRIATE MODEL OF REALITY

If we relate the notion of *optimum effectiveness* to the costs of the whole complex of the measuring and computing work, the assumption of the Gaussian error distribution in general cases was indeed justified for more than a century. *The change of the optimum assumption in time*:

In most cases *the assumption of Gaussian distribution* (the principle of least squares) *gives optimum effectiveness*; nevertheless, for the most frequently occurring distributions the least squares technique has, in the statistical sense, often poor efficiency

Generally, optimum effectiveness can be achieved by algorithms fitting the real distribution; *the assumption of Gaussian error distribution increasingly results in an unacceptably low effectiveness*

New points of view, new notions and new approaches in statistics:

circa from 1960

Development of robust and resistant methods, beginning of the elimination of harmful dogmas [TUKEY, HUBER]

Table I. Block diagram showing the most important statistical principles, statements concerning the relation between real probability distributions and the Gaussian one, and illustrating the changes in the most economically usable model of probability distributions

I. táblázat. Időbeázlat a legfontosabb statisztikai elvekre, a gyakorlatban előforduló valószínűségeloszlások ill. a Gauss-eloszlás viszonyát illető véleményekre és a leghatékonyabban alkalmazható modelleloszlásban nemrégiben bekövetkezett változásokra vonatkozóan

Таблица I. Хронологическая схема главнейших принципов статистики, высказываний относительно соотношения встречающихся на практике вероятностных распределений с гауссовским распределением, а также изменений в моделях распределения, наиболее эффективных происшедших в недавнем прошлом.

Principle:

Basic idea:

Maximum likelihood
The type of actual probability distribution *is known* (the analytical form of the density function $f(x)$ is given); based on the sample $x_1, ..., x_i, ..., x_n$ (measured values) that T is accepted as the appropriate value which results in *maximal probability of the sample*

Minimization of the *I*-divergence
The unknown probability distribution (characterized by the density function $f(x)$) is *substituted* (modelled) by the distribution $g(x)$ with given analytical form; the loss of information is measured by the so-called *I*-divergence, and that T is accepted which *minimizes the loss of information*

The appropriate value according to the basic idea is that T value which satisfies the corresponding equation:

| (some known logical steps)

$$\sum_{i=1}^{n} \frac{\dfrac{\partial f(x_i; T)}{\partial T}}{f(x_i; T)} = 0$$

(differentiation and application to samples)

$$\sum_{i=1}^{n} \frac{\dfrac{\partial g(x_i; T)}{\partial T}}{g(x_i; T)} = 0$$

If the type of actual distribution and that of the substituting distribution are identical (that is, $g = f$), both principles prescribe *the same procedure for the calculation of T*. (The parameter of location T and the parameter of scale S are calculated simultaneously in practice and this may cause differences in the procedures derived from the two principles, and in their practical advantages, too. It is further assumed, for the sake of simplicity, that S is known.) If the distribution of errors is characterized by a bell-shaped curve (the Gauss-curve is only one of the infinite possibilities) then the density function can be written as

$$g\left(\left[\frac{x-T}{S}\right]^2\right)$$

Consequently the equation written above is satisfied by a T value which is the final result of the iterative determination of weighted means calculated with weights given by

$$\varphi(x-T) = g'\left(\left[\frac{x-T}{S}\right]^2\right)\bigg/ g\left(\left[\frac{x-T}{S}\right]^2\right)$$

In practical terms, the principles prescribe weighted mean-calculations in an iterative manner:
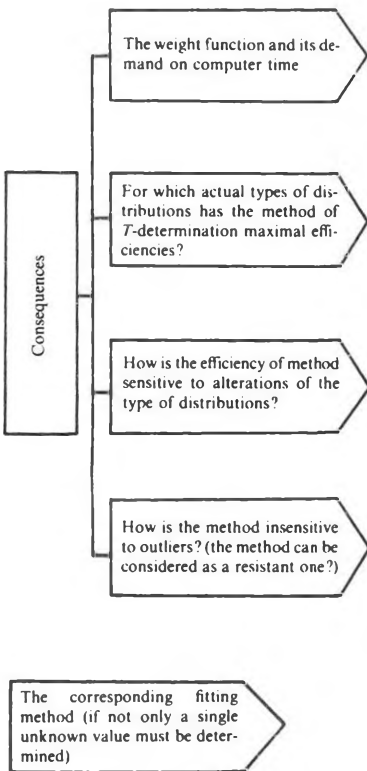
Accordingly,

$$T = \frac{\sum_{i=1}^{n} x_i \cdot \varphi(x_i - T)}{\sum_{i=1}^{n} \varphi(x_i - T)}$$

is the formula which must hold. The demand on computer time depends upon the analytical form of φ; i.e. upon the choice of the model distribution g. For simplicity, we assume that $S = 1$.

Alternatives for the choice of the model distribution g (with $T = 0$ if g is analytically given):

	Arbitrary	$f_a(x) = \dfrac{1}{c(a) \cdot (1 + x^2)^{a/2}}$	$f_G(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
The weight function and its demand on computer time	φ can only be calculated in a complicated manner with a *considerable demand on computer time*	$\varphi(x_i - T) = \dfrac{1}{1 + (x_i - T)^2}$; the calculation of these weights needs *minimal number of operations* among the real (i.e. not degenerated) alternatives for weights	$\varphi(x_i - T) \equiv 1$; the weighted mean simplifies into an ordinary arithmetic mean, consequently the iteration is also superfluous
For which actual types of distributions has the method of T-determination maximal efficiencies?	(The following questions can be answered only for concrete cases)	For distributions characterized by $f_a(x)$, these distributions (using different values for a) adequately model very different distributions occurring in practice	For only *one single* type of symmetrical probability distribution (namely for the Gaussian distribution characterized by f_G)
How is the efficiency of method sensitive to alterations of the type of distributions?		For a very large interval of a-values the efficiency is *to a great extent insensitive* to the type of distribution (i.e. the method is *robust*)	*The efficiency drops quickly* if the actual distribution differs from the Gaussian distribution (i.e. the method is not robust)
How is the method insensitive to outliers? (the method can be considered as a resistant one?)		For a very large interval of a-values *the procedure is resistant*	*The procedure is not resistant*, the result can be completely distorted in the presence of only few outliers
The corresponding fitting method (if not only a single unknown value must be determined)		*Adjustment according to the most frequent values* (the so-called *M*-fitting). The resulting hypersurface is determined by the points which have a tendency of cohesion (the procedure goes on without any regard to outliers)	*The least squares adjustment* (this is at the same time the limiting case of *M*-fitting if $a \to \infty$). The hypersurface (demonstrating the results) tends to situate itself so that no point should be too far from the surface (even if because of this, the surface removes itself from the densest lying points)

Consequences

Table II. Simplified scheme showing similarities between the maximum likelihood principle and the minimization of the *I*-divergence: both result in reweighted calculations. The correct choice of the weight function (or, rather of the model distribution) leads to robust, resistant and economical statistical procedures

II. táblázat. Egyszerűsített vázlat a legnagyobb valószínűség elve és az *I*-divergencia minimalizálása közötti hasonlóság bemutatására: bizonyos (általában teljesülő) feltételek mellett mindkettő iterative súlyozott számításokra vezet. A súlyfüggvény (ill. elsődlegesen a modelleloszlás) alkalmas megválasztásával robusztus, rezisztens és gazdaságos statisztikai eljárásokat kapunk

Таблица II. Упрощенная схема для иллюстрации сходства между принципом максимальной вероятности и минимализацией *I*-дивергенции: при определенных, обычно выполняющихся условиях оба приводят в расчетам, итеративно взвешенным. Подходящий выбор весовой функции, то-есть в первую очередь модели распределения, приводит к мощным, устойчивым и экономичным статистическим способам.

giving the analytical form of the weight function φ), all being superior in comparison with standard statistical methods in respect, for example, of the insensitivity to outliers, but not fulfilling the given set of demands. The reader will ascertain that algorithms developed on the grounds of the conception of the most frequent values fulfil all the demands of *Table III*.

CRITERIA OF EFFECTIVENESS AND OF THE GENERAL PRACTICAL APPLICABILITY	
PERSPICUITY:	the statistical procedure must also be heuristically understandable;
THEORETICAL BACKGROUND:	the procedure must be derived from the modern principles of statistics;
CORRESPONDING ADJUSTMENT:	the procedure, defined originally as a determination of the parameter of location, must be capable of immediate generalization to an adjustment (in case of more unknowns);
DISTRIBUTION-FAMILY FOR MODELLING:	a set of distributions must be available (that are sufficiently general and easily manageable) for adequate modelling of real distributions;
GREAT EFFICIENCY:	the statistical procedure must be as effective as possible for the distributions of the model-family (mentioned above)
SHORT COMPUTER-TIME:	The procedure must be as simple as possible but it must satisfy the principal demands (e.g. in respect of the efficiency);
ROBUSTNESS:	the efficiency must be sufficiently insensitive to the eventual alteration of the type of distribution;
RESISTANCE:	the procedure must be insensitive to outliers since they may occur, and for the determination of many parameters (i.e. in cases of adjustments) the conventional visual elimination of outliers is totally impossible (not only subjective and tedious).

Table III. Eight demands to be fulfilled simultaneously for estimation procedures applied in the earth sciences

III. táblázat. Nyolc követelmény, amelyeknek egyidejüleg kell teljesülniük a földtudományokban alkalmazott becslési eljárásokra

Таблица III. Восемь требований к способам оценки, применяемым в геологических науках.

Chapter 1
The notion of 'robustness' in mathematics and in practice

1.1 Probability distributions, parameters of location, estimates, distribution of estimates, the law of large numbers, limit distributions

The primary purpose of this monograph is to define algorithms for most frequent value calculations and to show how to apply this statistical conception. We shall refer to a number of notions and theorems of probability theory and statistics, without deriving or even accurately defining them. If required, all relevant notions not defined in detail here are to be found in CRAMÉR [1958], Box et al. [1978] and RÉNYI [1962].

However, we shall often comment on the known results in such a manner that the matter is put in another light: namely, it is impossible to speak about robustness without some criticism of hardened opinions. (The style of articles about robustness is therefore very often significantly different from that of mathematics in general.) The secondary purpose of this monograph is to help the reader—if necessary—to remove any preconceptions and even dogmata which are wide-spread even nowadays about probability theory and statistics.

It is therefore not advisable to recall the well known notions solely by giving a set of defining formulae, so we will make as if we would start from the very beginnings.

1.1.1 The 'expected value', as a parameter of location and its estimation

If we have to measure some quantity, by repeated measuring (n times) we will in general get differing values denoted by $x_1, x_2, ..., x_i, ..., x_n$. What can be regarded as the characteristic, or even 'true' value of the measured quantity in question? (True values can really exist, e.g. if the sum of angles in a triangle are geodetically measured; in other cases, however, the notion of true value can be without any meaning, e.g. if we measure the weight of grains of wheat. Also in the latter case, however, the weight can be regarded as a true value but in a different manner modified by unavoidable effects [BAULE 1963].) Which single value should be accepted as the most characteristic for the location of the data $x_1, x_2, ..., x_n$ demonstrated on a datum line?

The first proposition of BAULE [1963] is to accept that value which is characteristic for the gathered points ('häufigster Wert', i.e. most frequent value). If the mother distribution is symmetrical and unimodal (i.e. the density function has only one maximum), we can perhaps accept the arithmetic mean

$$E_n = \frac{1}{n} \sum_{i=1}^{n} x_i, \tag{1}$$

as an estimate of location, where

$$E = \int_{-\infty}^{\infty} x f(x) \, dx \tag{2}$$

is a possible definition of the parameter of location, and this gives (if the integral in Eq. (2) exists) the symmetry point. *Figure 3* shows that in the cases defined above we get estimates calculating E_n for such a parameter of location E which has for the interval $(E - \varepsilon, E + \varepsilon)$ *maximum probability* expressed by

$$\int_{E-\varepsilon}^{E+\varepsilon} f(x) \, dx \tag{3}$$

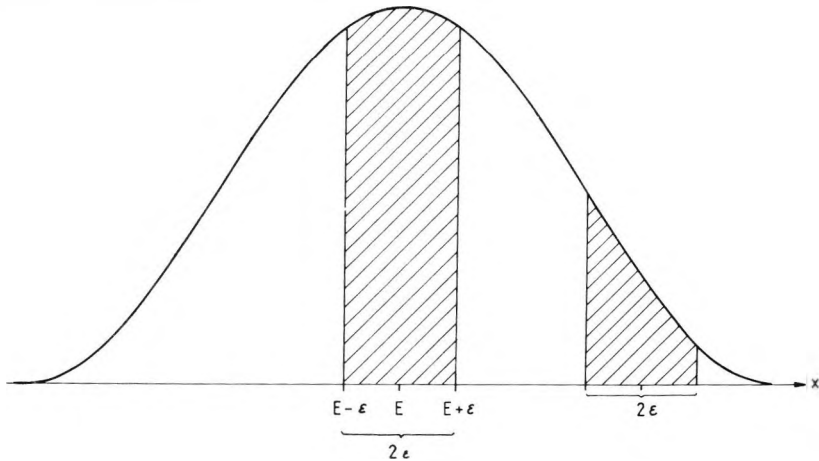in comparison to any other interval of the same length.

Fig. 3. In a symmetrical and unimodal case the expected value E gives the most probable value in the sense that the occurrence of values in the interval $(E - \varepsilon, E + \varepsilon)$ has maximum probability, i.e. for any other value V differing from E the interval $(V - \varepsilon, V + \varepsilon)$ is less probable

3. ábra. Szimmetrikus és egy maximumú esetben az E várható érték valóban a legvalószinübb értéket adja abban az értelemben, hogy az $(E - \varepsilon, E + \varepsilon)$ intervallumbeli értékek előfordulásának maximális a valószinüsége, azaz E-től különböző bármely V értékre a $(V - \varepsilon, V + \varepsilon)$ intervallumba esés kevésbé valószinü

Рис. 3. При симметричном распределении с одним максимумом математическим ожиданием E действительно определяется наиболее вероятное значение в том смысле, что вероятность значений, попадающих в интервал $(E - \varepsilon, E + \varepsilon)$ максимальна, то-есть попадание в интервал $(V - \varepsilon, V + \varepsilon)$, где V — любое отличное от E значение, менее вероятно.

Integrals in this monograph are to be understood as Lebesgue–Stieltjes integrals but a rather conventional notation is used (without $...dF(x), ...dF$ or $F(dx)$). In this sense Eq. (1) and many of the equations in this monograph containing summations seem to be superfluous because if we substitute the 'density function'

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} \delta(x - x_i) \tag{4}$$

in some integral formulae ($x_1, ..., x_i, ..., x_n$ is the sample and δ is the Dirac function), we get the summation formulae (in the present case Eq. (1) from Eq. (2)). It is well known, however, that not only estimates of this type exist.

The density function can really consist of Dirac functions by using digital equipment. In this case

$$\tilde{f}(x) = \Delta x \sum_{i=-\infty}^{\infty} f(i\Delta x) \cdot \delta(x - i\Delta x) \tag{5}$$

holds if Δx characterizes the last digit. By proper use of the equipment the Dirac functions are to such a degree close to one another that their envelope curve gives Δx-times the primarily continuous density function, see *Fig. 4*.
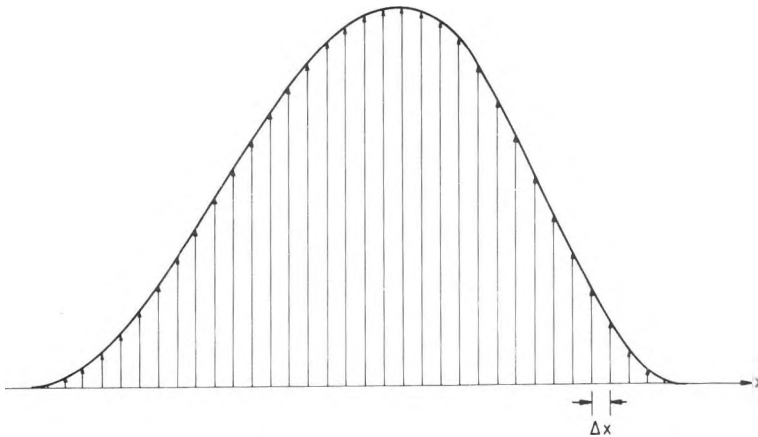

Fig. 4. Interconnection between a continuous probability distribution and its realization as a sequence of Dirac functions (if we use digital devices)

4. ábra. Kapcsolat folytonos eloszlás és Dirac függvénysorozattal való realizációja között (ha digitális műszerrel mérünk)

Рис. 4. Зависимость между непрерывным распределением вероятностей и его реализацией рядом Дирака-δ (при производстве измерений цифровой аппаратурой).

The maximum probability shown in *Fig. 3* seems to justify the technical term 'expected value' for E. Speaking about the expected value for a unimodal distribution, we really expect the clustering of our results around this value, but we can face a quite other situation, too. For example *Fig. 5* shows a lognormal density function, and the interval $(M - \varepsilon, M + \varepsilon)$ is much more probable than the interval of the same length around E:

$$\int_{M-\varepsilon}^{M+\varepsilon} f(x)\,dx > \int_{E-\varepsilon}^{E+\varepsilon} f(x)\,dx.$$

(It would be easy to show examples, too, where not only the relation sign $>$ but also $\gg$ would be justifiable.) The name 'expected value' for the value E can therefore be highly unjustifiable but we shall consistently use this expression as a technical term independently of its real meaning.
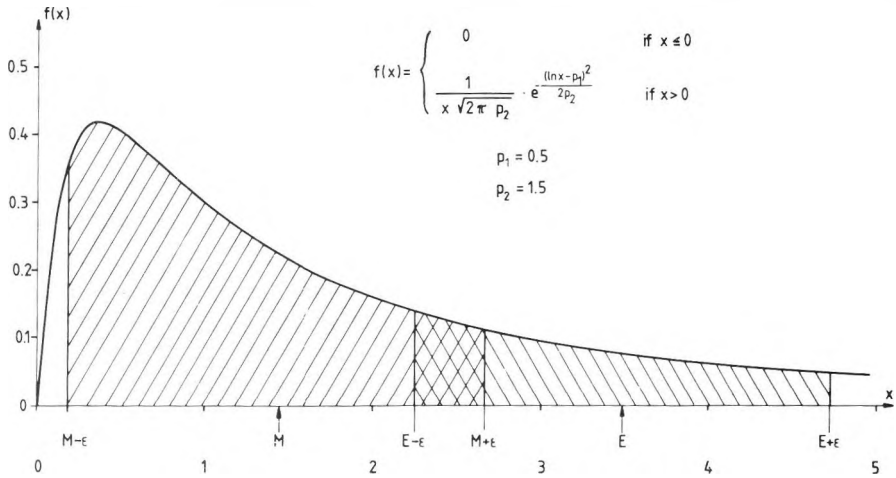
$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \dfrac{1}{x\sqrt{2\pi}\,p_2} \cdot e^{-\frac{(\ln x - p_1)^2}{2p_2}} & \text{if } x > 0 \end{cases}$$

$$p_1 = 0.5$$
$$p_2 = 1.5$$

Fig. 5. Example to show that the probability belonging to the interval $(E-\varepsilon,\ E+\varepsilon)$, i.e. around the expected value, can be much less probable than the probability of an interval of the same length around another value, e.g. around the so-called most frequent value (M)

5. ábra. Példa arra, hogy a várható érték körüli $(E-\varepsilon,\ E+\varepsilon)$ intervallumba esés valószínűsége sokkal kisebb lehet, mint más érték, pl. az M leggyakoribb érték körüli, ugyanolyan hosszúságú intervallumba esés valószínűsége

Рис. 5. Пример значительно более низкой вероятности попадания в интервал около математического ожидания $(E-\varepsilon,\ E+\varepsilon)$, нежели в интервал такой же длины около наиболее частого значения M.

If our task really needs the 'location of gathering', then it would be dangerous if we were to be misled by the suggestion of the name 'expected value': we see that there are cases in which Eq. (1) is for this purpose unusable. BAULE [1963] (on page 4) refers to the fact that e.g. in biology, another value is more suitable for characterizing the totality of data than the expression in Eq. (1): 'arithmetic mean lies near us because of custom but it is hereby by no means justified'. The objection belongs here just to the fact that arithmetic means do not always coincide with the most frequently occurring values.

1.1.2 Distribution of estimates, asymptotic variance, efficiency

Let us once more suppose that $f(x)$ is unimodal and symmetrical and that the integral in Eq. (2) exists: even in this case one cannot be certain that the arithmetic mean is effective enough to estimate the symmetry point. To what does the meaning of efficiency belong? Suppose that we have to compare two different estimation algorithms whose results are denoted by T' and T''. If we sample N times from the distribution $f(x)$ and each sample contains n data, we get equally $N\ T'_n$ and $N\ T''_n$ estimates; if $N \to \infty$ we get more and more

accurately the *distributions of the estimates* for both estimation procedures (and naturally for *n*). (To enhance the difference, we often speak about 'mother distribution' in the case of the original $f(x)$. The sample from a mother distribution does not categorically mean directly measured data: it can be a set of derived data, too. However, the principal thing in the statistical respect is that this is the 'raw material' of statistical procedures for estimating the parameter of location.)

Comparison of the variances shows which type of estimation is more advantageous. If *n* also tends to infinity, then *n* times the variance of the estimates give as the limiting value the so-called asymptotic variance (A'^2 and A''^2); the square roots are the asymptotic standard deviations (A' and A''). If $A' < A''$,

$$e = \left(\frac{A'}{A''}\right)^2 \tag{6}$$

is the *relative efficiency*; if A'^2 is the minimum asymptotic variance, Eq. 6 gives the efficiency of T'' in the absolute sense. The quantity $(1/e - 1) \cdot 100$ is economically very important: this gives the percentual surplus of data needed for the same accuracy, if we calculate according to the less effective procedure.

If we use the arithmetic mean to estimate the parameter of location, not only will the estimating procedure be the simplest possible but the asymptotic variance will also be very easy to calculate:

$$A_E^2 = \int_{-\infty}^{\infty} (x - E)^2 f(x)\,dx. \tag{7}$$

If we use sample medians (m) in the case of unimodal and symmetric distributions for estimating the parameter of location, i.e. the symmetry point, the asymptotic standard deviation (or briefly 'asymptotic scatter') A_m can also be calculated for this estimation procedure very simply:

$$A_m = \frac{1}{2f(0)}. \tag{8}$$

1.1.3 The law of large numbers, 'central' limit theorem

The expression in Eq. (7) is customarily used for the general characterization of the *mother distribution* itself (and is called simply 'variance', without any specification). This expression, however, characterizes the distribution $f(x)$ in one single and in addition in a very special point of view: is Eq. (1) (i.e. the simplest procedure for estimation) able to give acceptable approximations for the parameter of location for samples from a given $f(x)$ probability distribution?

It is a practically important demand that *the law of large numbers must be fulfilled*, i.e. if *n* is greater, the estimate is more accurate. (It is well known that 'the law of large numbers' means a series of theorems in probability theory, with

well defined premises and conclusions; we have the essence of these theorems condensed in just a few words.)

The limit theorem, known as 'central', is perhaps the most beautiful theorem of probability theory: according to this the probability distribution of E_n is of Gaussian type if n is large enough:

$$f(E_n; x) \approx \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-E)^2}{2\sigma^2}},$$ (9)

where

$$\sigma = A_E/\sqrt{n},$$ (9a)

i.e. also given is the asymptotic distribution of the estimates.

One would perhaps say that because of the well defined distribution of estimates of simple analytical form it is worth while to use arithmetic means as estimates. We shall see in Chapter 4, however, that many more advantageous estimates have Gaussian limit distribution, too. Why is just the theorem given by Eq. (9) called 'central'? If we decide arbitrarily to estimate *always* by arithmetic means, then Eq. (9) indeed plays a central role (and the expression in Eq. (7), too). By using other estimates, however, other formulae will play the central role, and therefore Eq. (9) will be ousted to the periphery together with the expression in Eq. (7).

We shall see, on the one hand, in connection with *Table IV* (in Chapter 5) that arithmetic means can behave differently if $A_E = \infty$. On the other hand, the finite value of A_E does not nessesarily mean at the same time that the use of arithmetic means as estimates is economical. The question of economy arises, however, concerning the results of more advantageous estimating methods too, and the answer depends on the probability distribution occurring in practice.

But what types of probability distributions really occur in practice? We can naturally suppose an arbitrarily chosen special type of distribution—but we can also acknowledge fairly as do ANDREWS et al. [1972], that 'we never know in practice what situation we face'.

It is commonplace to suppose Gaussian distribution which is also called 'normal' distribution. But why? A part of the truth is contained in the witty remark of Poincaré [1912]: 'mathematicians accept the normal distribution believing that this is the physical reality, physicists use normal distribution believing that this is a mathematical law'. It depends only on suitable premises that the occurring probability distribution really becomes a Gaussian one: if we deal exclusively only with such cases in which the superposition of a very large number of very small effects (fulfilling some mathematical demands) results in the statistical fluctuations, then the summing takes place simultaneously with the establishment of the phenomenon and therefore *by the fulfilling of these suppositions* the distribution can really be near the Gaussian one.

Mathematicians often reckon on the fulfillment of similar conditions. A typical citation is that of PRÉKOPA [1962]: 'If we want to apply the method of probability theory successfully, the conditions not taken into account must

F. Steiner

each be of relatively small effect.' The problem is that this is unacceptable for geophysicists and geologists (and, I would think, for many other disciplines, too). If a condition is neglected because of its very sporadic occurrence (and not because of its negligible effect), and this causes outliers in an actual situation, the statistical method must also be able in such cases to give reliable results. In one word: the statistical procedure must be *resistant*.

Class No.	No.	$f(x)$	ε	$n(\varepsilon)$	$\dfrac{c}{\sqrt{n(\varepsilon)}}$	σ	Is the law of large numbers fulfilled?	$\dfrac{n_E}{n_M}$
1	1	$\dfrac{1}{6.2691(\sqrt{x^2+1})^{1.4}}$	1.500	0.356	2.513	∞	no (its reverse is true)	∞
	2	$\dfrac{1}{4.5545(\sqrt{x^2+1})^{1.6}}$	1.272	0.426	1.949	∞	no (its reverse is true)	∞
	3	$\dfrac{1}{3.6791(\sqrt{x^2+1})^{1.8}}$	1.112	0.469	1.623	∞	no (its reverse is true)	∞
	4	$\dfrac{3\sqrt{3}}{8\pi(|x|^{3/2}+1)}$	1.299	0.373	2.128	∞	no (its reverse is true)	∞
2	5	$\dfrac{1}{\pi(x^2+1)}$	1.000	0.500	1.414	∞	no	∞
	6	$\dfrac{1}{2(|x|+1)^2}$	0.683	0.419	1.055	∞	no	∞
	7	$\begin{cases} \dfrac{1}{4} & \text{if } |x| \leq 1 \\ \dfrac{1}{4x^2} & \text{if } |x| > 1 \end{cases}$	1.240	0.561	1.655	∞	no	∞
	8	$\dfrac{1}{\pi}\left(\dfrac{\sin x}{x}\right)^2$	1.075	0.589	1.401	∞	no	∞
3	9	$\dfrac{1}{2.5056(\sqrt{x^2+1})^{2.4}}$	0.840	0.537	1.146	∞	yes	∞
	10	$\dfrac{1}{2.1348(\sqrt{x^2+1})^{2.8}}$	0.735	0.558	0.984	∞	yes	∞
	11	$\dfrac{3\sqrt{3}}{4\pi(|x|^3+1)}$	0.832	0.596	1.078	∞	yes	∞
4	12	$\dfrac{1}{1.8873(\sqrt{x^2+1})^{3.2}}$	0.662	0.573	0.875	2.236	yes	6 530
	13	$\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	0.925	0.631	1.165	1.000	yes	0.737
	14	$\begin{cases} \dfrac{1}{2} & \text{if } |x| \leq 1 \\ 0 & \text{if } |x| > 1 \end{cases}$	0.719	0.681	0.871	0.577	yes	0.439
	15	$\dfrac{1}{2}e^{-|x|}$	0.807	0.562	1.076	1.414	yes	1.726
	16	$\dfrac{2K_0^2(x)^*}{\pi^2}$	0.0447	0.364	0.0741	0.35	yes	22
	17	$\begin{array}{ccc} \uparrow 0.5 & & \uparrow 0.5 \\ -1 & 0 & +1 \end{array}$	1.732	0.750	2.000	1.000	yes	0.250
	18	$\dfrac{2}{\pi(x^2+1)^2}$	0.562	0.590	0.732	1.000	yes	1.866

* $K_0(x)$ is the so-called modified Bessel function, see e.g. JAHNKE-EMDE-LÖSCH: Tables of Higher Functions. McGraw-Hill. 1960

The conditions mentioned above (that result in a Gaussian error distribution) could be fulfilled in principle in the majority of cases but the real situation can be characterized by a short remark of TUKEY [1977]: 'When the underlying distribution, as always, is nongaussian...'. (Some similar citations have already been quoted in Table I). The attribute 'normal' for the Gaussian distribution type is therefore unjustified (and misleading) for mother distributions (the question *of estimates* of distribution-type must be dealt with separately, this is quite another question). Instead of 'normal' we shall subsequently speak about Gaussian distribution (as is also done in the second half of ANDREWS et al. [1972]).

1.2 The notion of robustness with regard to practice

There are some optimum procedures in classical statistics which suppose the occurrence of a given distribution type. One would think perhaps that the efficiency of such procedures is near the optimum if similar error distribution occurs. (E.g. the computation of arithmetic means is optimum for the Gaussian distribution. This sort of estimating is perhaps near the optimum, say, also for similar, i.e. for symmetrical and unimodal distributions.)

The article of TUKEY [1960] changed dramatically the opinion of appliers and of statisticians, too. Instead of following the original train of thought of Tukey, we show in the following that an estimation procedure which can be

Table IV. Some characteristics of dispersion for various types of distributions given by their density function $f(x)$. These characteristics are: dihesion (ε); asymptotic scatter of the most frequent values ($\varepsilon/\sqrt{n(\varepsilon)}$); asymptotic scatter of the arithmetic means (σ; the value σ^2 is known simply as 'variance' and is commonly but misleadingly used to characterize the original distributions themselves). The relative efficiencies are also given (n_E/n_M): how many times more data are needed to achieve asymptotically the same accuracy using arithmetic means (i.e. least squares techniques) than by calculating most frequent values

IV. táblázat. A diszperzió néhány jellemzője különböző $f(x)$ sűrűségfüggvényű eloszlásokra. A jellemzők a következők: dihézió (ε); a leggyakoribb értékek aszimptotikus szórása ($\varepsilon/\sqrt{n(\varepsilon)}$); a számtani átlagok aszimptotikus szórása (σ; a σ értéket egyszerűen „szórás"-nak nevezik és szokásosan — de félrevezető módon — maguknak az eredeti eloszlásoknak a jellemzésére használják). A relatív hatásfokok is adottak (n_E/n_M): hányszor annyi adat szükséges a számtani átlagok (azaz a legkisebb négyzetek elvének) alkalmazásakor a leggyakoribb érték-számítás pontosságával aszimptotikusan azonos pontosság eléréséhez

Таблица IV. Некоторые характеристики дисперсии для распределений с различными весовыми функциями $f(x)$: дигезия (ε), асимптотическая дисперсия наиболее частых значений ($\varepsilon/\sqrt{n(\varepsilon)}$), асимптотическая дисперсия среднеарифметических σ (часто называемая просто «дисперсией» и обычно, но ошибочно используемая для характеристики первичных распределений). Относительные эффективности n_E/n_M также заданы, то-есть задано, во сколько раз больше данных необходимо при использовании среднеарифметических, то-есть принципа наименьших квадратов, для обеспечения асимптотически идентичной точности, с таковой при расчете наиболее частых значений.

applied advantageously for a given type of distribution, can become completely unusable in the asymptotic sense if this distribution is changed arbitrarily small.

Let us define the following family of distributions:

$$f(x) = \begin{cases} 1, & \text{if} \quad |x| \leqq x_0 \\[2em] \dfrac{1}{1 + \left(\pi \dfrac{|x| - x_0}{1 - 2x_0} \right)^2}, & \text{if} \quad |x| > x_0. \end{cases} \qquad (0 < x_0 < 0.5) \qquad (10)$$

(All distributions are symmetrical to the origin.)

For the parameter value $x_0 = 0.4$ the density function is shown in *Fig. 6.*, the distribution function in *Fig. 7*. It is obvious that the nearer the parameter x_0 is to the value 0.5, the longer on the one hand is the interval for which the distribution is uniform and, on the other hand, the smaller is the role of the flanks, which decrease as const. x^{-2} for large values of $|x|$. In contradiction to this, not only is the asymptotic variance A_E^2 (Eq. (7)) of the arithmetic means infinite but also the integral in Eq. (2) is divergent, this means that even the expected value does not exist. Whereas the arithmetic mean gives a good estimate for the symmetry point of uniform distribution (with an asymptotic
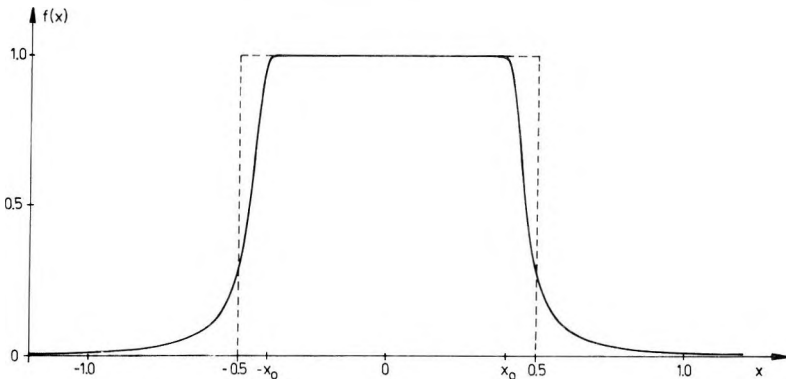

Fig. 6. Density function $f(x)$ given by Eq. (11) for some value of x_0. If $x_0 \to 0.5$, the distribution $f(x)$ is infinitely close to the uniform distribution (i.e. the distance of $f(x)$ from the uniform distribution can be arbitrarily small); in contrast, the variance of all $f(x)$ distributions is infinite, compared to the naturally finite variance value of the uniform distribution

6. ábra. A (11) egyenlettel adott $f(x)$ sürüségfüggvény valamely x_0 értékre. Ha $x_0 \to 0{,}5$, az $f(x)$ eloszlás végtelenül közel jut az egyenletes eloszláshoz (azaz az $f(x)$ távolsága az egyenletes eloszlástól akármilyen kicsiny lehet), ugyanakkor az $f(x)$-ek szórása mindig végtelen marad, szemben az egyenletes eloszlás nyilvánvalóan véges szórásával

Рис. 6. Плотностная функция $f(x)$, заданная уравнением 11 для произвольной x_0. При $x_0 \to 0{,}5$ распределение $f(x)$ неограниченно приближается к равномерному распределению (то-есть расстояние $f(x)$ от равномерного распределения может быть сколь угодно малым), в то же время дисперсия значений функции $f(x)$ все время остается бесконечной в противоположность явно конечной дисперсии равномерного распределения.

scatter of $A_E = 1/(2\sqrt{3})$) for the interval $[-0.5; +0.5]$, the same estimate can completely deteriorate in the asymptotic sense arbitrarily near this distribution.

It is important that the efficiency for a given estimation procedure should not depend seriously upon the underlying distribution type. Expressed in a single word: the estimation should be robust.

For what the meaning of this word is for engineers PIRKLE et al. [1982] is cited: robust estimators are 'those which perform well under a condition of nonnormality or of normality with outliers added as well as under conditions of normality'. The order of succession is interesting and instructive: the first point of view is the good behaviour at nonnormal distributions; the insensitivity of outliers is the second one; finally the acceptable performance at the Gaussian distribution is also demanded as cases may occur in which the actual distribution is near the Gaussian one.
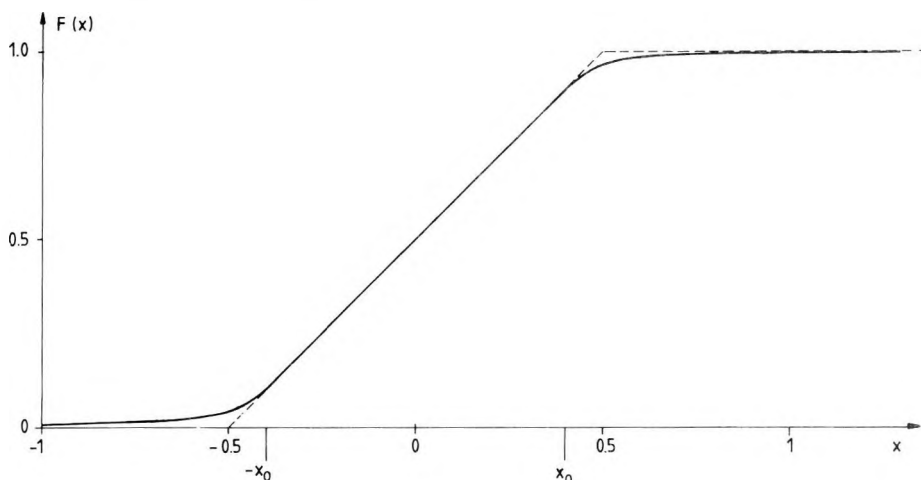

Fig. 7. Distribution function $F(x)$ belonging to the density function defined by Eq. (11) and shown in Fig. 6.

7. ábra. A 6. ábrán bemutatott $f(x)$ sürüségfüggvény $F(x)$ eloszlásfüggvénye

Puc. 7. Функция распределения $F(x)$ плотностной функции $f(x)$, представленного на рис. 6.

1.3 Mathematical definition of robustness

We have spoken about distributions lying near each other (e.g. the distribution given by Eq. (10) in the case of $x_0 \approx 0.5$, and the uniform distribution). In the mathematical respect, therefore, the first thing is to define the *distance between two probability distributions*.

There are some definitions for this distance. One is the *Hellinger-distance* being the L_2 norm of the difference function of the square roots of the density functions:

$$\|\sqrt{f(x)} - \sqrt{g(x)}\|. \tag{11}$$

It is clear that the Hellinger-distance of two distributions which have not too heavy tails (i.e., the density functions tend quickly enough to zero if $|x| \to \infty$), is determined mainly by those parts of the density functions which are characteristic for the gathering of the data. This means that the distance of the uniform distribution and the distribution defined by Eq. (10) tend to zero if $x_0 \to 0.5$. Another possibility is to accept the definition of the Prohorov-distance [HUBER 1981]. The definition is more complicated than Eq. (11) but there are some interesting and important theorems to this sort of distance. For appliers the most important statement is that a great Prohorov-distance characterizes the case if two distributions are also practically distinguishable.

The definition of the robustness for estimates is the following: *the estimation is robust if the distribution of estimates is a uniformly continuous functional of the mother distribution, i.e. to near lying mother distributions belong near lying distributions of estimates.* Consequently, arithmetic means are not robust estimates as our example has shown that to arbitrarily near lying mother distributions can belong on the one hand a distribution of estimates with finite asymptotic variance, and on the other hand, a distribution of estimates with infinite asymptotic variance (even the law of large numbers is not fulfilled in the second case if n is great).

Mathematicians refer to the definition of robustness mentioned above as a qualitative one. For quantitative investigations, the von Mises-derivative of the functional is mostly used. We shall get acquainted with this notion in Chapter 4, known as the 'influence curve' or IC-function; its excellent practical applicability will be clear in the second half of the monograph.

Chapter 2
Heuristic definition of the most frequent value and of the fitting according to the most frequent values

2.1 Most frequent values with a weight function of fixed parameter of scale

2.1.1 The parameter of scale (S)

Let us first remember the notion of the parameter of scale which will express primarily the degree of the tendency of gathering of our data. (This technical term originates from the fact that $\frac{1}{S} f\left(\frac{x}{S}\right)$ has the same distribution type as $f(x)$, only the unit of scale is changed.)

If T is the parameter of location, the density function has the general form

$$f(x) = f(T; S; x). \tag{12}$$

For example, the density function of the well known Cauchy distribution is

$$f_C(x) = \frac{1}{\pi} \frac{S}{S^2 + (x - T)^2}.$$ (13)

Figure 8 shows this density function for four (T, S)-pairs, namely for $T = 10$ and 20, and for $S = 2$ and 4. The parameter of location T here is the symmetry point and S gives—independently from T—information on whether data are expected to be more or less dense around the actual value of T.

We can justify with simple integration that in the case of the Cauchy distribution S equals the semi-interquartile range Q (known also as probable error). Q has an immediate probability interpretation and therefore it would be advantageous in this respect to demand $S = Q$ for all distributions, but the usual practice is to define such a parameter of scale S which results in a density function of simple analytical form.

The parameter of scale is not only a characteristic of the density functions but it also gives for other functions the value of the breadth in the sense that twice as great S means twice as bright function, etc. In the following this latter interpretation will be used for weight functions.
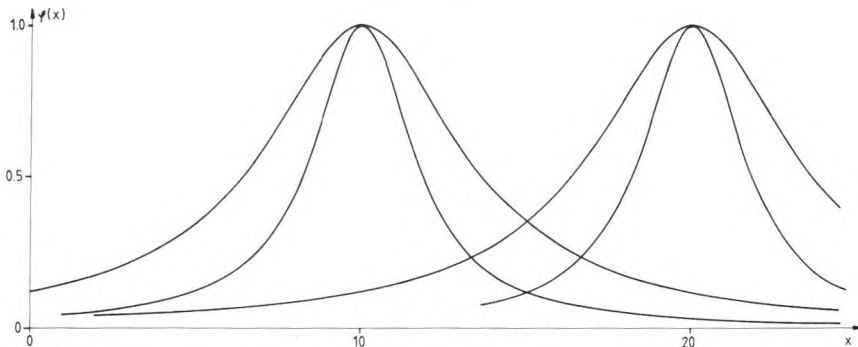

Fig. 8. Four density functions of Cauchy-type: different parameters of location define different symmetry points, different parameters of scale characterize different widths. (The four density curves are normalized to $f(T) = 1$)

8. ábra. Négy Cauchy-tipusú sűrűségfüggvény: különböző helyparaméterek különböző szimmetriapontokat definiálnak, különböző skálaparaméterek pedig különböző szélességeket. (A négy sűrűségfüggvény $f(T) = 1$-re van normálva)

Рис. 8. Четыре плотностных функций типа Коши: различными параметрами места определяются различные точки симметрии, а различными параметрами шкал — различные ширины. (Все четыре плотностных функции приведены к $f(T) = 1$.)

2.1.2 Outliers

If our data are visualized as points on the datum line, we should first glance to see whether or not we have one or more outliers. A Hungarian scientist wrote

[TÁRCZY-HORNOCH 1956] that 'outliers essentially differ from the other meas-
ured data'. It was customary at one time to cancel these data before calculating
the estimate (in the simplest case the arithmetic mean).

> The so-called deterministic distribution in KORN and KORN [1961] is, from the point of view
> the probability theory, undoubtedly a degenerate one. We shall use it now, however, in connection
> with a situation which is familiar to all readers. Let us suppose that the data and the operations
> to be carried out are typed onto a calculator; if— by repeating this procedure the result is the
> same—this is regarded as the 'true result'. If not, we repeat the procedure twice more and naturally
> accept that value which is obtained three times, completely neglecting the differing value: it would
> be absurd to accept the arithmetic mean of the four results. This remark was made as classic
> statistical literature is often somewhat contemptuous at the neglecting of one or a few items of data
> but for healthy minded practicians this is quite natural.

Is it possible somehow to 'legalize' from the viewpoint of probability theory
the cancelling of the outliers?

2.1.3 Weighted means

Most books written on probability theory and/or statistics prove the the-
orem that the weighted mean:

$$\frac{\sum_{i=1}^{n} q_i x_i}{\sum_{i=1}^{n} q_i} \tag{14}$$

calculated from $x_1, x_2, ..., x_i, ..., x_n$ having a probability distribution $f(x)$, gives
the most efficient estimate if all q_i weights are the same.

Supposing that the measuring conditions for all data are unaltered—dis-
regarding cases in which, because of more accurate equipment (or because of
a similarly obvious condition), it is really justifiable to take some data with a
greater weight into consideration—for engineers or other appliers this theorem
seems therefore rather trivial: why would it be justifiable, say, to take the
seventh value from a sample containing ten data with twice as great weight than
the other data?

Instead of *weights depending on the index* of the sample element, a weight
distribution derived statistically from all samples can be useful:

$$q_i = q_i(x_1, x_2, ..., x_n). \tag{15}$$

The question posed at the end of paragraph 2.1.2 can be answered with a
weighting system which gives zero (or nearly zero) weights to outliers.

The simple theoretical background of the cancellation of outliers is the fact
that the occurrence of a distant point x_k on the datum line is not yet probable:

$f(x_k) = 0$ or $f(x_k) \approx 0$. We can write according to the foregoing an interesting and instructive but in practice a not yet applicable formula, viz.

$$\frac{\sum\limits_{i=1}^{n} f(x_i) \cdot x_i}{\sum\limits_{i=1}^{n} f(x_i)}. \tag{16}$$

We are protected now against outliers (without accepting arbitrary, ad hoc criteria), and we also find the behaviour of this expression sympathetic near the symmetry point (supposing that $f(x)$ is unimodal and symmetrical), since the nearer the sample element to the symmetry point, the greater its weight.

In general, however, we know neither the type of the distribution nor the parameters T and S. If we knew all this information, even the sampling itself would be superfluous (to say nothing about the calculation of Eq. (16)). But using a weight function $\varphi(x)$ (not too far distant from $f(x)$) this function also cancels the outliers and gives the greatest weights to the most gathering values. (Chapter 8 shows that even the best weight function is not always the density function itself.)

From the point of view of computing techniques, the simplest possible choice of the weight function seems to be

$$\varphi(x) = \frac{\varepsilon^2}{\varepsilon^2 + (x - M)^2}. \tag{17}$$

Comparing this expression with Eq. (13), this is a weighting according to the Cauchy density function. (The theoretical background for using this weight function is dealt with in Chapters 7 and 8; a heuristic foundation is given in paragraph 2.4.3.)

2.1.4 Most frequent values calculated by weight functions of fixed parameter of scale

Let us suppose that ε is known in the weight function $\varphi(x)$. (The determination of ε is dealt with in paragraph 2.2.3.)

The value of M in Eq. (17) is unknown but it is a plausible demand that the symmetry point $T = M$ of the distribution $f(x)$ should be given as the weighted mean:

$$M = \frac{\sum\limits_{i=1}^{n} \varphi(x_i) \cdot x_i}{\sum\limits_{i=1}^{n} \varphi(x_i)}, \tag{18}$$

or substituting Eq. (17), as

$$M = \frac{\sum_{i=1}^{n} \dfrac{\varepsilon^2 x_i}{\varepsilon^2 + (x_i - M)^2}}{\sum_{i=1}^{n} \dfrac{\varepsilon^2}{\varepsilon^2 + (x_i - M)^2}}. \tag{18a}$$

The corresponding integral formula for the mother distribution $f(x)$ is clearly

$$M = \frac{\displaystyle\int_{-\infty}^{\infty} \dfrac{\varepsilon^2 x}{\varepsilon^2 + (x - M)^2} f(x)\,\mathrm{d}x}{\displaystyle\int_{-\infty}^{\infty} \dfrac{\varepsilon^2}{\varepsilon^2 + (x - M)^2} f(x)\,\mathrm{d}x}. \tag{18b}$$

The value M appears on both sides of these equations, they are therefore practically iteration formulae. If, in the j-th step of the iteration according to Eq. (18a) the result is denoted by M_j, in the following step

$$M_{j+1} = \frac{\sum_{i=1}^{n} \dfrac{\varepsilon^2 x_i}{\varepsilon^2 + (x_i - M_j)^2}}{\sum_{i=1}^{n} \dfrac{\varepsilon^2}{\varepsilon^2 + (x_i - M_j)^2}} \tag{18c}$$

is to be calculated. The iteration is stopped if there are only negligible changes in the M-value. In order to start the iteration, both the arithmetic mean and the sample median are equally acceptable.

The value M fulfilling Eq. (18b) is called the most frequent value of the probability distribution $f(x)$. This is evidently a parameter of location. The value M (or M_n) fulfilling Eq. (18a) is the most frequent value of the sample; this is an estimate for the just defined parameter of location. Obviously Eq. (18b) gives the symmetry point, if $f(x)$ is symmetrical, similarly to Eq. (2). In general cases, however, M can differ from E. The symbol M has its origin in the name 'most frequent value'. We have to show, however, that this name is really justified.

Substituting $M_j = \bar{M}$ and $M_{j+1} = \bar{M} + \varDelta M$ in Eq. (18c) we get for $\varDelta M$ the expression

$$\varDelta M = \frac{1}{\sum_{i=1}^{n} \dfrac{\varepsilon^2}{\varepsilon^2 + (x_i - \bar{M})^2}} \sum_{i=1}^{n} \frac{\varepsilon^2}{\varepsilon^2 + (x_i - \bar{M})^2} (x_i - \bar{M}). \tag{18d}$$

The sign of $\varDelta M$ is determined only by the second sum. This sum is, however, the sum of (plus or minus) the areas of the rectangles defined by the actual difference $(x_i - \bar{M})$ and the corresponding weight $\varphi(x_i)$. Two cases can occur:
1. the sum of areas of the rectangles right and left of $\bar{M}$ are equal, $\varDelta M = 0$,

therefore $M = \overline{M}$, the iteration is finished; 2. the sum of areas of the rectangles is say, on the left hand side of $\overline{M}$ greater than on the other side, $\Delta M < 0$, therefore the iteration brings $\overline{M}$ nearer and nearer the gathering of the data. Both alternatives are shown in *Fig. 9a* and *9b* for the sample -0.2; 0; $+0.2$;
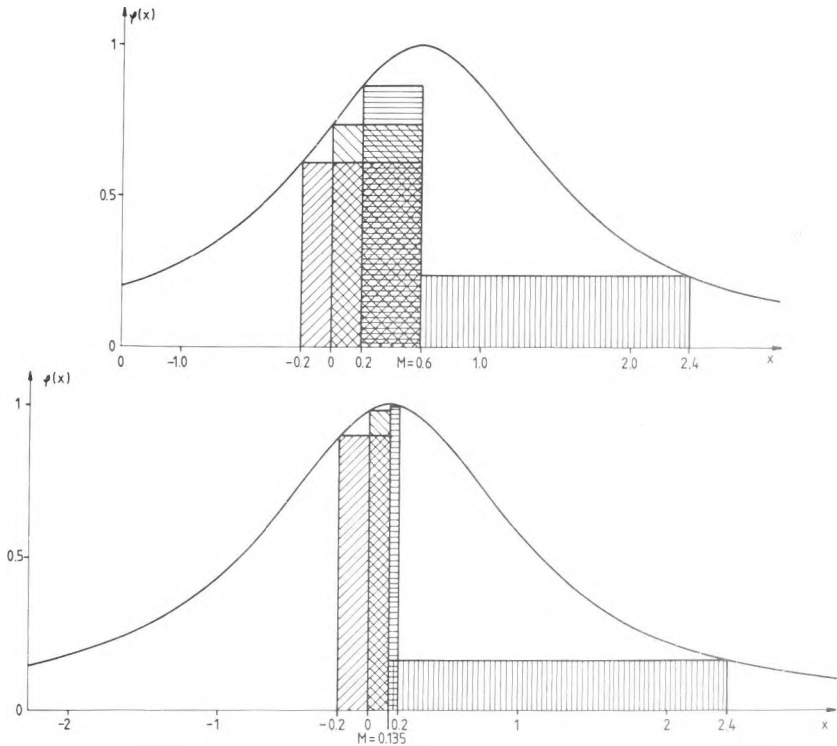

Fig. 9. Reweighted mean-calculations (with the weight function φ) approximate the value which can really be called the most frequent value. Namely, if the sum of the areas of rectangles defined by the elements of the sample left and right of the symmetry point of φ are not equal (see Fig. 9/a), the following iteration step will bring the value M_{j+1} nearer the cluster consisting of the majority of sample elements. The iteration is finished if the sums of the areas on both sides of the symmetry point are equal; this is the case for Fig. 9/b

9. ábra. A φ súlyfüggvénnyel történő iterált súlyozott átlagszámítások ahhoz az értékhez tartanak, amelyet joggal nevezhetünk leggyakoribb értéknek. Ha ugyanis a szimmetriaponttól jobbra és balra levő, a mintaelemek által definiált téglalapok területösszegei nem azonosak (l. pl. a 9/a ábrát), akkor a következő iterációs lépés M_{j+1}-et közelebb viszi a mintaelemek zöme által definiált tömörödési helyhez. Az iteráció leáll, ha a területösszegek egyensúlyba jutottak (9/b ábra)

Рис. 9. Средневзвешенные, рассчитанные путем итерации с весовой функцией φ, приближаются значению, по праву называемому наиболее частым значением. Ибо если суммы площадей прямоугольников, определяемых элементами примера, находящимися вправо и влево от точки симметрии, не равны между собой (см. рис. 9/a), то при следующем шаге итерации значение M_{j+1} приблизится к участку сгущения, определяемому преобладающей частью элементов примера. Итерация приостанавливается, если суммы площадей уравновесятся (см. рис. 9/b).

2.4. In this case 0.6 is the arithmetic mean, as the starting value of the iteration; the result of it is $M = 0.135$.

The author considers that the term 'most frequent value' is justifiable. It is certain that to an outlier belongs a long rectangle but its area is approximately proportional to $(x_i - \overline{M})^{-1}$. This means that in the overwhelming majority of cases the suppression of outliers is satisfactory.

2.2 Determination of the scale parameter of the weight function

2.2.1 The demand of maximum number of effective data around M for an interval of minimum length

Let us suppose at the beginning for the sake of simplicity that the value M in φ is known and our task is to determine the proper value of ε.

If our sample is—occasionally—symmetrical to some value (see *Fig. 10*) then the iteration according to Eq. (18a) gives with arbitrary (nonzero) value of ε the right M (i.e. the symmetry point). It is clear that the probability of a fully symmetrical sample is very small even for a symmetrical $f(x)$, the gathering
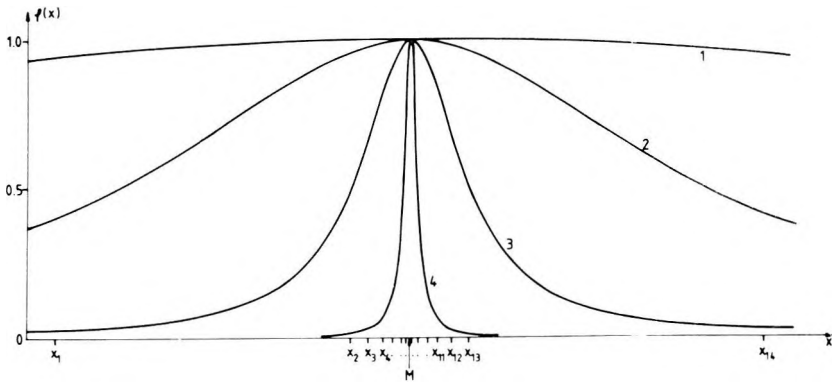
Fig. 10. Weight functions of different widths (i.e. with different scale parameters). If the weight function is too broad (1), the result of the most frequent value calculation will be similar to that of the arithmetic mean together with all its disadvantages. If the weight function is too narrow (4), some of the valuable data are practically neglected resulting in a decrease of efficiency

10. ábra. Különböző szélességű (azaz különböző skálaparaméterű) súlyfüggvények. Ha a súlyfüggvény túl széles (1), a leggyakoribb érték számítása az átlagképzéshez nagyon hasonló lesz, annak minden hátrányával. Ha a súlyfüggvény túl keskeny (4), az értékes adatok egy részét gyakorlatilag elhanyagoljuk, ami a hatásfok csökkenését eredményezi

Рис. 10. Весовые функции различной ширины, то-есть с различными «параметрами шкал». При слишком большой ширине весовой функции (1), расчет наиболее частой величины станет весьма сходным с определением среднего со всеми соответствующими недостатками. При слишком малой ширине весовой функции (4) часть значимых данных пренебрегается, что приводит к снижению эффективности.

of the values, however, will follow in the case of repeated sampling the law which is for simplicity represented by the sample of only 14 elements in Fig. 8. The following line of thought will also be the simplest if we make it on the ground of this sample.

If ε were too great, we would get as the weight function a curve which is similar to the curve 1 in Fig. 8: to every item of data belongs practically the same weight, the most frequent value according to Eq. (18a) has therefore approximately the same behaviour as has the arithmetic mean. For example, having also the outliers x_1 and x_{14} practically a weight near to the maximum because of the very great value of ε, in real situations—where the supposition of the symmetry of the outliers would be fully absurd—the most frequent value fulfilling Eq. (18a) can be deteriorated in the same manner as the arithmetic mean.

The other extreme causes drawbacks of a different nature (see curve 4 in Fig. 10): if the value ε is too small, not only will the outliers be neglected but also a part of the gathered values. We calculate as if we had less data. The result is a decrease in the accuracy as, with nearly all estimations, the increase in accuracy is proportional to $\sqrt{n}$.

We need a rational compromise but the arguments are stronger for not losing reliable values: we do not want to measure, say, twice as much data because of the too small value of ε. It is certain that the disturbing effect of the outliers should be also negligible.

The question arises as to which analytical expression gives a suitable measure for the number of data playing a really significant role by the computing of M according to Eq. (18a). The weights have values of nearly unity in the centre of the gathering, and decrease to approximately zero for distant values, i.e. we have no argument against the acceptance of the sum of weight as the 'number of effective data' denoted by $n_{\text{eff}}(\varepsilon)$:

$$n_{\text{eff}}(\varepsilon) = \sum_{i=1}^{n} \frac{\varepsilon^2}{\varepsilon^2 + (x_i - M)^2}. \tag{19}$$

Consequently, to be in accordance with the foregoing discussions, that value of ε is accepted which gives the maximum value for the expression

$$\frac{n_{\text{eff}}^2(\varepsilon)}{\varepsilon}. \tag{20}$$

This means that the demand

$$\frac{1}{\varepsilon} \left[\sum_{i=1}^{n} \frac{\varepsilon^2}{\varepsilon^2 + (x_i - M)^2} \right]^2 = \text{maximum} \tag{21}$$

is to be fulfilled.

If the distinction is really needed ε denotes the variable in Eq. (21), and ε_0 that maximum value which fulfils this condition. In other cases both are denoted by ε.

2.2.2 Cohesion and dihesion of the mother distribution

Getting ε_0 by solving Eq. (21), this value seems primarily to be the scale parameter of the weight function, which makes the practical calculation of the most frequent value unambiguous. Because for large n values the following equation holds

$$\frac{n_{\text{eff}}(\varepsilon)}{n} \approx \int_{-\infty}^{\infty} \frac{\varepsilon^2}{\varepsilon^2 + (x - M)^2} f(x)\,\mathrm{d}x \equiv n(\varepsilon), \tag{22}$$

the ε value satisfying the condition

$$\frac{n^2(\varepsilon)}{\varepsilon} = \text{max.} \tag{23}$$

can be accepted as *the parameter of scale for the mother distribution itself*. It is clear that the demand in Eq. (23) is the same as

$$\int_{-\infty}^{\infty} \frac{\varepsilon^{3/2}}{\varepsilon^2 + (x_i - M)^2} f(x)\,\mathrm{d}x = \text{maximum} \tag{24}$$

[STEINER 1973].

The ε value satisfying Eq. (24) characterizes the tendency of cohesion of data, if sampling is made, in the sense that small ε means considerable cohesion, and vice versa. Therefore we can define cohesion as

$$\kappa = \frac{1}{\varepsilon}. \tag{25}$$

On the basis of Eq. (25) the name of the scale parameter could be 'reciprocal cohesion', too, but the cumbersomeness of this name justifies the introduction of an arbitrarily constructed name. We refer to this characteristic as *dihesion* of the probability distribution.

It can be shown [CSERNYÁK 1973] that by fulfilling the condition in Eq. (24) the weight function $\varphi(x)$ is the most similar to $f(x)$. (Similarity is meant here in exactly the same sense as Bhattacharyya means the affinity between density functions [MATHAI and RATHIE 1975].

2.2.3 Practical computation of dihesion

Let us suppose that M is known; our task is to determine the dihesion ε. It can be shown [CSERNYÁK and STEINER 1980] that for the dihesion

$$\varepsilon \leqq \frac{\sqrt{3}}{2} [\max(x_i) - \min(x_i)] \tag{26}$$

holds (the limit case is realized by the U-distribution, see series number 17 in Table IV). Starting from the right hand side of Eq. (26) as the initial ε-value, we diminish the value of ε till the expression

$$\sum_{i=1}^{n} \frac{\varepsilon^{3/2}}{\varepsilon^2 + (x_i - M)^2} \tag{27}$$

reaches its maximum value. This ε is used in Eq. (17), i.e. in the weight function. And this value is at the same time an estimate for the dihesion of that distribution $f(x)$ from which the sample comes.

The procedure described above is not sufficiently fast. In practice the iteration

$$\varepsilon_{k+1}^2 = \frac{3 \sum\limits_{i=1}^{n} \dfrac{\varepsilon_k^4 (x_i - M)^2}{[\varepsilon_k^2 + (x_i - M)^2]^2}}{\sum\limits_{i=1}^{n} \dfrac{\varepsilon_k^4}{[\varepsilon_k^2 + (x_i - M)^2]^2}} . \tag{28}$$

is used giving an estimate for that ε which fulfils the following equation:

$$\varepsilon^2 = \frac{3 \displaystyle\int_{-\infty}^{\infty} \dfrac{\varepsilon^4 (x - M)^2}{[\varepsilon^2 + (x - M)^2]^2} f(x)\, dx}{\displaystyle\int_{-\infty}^{\infty} \dfrac{\varepsilon^4}{[\varepsilon^2 + (x - M)^2]^2} f(x)\, dx} . \tag{28a}$$

The ε satisfying Eq. (28a) fulfils the primary condition (24), too. Putting $M = 0$ in Eq. (24) (for the sake of simplicity), and differentiating according to ε, we get

$$\int_{-\infty}^{\infty} \frac{3/2 \cdot \varepsilon^{1/2} (\varepsilon^2 + x^2) - 2\varepsilon\varepsilon^{3/2}}{[\varepsilon^2 + x^2]^2} f(x)\, dx = 0.$$

After some regrouping the result is

$$3 \int_{-\infty}^{\infty} \frac{x^2}{[\varepsilon^2 + x^2]^2} f(x)\, dx = \varepsilon^2 \int_{-\infty}^{\infty} \frac{1}{[\varepsilon^2 + x^2]^2} f(x)\, dx.$$

Expressing ε^2 and substituting the original $(x - M)^2$ instead of x^2, we can ensure that Eq. (28a) really defines the dihesion and consequently iteration (28) gives an estimate for it. (Additional speeding of the iteration is applied, too, in the know-how of the University of Miskolc).

If we calculate integrals of such types with asymptotically zero and continuous integrands (see e.g. in Eq. (28a) the simplest way is approximation with the sum of areas of rectangles defined by an equidistantly graduated abscissa and the actual values of the integrand, resulting in a much greater accuracy than supposed (the causes are dealt with in STEINER and ZILAHI-SEBESS [1988] Appendix 5).

2.3 Simultaneous calculation of the most frequent value and the dihesion

The supposition of known M or that of known ε: both are in general unfulfilled suppositions. In practice, the solution is the following: a twofold iteration is to be carried out, i.e. ε^2 from Eq. (28) and M from Eq. (18c) are to be calculated alternatingly. The final definition belongs therefore to a *pair* of values: *the most frequent value (M) and dihesion (ε) of a probability distribution (defined by its density function $f(x)$) are those values which simultaneously fulfil Eqs. (18b) and (28a).*

If not a sample but $f(x)$ is given the just mentioned equations are also to be regarded as iteration rules. As the starting value for M the median and the so-called expected value are equally suitable, and as a first value of ε it is convenient to use 3.1 times the semi interquartile range (Q), because the relation

$$\varepsilon \leqq 3.0924\ Q \tag{29}$$

holds [Csernyák and Steiner 1980]. For samples mutatis mutandis the same is true: the sample median or the arithmetic mean is the first M-value, and if the sample semi-interquartile range is Q_{emp}, the first ε-value is 3.1 times Q_{emp}.

Using digital equipment, if the value '1' on the last digit was chosen so that it was too great the ε-iteration can tend to zero. This value has nothing to do with the real value of the dihesion and therefore the ε-iteration must be stopped if $\varepsilon < \varepsilon_{min}$ (ε_{min} can be given according to physical plausibilities). After finishing the M-iteration a comment is appropriate, viz., that ε_{min} was reached and used in the last M-steps. Another comment can be disclosed if the number of iteration steps is a priori fixed.

It is obvious that the calculation of an M value needs about two orders of magnitude more mathematical operations than the determination of an arithmetic mean. If the latter were less effective, however, a surplus of expensive field measurements would be needed; on the contrary, operations on computers are very cheap and their costs are decreasing rapidly nowadays.

The pair of values (M, ε) informs us about the main part of the distribution: the values of the random variable are with relatively great probability in the interval $(M - \varepsilon,\ M + \varepsilon)$. Therefore—as figure 5 shows—the name 'expected value' would be much more justifiable for M than for E, but we shall use this term for E with the same consistency as the whole literature of statistics.

2.4 M-fitting: adjustment according to the most frequent values

2.4.1 Minimum condition fulfilled in each step of the M-iteration

It is easy to verify by differentiation that the expression of M given by Eq. (18a) fulfils the following condition:

$$\sum_{i=1}^{n} \varepsilon^2 \ln (\varepsilon^2 + (x_i - M)^2 = \text{min.} \tag{30}$$

In the same way it can be verified that Eq. (18c) corresponds to the condition

$$\sum_{i=1}^{n} \frac{\varepsilon^2}{\varepsilon^2 + (x_i - M_j)} (x_i - M_{j+1})^2 = \text{min.,} \tag{31}$$

where M_j and ε are known values. This is, from the computing technical point of view, equivalent to

$$\sum_{i=1}^{n} q_i (x_i - M)^2 = \text{min.} \tag{32}$$

(resulting in expression 14) as the weights are known in every iteration step. The condition in Eq. (32), however, is the special case of

$$\sum_{i=1}^{n} q_i \cdot (x_i - T(\mathbf{p}; \mathbf{y}_i))^2 = \text{min.,} \tag{33}$$

where a parameter vector $\mathbf{p} = p_1, p_2, ..., p_j, ..., p_J$ is to be determined on the grounds of n ($n > J$) data x_i measured at given values $\mathbf{y}_i = y_{i1}, y_{i2}, ..., y_{iN}$ of the independent vector variable $\mathbf{y} = y_1, y_2, ..., y_N$; the analytical form of T is a priori given.

2.4.2 Adjustment according to the most frequent values

In Eq. (33) the weight q_i is well defined according to the conception of the most frequent value; then instead of $(x_i - M)^2$ in φ we obviously have to write $(x_i - T(\mathbf{p}_k; \mathbf{y}_i))^2$, where $\mathbf{p}_k$ is the result of the previous iteration step. We have then to fulfil

$$\sum_{i=1}^{n} \frac{\varepsilon^2}{\varepsilon^2 + (x_i - T(\mathbf{p}_k; \mathbf{y}_i))^2} (x_i - T(\mathbf{p}_{k+1}; \mathbf{y}_i))^2 = \text{min.,} \tag{34}$$

for every iteration step; i.e. the program can be constructed on the basis of the weighted mean squares program, which is for different T-s ready available in program libraries.

Often is T a polinomial, or more generally

$$T(\mathbf{p}; \mathbf{y}) = p_1 T_1(\mathbf{y}) + p_2 T_2(\mathbf{y}) + ... + p_J T_J(\mathbf{y}), \tag{35}$$

with T_j-s of given analytical form. In this case Eq. (35) leads to a linear algebraic equation system—similarly as does the principle of least squares.

Solving this equation system we have to carry out

$$\varepsilon_{m+1}^2 = \frac{3 \sum_{i=1}^{n} \dfrac{\varepsilon_m^4 d_i^2}{[\varepsilon_m^2 + d_i^2]^2}}{\sum_{i=1}^{n} \dfrac{\varepsilon_m^4}{[\varepsilon_m^2 + d_i^2]^2}} \tag{36}$$

with

$$d_i = (x_i - T(\mathbf{p}_{k+1}; \mathbf{y}_i)) \tag{36a}$$

($\mathbf{p}_{k+1}$ is already known). Evidently Eq. (28) is a special case of Eq. (36).

The parameter vector $\mathbf{p}$ of the same T function can be determined also fulfilling the demand of the least squares principle

$$\sum_{i=1}^{n} (x_i - T(\mathbf{p}; \mathbf{y}_i))^2 = \text{min.} \tag{37}$$

on the grounds of the same measuring data x_i. In general the results of Eq. (37) and that of the M-fitting do not agree, therefore these results are denoted as $\mathbf{p}_E$ and $\mathbf{p}_M$, respectively. The index E refers to the expected value (as to the result of the least squares adjustment in the simplest case); for brevity, we shall refer to the adjustment according to the principle of least squares as E-fitting.

2.4.3 Heuristic comparison of the E-fitting and the M-fitting

Both fitting techniques have the same starting point: that $\mathbf{p}$ is declared as 'true' (i.e. most probable), to which the maximum 'reliability' of the measured values x_i belongs. The difference lies in the different characterization of the 'reliability'.

The classical train of thought measures the reliability with the quantity

$$(x_i - T(\mathbf{p}; \mathbf{y}_i))^2 \tag{37a}$$

in the sense that x_i is more reliable if this expression is smaller.

From the point of view of physics (and also of common sense) it is hardly acceptable that the reliability tends to infinity if expression (37a) tends to zero. The simplest solution for solving the problem is if we accept

$$\varepsilon^2 + (x_i - T(\mathbf{p}; \mathbf{y}_i))^2 \tag{38}$$

instead of (37a) where ε is a characteristic value for the absolute value of errors. If the error of an x_i (i.e. the absolute value of the difference $(x_i - T(\mathbf{p}; \mathbf{y}_i))$ is $\varepsilon/10$, and that of an x_k is $\varepsilon/33$, both are *equivalently* 'very good' values according to expression (38) (differing only by 1% and 1‰, respectively, from the minimum value of (38)). In flat contradiction to this, expression (37a) qualifies (without any real ground) x_k as a much better value than x_i: accepting (37a), x_k seems more accurate by an order of magnitude than x_i.

The second step is to characterize the reliability of the *whole set* of data. This is done by the method of least squares by summarizing the characteristics of the separate x_i-s given in the form of (37a), resulting naturally in a very high sensitivity to outliers. To avoid similar effects and to remain as simple as possible, we consider the *product* of the characteristics given in (38) as the characteristic value of the whole set of data and the minimum value of this product is demanded:

$$\prod_{i=1}^{n} [\varepsilon^2 + (x_i - T(\mathbf{p}; \mathbf{y}_i))^2] = \min. \tag{38a}$$

After logarithmization we get the condition

$$\sum_{i=1}^{n} \ln [\varepsilon^2 + (x_i - T(\mathbf{p}; \mathbf{y}_i))^2] = \min. \tag{38b}$$

which is to be regarded as the generalized form of the condition given in Eq. (30).

T is often given by Eq. (35) (with T_j-s of known analytical form), and after differentiation according to all p_j-s we get

$$\sum_{i=1}^{n} \frac{\varepsilon^2}{\varepsilon^2 + (x_i - T(\mathbf{p}; \mathbf{y}_i))^2} T_1(\mathbf{y}_i) \left[p_1 \cdot T_1(\mathbf{y}_i) + p_2 \cdot T_2(\mathbf{y}_i) + \ldots + p_J \cdot T_J(\mathbf{y}_i) - x_i \right] = 0$$

$$\sum_{i=1}^{n} \frac{\varepsilon^2}{\varepsilon^2 + (x_i - T(\mathbf{p}; \mathbf{y}_i))^2} T_2(\mathbf{y}_i) \left[p_1 \cdot T_1(\mathbf{y}_i) + p_2 \cdot T_2(\mathbf{y}_i) + \ldots + p_J \cdot T_J(\mathbf{y}_i) - x_i \right] = 0$$

$$\vdots \tag{39}$$

$$\sum_{i=1}^{n} \frac{\varepsilon^2}{\varepsilon^2 + (x_i - T(\mathbf{p}; \mathbf{y}_i))^2} T_j(\mathbf{y}_i) \left[p_1 \cdot T_1(\mathbf{y}_i) + p_2 \cdot T_2(\mathbf{y}_i) + \ldots + p_J \cdot T_J(\mathbf{y}_i) - x_i \right] = 0.$$

The unknowns (the p_j-s) are also in the denominator; the iterative solution uses here in every step the parameter vector obtained in the preceding iteration step. As also Eq. (36) must be fulfilled, we have got in a very simple heuristic way essentially the same result as in paragraph 2.4.2 by generalizing the most frequent value calculations.

The M-fitting gives a hypersurface defined by the densest lying points, neglecting outliers. The result of the E-fitting, on the contrary, can be heavily influenced by the latter ones.

2.4.4 Weighted adjustment according to the most frequent values

The q_i weights given a priori express the fact that the measured data are, e.g. because of the different accuracy of the equipment used, not of equal worth. If, for example, $q_{n+1} = 2$ for x_{n+1}, but for all other data $q_i = 1$ holds

$(i = 1, ..., n)$, then a possible solution is to take the last summand twice into account both in Eq. (37) and in Eq. (38b) but with the weight value 1. Consequently, we can interpret a q_i weight as a number of equal data (even if q_i is not an integer).

Equation (38b) can therefore be transcribed rather mechanically to

$$\sum_{i=1}^{n} q_i \ln \left[\varepsilon^2 + (x_i - T(\mathbf{p}; y_i))^2 \right] = \min. \tag{40}$$

and the modified form of the iteration formula for calculating the dihesion (Eq. (36)) is

$$\varepsilon_{k+1}^2 = \frac{3 \sum_{i=1}^{n} \dfrac{q_i \varepsilon_k^4 d_i^2}{[\varepsilon_k^2 + d_i^2]^2}}{\sum_{i=1}^{n} \dfrac{q_i \varepsilon_k^4}{[\varepsilon_k^2 + d_i^2]^2}}. \tag{41}$$

(the d_i-values are defined in Eq. (36a)).

Chapter 3
Practical examples

Prior to more (but unavoidable) theoretical chapters let us show on some practical examples that M-fitting can really give quite different results from E-fitting (i.e. adjustment according to the principle of least squares). A general statement that M-fitting is always better than E-fitting would not be justifiable even by a very great number of examples – moreover this statement is not true as though rarely a probability distribution can occur in the very neighbourhood of the Gaussian one and for treating this case the E-fitting is better (assuming that no outliers can occur). Only in Chapter 8 will it be absolutely clear in which sense is the M-fitting 'better'.

3.1 Example for estimating the location

When calculating the most frequent value M of the sample $x_1, x_2, ..., x_n$ and its arithmetic mean, significant or even great differences are often found, and it is not easy to decide which is nearer the true value. In view of this our example shows a case for which the true value is known.

Figure 11 shows 25 resistivity values measured in a coal mine with the same electrode arrangement but at different locations; the geometrical arrangement relative to the coal seam was also always the same. Here, the author wishes to thank Dr. Á. Gyulai, University of Miskolc, Geophysical Department, for the data. The true value is known from many foregoing series of measurements

carried out for the same coal seam and with the same electrode arrangement. This value is 0.151 ohm; we see in Fig. 11 that the difference of the arithmetic mean (E_n) from the true value is twice as great as that of the most frequent value (M). The asymmetric distribution of the data is determined by the fact that tectonic disturbances decrease this specially measured resistance, and the extent of this decrease provides mining engineers with information on the nature, distance, etc. of the tectonic disturbance [Csókás et al. 1979]. Any increase in accuracy here is obviously of great practical importance.
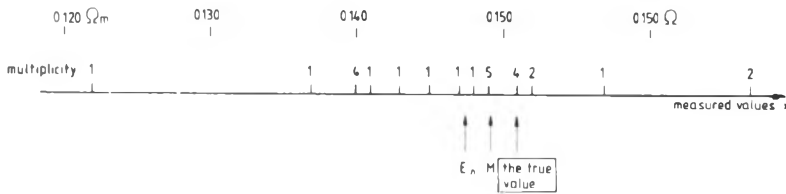
Fig. 11. Results of in-mine resistivity measurements with special electrode arrangement. The most frequent value M is significantly nearer the true value than the arithmetic mean (E_n)

11. ábra. Speciális elektróda elrendezéssel mért bányabeli elektromos ellenállásmérések eredményei. Az M leggyakoribb érték jelentősen közelebb van a helyes értékhez, mint az E_n számtani átlag

Рис. 11. Результаты подземных измерений электросопротивления со специальной установкой. Наиболее частое значение M значительно ближе к правильному значению, нежели среднеарифметическое E.

3.2 Fitting of straight lines

We often assume a linear connection between two sets of measured values. In this section two examples are treated.

3.2.1 Connection between two mineral contents at different depth intervals

In this paragraph we deal with measuring data originating from a mineral mine (the author is indebted to K. Mészáros, geologist, for the data). Mészáros' paper, read at the University of Miskolc in 1971, dealt with the same material, and it indicates that the investigation of such links can give valuable information for solving geological problems.

The two mineral contents are denoted by x and y *(Fig. 12)*. Full lines indicate the results of M-fitting, dotted lines those of the E-fitting. Figure *12/a* and *b* belong to neighbouring depth intervals.

In Fig. 12/a there is only a negligible difference between the two straight lines. In contrast, in Fig. 12/b the position of the dotted line (i.e. the result of the E-fitting) is strongly influenced by two outliers; the result of the M-fitting is similar in this case to that in Fig. 12/a, indicating proportionality between x and y.
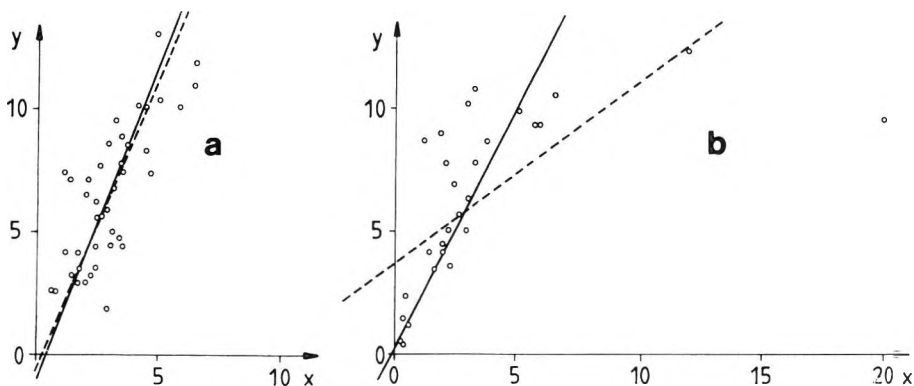
Fig. 12. Interconnection of two mineral contents (denoted by x and y) in neighbouring depth intervals. Full lines show the results of the adjustment according to the most frequent value, dotted lines correspond to least squares results. These latter may be heavily influenced by outliers (12/b shows that two outliers are enough to distort the results). Full lines indicate proportionality in both cases

12. ábra. Kétféle fémtartalom (x-szel és y-nal jelölve) szomszédos mélységszakaszokra. Folytonos egyenesek mutatják a leggyakoribb érték szerinti kiegyenlítés eredményeit, a szaggatottak a legkisebb négyzetek elve alapján kapott eredményeket. Az utóbbiakat nagymértékben befolyásolhatják kieső adatok: a 12/b ábra mutatja, hogy két ilyen adat elég ahhoz, hogy teljesen eltorzítsa a legkisebb négyzetes eredményeket, míg a folytonos vonallal rajzolt egyenesek mindkét esetben arányosságra utalnak

Рис. 12. Содержания двух различных металлов (обозначенные x и y) по соседним интервалам глубин. Сплошными прямыми показаны результаты выравнивания по наиболее частым значениям, а пунктирными – результаты, полученные по принципу наименьших квадратов. На последние сильное влияние оказывают выпадающие данные: по рис. 12/b видно, что достаточно двух таких данных, чтобы полностью исказить результаты с наименьшими квадратами, в то время как сплошными прямыми в обоих случаях отмечается пропорциональность.

3.2.2 Telluric straight lines

LANDY and LANTOS [1982] carried out systematic investigations concerning telluric straight lines. Their statistical conclusions are very important (it is of no importance that the telluric method is no longer regarded as modern). The main content of the cited article is given in the following.

The well known connection between the telluric vector of the measuring station (u, v) and that of the basic station (x, y) is written [NEMESI 1963] in the form:

$$\frac{u}{y} = a\frac{x}{y} + b$$

and   (42)

$$\frac{v}{x} = d\frac{y}{x} + c.$$

The tensor consisting of *a*, *b*, *c* and *d* depends on the geoelectric parameters below both stations (in the two layer model; if the specific electric conductivity is negligible in the second layer, the tensor gives relative depth values). Equation (42) shows that the value pairs $(u/y, x/y)$ define a straight line (as do the pairs $(v/x, y/x)$), resulting in two tensor components.

The points in *Fig. 13* represent measured values; the *M*-fitting (full line) and the *E*-fitting give nearly the same result.

This situation, however, is rather exceptional. The custom sanctified both by time and success was the following: an experienced geophysicist had cancelled (with a degree of subjectivity) the outliers before performing *E*-fitting. This procedure was called 'interpretation by hand'.
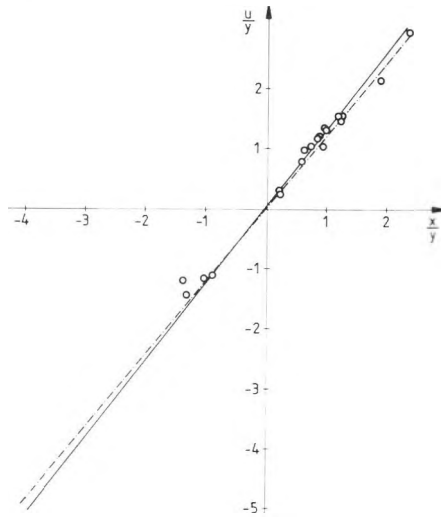
Fig. 13. Interpretation of telluric measurements. Both fittings give nearly the same result

13. ábra. Tellurikus mérések értelmezése: mindkét kiegyenlítés közel azonos eredményre vezet

Рис. 13. Интерпретация результатов измерений методом теллурических токов: выравнивание обоими методами приводит к одному и тому же результату.

Figure 14 shows a typical case. The result of *M*-fitting of *all* points (full line) is nearly the same as the result of *E*-fitting *without* the cancelled points. The *E*-fitting of all points gives an axial section (see Fig. 14) which results in a completely unrealistic depth value.

LANDY and LANTOS [1982] deal with many such cases for which classical procedures started to become uncertain; as well as the 'interpretation by hand' both *M*-fitting and *E*-fitting were performed without cancelling the outliers. The question arises as to which result can be regarded as the best. The answer is not quite unambiguous but because of the potential character of the telluric field we can accept that result as the most probable which shows minimum differences in comparison with those of the neighbouring stations. Accepting this

criterion, the consequence was the following, comparing the results of the M-fitting with those of the 'interpretation by hand':

the results of the M-fitting are better:	44%
the results of the 'interpretation by hand' are better:	22%
the differences are negligible:	22%
neither the result of M-fitting, nor the result of the 'interpretation by hand' is acceptable:	12%

Shortly speaking, the M-fitting was in the overwhelming majority of these problematic cases successful whereas the E-fitting gave acceptable results only
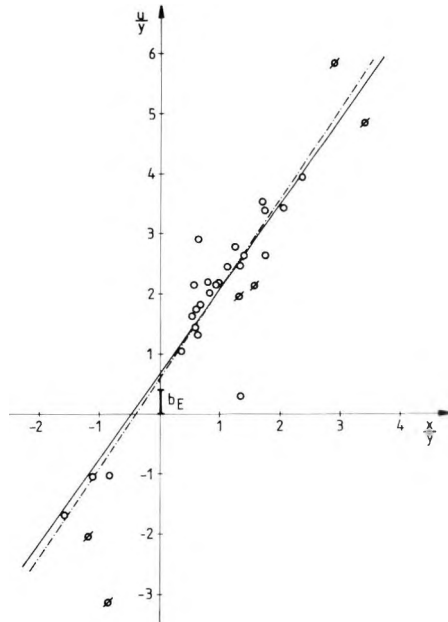
Fig. 14. Interpretation of telluric measurements. The M-fitting of all points (full line) gives nearly the same result as the E-adjustment without the crossed circles (dotted line). The E-fitting of all points gives the intercept value denoted by b_E, resulting in a completely unreal depth to basement

14. ábra. Tellurikus mérések értelmezése. Az összes pontot figyelembe vevő M-kiegyenlítés (folytonos vonal) közel azonos eredményre vezet, mint az áthúzott körök nélküli legkisebb négyzetes kiegyenlítés (szaggatott vonal). Az összes pont E-kiegyenlítése a b_E-vel jelölt tengelymetszetet szolgáltatja, amely viszont teljesen irreális alaphegység-mélységre vezet

Рис. 14. Интерпретация результатов измерений методом теллурических токов. Выравнивание M (сплошная линия), при котором во внимание принимаются все точки, приводит примерно к тому же результату, что и выравнивание по способу наименьших квадратов (пунктирная линия) без прочеркнутых кругов. Выравнивание E по всем точкам приводит к осевому сечению, обозначенному b_E, которое дает совершенно нереальную глубину залегания фундамента.

in 33% of these cases. Acceptability is not always sufficient: we strive occasionally after greater accuracy, too. This is reached if *n* is great enough — let us recall the law of large numbers.

LANDY and LANTOS [1982] show examples of how many points define with the prescribed accuracy the tensor components *a* and *b*. The answer depends significantly upon the type of fitting chosen.

A typical example is shown in *Fig. 15*. Short horizontal straight lines on the right hand side of the figure indicate the 'true' *a* and *b* values obtained with very careful manual interpretation, taking all available points into consideration. The values on the abscissa are the numbers of those points which were taken into consideration by carrying out both *M*- and *E*-fittings (full and dotted lines).
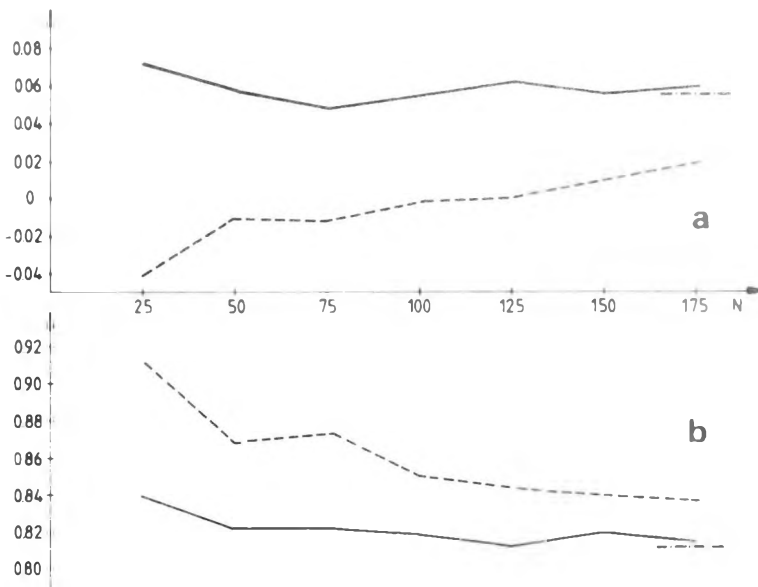
Fig. 15. Two examples for the interpretation of telluric measurements. The same number of data can give much more accurate results calculating with the most frequent value (full lines) than with the least squares method (dotted lines). The true values are indicated on the right side with short horizontal lines

15. ábra. Két példa tellurikus mérések értelmezésére: azonos adatszám sokkal pontosabb eredményre vezethet a leggyakoribb értékek szerint számolva (folytonos vonalak), mint a hagyományos legkisebb négyzetes módszer (szaggatott vonalak). A helyes értékeket a jobb szélen rövid vízszintes egyenes szakaszok jelzik

Рис. 15. Два примера интерпретации результатов измерений методом теллурических токов. При том же количестве данных расчет по наиболее частым значениям приводит ко значительно более точным результатам (сплошные линии), нежели расчет по традиционному методу наименьших квадратов (пунктир). Правильные значения обозначены на правой стороне короткими горизонтальными отрезками прямых.

Beyond $n = 60$ or 80, neither a nor b varies significantly if M-fitting was made (in Fig. 15 the fluctuation of both quantities is about 1% if $n > 80$). In the contrary, dotted lines show that twice as great n ($n \approx 150$) is still not enough to stabilize the values a and b, to say nothing about the significant differences from the true values even in this n-region. The author considers it superfluous to enhance the importance of the economical aspects by a prescribed accuracy —or that of the aspects of the surplus of geological information to a fixed value of n.

3.3 Quadratic fitting with two variables: the determination of the magnetic normal field in Hungary

Let us deal first, in a general way, with the quadratic M-fitting with two variables.

3.3.1 Iteration steps of quadratic M-fitting with two variables

If the T-function has the analytical form

$$T(\mathbf{p}; \mathbf{y}) = p_1 + p_2 y_1 + p_3 y_2 + p_4 y_1^2 + p_5 y_1 y_2 + p_6 y_2^2,$$

then Eq. (39) becomes the following:

$$p_1 \sum_{i=1}^{n} \varphi_i + p_2 \sum_{i=1}^{n} y_{1i}\varphi_i + p_3 \sum_{i=1}^{n} y_{2i}\varphi_i + p_4 \sum_{i=1}^{n} y_{1i}^2\varphi_i + p_5 \sum_{i=1}^{n} y_{1i}y_{2i}\varphi_i +$$

$$+ p_6 \sum_{i=1}^{n} y_{2i}^2\varphi_i = \sum_{i=1}^{n} x_i\varphi_i$$

$$p_1 \sum_{i=1}^{n} y_{1i}\varphi_i + p_2 \sum_{i=1}^{n} y_{1i}^2\varphi_i + p_3 \sum_{i=1}^{n} y_{1i}y_{2i}\varphi_i + p_4 \sum_{i=1}^{n} y_{1i}^3\varphi_i + p_5 \sum_{i=1}^{n} y_{1i}^2 y_{2i}\varphi_i +$$

$$+ p_6 \sum_{i=1}^{n} y_{1i}y_{2i}^2\varphi_i = \sum_{i=1}^{n} x_i y_{1i}\varphi_i$$

$$p_1 \sum_{i=1}^{n} y_{2i}\varphi_i + p_2 \sum_{i=1}^{n} y_{1i}y_{2i}\varphi_i + p_3 \sum_{i=1}^{n} y_{2i}^2\varphi_i + p_4 \sum_{i=1}^{n} y_{1i}^2 y_{2i}\varphi_i + p_5 \sum_{i=1}^{n} y_{1i}y_{2i}^2\varphi_i +$$

$$+ p_6 \sum_{i=1}^{n} y_{2i}^3\varphi_i = \sum_{i=1}^{n} x_i y_{2i}\varphi_i$$

$$p_1 \sum_{i=1}^{n} y_{1i}^2 \varphi_i + p_2 \sum_{i=1}^{n} y_{1i}^3 \varphi_i + p_3 \sum_{i=1}^{n} y_{1i}^2 y_{2i} \varphi_i + p_4 \sum_{i=1}^{n} y_{1i}^4 \varphi_i + p_5 \sum_{i=1}^{n} y_{1i}^3 y_{2i} \varphi_i +$$

$$+ p_6 \sum_{i=1}^{n} y_{1i}^2 y_{2i}^2 \varphi_i = \sum_{i=1}^{n} x_i y_{1i}^2 \varphi_i$$

$$p_1 \sum_{i=1}^{n} y_{1i} y_{2i} \varphi_i + p_2 \sum_{i=1}^{n} y_{1i}^2 y_{2i} \varphi_i + p_3 \sum_{i=1}^{n} y_{1i} y_{2i}^2 \varphi_i + p_4 \sum_{i=1}^{n} y_{1i}^3 y_{2i} \varphi_i +$$

$$+ p_5 \sum_{i=1}^{n} y_{1i}^2 y_{2i}^2 \varphi_i + p_6 \sum_{i=1}^{n} y_{1i} y_{2i}^3 \varphi_i = \sum_{i=1}^{n} x_i y_{1i} y_{2i} \varphi_i$$

$$p_1 \sum_{i=1}^{n} y_{2i}^2 \varphi_i + p_2 \sum_{i=1}^{n} y_{1i} y_{2i}^2 \varphi_i + p_3 \sum_{i=1}^{n} y_{2i}^3 \varphi_i + p_4 \sum_{i=1}^{n} y_{1i}^2 y_{2i}^2 \varphi_i + p_5 \sum_{i=1}^{n} y_{1i} y_{2i}^3 \varphi_i +$$

$$+ p_6 \sum_{i=1}^{n} y_{2i}^4 \varphi_i = \sum_{i=1}^{n} x_i y_{2i}^2 \varphi_i \qquad (43)$$

where

$$\varphi_i = \frac{\varepsilon^2}{\varepsilon^2 + (x_i - p_1 - p_2 y_{1i} - p_3 y_{2i} - p_4 y_{1i}^2 - p_5 y_{1i} y_{2i} - p_6 y_{2i}^2)^2} . \qquad (43a)$$

With $\varphi_i \equiv 1$ the equation system is naturally the same as by the E-fitting. Using this to start the procedure, the results are denoted by $p_1^{(1)}, p_2^{(1)}, ..., p_6^{(1)}$. The first d_i set will consequently be calculated as

$$d_i = x_i - p_1^{(1)} - p_2^{(1)} y_{1i} - p_3^{(1)} y_{2i} - p_4^{(1)} y_{1i}^2 - p_5^{(1)} y_{1i} y_{2i} - p_6^{(1)} y_{2i}^2. \qquad (44)$$

The starting value of ε according to Eq. (26) is

$$\varepsilon_{st} = \frac{\sqrt{3}}{2} [\max(d_i) - \min(d_i)]. \qquad (44a)$$

The iteration according to Eq. (36) is to be made several times; with its result ε and with the values $p_1^{(1)}, ..., p_6^{(1)}$ the weights φ_i are to be calculated according to Eq. (43a). Putting the φ_i values in Eq. (43), the linear algebraic equation system is solved once more. The results are denoted by $p_1^{(2)}, ..., p_6^{(2)}$; the new d_i-set is to be calculated according to

$$d_i = x_i - p_1^{(2)} - p_2^{(2)} y_{1i} - ... - p_6^{(2)} y_{2i}^2. \qquad (44b)$$

With the last ε and with the new d_i-values several step are to be made according to Eq. (36). With the obtained value of ε and with the parameters $p_1^{(2)}, ..., p_6^{(2)}$, the φ_i-values must once more be calculated, the equation system is solved resulting in the values $p_1^{(3)}, ..., p_6^{(3)}$, etc. The procedure is stopped if the p_j values vary only insignificantly.

3.3.2 The example of the magnetic normal field in Hungary

In this paragraph an example given by STEINER [1980] is discussed.

The normal field of all magnetic components is described in Hungary with quadratic polynomials of two variables. The first one is the latitude φ minus a fixed value $\varphi_0 = 45.5°$, denoted by $\Delta\varphi$; the second one is the longitude λ minus $\lambda_0 = 16°$ (both are measured in minutes). The vertical component of the normal magnetic field (denoted by Z_n and measured in γ-s) has then the analytical form:

$$Z_n(\varphi_0 + \Delta\varphi, \lambda_0 + \Delta\lambda) = Z_0 + a\,\Delta\varphi + b\,\Delta\lambda + c(\Delta\varphi)^2 + d\,\Delta\varphi\,\Delta\lambda + e(\Delta\lambda)^2. \quad (45)$$

Obviously

$$Z_0 = p_1, p_2 = a, ..., p_6 = e.$$

On the basis of 296 data [ACZÉL and STOMFAI 1968] both M-fitting and E-fitting was carried out. The results are the following:

	M(296)	E(296)	
Z_0	+ 40835.9	+ 40829.7	
a	+ 10.6441	+ 10.8892	
b	+ 1.26563	+ 1.24046	(46)
c	− 0.00570512	− 0.00268327	
d	− 0.000726043	− 0.00412334	
e	+ 0.000555918	+ 0.00132706	

Which parameter vector represents more closely the physical reality? From the geomagnetic point of view, the answer to this question would need a long discussion of very different nature in comparison with the theme dealt with in the present monograph.

A possible solution of this problem is to carry out both fittings for only half of the data, as a first step. The results are given below:

	M(148)	E(148)	
Z_0	+ 40821.8	+ 40793.6	
a	+ 10.7667	+ 11.4433	
b	+ 1.43344	+ 1.53378	(47)
c	− 0.00612962	− 0.00640111	
d	− 0.00101322	− 0.00354293	
e	+ 0.000022289	+ 0.000204031	

The second step is to calculate the differences: in what measure do the parameters depend upon n? As the quotients of the differences show,

	ΔM (296–148)	ΔE (296–148)	$\Delta E/\Delta M$	
Z_0	14.1	36.1	2.56	
a	0.1626	0.5541	4.52	
b	0.1678	0.2933	1.75	(47a)
c	0.0004245	0.003718	8.76	
d	0.0002872	0.0005804	2.02	
e	0.0005336	0.001123	2.10	

the dependence on n in the case of the M-fitting is about half or less than that by the E-fitting, i.e. for the M-fitting fewer data are needed to find the regularity. (This important result is analogous to that dealt with in connection with Fig. 15.)

FERENCZY [1980] has investigated in detail the application of M-fitting to magnetotelluric problems. The results are very interesting but their discussion even in a very shortened form is beyond the scope of this monograph.

3.4 Weighted adjustments

3.4.1 General remarks

It was seen in Chapter 2 that M-fittings can be carried out as a twofold iteration, one of them being—mathematically—in every step a weighted E-fitting. It is perhaps not superfluous to underline as a first general remark that M-fitting is *not* a weighted E-fitting: in the latter case the weights are a priori known; on the other hand, the weights used in M-fitting are dependent upon the sample, these weights are also results of the statistical procedure. Analogously, an irrational number remains irrational even if it is given as the limit of a series of rational numbers.

The second remark: if we can choose arbitrarily the values of an independent variable (or more independent variables), we often prefer equidistancy; e.g. field measurements of profiles are often made equidistantly, and if gravity mapping is carried out, measuring stations are mostly points of a quadratic grid. Figures 12–14 have shown cases in which the values of the abscissa are not equidistant: in these cases it would be absurd to suppose that equidistance is to be realized because these values are not chosen arbitrarily. Although the fitting has given very reliable results, the question arises: Is it not highly problematic when—perhaps as a consequence of some type of recording—one x_1-value occurs, say, 10,000 times more frequently than another x_2-value? This situation can really cause difficulties as the measured value at x_2 can be neglected even if this x_2-value represents an important x-interval in the special problem studied. The solution can be a weighting, in its original sense—but the appropriate system of this weighting depends upon the investigated problem itself. Consequently the attribute 'a priori' can have a relative sense, too.

3.4.2 Determination of the q(H) function by means of calculating dihesions

Let us suppose the simplest case, namely $T = \text{const}$. In some cases the q_i-values are trivially known: equipment number one is twice as accurate as equipment number two. A more frequently occurring case is, however, that we know that q_i (the a priori weight of x_i) definitely depends upon another known quantity, say H—but the dependence is not given even in a general analytical form. (H can have the meaning of temperature, pressure, etc.—by all means a measurable quantity.) In this case we can write $q(H_i)$ instead of q_i; $q(H)$ denotes the functional dependence not yet known.

To determine this $q(H)$ function, let the first step be an ordering: after this $H_i \geq H_l$ holds if $i > l$. The second step is the appropriate choice of two integer numbers: of m and r, in that manner that the relations $n \gg m$ and $m \geq r$ must be fulfilled. As third step the ε_j dihesions are to be determined for all x_i subsets given by their indices as follows:

$$\begin{array}{c}
1, \ldots, m \\
1 + r, \ldots, m + r \\
1 + 2r, \ldots, m + 2r \\
\vdots \\
1 + (J-1)r, \ldots, m + (J-1)r.
\end{array}$$

To get a reliable ε_j value it is obvious that m cannot be too small.

Let us denote the maximum ε_j value by $\varepsilon_{\max}$; this characterizes the 'worst' x_i subset. The minimum weight 1 (which could be chosen arbitrarily) belongs therefore to the mean value of those H_i-s which are defined with the indices of this 'worst' subset.

The already known meaning of the a priori weight allows us to interpret an arbitrary other subset as if its elements x_i had been calculated on the grounds of q_i pieces of 'worst' x-values by means of some estimation (for the sake of simplicity, say, as arithmetic means). Because of the well known '$1/\sqrt{n}$-law' of estimates we can write

$$\frac{\varepsilon_{\max}}{\sqrt{q_j}} = \varepsilon_j; \tag{48}$$

finally, the weight q_j of the mean value of H_i-s corresponding to the x_i subset in question is given by

$$q_j = \frac{\varepsilon_{\max}^2}{\varepsilon_j^2}. \tag{48a}$$

The (H_j, q_j) points define principally the function $q(H)$, but other secondary points of view can be considered, too. Commonly simple analytical forms are chosen for $q(H)$.

3.4.3 Gravity example demonstrating a priori weight calculations and to show the differences between results of weighted and unweighted fittings of both kinds

The well known formula for the elevation correction of a gravity measurement is

$$g_P^{(corr)} = g_P^{(m)} + kh_P \qquad (49)$$

where h_P is the height above sea level of the measuring point P; $g_P^{(m)}$ is the measured g-value for the point P, other (normal and topographical) corrections also included; $g_P^{(corr)}$ is the Bouguer anomaly. The value k depends in a simple known form upon the mean density; this latter also being unknown, it is easier to regard k as the unknown which is to be determined.

STEINER [1959] has transcribed the well known graphic method of NETTLETON [1939] to a numerical method. The fundamental idea of Nettleton is that that the topographical fluctuations must not appear in the Bouguer map, neither in the positive nor in the negative sense. Consequently the $g^{(corr)}$-values at the corners of an elementary quadrangle of the measuring grid define nearly a plain —supposing that the grid was dense enough—and if the elementary quadrangle is a parallelogram (see *Fig. 16*) then the following equation is approximately fulfilled:

$$G^{(corr)} \equiv g_1^{(corr)} + g_3^{(corr)} - g_2^{(corr)} - g_4^{(corr)} = 0. \qquad (50)$$

Analogously to $G^{(corr)}$ we introduce $G^{(m)}$ and H:

$$G^{(m)} \equiv g_1^{(m)} + g_3^{(m)} - g_2^{(m)} - g_4^{(m)} \qquad (50a)$$

$$H \equiv h_1 + h_3 - h_2 - h_4. \qquad (50b)$$
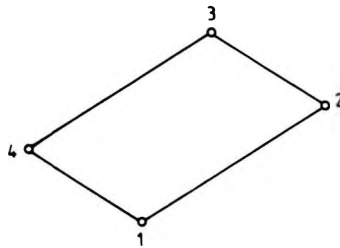
Fig. 16. Four neighbouring points of a gravimetric field measurement, lying nearly at the corners of a parallelogram

16. ábra. Graviméteres terepmérés közelítőleg parallelogramma csúcspontjain fekvő négy szomszédos mérőpontja

Рис. 16. Четыре соседних точки измерений гравиразведки, находящихся примерно на вершинах параллелограммы.

Taking Eq. (49) into account, it is obvious that the following also holds:

$$G^{(corr)} = G^{(m)} + kH. \tag{51}$$

In the case of general quadrangles the expressions in Eqs. (50), (50a) and (50b) are more complicated but from Eq. (49), Eq. (51) follows in just the same manner [STEINER 1959]. Since $G^{(corr)} = 0$ according to Eq. (50),

$$k_j = -\frac{G_j^{(m)}}{H_j} \tag{52}$$

gives a value for k on the grounds of the data of the i-th elementary quadrangle.

If on the whole measuring area the quadrangles are constructed (see in *Fig. 17* a small part of a map), Eq. (52) gives for every quadrangle a k_j-value. These data have very different weights as H_j may also be very near zero and in this case the k_j-value in Eq. (52) is determined only by statistical fluctuations.

By appropriate numbering of the apices of the elementary quadrangles all H_i-s have positive values and the method dealt with in item 3.4.2 can directly be applied; the dependence of the weight q on H is evident.

In the following a Hungarian example will be treated [STEINER 1982]. The number of quadrangles was 470; $m = 100$ and $r = 25$ were used (see paragraph 3.4.2). Some dihesions of k_j-values are given to a mean value H_i (this latter index belongs to the subsets): $\varepsilon_1 = \varepsilon_{max} = 0.2102$ mGal/m to $H_1 = 1.627$ m; $\varepsilon_2 = 0.1276$ mGal/m to $H_2 = 2.531$ m; $\varepsilon_3 = 0.0851$ mGal/m to $H_3 = 3.448$ m, etc; the weights according to our agreement (i.e. that the 'worst' subset has the weight 1,) and according to Eq. (48a) are the following: $q_1 = 1$; $q_2 = 2.7$; $q_3 = 6.1$, etc.
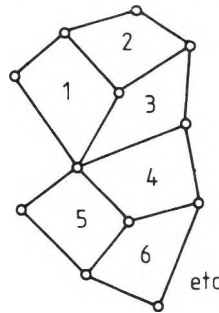
Fig. 17. Quadrangles defined by measurement points

17. ábra. Mérési pontok által definiált négyszögek

Рис. 17. Четырехугольники, определяемы точками измерений.

The points in *Fig. 18* correspond to the pairs (H_i, q_i). Figure 18 shows that q can be regarded as zero in the interval $0 \leq H \leq 2.5$, and the points define for $H > 2.5$ the straight line $q = 3.43 \ (H - 2.5)$. An additional point of view is, however, that too great weights are not wanted: in this case the final value of k would be determined by only a few quadrangles of very great H_j. But great H_j-s can occur also in cases if the quadrangle is too great (because of the difficulties of field measurements) and in this case $G^{(corr)} = 0$ may be wrongly approximated. The fixation of an upper limit for the q-values therefore seems to be unavoidable; it is certain that our choice $q_{max} = 100$ (see Fig. 18) is somewhat arbitrary but the most important cause of the weighting is the appropriate suppression of the k-values of poor quality—and this does not depend on the q_{max}-value.
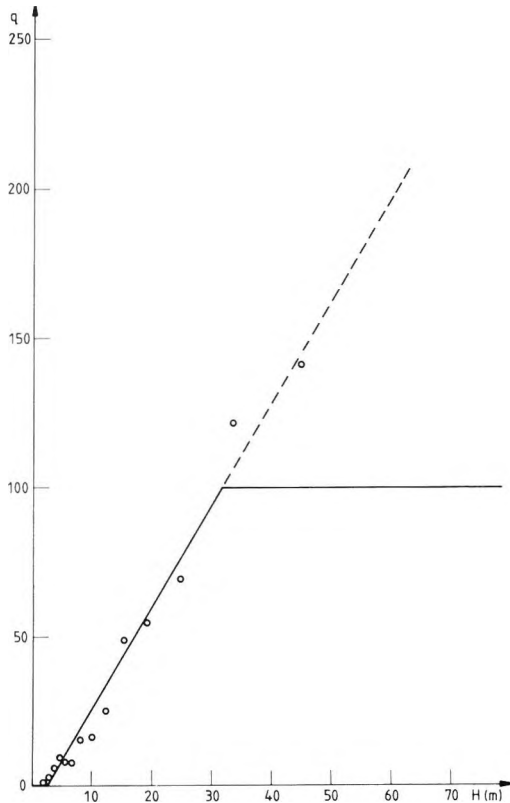
Fig. 18. Weight function to calculate both weighted M- and weighted E-fittings

18. ábra. Súlyfüggvény súlyozott M-, és súlyozott E-kiegyenlítés végrehajtásához

Рис. 18. Весовая функция для выполнения выравниваний по взвешенным M и по взвешенным E.

Weighted fittings were made therefore with the $q(H)$ function:

$$q(H) = \begin{cases} 0 & , & \text{if} & H \leq 2.5 \\ 3.43\,(H-2.5), & \text{if} & 2.5 < H \leq 31.65 \\ 100 & , & \text{if} & H > 31.65. \end{cases} \tag{53}$$

From the 470 k_j data there were chosen 10 values randomly 200 times, as well as 40 values 200 times; for these 400 samples weighted and unweighted means as well as most frequent values were determined. Eight frequency diagrams were constructed for the four types of estimates for $n = 10$ and $n = 40$. Sample medians and sample quartiles of both types (upper and lower) were determined, see *Fig. 19;* the differences of the latter ones divided by 2 are equal to the probable error.
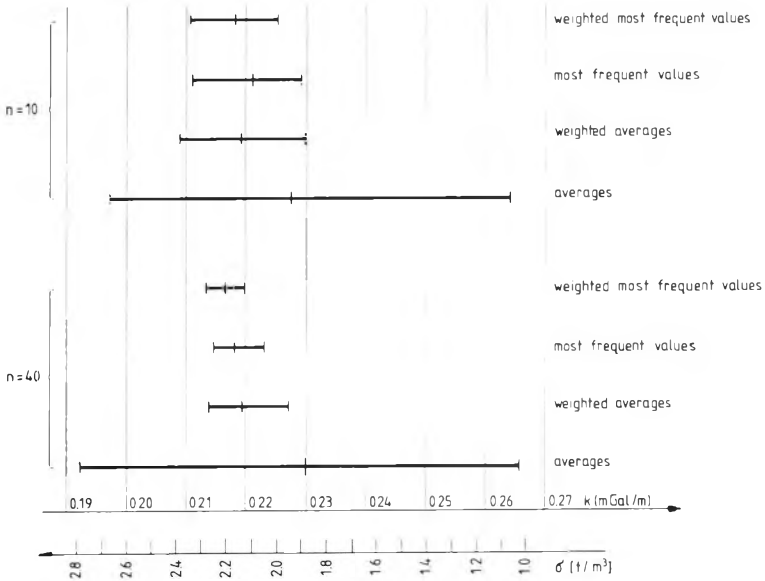
Fig. 19. Comparison of four statistical procedures: weighted and simple most frequent values and means, respectively. For $n = 10$ and $n = 40$ the medians and the interquartile ranges of the four sorts of estimates are given

19. ábra. Négy statisztikai eljárás eredményeinek összehasonlítása: súlyozott és súlyozatlan leggyakoribb értékek és átlagok. Adottak a négyfajta becslési módra a mintamediánok és interkvartilis terjedelmek, $n = 10$ és $n = 40$ esetére

Puc. 19. Сопоставление результатов по четырем статистическим способам: взвешенные и не взвешенные наиболее частые значения и средние. Для всех четырех способов оценки даются медианы и межквартильные интервалы для $n = 10$ и $n = 40$.

The consequences are the followings:

    a) the probable error of the arithmetic means does not depend on n (this property is well known for Cauchy-distributed data, we can therefore suppose that the k_j-set is not far from this distribution type);

b) arithmetic means and most frequent values significantly differ from one another: by $n = 40$ the whole interquartile interval of the most frequent values is far from the sample median of arithmetic means (and the density value of the latter is also geologically improbable in the area investigated);

c) as a consequence of the weighting, the probable error of the most frequent values diminishes (to achieve this visually not too great diminution a data surplus of 50% would be needed), consequently, if the accuracy is to increase, a weighted M-fitting is appropriate;

d) both alternatives of the most frequent values show a decrease of about 50% with regard to their probable errors, comparing the results for $n = 40$ with those for $n = 10$ (i.e. the law of $1/\sqrt{n}$ holds for both types of M-fitting);

e) the weighted means are much more accurate than the arithmetic means;

f) for both n-values the sample median of the weighted means is near the medians of weighted and unweighted most frequent values (and is far from the median of the arithmetic means);

g) the interquartile range of weighted means is significantly less for $n = 40$ than for $n = 10$, but the decrease is more moderate than would be prescribed by the $1/\sqrt{n}$-law;

h) it can be stated from the geophysical point of view that a demand to know the mean density σ with a probable error of 0.05 t/m^3 would need practically the whole data set (according to the $1/\sqrt{n}$-law), even if we calculate weighted most frequent values;

and, finally, a conclusion of fundamental importance:

i) for great n-values the probable error of unweighted most frequent values can be significantly less than that of the weighted means, i.e. it may occur that we 'automatically' get a more reliable result by calculating unweighted M-fitting than by weighted E-fitting (as is well known, the determination of weights is often a cumbersome procedure).

The discussion above has clearly shown that E-fitting and M-fitting are two different statistical procedures, each having a weighted version, too.

Chapter 4
Main definitions and theorems
of robust statistics

4.1 Some general remarks belonging to the theory of robust statistics

The first publications dealing with various estimating procedures, recently called as robust ones, were written by those who applied them [NEWCOMB 1886]. This is hardly surprising since practical tasks force those dealing with them to search for suitable methods that are fully adequate in terms of reality. The

developing of an exact theory, however, is naturally the territory of mathematicians and statisticians. (A well known analogous case is the introduction by Dirac of the notion 'delta function'. Now a mathematically correct theory of generalized functions is available to define correctly the Dirac delta function, and, moreover, it is possible on the basis of this theory even to criticize Dirac's original definition. See, for example, the entry 'Diracsches Functional' in NAAS and SCHMID [1965]).

It is usual to reckon the development of robust statistics beginning from the article of HUBER [1964] (who also wrote the first monograph on this topic: HUBER [1981]). The rapid development of computing possibilities has been one of the causes of the rapidly growing interest in this topic. In 1972 ANDREWS et al. published a great deal of information about robust statistics (this was not done in fully ordered form), and the increasing interest in the seventies is illustrated by our Fig. 2.

The theory of robust statistics deals with the so-called *M*-, *L*- and *R*-estimates, with the so-called minimax methods, etc. Our brief outline contains information nearly exclusively about the *M*-estimates (and about some notions of central importance in robust statistics, e.g. IC-curve, asymptotic variance). The reasons are the following: a) the Princeton study of various robust estimates [ANDREWS et al. 1972] has shown the advantages of *M*-estimates; b) we shall see that the most frequent value is an *M*-estimate, too.

4.2 Generalized maximum likelihood estimates (M-estimates)

4.2.1 Generalization of the maximum likelihood method with an arbitrary ϱ-function

Let us suppose that only the location parameter is unknown. By a given distribution type $f(x)$ and on the grounds of a sample $x_1, ..., x_n$ the demand

$$\sum_{i=1}^{n} \ln f(x_i, T) = \text{maximum}, \tag{54}$$

or

$$\sum_{i=1}^{n} -\ln f(x_i, T) = \text{minimum} \tag{54a}$$

must be fulfilled: the resulting *T*-value is the maximum likelihood estimate of the location parameter. The generalization of HUBER [1964] is the following: we write ϱ instead of $-\ln f$ (ϱ is a differentiable function); in this case the condition

$$\sum_{i=1}^{n} \varrho(x_i, T) = \text{minimum} \tag{55}$$

results in an estimate *T* for the location parameter, even if the distribution type $f(x)$ is unknown. HUBER [1964] has denoted this by '*M*-estimate' since this is

the generalization of the maximum likelihood procedure. (Our symbol M from the name 'most frequent value' can be regarded as an eventual but lucky coincidence: we have already mentioned above that our algorithm is a special M-estimation.)

4.2.2 The ψ-function and its connection with ϱ and φ. The χ-function

It is more practical to solve

$$\sum_{i=1}^{n} \psi(x_i, T) = 0 \tag{56}$$

where

$$\psi(x, t) = \frac{\partial \varrho(x, T)}{\partial T}, \tag{56a}$$

than to use Eq. (55) in its original form. The most important characteristics are commonly expressed by the ψ-function; even the estimation method itself is often defined immediately by the ψ-function.

In that T is an estimate of location, we can write instead of the general Eq. (55)

$$\sum_{i=1}^{n} \varrho(x_i - T) = \text{minimum} \tag{56b}$$

and similarly instead of Eq. (56)

$$\sum_{i=1}^{n} \psi(x_i - T) = 0. \tag{56c}$$

Introducing

$$\varphi(x_i) = \frac{\psi(x_i - T)}{x_i - T} \tag{57}$$

with the remark that $\varphi(T)$ is to be calculated according to the rule of l'Hospital, we get

$$\sum_{i=1}^{n} \varphi(x_i) \cdot (x_i - T) = 0; \tag{58}$$

its solution is

$$T = \frac{\sum_{i=1}^{n} \varphi(x_i) \cdot x_i}{\sum_{i=1}^{n} \varphi(x_i)}. \tag{58a}$$

The right hand side of Eq. (58a) is the same as that of Eq. (18) (in the latter case φ was unambiguously defined by Eq. (17) and therefore M was written in

Eq. (18) instead of T). We see that the most frequent value is really an M-estimate.

In the case of the estimation of the scale parameter the condition

$$\sum_{i=1}^{n} \varrho(x_i, S) = \text{minimum} \tag{59}$$

is to be fulfilled (the analytical form of ϱ in Eq. (59) being naturally different from the ϱ used in location parameter determinations). The function χ is analogous to ψ as here

$$\sum_{i=1}^{n} \chi\left(\frac{x_i - T}{S}\right) = 0 \tag{60}$$

is to be fulfilled (with a known or estimated T-value). If both T and S are unknown, that pair of values is to be accepted which simultaneously fulfils Eqs. (58a) and (60). (The weight function depends in general upon the S value, too; let us recall the weight function in Eq. (17) with $S = \varepsilon$).

Which scale parameter is estimated by Eq. (60)? The answer is perhaps surprising but at the same time quite natural: the S defined by

$$\int_{-\infty}^{\infty} \chi\left(\frac{x - T}{S}\right) f(x)\,dx = 0 \tag{60a}$$

is estimated. T, however, also obtained by solving

$$\sum_{i=1}^{n} \psi\left(\frac{x_i - T}{S}\right) = 0 \tag{61}$$

(this is a slightly more detailed form of Eq. (56c)), is the estimate of T defined by

$$\int_{-\infty}^{\infty} \psi\left(\frac{x - T}{S}\right) f(x)\,dx = 0. \tag{61a}$$

The functions ψ and χ must have advantageous behaviour (in this sense they are not arbitrary); e.g. a fundamental practical demand concerning the ψ-functions is that they must give as T the symmetry point if $f(x)$ is symmetrical.

4.3 IC-functions and their calculation based on ψ- and χ-functions

4.3.1 Definition of the IC-function

If f is the original density function (supposed as a continuous one) and T is the location parameter of the distribution given in its argument, let us define

the IC-function as

$$IC(x, f, T) = \lim_{t \to 0} \frac{T[(1-t)f + t\delta(x)] - T[f]}{t}, \tag{62}$$

where δ is the Dirac function and t the probability of the (surplus) occurrence of the value x. (This definition was given by HAMPEL [1968]). In the argument of IC the symbol T refers to the estimation procedure studied. The symbol 'IC' itself refers to the synonym '*in*fluence *c*urve'.

In the theory of robust statistics the IC-function plays a central role. The practical applicability of it lies in the fact that if n is large then a surplus datum of value x modifies the estimate T with the value

$$\Delta T = \frac{IC(x, F, T)}{n}. \tag{62a}$$

The influence of outliers is therefore quantitatively given by the IC-function for the estimating procedure studied. But also if the distributions are 'clean', the IC-function informs us whether the value x has positive or negative, small or great influence on the 'formation' of the T-value for a given distribution f. This latter interpretation will play an important role in Section 4.4.

4.3.2 Calculation of IC-functions based on ψ- and χ-functions

Analogously to Eq. (62) the IC-function of every other statistical parameter can be defined too, e.g. also the IC-function of the scale parameter. In reality both quantities, T and S (characterized by the ψ- and χ-function), are simultaneously calculated therefore the IC-functions are also to be determined simultaneously for general distributions on the basis of the equation system:

$$IC(x, T) \int_{-\infty}^{\infty} \psi'\left(\frac{y-T}{S}\right) f(y)\, dy +$$

$$+ IC(x, S) \int_{-\infty}^{\infty} \psi'\left(\frac{y-T}{S}\right) \frac{y-T}{S} f(y)\, dy = S\psi\left(\frac{x-T}{S}\right)$$

$$\tag{63}$$

$$IC(x, T) \int_{-\infty}^{\infty} \chi'\left(\frac{y-T}{S}\right) f(y)\, dy +$$

$$+ IC(x, S) \int_{-\infty}^{\infty} \chi'\left(\frac{y-T}{S}\right) \frac{y-T}{S} f(y)\, dy = S\chi\left(\frac{x-T}{S}\right)$$

If the distribution is symmetrical, the IC-functions for T and S can be given separately. For simplicity, it is supposed that the symmetry point is at the origin. The IC-formulae have then the form (see Eq. (63)):

$$IC(x, F, T) = \frac{S}{\displaystyle\int_{-\infty}^{\infty} \psi'\left(\frac{y}{S}\right) f(y)\,dy} \, \psi\left(\frac{x}{S}\right) \tag{64}$$

and

$$IC(x, F, S) = \frac{S}{\displaystyle\int_{-\infty}^{\infty} \chi'\left(\frac{y}{S}\right) \frac{y}{S} f(y)\,dy} \, \chi\left(\frac{x}{S}\right). \tag{65}$$

Equation (64) has a greater practical importance. The denominator being a constant value, Eq. (64) expresses the fact that the ψ-function is proportional to the IC-function. This statement gives the most obvious meaning of the ψ-function.

The effectiveness of the estimation of location parameters is of crucial importance for practitioners. Therefore a short immediate proof of Eq. (64) is given below. Let us suppose that $S = 1$.
Substituting the modified distribution

$$g(x) = (1 - t)f(x) + t\delta(x - x_0)$$

into Eq. (61a), we get

$$(1 - t) \int_{-\infty}^{\infty} \psi(x - T)f(x)\,dx + t(x_0 - T) = 0.$$

Differentiating this according to t, we can express $\dfrac{dT}{dt}$ in the form:

$$\frac{dT}{dt} = \frac{\psi(x_0 - T) - \displaystyle\int_{-\infty}^{\infty} \psi(x - T) \cdot f(x)\,dx}{(1 - t) \cdot \displaystyle\int_{-\infty}^{\infty} \psi'(x - T)f(x)\,dx + t \cdot \psi'(x - T)}$$

Taking Eq. (62) into consideration, it is clear that $\dfrac{dT}{dt}$ equals IC, if $t \to 0$. The second expressions both in the numerator and in the denominator then become zero (in the first case as a consequence of Eq. (61a), in the second case trivially). As x_0 was arbitrary, we can write simply x, or, if $S \neq 1$, x/S the equivalence to Eq. (64) is then obvious (for $T = 0$).

(Putting $g(x)$ defined above in Eqs. (60a) and (61a) and differentiating according to t, we also get the general equation system (63)).

4.4 Calculation of asymptotic variances of estimates

4.4.1 A simple heuristic proof of the general formula for the asymptotic variance

Let us suppose that the continuous density function $f(x)$ characterizes a general (i.e. not necessarily symmetrical) probability distribution. The probability of the event that the random variable has values around x in the short Δx interval is $f(x)\Delta x$. If Δx decreases so that no distinction is possible by the given measurement technique in that interval, $f(x)$ can be substituted by a Dirac function of $\Delta = f(x)\Delta x$ at point x *(Fig. 20)*. (If $\Delta \ll 1$ were not fulfilled, Δx is to be diminished until this relation is satisfied. The value x in the inteval Δx is to be situated in such a manner that the probability for greater or less values than x are equally probable.)
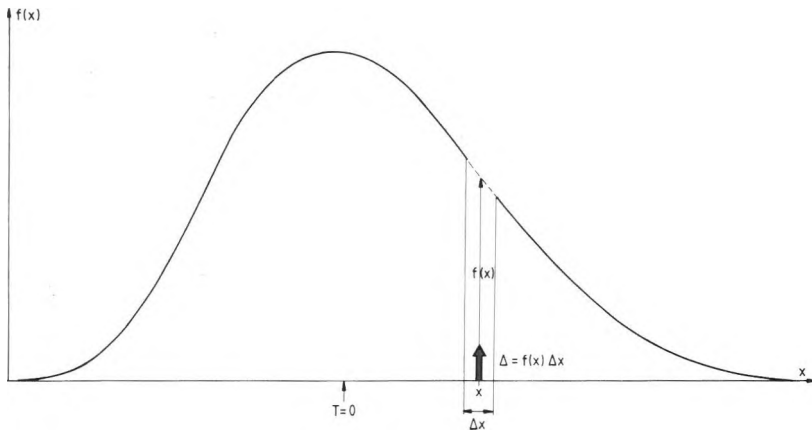
Fig. 20. A short Δx interval can be represented by a Dirac function instead of the continuous line of $f(x)$

20. *ábra.* Egy rövid, Δx-hosszúságú intervallum Dirac-függvénnyel helyettesíthető, az $f(x)$ folytonos sűrűségfüggvény megfelelő szakasza helyett

Рис. 20. Короткий интервал длиною Δx заменим о Дирака вместо соответствующего отрезка непрерывной плотностной функции $f(x)$.

The Dirac delta function in x 'participates' with the value

$$\Delta T = IC(x, f, T) \cdot \Delta \tag{66}$$

in the 'formation' of the T value (see Eq. (62)). If Δ is really very small, the whole x-axis can be divided into small intervals all having the same probability. If the number of intervals is n then $\Delta = 1/n$ and Eq. (66) is the same as Eq. (62a).

If n is large, after a well known theorem of Glivenko the frequency diagram constructed on the basis of the sample is very near $f(x)$. In other words, the

distance (say, the Prohorov-distance) between the sample—as a 'density function' consisting of Dirac delta functions each of the measure $1/n$—and the original $f(x)$ density function is negligible. Therefore the series of $n = 1/\Delta$ pieces of substituting Dirac delta functions (defined above at x-locations according to $f(x)$) is regarded as a sample (a so-called 'ideal sample'). This is justified by the fact that we want to derive an asymptotic result for a joint effect of the sample elements on the resulting T.

Let us denote the deviation caused by the j-th sample element by ΔT_j, whose sample variance $D^2_{n,j}$ is

$$D^2_{n,j} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\mathrm{IC}(x_i)}{n} \right)^2 = \frac{1}{n^3} \sum_{i=1}^{n} [\mathrm{IC}(x_i)]^2, \tag{67}$$

Every value of the x-series $x_1, \ldots, x_n$, defined above as the j-th sample element, can occur with equal probability for a given sampling.

Obviously the fluctuations of T are not caused solely by a single sample element. As the effects of measure ΔT_j are summarized and since the variance given in Eq. (67) is clearly independent of j, applying a well known theorem for the variance of the sum of equally distributed random variables we get the variance D^2_n of T (estimate of location) as

$$D^2_n = n \cdot D^2_{n,j} = \frac{1}{n^2} \sum_{i=1}^{n} [\mathrm{IC}(x_i)]^2 = \frac{1}{n} \sum_{i=1}^{n} [\mathrm{IC}(x_i)]^2 \frac{1}{n}. \tag{68}$$

The probability $1/n$ equals $f(x) \cdot \Delta x$ (being not only $f(x)$ but also Δx dependent on x); consequently the integral form of Eq. (68) is the following:

$$D^2_n = \frac{1}{n} \int_{-\infty}^{\infty} [\mathrm{IC}(x)]^2 f(x) \, \mathrm{d}x. \tag{69}$$

As the definition of the asymptotic variance A^2 is given by

$$A^2 = n \cdot D^2, \tag{70}$$

we obtain the following very important formula for the asymptotic variance of estimates:

$$A^2 = \int_{-\infty}^{\infty} [\mathrm{IC}(x)]^2 f(x) \, \mathrm{d}x. \tag{71}$$

The heuristic train of thought given above shows at the same time that the estimates have Gaussian distribution (if $n \to \infty$) as the fluctuation of estimates is the sum of n equal random variables and therefore (if A^2 exists) the premises of the so-called central limit theorem are fulfilled. (The exaxt proofs of these statements are to be found in HUBER [1981] but the principal content is much more enhanced for appliers by the heuristic way followed above.) The attribute 'central' is therefore just as much justified for distributions of estimates as it is misleading for mother distributions.

4.4.2 Calculation of asymptotic variances of M-estimates

Substituting the expression of the IC-function given by Eq. (64) into Eq. (71), we get for symmetrical distributions (having their symmetry points in the origin) the asymptotic variance of the estimates of location parameters:

$$A^2(T, S) = \frac{\displaystyle\int_{-\infty}^{\infty} S^2 \psi^2 \left(\frac{x}{S}\right) f(x)\, dx}{\left[\displaystyle\int_{-\infty}^{\infty} \psi'\left(\frac{x}{S}\right) f(x)\, dx\right]^2}. \tag{72}$$

In the argument of A^2, T denotes the method of the estimation itself. If the estimation of T is carried out simultaneously with the estimation of the scale parameter, in Eq. (72) S denotes the other result of the twofold iteration (fulfilling Eq. (60a)). The symbol S can be regarded, however, as an 'independent variable', too: Eq. (72) belongs in this case to a simple iteration carried out with constant S-value. This interpretation is evidently important in looking for that value of S which results in a minimum asymptotic variance.

Substituting Eq. (65) in Eq. (71) we get the asymptotic variance of the scale parameter, too:

$$A^2 = \frac{\displaystyle\int_{-\infty}^{\infty} S^2 \chi^2 \left(\frac{x}{S}\right) f(x)\, dx}{\left[\displaystyle\int_{-\infty}^{\infty} \chi'\left(\frac{x}{S}\right) \frac{x}{S} f(x)\, dx\right]^2}. \tag{73}$$

As in the overwhelming majority of cases we have to determine T and the fluctuation of S causes in general little fluctuation of T, Eq. (73) has only secondary importance.

Utilizing the formula for the asymptotic variance of location estimates (Eq. (72)) it is possible to prove shortly the famous and practically important Cramér-Rao inequality, too.

If both sides of the well known relation

$$[\psi(x)\, f(x)]' = \psi'(x)\, f(x) + \psi(x)\, f'(x)$$

are integrated from $-\infty$ to $+\infty$, the left hand side gives zero as we have to take the values of the expression inside the square parantheses at $x = -\infty$ and $x = \infty$ and both are clearly zero (since $f(x)$ is a density function). The integrals on the right hand side are consequently equal in the absolute sense, and therefore the following also holds, viz.

$$\left[\int_{-\infty}^{\infty} \psi(x)\, f'(x)\, dx\right]^2 = \left[\int_{-\infty}^{\infty} \psi'(x)\, f(x)\, dx\right]^2$$

The integrand on the left hand side is the product of the square roots of the functions $\psi^2(x)\, f(x)$ and $\left(\dfrac{f'(x)}{f(x)}\right)^2 f(x)$. According to the SCHWARZ inequality it holds (also taking into account the equation above) that

$$\int\limits_{-\infty}^{\infty} \psi^2(x)\, f(x)\, \mathrm{d}x \int\limits_{-\infty}^{\infty} \left[\frac{f'(x)}{f(x)}\right]^2 f(x)\, \mathrm{d}x \geqq \left[\int\limits_{-\infty}^{\infty} \psi'(x)\, f(x)\, \mathrm{d}x\right]^2.$$

Writing this in the form

$$\frac{\int\limits_{-\infty}^{\infty} \psi^2(x)\, f(x)\, \mathrm{d}x}{\left[\int\limits_{-\infty}^{\infty} \psi'(x)\, f(x)\, \mathrm{d}x\right]^2} \geqq \frac{1}{\int\limits_{-\infty}^{\infty} \left[\frac{f'(x)}{f(x)}\right]^2 f(x)\, \mathrm{d}x}$$

on the left hand side we have the asymptotic variance of the estimation of the location parameter working actually with a given ψ-function (this function defines the estimation procedure itself), and on the right hand side is the well known Cramér-Rao bound depending only upon $f(x)$.

4.5 Some additional aspects and remarks on the theory of robust estimates

4.5.1 The breakdown point

The resistance of an estimation procedure means that some outliers negligibly influence the estimate. The question arises, however, what percentage of the data is tolerable as an outlier without totally spoiling the estimation (i.e. without it resulting in an estimate of location which no longer adequately characterizes the gathering of the data).

Instead of referring to some theoretical discussion of the question let us cite in brief an investigation method of ad hoc type [KERÉKFY 1978]. The samples of n elements are to be constructed in the following way: $(n - j)$ data come randomly from standard Gaussian distribution, the other data of the sample are $1 \cdot 100, 2 \cdot 100, \ldots, (j - 1) \cdot 100, j \cdot 100$. If j/n is constant, we can calculate by increasing n an asymptotic value for the estimation procedure in question. If this value is greater than 3, the estimation is totally spoilt. Usually that j/n value is given (in per cent) as the 'breakdown point', to which the limit value of 3 belongs.

The table of KERÉKFY [1978] belonging to this contains limiting j/n values in the sense mentioned above (i.e. breakdown points) for 15 estimation types. This maximum content of outliers can also be 50%, too—but this is only a statistical aspect; in actual cases the applier can be much more rigorous (tolerating only, say, maximum 10%). On the other hand, special problems (e.g. in astronomy) would tolerate also even greater than 50% of outliers to find a characteristic gathering of the data.

The calculation of arithmetic means as estimates has a very special behaviour: its breakdown point of 0% means that this method can be totally spoilt by the occurrence of an arbitrary small percentage of outliers.

4.5.2 Some robust procedures

Some simple procedures often occur in the investigations of robust statistics:

a) the sample median;

b) the α-trimmed mean (from the ordered sample the smallest and the greatest data are cancelled, both being α% of the whole amount of the data, and the arithmetic mean of the remaining is calculated);

c) the α-winsorized mean (is calculated similarly to the α-trimmed mean but the smallest data of the sample are replaced by the first remaining item of data instead of cancelling them, the greatest data are replaced by the last remaining item of data, before the arithmetic mean is calculated);

d) the trimean of TUKEY (half of the sum of the sample median and of the arithmetic mean of the sample quartiles);

e) skipped methods usually calculate arithmetic means after cancelling data which are outside an interval, say, 1.5 or 2 times the sample interquartile range;

f) the Hodges–Lehmann estimate (in the first step arithmetic means of pairs of data are calculated, composing these pairs in every possible manner; in the second step the sample median of these n^2 means is determined).

These often cited 'classical' procedures are not the most effective ones. Newer and more sophisticated methods are unfortunately also often ad hoc procedures [see e.g. 'HAMPEL's estimate of 14 April 1971', 'Hogg proposal in letter to John TUKEY' in ANDREWS et al. 1972]. The applier needs, however, procedures which fulfil the demands given in Table III.

As for the nomenclature: the methods in points a)—e) above are so-called L-estimates (linear combinations of order statistics), and in f) example was given for R-estimates (estimates derived from rank tests). The fundamental notion for L- and R-estimates is also the already known IC-function (clearly no ψ-function exists for L- and R-estimates). Equation (71) also holds for these cases (for our heuristic proof the type of estimation is of no importance).

To enable eventual quick comparisons, firstly the formula for asymptotic variance of Hodges–Lehmann estimates is given:

$$A^2 = \frac{1}{12 \cdot \left[\int\limits_{-\infty}^{\infty} f^2(x)\,dx \right]^2},$$

secondly the expression of the IC-function of α-trimmed means for symmetrical distributions:

$$IC(x) = \begin{cases} \dfrac{\text{sign}(x) \cdot F^{-1}(1-\alpha)}{1-2\alpha}, & \text{if} \quad |x| > F^{-1}(1-\alpha) \\[2ex] \dfrac{x}{1-2\alpha}, & \text{if} \quad |x| \leqq F^{-1}(1-\alpha) \end{cases}$$

[see e.g. in HUBER 1981].

4.5.3 The IC-function of arithmetic means

If it is $\varrho(.) = (.)^2$ in Eq. (56b) (least squares method for $T = \text{const.} = E$) then $IC(x) = x$ (according to Eqs. (56a) and (64)). The great influence of outliers (leading to the breakdown point of 0%) is evident. ANDREWS et al. [1972] do not indicate that any single estimation method from the investigated ones is the best, but they declare—no wonder—that the arithmetic mean is the worst one.

Substituting $IC(x) = x$ in Eq. (71) we get $\int\limits_{-\infty}^{\infty} x^2 f(x)\,dx$. (Remember that the location parameter was assumed to be zero, for the sake of simplicity.) This expression is usually called 'the variance' and is commonly used for characterizing the original (mother) distribution $f(x)$ itself although this quantity characterizes only the dispersion of *estimates* (giving the asymptotic variance of them) for the very special case *if* arithmetic means are calculated as estimates. No wonder that 'the variance' σ^2 can have infinite values, too, even if the mother distribuion $f(x)$ shows an expressed maximum of small dispersion (i.e. the interquantile ranges are small); '$\sigma^2 = \infty$' enhances only the fact that the mean

square technique is either unusable (if also the first moment is infinite) or its use is not effective (in cases of finite first moments).

Summerizing the statements above: 'the variance' σ^2 does not characterize the original (mother) distribution $f(x)$ itself; only if two distributions show the same type (and the second moments are finite) can have σ a relative meaning concerning the original distributions in the sense that σ_2/σ_1 gives the ratio of the scale parameters in question (S_2/S_1). It is misleading if we accept any kind of more general meaning of σ^2 concerning mother distributions (unfortunately most handbooks written on these topics do this).

Chapter 5
Characteristics and formulae of the most frequent values

It seemed to be advantageous to deal in a separate chapter with the application of the theory of robust statistics to the most frequent value calculations.

5.1 Characteristic functions and asymptotic variance

5.1.1 The ψ- and χ-functions by calculating most frequent value and dihesion, respectively

According to Eq. (17) the weight function has the form

$$\varphi(.) = \frac{\varepsilon^2}{\varepsilon^2 + (.)^2}. \tag{74}$$

Using standardized variables, we need only write x in the argument instead of $(x - M)$. At the same time, however, standardization means that $\varepsilon = 1$. The weight function is then

$$\varphi(x) = \frac{1}{1 + x^2} \tag{75}$$

where x is the standardized variable.
According to Eq. (57) ψ is the weight function φ multiplied by the variable, consequently

$$\psi(x) = \frac{x}{1 + x^2} \tag{76}$$

holds.

The $\chi(x)$ function belonging to the determination of the scale parameter, in our case the dihesion ε, is the following:

$$\chi(x) = \frac{3x^2 - 1}{(x^2 + 1)^2}.$$ (77)

Proof. Substitute $\dfrac{x - M}{\varepsilon}$ instead of x in Eq. (77). For samples Eq. (60)—with $S = \varepsilon$—must hold,

i.e. the equation

$$\sum_{i=1}^{n} \frac{3\left(\dfrac{x_i - M}{\varepsilon}\right)^2 - 1}{\left[\left(\dfrac{x_i - M}{\varepsilon}\right)^2 + 1\right]^2} = 0,$$ (78)

is valid. Written in another form

$$\varepsilon^2 = \frac{3 \displaystyle\sum_{i=1}^{n} \frac{(x_i - M)^2}{[\varepsilon^2 + (x_i - M)^2]^2}}{\displaystyle\sum_{i=1}^{n} \frac{1}{[\varepsilon^2 + (x_i - M)^2]^2}}$$ (78a)

holds, which formula is clearly fulfilled with the same ε as the iteration formula given in Eq. (28a) for the determination of the dihesion.

5.1.2 The IC-function for the most frequent values

Writing $S = \varepsilon$, supposing $T = M = 0$, and substituting the ψ-function given in Eq. (76) and its derivate into the general expression of IC given in Eq. (64), we get

$$IC(x, F, M) = \frac{1}{\displaystyle\int_{-\infty}^{\infty} \frac{\varepsilon^2 - y^2}{(\varepsilon^2 + y^2)^2} f(y) \, dy} \cdot \frac{x}{\varepsilon^2 + x^2}$$ (79)

for symmetrical distributions. This is for $|x| > \varepsilon$ a decreasing function in the case of its absolute values. The resistance to outliers is guaranteed to a degree defined by const./$|x|$ for large values of $|x|$; this degree is satisfactory in the overwhelming majority of cases.

5.1.3 Asymptotic variance of most frequent values

Substituting the expression of the IC-function given in Eq. (79) in the generally valid Eq. (71), we get for distributions being symmetrical to the origin (and for arbitrary ε) the asymptotic variance of the most frequent values in the form:

$$A^2(M, \varepsilon) = \frac{\displaystyle\int_{-\infty}^{\infty} \frac{x^2}{(\varepsilon^2 + x^2)^2} f(x) \, dx}{\left[\displaystyle\int_{-\infty}^{\infty} \frac{\varepsilon^2 - x^2}{(\varepsilon^2 + x^2)^2} f(x) \, dx\right]^2}.$$ (80)

If ε is not arbitrary, we have to take into account Eq. (28) which defines the dihesion in an iterative manner. After some rearranging we get (with $M = 0$)

$$\int_{-\infty}^{\infty} \frac{x^2}{(\varepsilon^2 + x^2)^2} f(x)\, dx = \frac{1}{3} \int_{-\infty}^{\infty} \frac{\varepsilon^2}{(\varepsilon^2 + x^2)^2} f(x)\, dx. \tag{80a}$$

Using this relation several times in both directions, we get a very simple formula for $A^2(M, \varepsilon)$:

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

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

$$= \frac{1}{\displaystyle\int_{-\infty}^{\infty} \frac{4/3 \cdot \varepsilon^2}{(\varepsilon^2 + x^2)^2} f(x)\, dx} =$$

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

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

$$= \frac{1}{\displaystyle\int_{-\infty}^{\infty} \frac{1}{\varepsilon^2 + x^2} f(x)\, dx} = \frac{\varepsilon^2}{n(\varepsilon)}. \tag{81}$$

We have seen (see Eq. (22)) that $n(\varepsilon)$ gives the ratio of the number of effective data divided by n. The result is therefore not only simple but also immediately acceptable.

A statement of principal importance is [CSERNYÁK and STEINER 1983b] that *the expression $A^2 = \varepsilon^2/n(\varepsilon)$ is always finite*, i.e. the law of $1/\sqrt{n}$ for the increase of the accuracy is proved for all symmetrical distributions. For asymmetrical distributions is such a theoretical result not yet known but the Monte Carlo investigations of extreme asymmetrical cases has also given satisfactory results (see paragraph 5.2.2).

5.2 Examples for comparing the asymptotic behaviour of the most frequent value with that of the arithmetic mean

5.2.1 Symmetrical examples for $n \to \infty$

The density functions of 18 distributions that are symmetrical to the origin are contained in *Table IV*. For each distribution there are given the dihesion ε, the $n(\varepsilon)$–value (see Eq. (22)), the asymptotic scatter (the square root of the asymptotic variance) of the most frequent values $(\varepsilon/\sqrt{n(\varepsilon)})$, the asymptotic scatter of the arithmetic means (σ, often known as 'the scatter'), information about the fulfilment of the law of large numbers for arithmetic means (as already shown above, for the most frequent values the fulfilment is guaranteed in the 'best form', i.e. the accuracy always increases with $1/\sqrt{n}$), and, finally, the ratio of the asymptotic variances in question (the relative efficiency of the most frequent value calculations referring to that of the arithmetic means), denoted as n_E/n_M since this ratio gives at the same time (in an asymptotic sense) the ratio of numbers of data needed to achieve the same accuracy. If the integral $\int_{-\infty}^{\infty} x^2 f(x)\, dx$ is divergent, in the column for σ the sign ∞ is written.

The classification of the 18 distributions in four groups was made based on the different asymptotic behaviour of arithmetic means. These four possibilities are: 1) the reverse of the law of large numbers is fulfilled, i.e. the greater the number of data (n) the less the accuracy; 2) the law of large numbers is not fulfilled and the accuracy is independent of n; 3) the law of large numbers is fulfilled but the estimation of location has poor efficiency (being $\sigma = \infty$); 4) the law of large number is fulfilled, the accuracy increases with $1/\sqrt{n}$.

If $n_E/n_M < 1$, the use of arithmetic means is more economical than the estimating of location with the most frequent values. Distributions of such type (see numbers 13, 14 and 17 in Table IV) are, however, sterile in the sense of CSERNYÁK [1984], i.e. in those cases the flanks of the distributions either completely vanish (14 and 17) or are smaller than would be a real model of distributions mostly occurring in the earth sciences. The Gaussian distribution (number 13) is an example of unreal small flanks, at least from the viewpoint

of geophysics and geology; a surplus of 36% is needed for the same accuracy, using most frequent values instead of arithmetic means (the latter ones are well known optimal estimates in the case of Gaussian distribution). After suitable generalization of the concept of the most frequent values—given in Chapter 8— it is possible to approximate arbitrarily the efficiency of 100% even for Gaussian type distributions (if—rather exceptionally—approximately such a type occurred in the practice).

5.2.2 Monte Carlo results for symmetrical and asymmetrical distributions

Although the case of symmetrical distribution was dealt with in the previous paragraph, it nonetheless seams useful to show on some examples what the reverse fulfilment of the law of large numbers and the non-dependence on *n* mean, because these cases are not commonly known. Two asymmetrical examples are treated as well.

The sampling (with *n* = 10 and 100) was made randomly 200 times for all 5 distributions investigated and for each sample the arithmetic mean *E* and the most frequent value *M* were determined. Four frequency diagrams were constructed for each distribution to show the results in such a manner that the

$$f(x) = \frac{3\sqrt{3}}{8\pi} \cdot \frac{1}{|x|^{3/2}+1}$$
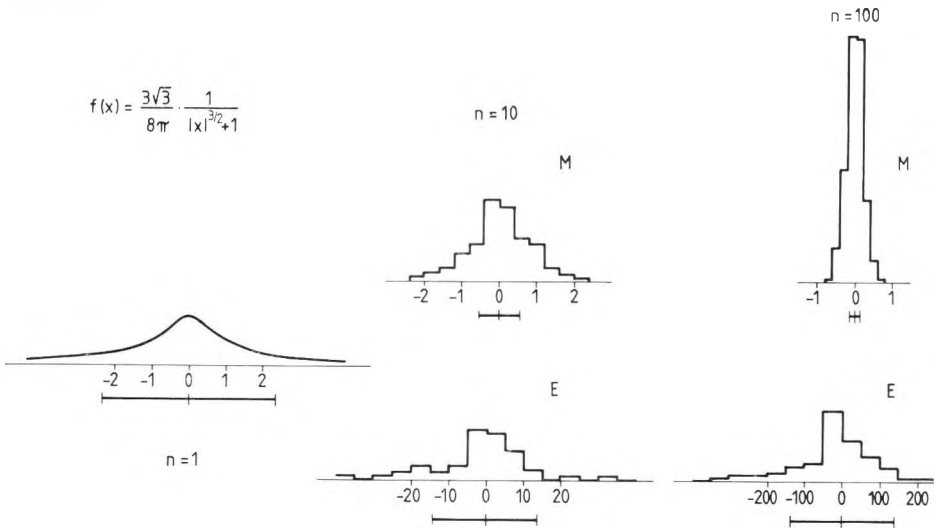
Fig. 21. No 4 probability distribution type of Table IV. It is noteworthy that arithmetic means become less accurate with increasing *n* (see scale of abscissae of *E*-diagrams) whereas most frequent values become more and more accurate

21. ábra. A IV. táblázat 4. eloszlás-típusa. Az átlagok pontatlanabbakká válnak *n* növekedésével (1. az *E*-diagramok abszcissza-skáláit), míg a leggyakoribb értékek egyre pontosabbak

Рис. 21. Функция распределения 4. по таблице IV. При возрастании *n* средние становятся неточными (см. шкалы абсцисс диаграмм *E*), в то время как наиболее частые значения — все более точными.

Figures 21–25. Results of a Monte Carlo study. The frequency diagrams are given for sample sizes $n = 10$ and $n = 100$, and for arithmetic means (E) and most frequent values (M). The mother distribution is given by its density function (both analytically and by showing the curve $n = 1$). Medians and interquartile ranges are marked in all cases

21–25. ábra. Egy Monte Carlo vizsgálat sorozat eredményei. A gyakorisági diagramok $n = 10$-re és 100-ra adottak mind az E számtani átlagokra, mind az M leggyakoribb értékekre. Az anyaeloszlás sűrűségfüggvényével adott (mind analitikusan, mind az $n = 1$ jelű görbével). Mindegyik esetben bejelöltük a mediánokat és az interkvartilis terjedelmet

Рис. 21–25. Результаты исследований методом Монте Карло. Частотные диаграммы приводятся для $n = 10$ и $n = 100$ как для среднеарифметических E, так и для наиболее честых значений M. Материнская функция задана плотностной функцией, как аналитически, так и кривой «$n = 1$». Во всех случаях обозначены медианы и межквартильные интервалы.
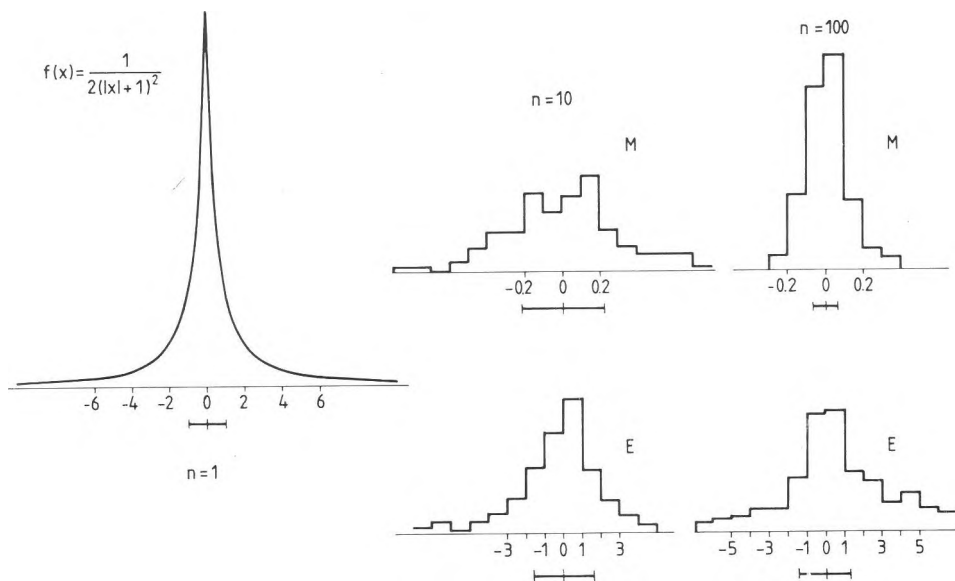
Fig. 22. No 6 probability distribution type of Table IV. It is noteworthy that the accuracy of the arithmetic means does not seem to depend on n (the interquartile range is almost the same at $n = 10$ and $n = 100$,) whereas the accuracy of the most frequent values increases with $\sqrt{n}$

22. ábra. A IV. táblázat 6. eloszlás-típusa. Úgy tűnik, az átlagok pontossága nem függ n-től (az interkvartilis terjedelem csaknem azonos $n = 10$-nél és 100-nál), míg a leggyakoribb értékek pontossága $\sqrt{n}$-nel növekszik

Рис. 22. Функция распределения 6. по таблице IV. Точность средних, видимо, не зависит от n (межквартильные интервалы почти одни и те же при $n = 10$ и $n = 100$), в то время как точность наиболее частых значений увеличивается с $\sqrt{n}$.

$$f(x) = \frac{3\sqrt{3}}{4\pi} \cdot \frac{1}{|x|^3 + 1}$$
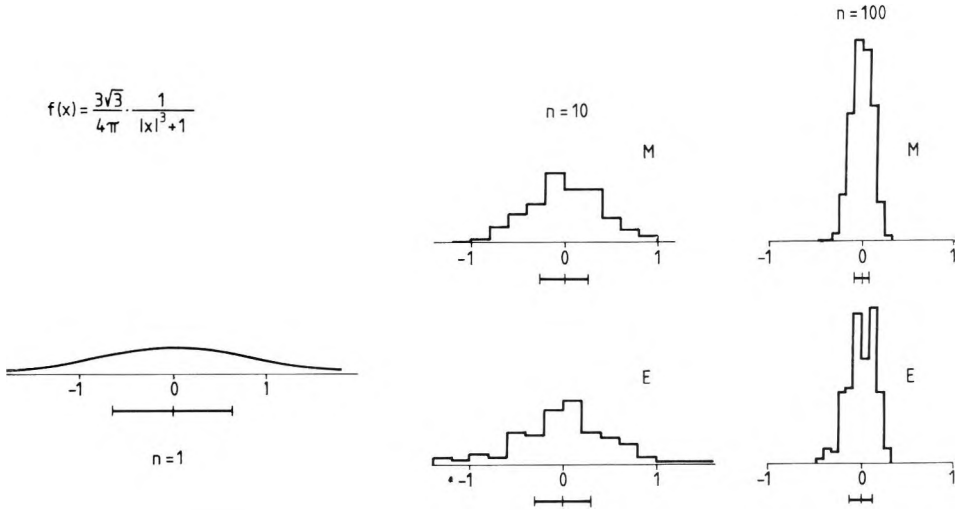
Fig. 23. No 11 probability distribution type of Table IV. The behaviour of the arithmetic means seems to be statisfactory in the investigated sample sizes, although the value of σ^2 (i.e. the asymptotic variance of the arithmetic means) is infinite in this case

23. ábra. A IV. táblázat 11. eloszlás-típusa. Az átlagok viselkedése kielégitőnek látszik a vizsgált mintaelemszámoknál, noha a σ^2 értéke (azaz a számtani átlagok aszimptotikus szórása) ebben az esetben végtelen

Рис. 23. Функция распределения 11. по таблице IV. Поведение средних кажется удовлетворительным, хотя значение σ^2 (тоесть асимптотическая дисперсия среднеарифметических) и бесконечно в этом случае.

surface under the frequency diagram is always the same (see *Figs. 21–25*). The density function itself is also given in each figure ($n = 1$).

Attention should be given to the scale of the abscissae: in the first and last example the E-diagrams clearly show the reverse fulfilment of the law of large numbers. To the contrary, M-diagrams show that the most frequent values also behave according to the $1/\sqrt{n}$-rule in these extreme cases.

The last example can be treated analytically, too. This type of distribution is called Smirnov-distribution [GNEDENKO and KOLMOGOROV 1949], the density function of it being

$$f(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2x}} \cdot x^{-\frac{3}{2}}, & \text{if} \quad x > 0 \\ 0, & \text{if} \quad x \leq 0 \end{cases} \tag{82}$$

The characteristic function has the form

$$\exp\left\{ -\sqrt{t}\left[1 - j\,\frac{t}{|t|}\,\text{tg}\,\frac{\pi}{4} \right] \right\} \tag{82a}$$

(j is the unit imaginary number). According to the well known convolution theorem the sum of n elements has the characteristic function

$$\exp\left\{-n\sqrt{t}\left[1-j\frac{t}{|t|}\,\text{tg}\,\frac{\pi}{4}\right]\right\},\tag{82b}$$

i.e. the arithmetic mean is also Smirnov-distributed with the density function

$$f^{(n)}(x)=\begin{cases}\dfrac{1}{n\sqrt{2\pi}}\left(\dfrac{x}{n}\right)^{-\frac{3}{2}}\cdot e^{-\frac{1}{2\left(\frac{x}{n}\right)}}, & \text{if}\quad x>0\\[4mm]0, & \text{if}\quad x\leq 0\end{cases}\tag{82c}$$
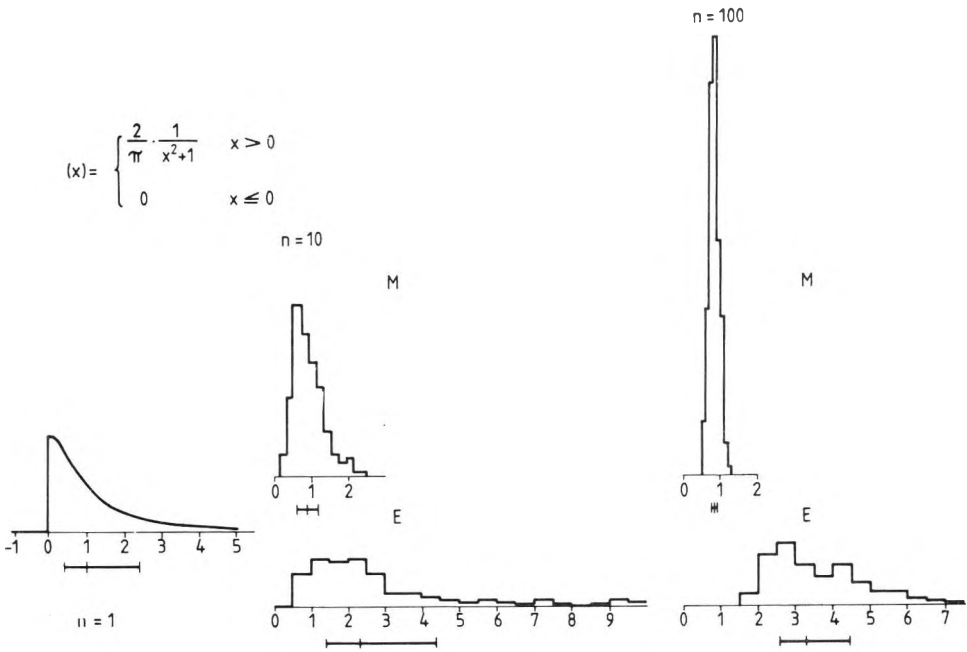
Fig. 24. Probability distribution type with density function given in the figure. With increasing n arithmetic means do not provide characteristic values for the mother distribution, in contrast to the situation with most frequent values

24. ábra. A számtani átlagok növekvő n-nel nem adnak az anyaeloszlást egyre pontosabban jellemző értéket, szemben a leggyakoribb értékek megnyugtató viselkedésével

Рис. 24. При возрастании среднеарифметические не приближаются все точнее к величине, характеризующей исходное распределение в противоположность хадежному поведению наиболее частых значений.
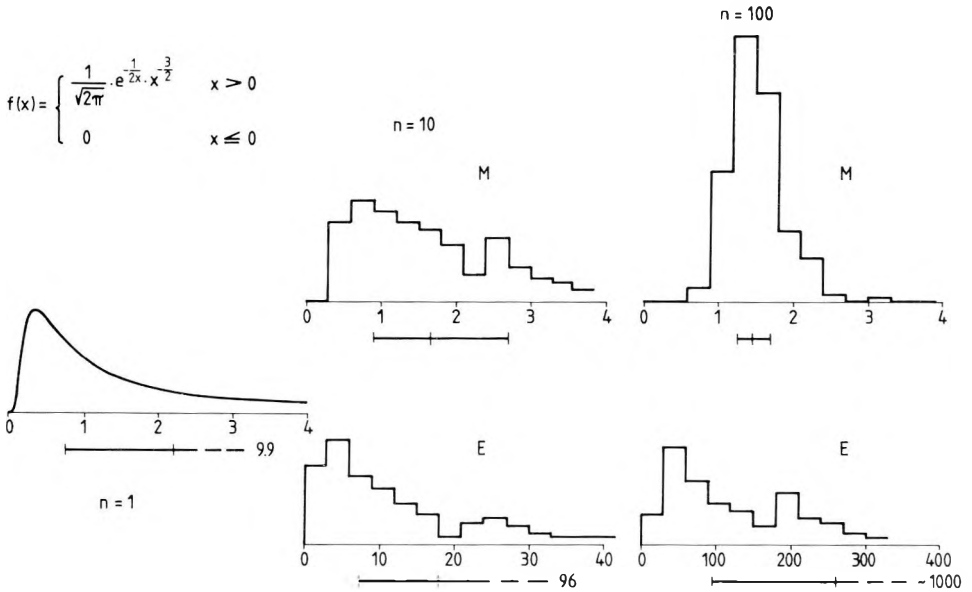
Fig. 25. Probability distribution type with density function given by Eq. (82) (Smirnov-type distribution). The sample interquartile range increases rapidly with increasing *n* (see scale of abscissae of *E*-diagrams), to say nothing about the enormous right-ward trend of the median. On the contrary, the accuracy of most frequent values increases with increasing *n*

25. ábra. A (82) egyenlet szerinti sűrűségfüggvénnyel megadott eloszlás-tipus (Szmirnov-eloszlás). Az empirikus interkvartilis terjedelem gyorsan növekszik *n* növekedésével (1. az *E*-diagramok abszcissza-skáláit), nem is beszélve a medián nagymértékű jobbra tolódásáról. Ezzel szemben a leggyakoribb értékek egyre pontosabbak az *n* növekedésével

Рис. 25. Тип распределения (распределение Смирнова), заданный функцией плотности по уравнению (82). Эмпирическая интерквартильная широта быстро возрастает при возрастании *n* (см. шкалы по сои абсцисс на *E*-диаграммах), не говоря уже о значительном смещении медианы вправо. В противоположность этому, наиболее частые значения все более точны при возрастании *n*.

This means that not only the interquartile range of arithmetic means increases proportionally with *n* but also their mode, median, etc. *(Fig. 26)*. Consequently, the distribution of arithmetic means has nothing more to do with the original (or mother) distribution. The grounds of this phenomenon lie in the calculation of arithmetic means as an estimation procedure itself. There are not *an sich* 'wrong' distributions: most frequent values also give in the case of the Smirnov-distribution with increasing *n* more and more accurately a value which characterizes the mother distribution itself (Fig. 25).
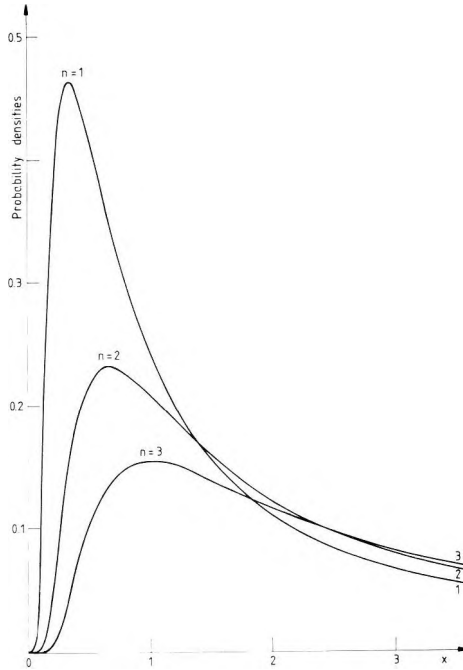
Fig. 26. Arithmetic mean of *n* values from a Smirnov distribution is also Smirnov-distributed but the density function is enlarged *n* times

26. ábra. Szmirnov-eloszlású *n* db érték számtani átlagának szintén *Szmirnov*-eloszlása van, csak ez utóbbi *n*-szeresen nyújtott

Рис. 26. Распределение среднеарифметических из *n* значений со смирновским распределением имеет также смирновский характер, но растянут по длине в *n* крат.

Chapter 6
Two families of probability distributions

6.1 Modelling

Vercors, in his excellent short novel entitled 'Les animaux dénaturés', wrote: 'As a general rule, investigation, experimentation and observation increase the uncertainty.' This feather-weight (and naturally exaggerated) statement with a considerable dose of salts can become a troubling reality if we hang fast to our accustomed statistical procedures: we have seen in the preceding chapter that expensive surplus data can really increase the uncertainty if we calculate according the principle of least squares.

This classical principle 'works' effectively only for 'nicely shaped' distributions (e.g. in the very neighbourhood of the Gaussian). For our investigations the broadest possible families of distributions are necessary for modelling the actually occurring error distributions.

The author hopes that the reader is not yet bound too tightly to the classical conception. Even in the early seventies, based on different approaches the following classification of statisticians was still valid (from the point of view of real distributions, cited from ANDREWS et al. [1972], page 128):

— 'Messrs. One-toe-in-the-water, where contamination is admitted, but only gentle contamination...'
— 'the wary classicist, who takes pure Gaussian as the conventional situation...'
— the realist.

The notion 'contamination' itself has come to be somewhat old-fashioned as a lot of actually occurring distributions are primarily long-tailed and not because of some secondary effect. (This technical term is only justifiable if the model-distribution is required to model some percentage of outliers, too.)

6.2 Generalized supermodel of Tukey

A family of probability distribution types defined for modelling real distributions is called for short 'supermodel'. A single model distribution-type of a supermodel can then be defined by a single value of the type-parameter, or by two values if the supermodel has two parameters.

One of the latter types of supermodel is the $f_T(p, \sigma_c; x)$ family of distribution types, known as the generalized supermodel of TUKEY:

$$f_T(p, \sigma_c; x) = (1-p) \cdot f_G(1; x) + p \cdot f_G(\sigma_c; x) \qquad (p < 0.5) \tag{83}$$

where $f_G(\bar{\sigma}; x)$ is the Gaussian distribution:

$$f_G(\bar{\sigma}; x) = \frac{1}{\bar{\sigma}\sqrt{2\pi}} e^{-\frac{x^2}{2\bar{\sigma}^2}}. \tag{84}$$

In σ_c the index refers to the 'contamination', as a reminder of the approach mentioned above.

As f_T supermodel has two parameters (p and σ_c), it is convenient to show on the (p, σ_c) plane just the same quantities as in Table IV (dihesion ε, $n(\varepsilon)$, $\varepsilon/\sqrt{n(\varepsilon)}$, σ and n_E/n_M, see *Fig. 27–31*; the last of these quantities is calucalted obviously as $\sigma^2 n(\varepsilon)/\varepsilon^2$).
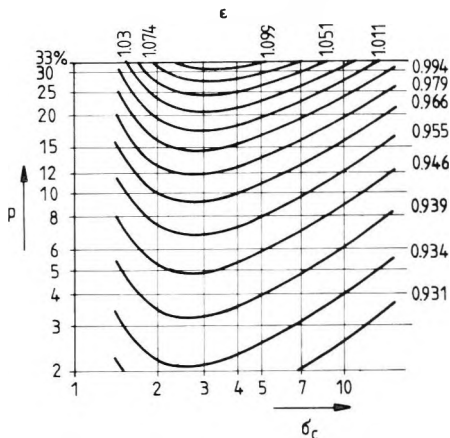
Fig. 27. Dihesions (ε) for Tukey's generalized supermodel of two parameters (p and σ_c)

27. ábra. Dihézió-értéke'. (ε) az általánosított Tukey-féle kétparaméteres (p és σ_c) szupermodellre vonatkozóan

Рис. 27. Значения дигезии (ε) в отношении обобщенной (p и σ_c) супермодели Тьюки.
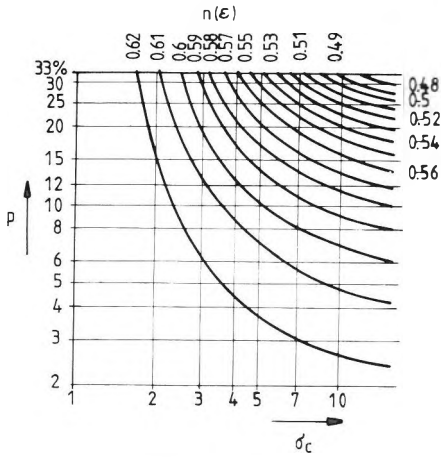
Fig. 28. Ratio of effective and all number of data (denoted by $n(\varepsilon)$) for the generalized supermodel of Tukey

28. ábra. Az effektív adatszám és a teljes adatszám $n(\varepsilon)$-nal jelölt aránya az általánosított Tukey-féle szupermodellre vonatkozóan

Рис. 28. Соотношение $n(\varepsilon)$ эффективного количества данных и полного количества данных в отношении обобщенной супермодели Тьюки.
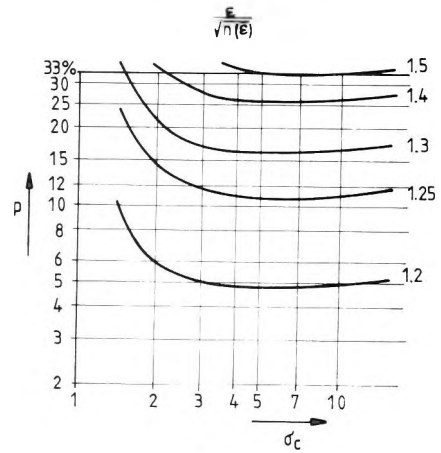

Fig. 29. Asymptotic scatter of most frequent values $(\varepsilon/\sqrt{n(\varepsilon)})$ for the generalized supermodel of Tukey

29. ábra. Leggyakoribb értékek aszimptotikus szórása $(\varepsilon/\sqrt{n(\varepsilon)})$ az általánosított Tukey-féle szupermodellre vonatkozóan

Рис. 29. Асимптотическая дисперсия наиболее частых значений $(\varepsilon/\sqrt{n(\varepsilon)})$ в отношении обобщенной супермодели Тьюки.
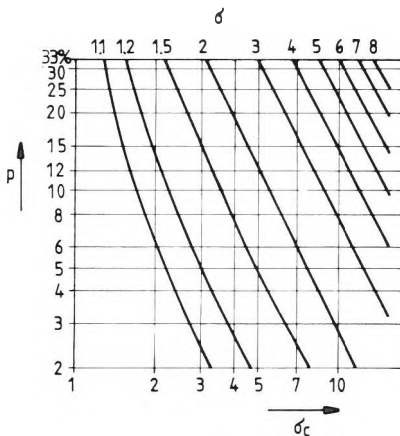
Fig. 30. Asymptotic scatter of the arithmetic means ('the scatter', σ) for the generalized supermodel of Tukey

30. ábra. Az átlagértékek aszimptotikus szórása („a szórás", σ) az általánosított Tukey-féle szupermodellre vonatkozóan

Рис. 30. Асимптотическая (обозначенная через σ) дисперсия средних значений в отношении супермодели Тьюки.
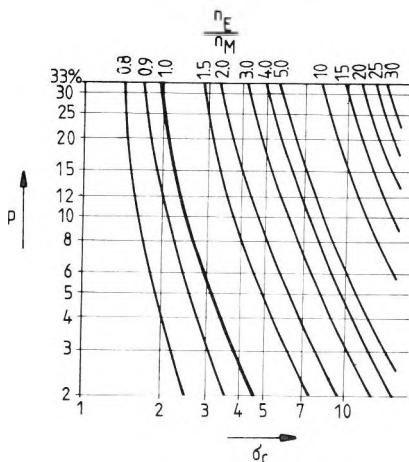
Fig. 31. Relative efficiency (n_E/n_M) of the calculation of most frequent values as estimates of location compared with the arithmetic means for the generalized supermodel of Tukey

31. ábra. A leggyakoribb érték-számítás, mint helyparaméter-becslés relatív hatásfoka (n_E/n_M), az átlagképzéshez viszonyítva, az általánosított Tukey-féle szupermodellre vonatkozóan

Рис. 31. Относительная эффективность расчета наиболее частых значений (n_E/n_M) в качестве оценки пространственных параметров в сравнении с расчетом средных, в отношении обобщенной супермодели Тьюки.

The thick line in Fig. 31 defines that subset of f_T for which E- and M-estimates are equally effective; for f_T-distributions on the left hand side the E-estimation is more advantageous (if the absence of outliers is absolutely guaranteed). and the opposite is valid for the f_T-distributions on the right hand side. The worst case for the M-estimation is the Gaussian distribution ($p = 0$): we have seen earlier that surplus data of 36% are necessary to achieve the same accuracy as has the optimum algorithm for this type of distribution (i.e. the mean square calculation).

Mention has already been made that the very solution inside the conception of most frequent calculations is given by generalizing the original algorithm,—see Chapter 8,—and in this way a resistant procedure can be given having an efficiency arbitrary near to 100% even in the case of a 'clean' Gaussian distribution. On the other hand, the least squares algorithm may need not just twice as many but possibly 5 times as many, 10 times as many (or even more) data to achieve the same accuracy as the M-estimation (the maximum value of n_E/n_M in Fig. 31 is 30 for the illustrated part of the (p, σ_c)-plane). It is certain that the finite σ-value of all f_T distributions guarantees for the whole supermodel the asymptotic fulfilment of the '$1/\sqrt{n}$-law' also in the case of least squares calculations —but the efficiency (from the point of view of the appliers) may not only be unsatisfactory but also absolutely unacceptable.

6.3 The supermodel $f_a(x)$

6.3.1 Some remarks on the behaviour of arithmetic means for small samples

We shall deal in this section with small samples, too. As an introduction to this topic, the present paragraph will show the interesting behaviour of arithmetic means of small samples from f_T-distributions. Namely, we need not in these cases carry out Monte Carlo computations since the density function of the arithmetic means of such samples can also be given analytically:

$$f_T^{(n)}(p, \sigma_c; x) = \sum_{k=0}^{n} \binom{n}{k} (1-p)^{n-k} p^k g_G(\sqrt{(n-k) + k\sigma_c^2}; x).\tag{85}$$

As $\exp\left[-(\bar{\sigma}t)^2/2\right]$, the characteristic function of the Gaussian distribution given in Eq. (84) is well known, it is simple to verify Eq. (85) on the grounds of Eq. (83) and applying the convolution theorem. If Q_n is the probable error (semi-interquartile range) of arithmetic means of samples consisting of n elements and Q_1 denotes the probable error for the mother distribution itself, Q_n/Q_1 is a measure of how the accuracy increases with n.

It is convenient to choose $1/\sqrt{n}$ for the abscissa since in this case the fulfilment of the $1/\sqrt{n}$-law at finite n-values is indicated by the fact that points, representing Q_n/Q_1-values, are situated on a straight line through the origin.

The examples in *Figs. 32* and *33* belong to the parameter-pair of ($p = 0.08$; $\sigma_c = 10$), and to that of ($p = 0.2$; $\sigma_c = 150$), respectively. The straight lines in these figures connecting the origin and the $(1; 1)$ point would indicate that the $1/\sqrt{n}$-law is for all finite n-values strictly fulfilled. Dotted straight lines indicate the asymptotic behaviour connecting the origin with the point

$$(1; 0.6745 \cdot \sqrt{(1-p) + p\sigma_c^2}).\tag{86}$$

Instead of analysing in detail Figs. 32 and 33, we only draw some inferences.
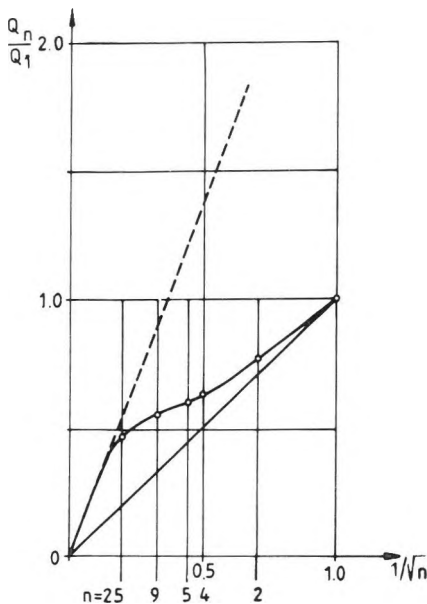
Fig. 32. Decrease of the sample interquartile range of arithmetic means by increasing n for a model distribution of Tukey type. Dotted line shows the asymptotic behaviour

32. ábra. Számtani átlagok empirikus interkvartilis terjedelmeinek csökkenése növekvő n-nel, egy Tukey-féle modelleloszlás esetén. A szaggatott egyenes az aszimptotikus viselkedést mutatja

Рис. 32. Убывание эмпирических межквартильных интервалов среднеарифметических с возрастанием n для случая распределения согласно модели Тьюки. Пунктиром демонстрируется асимптотическое поведение.
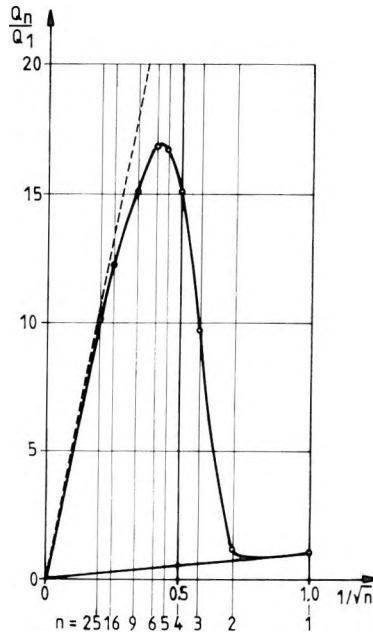
Fig. 33. For some of the generalized Tukey-distributions the sample interquartile ranges of arithmetic means with increasing *n* have a maximum value, before reaching that domain of *n*-values where the asymptotic behaviour becomes dominant (see the dotted line)

33. ábra. Az általánosított Tukey-eloszlások egy részénél a számtani átlagok empirikus interkvartilis terjedelmeinek növekvő *n*-nel előbb egy kifejezett maximumuk van, mielőtt elérnék az *n* értékek azon tartományát, ahol már az aszimptotikus viselkedés dominál (utóbbira nézve l. a szaggatott egyenest)

Рис. 33. В некоторых из обобщенных распределений Тьюки эмпирические межквартильные интервалы среднеарифметических значений при возрастании *n* сначала обнаруживают четкий максимум перед достижением того диапазона значений *n*, где уже преобладает асимптотическое поведение (в отношении последнего см. пунктир).

First a curiosity in Fig. 33 should be mentioned: arithmetic means can show on a finite *n*-interval the reverse fulfilment of the law of large numbers, even if σ is finite. (We shall see in Fig. 39 that most frequent values behave in accordance with the straight lines between the origin and the (1; 1) point.)

The density function of arithmetic means has the following form in the general case:

$$f^{(n)}(x) = n[f(nx)]^{n*} \tag{87}$$

where the symbol '*n**' means that convolution is to be carried out *n* times. Figures 32 and 33 can also be seen from this point of view: by increasing *n* the role of tails is increasingly emphasized (this is a consequence of one of the well known properties of the convolutions). the gathering itself can no longer have a significant influence on the probable error of arithmetic means.

6.3.2 Definition and characterization of the supermodel $f_a(x)$

If a is the type-parameter

$$f_a(x) = \frac{1}{c(a)[\sqrt{x^2+1}]^a} \qquad a > 1 \tag{88}$$

defines the density function $f_a(x)$ where

$$c(a) = \frac{\sqrt{\pi}\,\Gamma\left(\dfrac{a-1}{2}\right)}{\Gamma\left(\dfrac{a}{2}\right)} \tag{88a}$$

(Γ denotes the well known Γ-function). Some of these (to the origin symmetrical) density functions are shown for $x \geqq 0$ in *Fig. 34*. The corresponding distribution functions have the following form for integer values of $a > 1$:
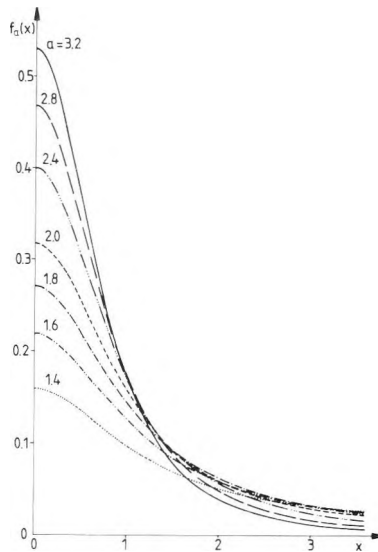
Fig. 34. Some probability density functions (because of symmetry shown for $x > 0$ only) from supermodel $f_a(x)$

34. ábra. Néhány $f_a(x)$-szupermodellbeli valószínűségeloszlás sűrűségfüggvénye (a szimmetria miatt elég pozitív x-ekre korlátozódni)

Рис. 34. Плотностные функции некоторых из вероятностных распределений в супермодели $f_a(x)$ (из-за симметрии достаточно ограничиться положительными x-ами).

$$F_2(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x;$$

$$F_3(x) = \frac{1}{2} + \frac{x}{2\sqrt{1+x^2}};$$

$$F_4(x) = F_2(x) + \frac{1}{\pi}\,\frac{x}{1+x^2}; \qquad\qquad\qquad\qquad (88b)$$

$$F_a(x) = F_{a-2}(x) + \frac{1\cdot 3\cdot\ldots\cdot(a-4)}{2\cdot 4\cdot\ldots\cdot(a-3)}\,\frac{1}{2}\,\frac{x}{(1+x^2)^{\frac{a-2}{2}}} \qquad \text{(if } a \text{ is odd and } a \geqq 5)$$

$$F_a(x) = F_{a-2}(x) + \frac{2\cdot 4\cdot\ldots\cdot(a-4)}{3\cdot 5\cdot\ldots\cdot(a-3)}\,\frac{1}{\pi}\,\frac{x}{(1+x^2)^{\frac{a-2}{2}}} \qquad \text{(if } a \text{ is even and } a \geqq 6).$$

Table V shows for some a values (i.e. for some distribution types) the numerical values of $c(a)$, of the semi-interquartile range Q, of the dihesion ε, of $n(\varepsilon)$, and of the asymptotic scatter $\varepsilon/\sqrt{n(\varepsilon)}$.

a	$c(a)$	Q	ε	$n(\varepsilon)$	$\dfrac{\varepsilon}{\sqrt{n(\varepsilon)}}$
(1.2	11.3231	17.1756	1.8719	0.2370	3.8449)
1.4	6.2687	3.1231	1.5035	0.3566	2.5176
1.6	4.5544	1.7219	1.2732	0.4260	1.9608
2.0	3.1416	1.0000	1.0000	0.5000	1.4142
2.5	2.3963	0.7125	0.8120	0.5438	1.1011
3	2.0000	0.5774	0.6974	0.5669	0.9262
4	1.5708	0.4416	0.5616	0.5900	0.7312
5	1.3333	0.3704	0.4819	0.6012	0.6215
6	1.1781	0.3250	0.4282	0.6078	0.5493
10	0.8590	0.2342	0.3149	0.6190	0.4003
40	0.4040	0.1090	0.1492	0.6296	0.1880
100	0.2526	0.0680	0.0937	0.6331	0.1177

Table V. Characteristics of some distribution-types of supermodel $f_a(x)$
a — type-parameter; $c(a)$ — normalization factor; Q — semi interquartile range; ε — dihesion;
$\varepsilon/\sqrt{n(\varepsilon)}$ — asymptotic scatter of the most frequent value

V. táblázat. Az $f_a(x)$ szupermodell néhány eloszlástípusának jellemző adatai
a — tipusparaméter; $c(a)$ — normálási faktor; Q — interkvartilis félterjedelem; ε — dihézió;
$\varepsilon/\sqrt{n(\varepsilon)}$ — a leggyakoribb érték aszimptotikus szórása

Таблица V. Некоторых характерные параметры типов распределения супермодели $f_a(x)$
a — типовой параметр; $c(a)$ — фактор нормирования; Q — межквартильный
полуинтервал; ε — дигезия; $\varepsilon/\sqrt{n(\varepsilon)}$ — асимптотическая дисперсия наиболее частого
значения.

The types are defined in Eq. (88) in its simplest analytical form. For arbitrary location- and scale parameter (T and S) we obviously have

$$g_a(x) = \frac{\Gamma\left(\dfrac{a}{2}\right)}{\Gamma\left(\dfrac{a-1}{2}\right)\sqrt{\pi}} \cdot \frac{S^{a-1}}{(\sqrt{S^2 + (x-T)^2})^a} \tag{88c}$$

The distributions 1, 2, 3, 5, 9, 10 and 12 in Table IV are clearly $f_a(x)$-distributions. There are $f_a(x)$-distributions in all four classes of Table IV; by contrast, however, all f_T distributions belong to the fourth class (σ being always finite for all f_T-s). In other words, the $f_a(x)$-family of probability distributions is able to model many more types of distributions occurring in geophysics and geology, although it has only one type-parameter. This latter fact results in a more convenient visualization of effectivites, of dihesions, etc., than in the case of f_T (when systems of isolines are needed, see Figs. 27–31).

It is easy to verify that for $a > 3$

$$\sigma^2 = \frac{1}{a-3} \tag{89}$$

holds. If $a \to \infty$, $f_a(x)$ tends to be a Gaussian distribution; if $a = 2$ we get the Cauchy distribution. For integer values of a, the $f_a(x)$ distributions are Student's t-distributions with $(a-1)$ degrees of freedom. It would therefore be possible to refer to the the supermodel $f_a(x)$ also as the class of generalized Student's t-distributions, as in fact, was done by HAJAGOS [1985a]. It is pointed out, however, that to a certain extent this name can be misleading since Student's t-distributions are introduced in close connection with the Gaussian distribution, but the use of the $f_a(x)$-distributions for modelling has nothing to do with this connection.

With $f_a(x)$-distribution a very great variety of flanks can be modelled: for great values of $|x|$ the density function approximates the expression const. $|x|^a$. To show two examples, in *Fig. 35* to $(a = 1.4; S = 0.1)$ and to $(a = 3.2; S = 1)$ the density functions $\tilde{f}(x)$ and $\bar{f}(x)$ are drawn. Which is the more advantageous from the practical point of view? Since the probable error $\tilde{Q}$ is less than that of $\bar{Q}$, $\tilde{f}(x)$ can be declared as the better one. And really: also the most frequent values have a lower asymptotic scatter in the case of $\tilde{f}(x)$ (0.251) than in the case of $\bar{f}(x)$ (0.875; see also Table IV). The fact that for arithmetic means as estimates for $\tilde{f}(x)$ the reverse of the law of large numbers is fulfilled, clearly qualifies the estimation procedure itself and not the distribution. Since 'the variance' for $\bar{f}(x)$ is finite, in this respect $\bar{f}(x)$ would be much more advantageous than $\tilde{f}(x)$, contradicting not only the numerical comparisons made above, but also our first impression on looking at the curves in Fig. 35. It is again verified that 'the variance' is not the characteristic of the uncertainty contained in the original (mother) distribution itself. This statement will also be justified in respect of the entropies, see Chapter 7.

The question arises as to whether or not a-values very near to 1 do model real distributions. The density function for $a = 1.2$ *(Fig. 36)* shows that the upper quartile Q is already far from the actual gathering of the values (the interval of the latter is measured by $f^{-1}[f(0)/2]$). Some (e.g. meteorological) data can show similar distributions, too; in geophysics, however, the quotient $f^{-1}[f(0)/2]/Q$ seems too small to be $f_{1.2}(x)$, a model of a real distribution. To be able to judge in a simple way the reality of a given $f_a(x)$ for a given case from this point of view, in *Fig. 37* the values of this quotient are shown versus $1/(a-1)$.

How are the $f_a(x)$-distributions able to model the reality? The answer can be given in a convenient way if an analytical expression for the density function based on the measuring data is already given. NEWCOMB [1886] had found for his astronomical data the expression:

$$f(x) = c_1 e^{-\left(\frac{x}{10}\right)^2} + c_2 e^{-\left(\frac{x}{18}\right)^2} + c_3 e^{-\left(\frac{x}{36}\right)^2},$$

with given c_i-s; this $f(x)$ curve is shown in *Fig. 38*. Points in the same figure belong to an $f_a(x)$ characterized by $a = 5.3$ and $S = 21.1$.

Although the excellent modelling in this case on the grounds of *Fig. 38* is obvious, this is one single example, and for mathematicians it is often the case that even a great many examples are not convincing. Let us remember, however, that in Chapter 3 we have seen Cauchy distributed data, too. In other words, $f_a(x)$ with $a = 2$ can also be a suitable model in actual cases, and according to JEFFREYS [cited in KERÉKFY 1978], very careful measurements carried out under undisturbed conditions can result in Student's *t*-distributions of 5–9 degrees of freedom, i.e. in $f_a(x)$-distributions characterized by $6 < a < 10$ (this is the so-called Jeffreys interval). Exceptionally, distributions characterized by $a > 10$ also occur: e.g. the example given by LINNIK [1961] for the Gaussian distribution(!) can be modelled by approximately f_{20}-type ($12 < a < 22$; a more accurate determination of type is not possible in this case because of the small number and the grouping of the data). It can therefore be expected that the members of the $f_a(x)$ family can, under widely varying conditions, adequately model the actual distributions.
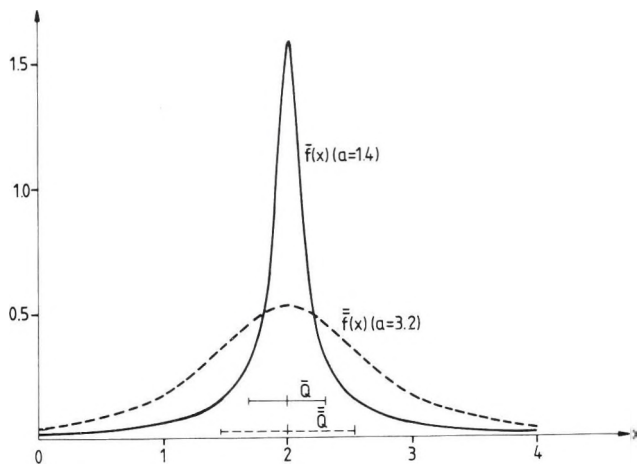
Fig. 35. Density functions $\bar{f}(x)$ ($f_a(x)$ for $a = 3.2$ with $S = 1.0$) and $\bar{\bar{f}}(x)$ ($f_a(x)$ for $a = 1.4$ with $S = 0.1$ as scale parameter). The asymptotic scatter of the most frequent value is 0.251 for $\bar{f}(x)$ and 0.875 for $\bar{\bar{f}}(x)$, while the arithmetic mean as an estimate is inapplicable for $\bar{f}$ (see the first distribution in Table IV). It is completely misleading to consider 'the variance' σ^2 as a characteristic of the distribution itself

35. ábra. Az $\bar{f}(x)$-szel jelölt sürüségfüggvény ($f_a(x)$ $a = 3,2$-nél $S = 1,0$-val) és az $\bar{\bar{f}}(x)$ (amely $S = 0,1$-es skálaparaméterü, $a = 1,4$-hez tartozó $f_a(x)$). A leggyakoribb értékek aszimptotikus szórása 0,251 $\bar{f}(x)$-re és 0,875 $\bar{\bar{f}}(x)$-re, a számtani átlag azonban becslésként használhatatlan $\bar{f}$-ra (l. a IV. táblázat 1. sorszámú eloszlását). Mindenféleképpen félrevezetö a „szórást" az eredeti eloszlás jellemzöjeként elfogadni

Рис. 35. Плотностные функции $\bar{f}(x)$ ($f_a(x)$ при $a = 3,2$ с $S = 1,0$) и $\bar{\bar{f}}(x)$ ($f_a(x)$ при $a = 1,4$ с параметром шкалы $S = 0,1$). Асимптотическая дисперсия наиболее частых значений составляет 0,251 для $\bar{f}(x)$ и 0,875 для $\bar{\bar{f}}(x)$, но среднеарифметическое не может быть использовано в качестве оценки $\bar{f}(x)$ (см. распределение № 1 таблицы IV). В любом случае было бы ошибочно принять «дисперсию» в качестве характеристики первичного распределения.
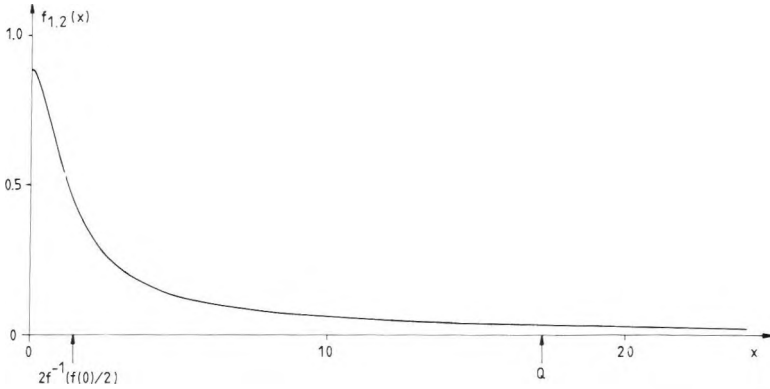
Fig. 36. If a is very close to 1 the interquartile range of the $f_a(x)$ distribution is very long compared with the interval of maximum probabilities characterized by the value of $2f^{-1}(f(0)/2)$

36. ábra. Ha a túl közel van 1-hez, az $f_a(x)$ eloszlás interkvartilis terjedelme nagyon nagy lesz a legnagyobb valószinüségü értékek $2f^{-1}(f(0)/2)$-vel mért intervallumhosszához viszonyítva

Рис. 36. Если a близко к 1, то межквартильный интервал распределения $f_a(x)$ становится слишком большим относительно длины интервала наиболее частых значений, измеренной соотношением $2f^{-1}(f(0)/2)$.
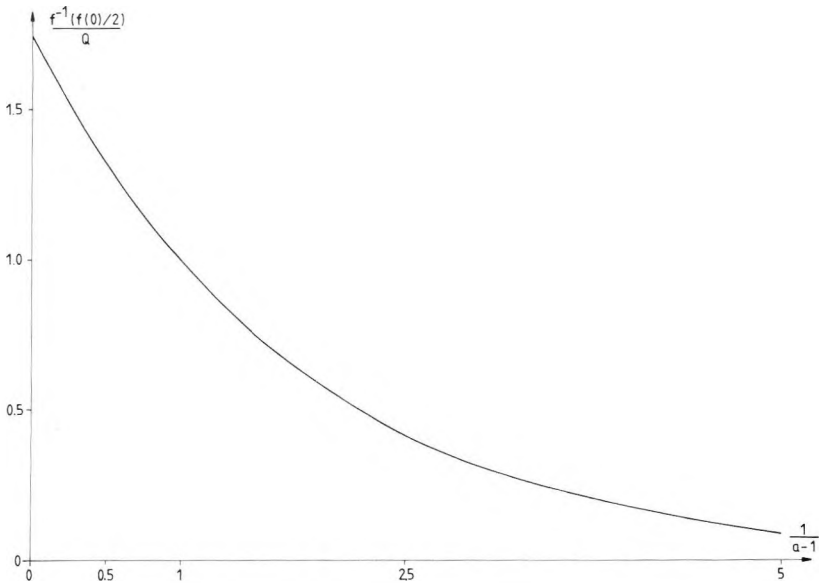
Fig. 37. Ratio of two lengths: that of the interval of maximum probabilities and that of the interquartile range, versus $1/(a-1)$ for supermodel $f_a(x)$

37. ábra. A maximálisan valószinü értékek intervallumhosszának és az interkvartilis terjedelemnek az aránya, $1/(a-1)$ függvényében, az $f_a(x)$ szupermodellre

Рис. 37. Отношение длины интервала наиболее вероятных значений к длине межквартильного интервала как функция $1/(a-1)$ для супермодели $f_a(x)$.
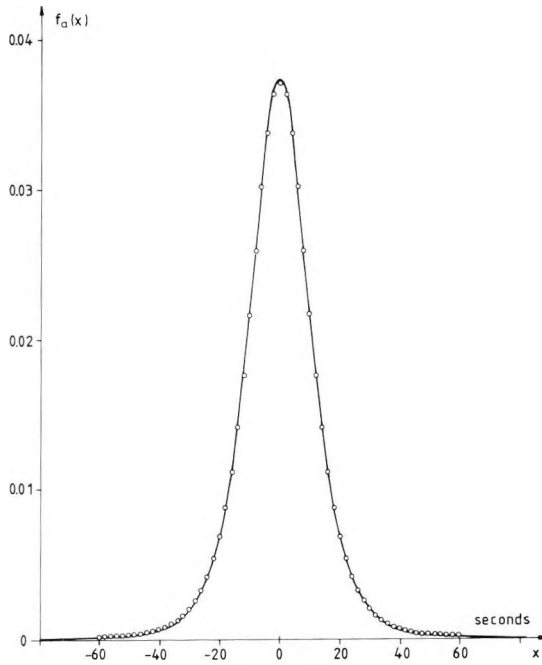
Fig. 38. Line: density function given by Newcomb for his experimental data sets; circles: values of $f_a(x)$ for the type parameter $a = 5.3$ and for the scale parameter $S = 21.1$

38. ábra. Folytonos vonal: Newcomb által megadott sürüségfüggvény, amely jól leírja tapasztalati adatrendszereinek eloszlását; nullkörök: az $a = 5,3$ tipusparaméterhez és az $S = 21,1$ skálaparaméterhez tartozó $f_a(x)$ értékek

Рис. 38. Сплошная линия: плотностная функция Ньюкома, с достаточной точностью описывающая распределение систем эмпирических данных; кружочки: значения $f_a(x)$, соответствующие типовому параметру $a = 5,3$ и параметру шкалы $S = 21,1$.

6.3.3 Investigation of finite samples from different $f_a(x)$ distributions

The investigation of finite samples coming from f_T-distributions in respect of their arithmetic means was convenient: Eq. (85) has given the density function of arithmetic means explicitly. In order to investigate most frequent values and arithmetic means for $f_a(x)$ distributions Monte Carlo calculations are necessary.

The distribution $f_a(x)$ for $a = 1.4$; 1.6; 1.8; 2.0; 2.4; 2.8 and 3.2 was randomly sampled 200 times for $n = 4$; 9; 16; 25; 36; 49 and 64, the most frequent value and arithmetic mean for each sample calculated, and the semi-interquartile range Q_n for each set of estimates determined. The results are demonstrated in the $(1/\sqrt{n}; Q_n/Q_1)$ system of coordinates (where $Q_1 = Q$ is the semi-interquartile range of the mother distribution), similarly to *Figs. 32* and *33*.

The results obtained for most frequent values are shown by vertical straight sections in *Fig. 39*, and in detail in *Fig. 40*. The straight line between the origin and the point $(1; 1)$ indicates the most regular behaviour: if the mother distribution were the Gaussian one, the Q_n/Q_1 values of arithmetic
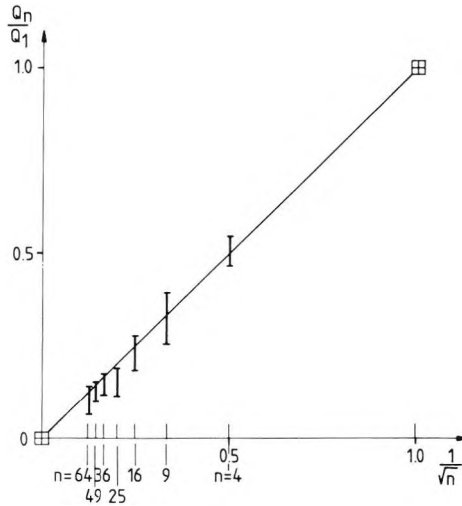
Fig. 39. Results of a Monte Carlo study. If most frequent values are used, the interval of very small n-values also shows asymptotic behaviour

39. ábra. Egy Monte Carlo vizsgálat eredményei. Ha leggyakoribb értékeket használunk a becsléshez, a nagyon kis n-értékek tartománya is az aszimptotikus viselkedést mutatja

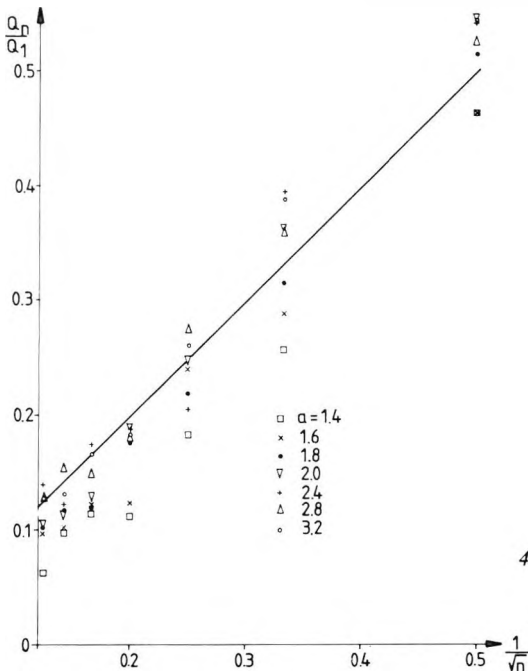Рис. 39. Результаты исследования методом Монте-Карло. При использовании наиболее частых значений в оценках, даже область очень малых значений n обнаруживает асимптотическое поведение.

Fig. 40. Results of Fig. 39 enlarged

40. ábra. A 39. ábrán bemutatott eredmények kinagyítva

Рис. 40. Результаты, представленные на рис. 39, в увеличенном виде.

means would fluctuate around this line. A very important conclusion made on the basis of *Figs. 39* and *40* is that most frequent values behave in the most regular and, at the same time, in a practically most advantageous manner.

What about arithmetic means calculated from the same samples? *Figure 41* for $n = 64$ is very instructive: Q_n/Q_1 values increase with descending a so rapidly that a logarithmic scale was necessary on the ordinate. The 'most regular' behaviour would correspond clearly to the value 0.125, but even at $a = 3.2$, where the variance is already finite, Q_n/Q_1 is significantly greater than this value.

The structure of *Fig. 42* is the same as that of *Fig. 40*, only the Q_n/Q_1 values in *Fig. 42* belong to the arithmetic means. For $a = 2$ the Q_n/Q_1 values fluctuate around the value 1—but it is commonplace that the accuracy of arithmetic means at the Cauchy distribution does not depend on n.

For $a = 2.4$; 2.8; 3.2 and 3.6 the law of large numbers is fulfilled but in very different manner. The behaviour of arithmetic means even at $a = 3.6$ does not reach the 'most regular behaviour', indicated by the straight line.

All points are above the straight line between $(0; 0)$ and $(1; 1)$ in *Fig. 42*—and this means that the use of arithmetic means is economically disadvantageous. An example: to $n = 64$ belongs $Q_n/Q_1 = 0.41$ at $a = 2.4$ (upper dotted curve)—this value would belong to $n = 6$ if the most regular behaviour were to occur ($1/\sqrt{6} = 0.41$). This means that data of more than an order of magnitude are necessary to achieve the same accuracy.

In *Fig. 43* the Q_n/Q_1 quotients versus $\sqrt{n}$ are demonstrated. As can clearly be seen the reverse fulfilment of the law of large numbers at $a = 1.8$, 1.6 and 1.4 is not only an asymptotic statement but also valid for small samples. Let us recall (*Figs. 39* and *40*) that most frequent values showed in these cases, too, the most regular (and at the same time most advantageous) behaviour.
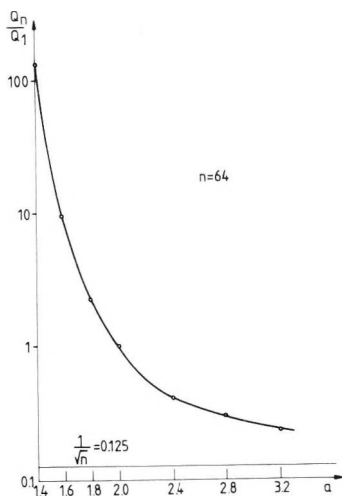
Fig. 41. Results of a Monte Carlo study. The different behaviour of the arithmetic means as estimates for various values of the type parameter a is shown for a fixed sample measure ($n = 64$). The asymptotic value is 0.125 belonging to $a \to \infty$ (this corresponds to Gaussian distribution)

41. ábra. Egy Monte Carlo vizsgálat eredményei. Fixnek felvett mintaterjedelemre ($n = 64$) adjuk meg a számtani átlagoknak az a tipusparamétertől függő viselkedését. Az aszimptotikus érték 0,125 (ez az érték felel meg a Gauss-eloszlásnak)

Рис. 41. Результаты исследования методом Монте-Карло. Поведение среднеарифметических, зависящее от типового параметра a сильно варьирующее, даются для объема данных ($n = 64$), принятого в качестве постоянной. Величина асимптотического значения — 0,125, соответствующее гауссовскому распределению.
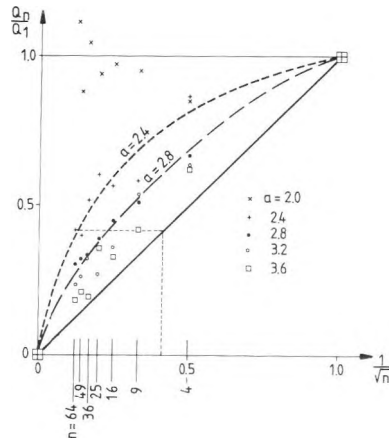
Fig. 42. Results of a Monte Carlo study. For $a = 2.4$, 2.8, 3.2 and 3.6, i.e. for some $f_a(x)$ distributions for which also the arithmetic mean follows the law of large numbers, the different types of fulfilments are shown. (Monte Carlo results are also shown for $a = 2$, i.e., also for the Cauchy distribution to demonstrate the measure of statistical fluctuation of this study)

42. ábra. Egy Monte Carlo vizsgálat eredményei. A nagy számok törvényének teljesülési módjai a számtani középértékekre, az a tipusparaméter 2,4; 2,8; 3,2 és 3,6 értékeire. Feltüntettük az $a = 2$, azaz a Cauchy-eloszlás esetére vonatkozó Monte Carlo eredményeket is, a vizsgálatot jellemző statisztikus ingadozás mértékének érzékeltetésére

Рис. 42. Результат исследования методом Монте-Карло. На рис. представлены различные способы выполнения закона больших чисел в отношении среднеарифметического для значений типового параметра a в 2,4; 2,8; 3,2 и 3,6 (на рис. приводятся также и результаты Монте-Карло для $a = 2$, то-есть для распределения Коши, с целью дать представление о степени статистических флюктуаций, характерных для проведенных исследований).
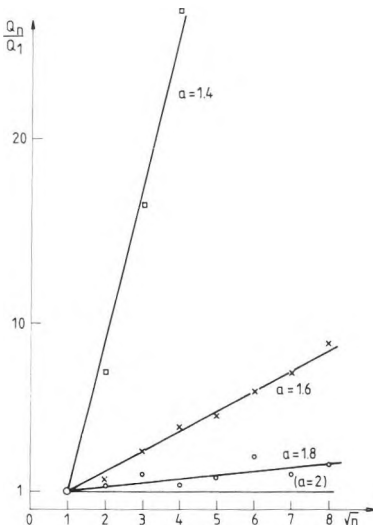
Fig. 43. Results of a Monte Carlo study. The astonishing behaviour of the arithmetic means: these estimates show for $a = 1.4$, 1.6 and 1.8 that the law of large numbers is inversely fulfilled

43. ábra. Egy Monte Carlo vizsgálat eredményei. A számtani átlagok meghökkentő viselkedése $a = 1,4$; 1,6 és 1,8 esetén: a nagy számok törvénye éppen fordítottan teljesül

Рис. 43. Результат исследования методом Монте-Карло. Странное поведение среднеарифметических при $a = 1,4$; 1,6 и 1,8: закон больших чисел выполняется в обратном смысле.

Chapter 7
Deduction of formulae defining most frequent value and dihesion, based on information theory

7.1 Entropy and I-divergence

7.1.1 The definition of entropy

The entropy of a probability distribution—defined by its density function $f(x)$—is given by the formula

$$H(f) = \int\limits_{-\infty}^{\infty} f(x) \cdot \log_2 \frac{1}{f(x)} \, \mathrm{d}x \tag{90}$$

[REZA 1961]. We shall deal in the following only with distributions having positive densities for all x-values. For these distributions the definition in Eq. (90) is without any comment correct; in general cases the integration is to be carried out only for x-domains where $f(x) > 0$ holds.

The entropy does not depend upon the location parameter (see Eq. (90)); $H(f)$ is only a characteristic of the dispersion of the mother distribution itself. The entropy of all $f_a(x)$ distributions—defined in Eq. (88)—is finite [HAJAGOS 1982].

For distributions having finite variance the entropy is also finite [REZA 1961]. The opposite statement, however, does not hold, e.g. the entropy of the Cauchy distribution characterized by the probable error Q (see Eq. (13) with $S = Q$) is given by the expression

$$H(f_C) = \log_2 (4\pi Q), \tag{91}$$

although not only the second but also the first moment is infinite.

Some remarks should be made in the following, regarding the definition in Eq. (90) from the viewpoint of the applier.

The entropy of the Gaussian distribution (see Eq. (84)) is given by

$$H(f_G) = \log_2 \left[\sigma \sqrt{2\pi e} \right] \tag{92}$$

(where e is the base of the natural logarithm). In both Eqs. (91) and (92) the parameter of scale is in the argument of the logarithm function – and this fact does not correspond to our demand that the characteristic of the dispersion must be proportional to the parameter of scale. The solution is simple: the characteristic defined by

$$B(f) = 2^{H(f)} \tag{93}$$

fulfils our demand, in addition, it does this in such a manner that relations, minimum or maximum properties being valid for H also remain valid for B.

As an example, proof should be given as to whether or not B is an appropriate characteristic for the dispersion of the mother distribution. A Gaussian and a Cauchy type distribution of the same probable error are compared (i.e. $\bar{\sigma} = Q/0.6745$):

$$\frac{B(f_C)}{B(f_G)} = \frac{4\pi Q}{\frac{Q}{0.6745}\sqrt{2\pi e}} = 2 \cdot 0.6745 \sqrt{\frac{2\pi}{e}} = 2.051. \tag{93a}$$

We can accept as real that B is twice as great in the case of the Cauchy distribution than for the Gaussian one because of the much heavier flanks.

The entropy defined in Eq. (90) – or B in Eq. (93) – fulfils the requirements of practice, in the case of heavy flanks, but H (and B) may also be infinite. Let us cite an example given by Csernyák [1982]:

$$h(x) = \begin{cases} 1 & |x| < 2 \\ \dfrac{1}{|x| \cdot \log_2^2 x} & |x| \geq 2. \end{cases} \tag{94}$$

(It is not superfluous to mention that Monte Carlo computations of this distribution also showed in this extreme case the regular increase of the accuracy of the most frequent values, i.e. the fulfilment of the $1/\sqrt{n}$ rule for the M values.)

According to an interesting theorem of information theory, Gaussian distribution is characterized by the maximum entropy among all distributions having the same variance. At first glance it seems astonishing that the measure of disordination has its maximum value just at the Gaussian distribution having small flanks. But we have seen earlier that the variance (or its square root, the scatter) does not adequately characterize the original (mother) distribution being extremely sensitive to the 'most far parts' of the flanks.

It seems to be appropriate to show an example. In *Fig. 44* two density functions (f_G to $\bar{\sigma} = 3$ and f_T to $p = 0.08$ and $\sigma_c = 10$) are shown, both clearly having the same variance. (For f_T defined in Eq. (83) the variance is to be calculated according to $\sigma^2 = 1 + p[\sigma_c^2 - 1]$.) The cited theorem shows the entropy of f_G to be greater than that of f_T – and we also feel according to our (subjective) judgement after a visual comparison of the curves in Fig. 44 that f_G contains more uncertainty than f_T does. Consequently, the variance (being extremely sensitive to the 'most far' parts of the flanks) is *always* unable to characterize the real uncertainty contained in the distribution (not only in cases when the variance is infinite).
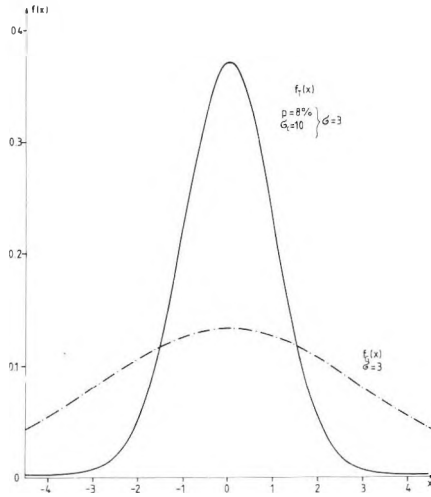
Fig. 44. Density function of a model-distribution of Tukey-type $f_T(x)$ and that of a Gaussian distribution $f_G(x)$ having the same variance (i.e. asymptotic variance of arithmetic means as estimates). The entropy of this latter distribution is greater than that of $f_T(x)$ ($f_G(x)$ is much more dispersed in its main part than is $f_T(x)$). The variance does not characterize the dispersion of the original distribution

44. ábra. Egy Tukey-tipusú sűrűségfüggvény $f_T(x)$, valamint az azzal azonos szórású Gauss-eloszlás $f_G(x)$ sűrűségfüggvénye. Az utóbbi eloszlás entrópiája nagyobb mint az $f_T(x)$-é, amit az $f_G(x)$ lényegesen szétterültebb volta vizuálisan is kézenfekvővé tesz. Látható, hogy „a szórás" nem az eredeti eloszlás jellemzöje

Рис. 44. Плотностная функция типа Тьюки $f_T(x)$ и гауссовского распределения $f_G(x)$ с той же дисперсией. Энтропия последнего распределения существенно больше, нежели для $f_T(x)$, что визуально иллюстрируется значительно более расширенной формой $f_G(x)$. Видно, что «дисперсия» не является характеристикой исходного распределения.

7.1.2 Substituting distributions, I-divergence

The notion of 'relative information', commonly known as '*I*-divergence' was introduced by KULLBACK [1959] and PEREZ [1967]. The *I*-divergence can be interpreted as the loss of information, too.

Let us suppose that $f(x)$ is the density function of the actually occurring but analytically unknown probability distribution; $f(x)$ being unknown, we treat our case as if the density function would be the analytically known $g(x)$. In this case $g(x)$ is called the substituting distribution, and the *I*-divergence is defined as

$$I_g(f) = \int_{-\infty}^{\infty} f(x) \log_2 \frac{f(x)}{g(x)} \, dx \tag{95}$$

(supposing that $g(x) > 0$ is fulfilled for all x values).

It can immediately be seen that in the case of $g(x) = f(x)$, $I_g(f) = 0$ is valid. It can be proved [VINCZE 1953] that $I_g(f)$ cannot be negative. If $g(x)$ is near $f(x)$, $I_g(f)$ is a small value, and the opposite is fulfilled if $g(x)$ is far from $f(x)$. In this sense $I_g(f)$, from the point of view of information theory, can be regarded as a distance of g measured from f (although all the usual criteria demanded for 'distance' are not fulfilled.)

It can be proved [HAJAGOS 1982] that when g is the Gaussian distribution (denoted here as G) with the scale parameter σ, and the variance $\bar{\sigma}^2$ of f is finite being $\sigma^2 = \bar{\sigma}^2$, then

$$I_G(f) = H(G) - H(f) \tag{96}$$

holds. This formula shows the connection between the notions of entropy and I-divergence: the distance of the Gaussian distribution from f_T, both shown in Fig. 44, is just the difference of their entropies: $H(G) - H(f_T)$. (It should be noted that the connection is not always so simple.)

It is at first glance astonishing that in cases when f has an infinite variance, the distance of G (i.e. of the Gaussian distribution) from f is infinite (independently of the value of the scale parameter). This is, however, in close connection with the already known fact (see also paragraph 7.7.2) that least squares procedures may be completely unusable also in well treatable cases; this circumstance is adequately expressed by the infinite value of this distance.

It can be shown [HAJAGOS 1982 and CSERNYÁK 1982] that when the substituting distribution is the Cauchy distribution, the I-divergence is finite for the whole $f_a(x)$ supermodel (and for all distributions having not heavier flanks than those of the $f_a(x)$ distributions). It seems therefore more advantageous—from the viewpoint of information theory and in respect of general applicability—to choose the Cauchy distribution as the substituting distribution instead of the Gaussian one.

7.2 Minimization of I-divergence

The practical way of using the notion of the I-divergence is the following. Substituting the actual but unknown distribution f by g of known analytical form, that pair of T and S values playing a role in g is accepted as characteristics of f, that minimizes the loss of information (i.e. the I-divergence).

7.2.1 General formulae

We write in the following $g(x; T)$ and $I_g(f; T)$—instead of $g(x)$ and $I_g(f)$—to emphasize the fact that our primary task is the determination of T. Let us suppose that $g(x; T)$ is symmetrical, is to be differentiated according to T, and that the integration and differentiation twice according to T are inter-

changeable. The distance of g from f will be minimum if the following conditions are fulfilled:

$$\frac{dI_g(f;T)}{dT} = 0 \tag{97}$$

and

$$\frac{d^2 I_g(f;T)}{dT^2} > 0, \tag{97a}$$

i.e. by the expression for $I_g(f;T)$ given in Eq. (95) the equation

$$\int_{-\infty}^{\infty} \frac{\partial g(x;T)}{\partial T} \cdot \frac{f(x)}{g(x;T)} \, dx = 0 \tag{98}$$

and the relation

$$\int_{-\infty}^{\infty} \left[\frac{\partial g(x;T)}{\partial T} \cdot \frac{1}{g(x;T)} \right]^2 f(x) \, dx - \int_{-\infty}^{\infty} \frac{\partial^2 g(x;T)}{\partial^2 T} \cdot \frac{f(x)}{g(x;T)} \, dx > 0. \tag{98a}$$

must hold. Equation (98a) is surely fulfilled if

$$\int_{-\infty}^{\infty} \frac{\partial^2 g(x;T)}{\partial^2 T} \cdot \frac{f(x)}{g(x;T)} \, dx = 0. \tag{99}$$

holds.

The simultaneous fulfilment of Eqs. (98) and (99) results in a T and S value which guarantees the minimum I-divergence.

7.2.2 Minimization of the loss of information if the substituting distribution is Gaussian

We need now the general form of the Gaussian distribution:

$$g(x;T) = \frac{1}{\bar{\sigma}\sqrt{2\pi}} e^{-\frac{(x-T)^2}{2\bar{\sigma}^2}} \tag{100}$$

Substituting this in Eq. (98) we get

$$\int_{-\infty}^{\infty} (x-T)f(x) \, dx = 0 \tag{101}$$

or in another form

$$T = \int_{-\infty}^{\infty} x f(x) \, dx; \tag{101a}$$

on the right hand side is the well known 'expected value' E.

Substituting Eq. (100) in Eq. (99), we get

$$\int_{-\infty}^{\infty} \left[\frac{(x-T)^2}{\bar{\sigma}^2} - 1 \right] f(x) \, dx = 0, \tag{102}$$

or in another form

$$\bar{\sigma}^2 = \int_{-\infty}^{\infty} (x-E)^2 f(x) \, dx. \tag{102a}$$

As the most advantageous pair of values in g we have the expected value and the variance (scatter) of $f(x)$, if the substituting distribution is the Gaussian one.

7.2.3 Minimization of the loss of information by substituting the Cauchy distribution

The general form of the Cauchy distribution is now the substituting distribution:

$$g(x; T) = \frac{1}{\pi} \frac{Q}{Q^2 + (x-T)^2}. \tag{103}$$

The form of Eq. (98) becomes

$$\int_{-\infty}^{\infty} \frac{x-T}{Q^2 + (x-T)^2} f(x) \, dx = 0, \tag{104}$$

but this can rearranged clearly into

$$T = \frac{\displaystyle\int_{-\infty}^{\infty} \frac{x}{Q^2 + (x-T)^2} f(x) \, dx}{\displaystyle\int_{-\infty}^{\infty} \frac{1}{Q^2 + (x-T)^2} f(x) \, dx}. \tag{104a}$$

Writing M instead of T, this is the defining formula for the most frequent value.

Putting Eq. (103) in Eq. (99), we get

$$\int_{-\infty}^{\infty} \frac{3(x-M)^2 - Q^2}{[Q^2 + (x-M)^2]^2} f(x)\, dx = 0 \tag{105}$$

which can also be written as

$$Q^2 = \frac{3 \int_{-\infty}^{\infty} \frac{(x-M)^2}{[Q^2 + (x-M)^2]^2} f(x)\, dx}{\int_{-\infty}^{\infty} \frac{1}{[Q^2 + (x-M)^2]^2} f(x)\, dx}. \tag{105a}$$

If ε is written instead of Q, this is the defining formula for the dihesion.

Obviously the deduction of the formulae for the most frequent value and dihesion is the same in all steps as the deduction of the formulae for the expected value and the scatter (variance), only the substituting distributions differ from each other.

7.3 Short remark on the maximum likelihood principle

Clearly we get the formulae for estimating E, σ, M and ε on the grounds of the sample $x_1, \ldots, x_i, \ldots, x_n$, if we put $f(x) = \frac{1}{n} \sum_{i=1}^{n} \delta(x - x_i)$ into the Eqs. (101a), (102a), (104a) and (105a):

$$E_n = \frac{1}{n} \sum_{i=1}^{n} x_i; \tag{101b}$$

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - E_n)^2; \tag{102b}$$

$$M_n = \frac{\sum_{i=1}^{n} \frac{x_i}{\varepsilon_n^2 + (x_i - M_n)^2}}{\sum_{i=1}^{n} \frac{1}{\varepsilon_n^2 + (x_i - M_n)^2}}; \tag{104b}$$

$$\varepsilon_n^2 = \frac{3 \sum_{i=1}^{n} \frac{(x_i - M_n)^2}{[\varepsilon_n^2 + (x_i - M_n)^2]^2}}{\sum_{i=1}^{n} \frac{1}{[\varepsilon_n^2 + (x_i - M_n)^2]^2}}. \tag{105b}$$

The maximum likelihood principle starts from the sample and therefore its primarily given formulae are expressed by sums. There is a difference of principal importance, viz: it is supposed that the actual error distribution is exactly known. The author means that modelling (substituting) is a more suitable procedure in practice because geophysicists and geologists never know precisely the type of distribution in advance.

What are the differences between the respective formulae obtained by applying the maximum likelihood principle and that of the minimization of the *I*-divergence?

The answer maybe surprising.

The maximum likelihood principle results in the same equations for E_n, σ_n, and M_n (Eqs. (101b), (102b) and (104b)) but instead of Eq. (105b) we get

$$
\bar{\varepsilon}_n^2 = \frac{\displaystyle\sum_{i=1}^{n} \frac{(x_i - M_n)^2}{\bar{\varepsilon}_n^2 + (x_i - M_n)^2}}{\displaystyle\sum_{i=1}^{n} \frac{1}{\bar{\varepsilon}_n^2 + (x_i - M_n)^2}}.
\tag{106}
$$

Is, perhaps, the application of this formula more advantageous for determining the scale parameter than Eq. (105b)? Lack of space prevents this question from being discussed in detail but two references are necessary. Firstly, CSERNYÁK and STEINER [1985c] showed that Eq. (105b) defines a resistant procedure but Eq. (106) is sensitive to outliers. Secondly, by the simultaneous fulfilment of Eqs. (106) and (104b) the asymptotic variance of *M* can be infinite, too, even for a symmetrical distribution [CSERNYÁK und STEINER 1985a]; on the contrary, for symmetrical distributions the usual way of calculating most frequent values (i.e. by using Eqs. (105b) and (104b)) is always characterized by a finite asymptotic variance.

Chapter 8
Generalized most frequent values

8.1 Most frequent value being optimum for a pair of distribution types

8.1.1 Dependence of efficiencies upon the scale parameter

It was shown in the section (7.3) that the most frequent value can also be regarded as the maximum likelihood estimate of the symmetry point of the Cauchy distribution (in a similar way to the arithmetic mean being the maximum likelihood estimate of the symmetry point of the Gaussian distribution). This type of estimate, however, gives an estimate of minimum asymptotic variance [CRAMÉR 1958]; the conditions needed are fulfilled for the whole $f_a(x)$ supermodel. Consequently, investigations can also be made on the absolute efficiencies.

Our goal here is to find a factor k so that using k times the dihesion as the parameter of scale ($S = k\varepsilon$), this algorithm should result in just the same efficiency both for the Gaussian and for the Cauchy type distributions. The importance of this question is obvious as we often face situations when no a priori knowledge on the distribution type of error is available.

Our method is the following. As Eq. (80) is valid for arbitrary scale parameters, we calculate both for the Gaussian and for the Cauchy type distribution the efficiencies as a function of k according to the formula

$$e(k) = \frac{A^2_{min.}}{A^2(M, k\varepsilon)} \, 100\%. \tag{107}$$

The results are visualized in *Fig. 45*. For the Cauchy distribution the maximum efficiency clearly belongs to $k = 1$; on the other hand, for the Gaussian distribution the maximum efficiency is only asymptotically reached for $k \to \infty$ (approximating the case of equal weights, i. e. the case of arithmetic means).

The intersection point of the curves in Fig. 45 gives the answer to our primary problem: approximately $k = 1.9$ defined that procedure which gives the same efficiency for both distribution types in question. We can ascertain that this common efficiency is greater than 90%. (Practically the same can be found if we use the round number $k = 2$.)
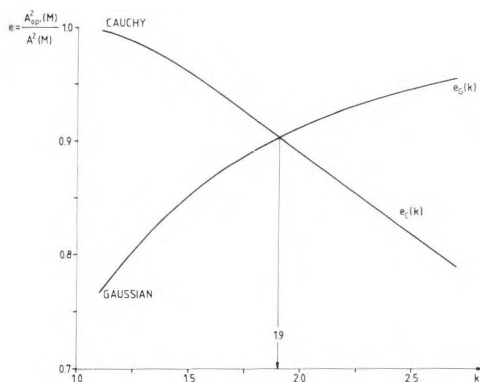
Fig. 45. Efficiencies (e) as a function of k (applied in the generalized most frequent value calculations), for two distribution types. Using $k = 1.9$, the efficiencies of both Cauchy and Gaussian distributions are the same (and greater than 90%)

45. ábra. Hatásfokok két eloszlásra a k függvényében (ez az általános leggyakoribb értékek számitásánál alkalmazott faktor). Ha $k = 1,9$-cel számolunk, a Cauchy- és a Gauss-eloszlásra azonos (mégpedig 90%-nál nagyobb) hatásfokot kapunk

Рис. 45. Эффективности двух распределений как функция k, являющегося фактором, применяемым в расчете обычных наиболее частых значений. При производстве расчетов с $k = 1,9$ эффективность распределений Коши и Гаусса оказывается одинаковой, а именно, свыше 90%

8.1.2 *Short discussion of the estimate* M'

Using 1.9ε instead of ε, the result is in general different from M. Denoting this new estimate by M', its definition is clearly

$$M' = \frac{\displaystyle\int_{-\infty}^{\infty} \frac{x}{(1.9\varepsilon)^2 + (x - M')^2}\, f(x)\, \mathrm{d}x}{\displaystyle\int_{-\infty}^{\infty} \frac{1}{(1.9\varepsilon)^2 + (x - M')^2}\, f(x)\, \mathrm{d}x}. \tag{108}$$

The isolines of relative efficiencies—similarly as was shown for E and M in Fig. 31—are given in *Fig. 46* for E and M' and for the $f_T(x)$ supermodel. The relative efficiencies are also given in *Fig. 47* for M and M', respectively, to compare these two variants of most frequent value procedures.
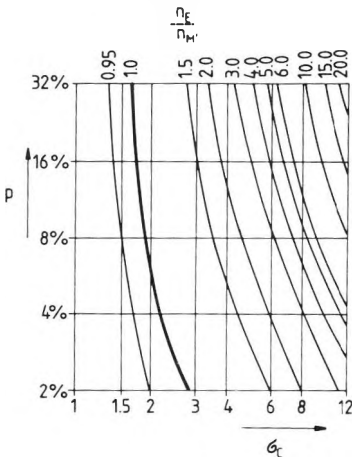
Fig. 46. Comparison of arithmetic means and generalized most frequent values (M' belongs to $k = 1.9$) on the supermodel of Tukey-type. (For notations, see caption of Fig. 31)

46. ábra. Számtani átlag és a $k = 1.9$-hez tartozó általánosított leggyakoribb érték (M') összehasonlítása a Tukey-féle szupermodellen (a jelöléseket illetően l. a 31. ábrát)

Рис. 46. Сопоставление среднеарифметического и обобщенного наиболее частого значения (M') при $k = 1.9$ в супермодели Тьюки (условные обозначения — как на рис. 31.).

Fig. 47. Comparison of M and M' on the supermodel of Tukey type

47. ábra. Az M és M' összehasonlítása a Tukey-féle szupermodellen .

Рис. 47. Сопоставление M с M' в супермодели Тьюки.
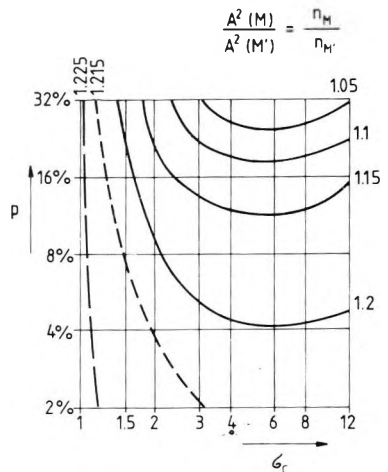
8.2 Optimum estimates for an arbitrary member of the distribution family $f_a(x)$

8.2.1 Formula for the weight function and the parameter of scale to be used. General most frequent values

Substituting an arbitrary element of the $f_a(x)$ supermodel in its general form given by Eq. (88c) into Eq. (98), i.e. minimizing the loss of information, we get the same, extremely simple form of weights to be used as by substituting the Cauchy distribution. The weight function coincides with the density function of the substituting distribution exclusively in the case of a Cauchy distribution [HAJAGOS 1985a]. (Consequently the weighting according the Cauchy density function is more effective for $a \neq 2$ than the weighting according to the actual $f_a(x)$ density function—see Eq. (16)—even if we were to know this density function in advance.)

Which value must be used as parameter of scale? The asymptotic scatter (Eq. (80)) is required to be minimum, and if $f_a(x)$ is the actual distribution, this is fulfilled by S satisfying

$$S^2 = \frac{(a+1) \int\limits_{-\infty}^{\infty} \dfrac{(x-M_k)^2}{[S^2+(x-M_k)^2]^2} f(x)\, dx}{\int\limits_{-\infty}^{\infty} \dfrac{1}{[S^2+(x-M_k)^2]^2} f(x)\, dx} \tag{109}$$

[HAJAGOS 1985a]. (Two remarks: substituting Eq. (88c) in Eq. (99), we get the same formula; in the case of $a = 2$ Eq. (109) is equivalent with Eq. (28).)

In Eq. (109) M_k is the generalized most frequent value estimated by the formula

$$M_k = \frac{\sum\limits_{i=1}^{n} \dfrac{(k\varepsilon)^2}{(k\varepsilon)^2+(x_i-M_k)^2} x_i}{\sum\limits_{i=1}^{n} \dfrac{(k\varepsilon)^2}{(k\varepsilon)^2+(x_i-M_k)^2}} \tag{110}$$

The value k in Eq. (110) can be defined as the quotient S/ε (the value S satisfies Eq. (109), ε satisfies Eq. (28)). The resistance of the procedure on the ground of Eq. (109), however, becomes more and more questionable at great values of a. It is much more practical to determine the dihesion and to multiply it by the k-value calculated according to the empirical formula:

$$k(a) = \sqrt{a-1} + \frac{1}{2\pi} \frac{\sqrt{a}-\sqrt{2}}{\sqrt{3}} \qquad (a > 2). \tag{111}$$

(The corresponding $k(a)$ curve is shown in *Fig. 48*). Consequently, if we know that the type of actual error distribution can be expected in the neighbourhood of $f_a(x)$ (to a given a-value), we use with $k = k(a)$ Eq. (110) and

$$\varepsilon^2 = \frac{3 \sum_{i=1}^{n} \dfrac{(x_i - M_k)^2}{[\varepsilon^2 + (x_i - M_k)^2]^2}}{\sum_{i=1}^{n} \dfrac{1}{[\varepsilon^2 + (x_i - M_k)^2]^2}} \tag{112}$$

(as a twofold iteration) to determine M_k, the general most frequent value. In practice, it is very important that *the program for determining the general most frequent value* (M_k) *differs only in a single multiplication from the program for determining M*.
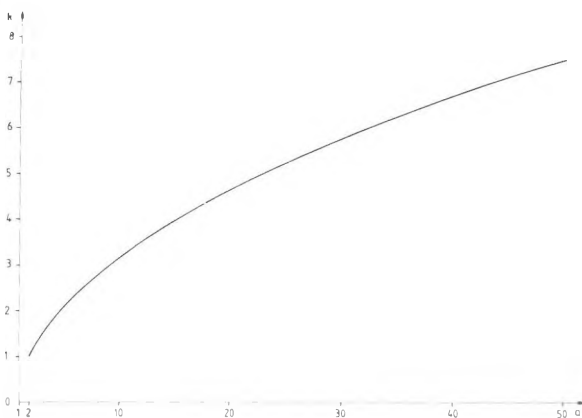
Fig. 48. Curve giving those k-values to be used to achieve an efficiency of 100% for an $f_a(x)$-distribution

48. ábra. A görbe azokat a k-értékeket adja meg, amelyekkel adott a-hoz tartozó $f_a(x)$-nél 100%-os hatásfok érhető el

Рис. 48. Значения k в виде кривой, с которыми при любом $f_a(x)$ для заданных a можно добиться эффективности в 100%.

We have no space here to discuss the problem of type determinations. (In the program library of the Geophysical Department of the University of Miskolc there are programs for this purpose.) One possibility in the $f_a(x)$ super-model should be mentioned: the sum counterpart of the integral formula given in Eq. (80) (see Eq. (142)) for estimating the asymptotic variance (for large n) can be minimized by suitable $S = k\varepsilon$ (see also Eq. (113)); divided by dihesion ε, the $k = k(a)$ value obtained defines the type parameter a, $k(a)$ being a monotonic function. The use of the well known 'χ^2 test', however, is dangerous (see Appendix III): this classical method, calculating on commonly used levels of significance, can lead to such great losses in efficiencies which are no longer acceptable.

8.2.2 Efficiencies for the supermodel $f_a(x)$. The estimates M'' and M'''. Some conclusions

Using $k = k(a)$ for a given a, the asymptotic variance can be calculated easily for the whole supermodel $f_a(x)$ as for these cases the following equation clearly holds (cf. Eq. (80))

$$A_k^2(a) = \frac{\displaystyle\int_{-\infty}^{\infty} \frac{x^2}{[(k\varepsilon)^2 + x^2]^2} f_a(x)\, dx}{\left[\displaystyle\int_{-\infty}^{\infty} \frac{(k\varepsilon)^2 - x^2}{[(k\varepsilon)^2 + x^2]^2} f_a(x)\, dx\right]^2} \tag{113}$$

On the other hand [HAJAGOS 1985a], the minimum variance for every $a > 1$ is given by

$$A_{\min}^2(a) = \frac{a+2}{a(a-1)}. \tag{114}$$

(In paragraph 4.4.2 the formula for the Cramér-Rao bound is given; substituting Eq. (88) in this expression, the result is Eq. (114).) Consequently, the absolute efficiencies of algorithms defined by k are to be calculated as

$$e_k(a) = \frac{A_{\min}^2(a)}{A_k^2(a)}. \tag{115}$$

The question arises as to which values of k should be used beyond the already chosen $k = 1.9$ and the original $k = 1$? *Figure 49* gives the square roots
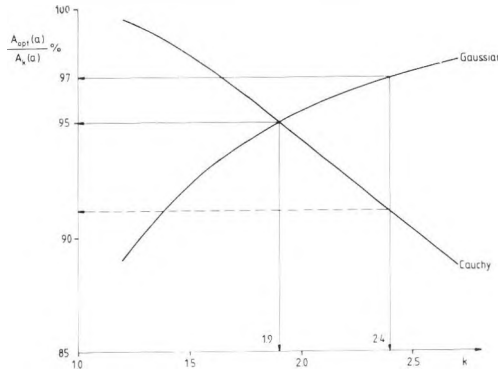
Fig. 49. Pair of curves (giving the quotient of accuracies) for defining M'' (i.e. M_k for $k = 2.4$)

49. ábra. Pontosságarányokat feltüntető görbepár az M'' definiálásához (M'' azonos M_k-val $k = 2.4$ esetén)

Рис. 49. Пара кривых с соотношениями точностей для определения M'', равным M_k при $k = 2.4$.

of the efficiencies on the ordinate, i.e. the curves show the relative accuracies, otherwise Fig. 49 is the same as Fig. 45. If we 'approve' only 3% in the loss of accuracy for the Gaussian distribution (instead of 5% in the case of M') then $k = 2.4$ is to be used, and the corresponding M_k is denoted by M''; the loss of accuracy in the case of the Cauchy distribution is about 9%. The corresponding loss in efficiency is 16.8%; if the Cauchy distribution very rarely occurs, this efficiency of 83.2% (or even less) is also acceptable. The M_k estimate working with $k = 2.8$ is optimum in the middle of the Jeffreys interval, i.e. at $a = 8$; its symbol is M'''.

The procedure determining M' is optimum at $f_a(x)$-distribution of $a = 4.4$; similarly the value $a = 6.2$ belongs to M''. These density functions are shown in *Figs. 50* and *51*, respectively; $f_{4.4}(x)$ has not too heavy flanks, and in the case
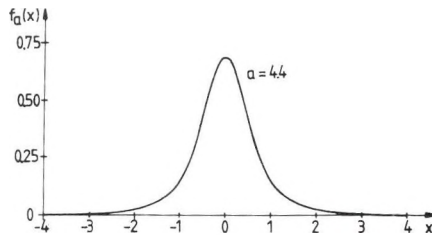
Fig. 50. Density function of $f_a(x)$ for $a = 4.4$, for which the M'-estimate (using $k = 1.9$) has an efficiency of 100%

50. ábra. Az $f_a(x)$ sűrűségfüggvény $a = 4,4$ esetén; erre az eloszlásra 100% a $k = 1,9$-del számított M' hatásfoka

Рис. 50. Плотностная функция $f_a(x)$ для $a = 4,4$; для данного распределения эффективность M', рассчитанного для $k = 1,9$, составляет 100%.
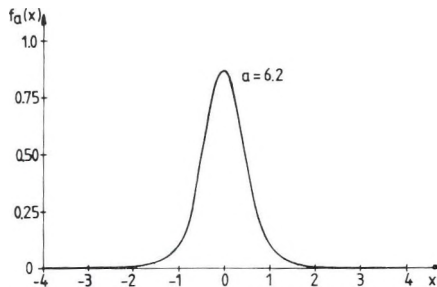
Fig. 51. Density function of $f_a(x)$ for $a = 6.2$, for which the M''-estimate (using $k = 2.4$) has maximum efficiency

51. ábra. Az $f_a(x)$ sűrűségfüggvény $a = 6,2$ esetén, amire az M''-becslés ($k = 2,4$) maximális hatásfokú

Рис. 51. Плотностная функция $f_a(x)$ для $a = 6,2$, при котором оценка M'' ($k = 2,4$) обладает максимальной эффективностью.

of $f_{6,2}(x)$ we can already speak about short flanks (shorter than those of the distribution of Newcomb's data, shown in Fig. 38).

Although *Figs. 52* and *53* have different abscissae (these being a and $1/(a-1)$, respectively) both refer to the distribution type. The ordinates are the same these being in both figures the relative accuracy A_{opt}/A_k in per cent for M, M', M'' and also for the arithmetic mean (E). For the last of these the A_{opt}/E value can be given in a very simple analytical form (see Eqs. (89) and (114)):

$$\frac{A_{opt}}{E} = 100 \sqrt{\frac{(a+2)(a-3)}{a(a-1)}}\% \quad (a>3). \tag{116}$$

This is the limit case if $a \to \infty$ and at the same time $k \to \infty$, i.e. least squares calculations can be regarded as general most frequent value calculations, too. The domain of great k-values, however, does not define robust procedures: as is clearly shown in Fig. 53 (see curve E), the relative accuracy (the square root of the efficiency) rapidly decreases already in the very neighbourhood of the Gaussian distribution. The other curves in Fig. 53., however, show excellent robustness the upper value being 95% (or even very near 100%) in a broad type-interval.
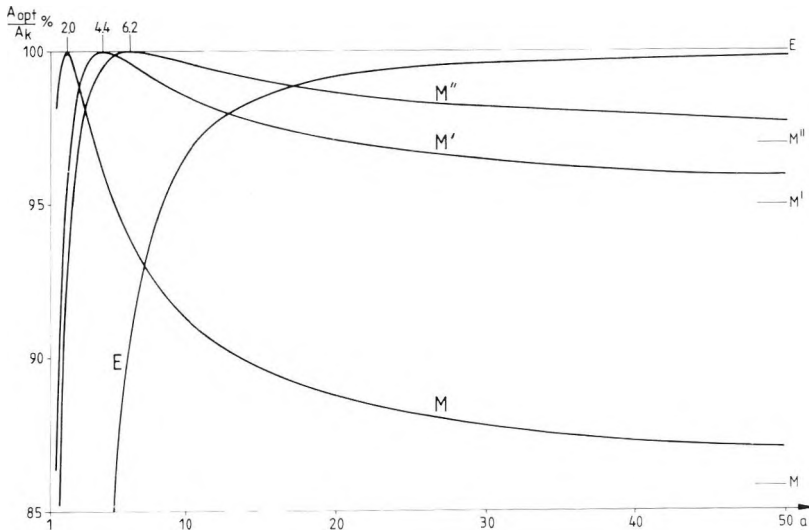
Fig. 52. Quotients of accuracies versus a (type-parameter), for arithmetic means (E) and different most frequent values

52. ábra. Pontosság arányok az a típusparaméter függvényében számtani átlagokra (E) és különböző leggyakoribb értékekre

Рис. 52. Соотношения точностей как функции типового параметра a для среднеарифметических (E) и различных наиболее частых значений.
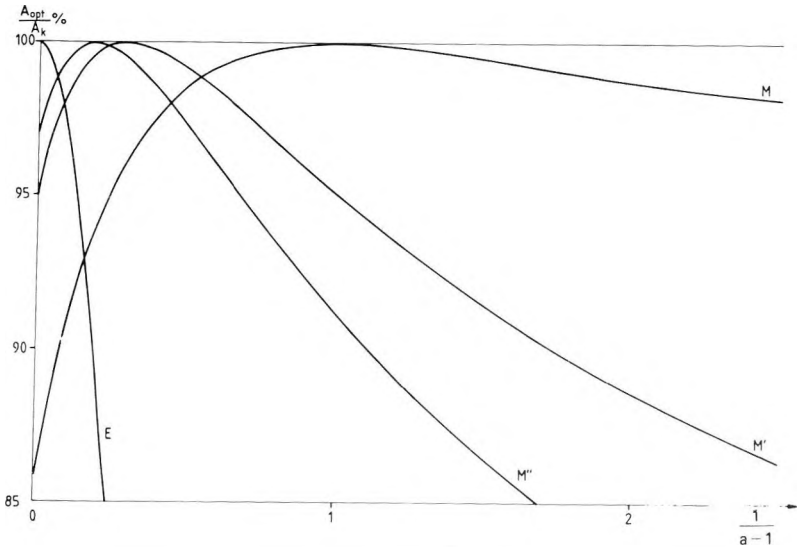
Fig. 53. Quotients of accuracies versus $1/(a-1)$ for different estimates

53. ábra. Pontosság arányok $1/(a-1)$ függvényében különböző becslésekhez

Рис. 53. Соотношения точностей как функции $1/(a-1)$ для различных оценок.

It is obviously more advantageous to use $1/(a-1)$ as the abscissa than simply a. (Its theoretical background is that type differences measured as differences of $1/(a-1)$ values are approximately proportional to other 'type differences' defined in mathematical statistics; CSERNYÁK [unpublished manuscript] deals with this problem. Using the quantity $1/(a-1)$ as the abscissa, the interval between 0 and 1 (i.e. between the Gaussian and Cauchy type) is the most important one, that part between 0.1 and 0.2 being by far the most interesting one. This is entitled 'Jeffreys interval' already mentioned as the a interval, this was defined by $6 < a < 10$).

Table VI shows efficiencies for five types of general most frequent values (including E for $k \to \infty$), and for the end points of the afore mentioned basic interval $0 < 1/(a-1) < 1$ (i.e. for Gaussian and Cauchy types). The author hopes that a detailed discussion is superfluous but a particular fact must by all means be mentioned to be able to make real comparisons. The Princeton study of robust estimates [ANDREWS et al. 1972] dealt with 68(!) different procedures applied to samples coming from a great variety of distribution types. *'No estimate in this study achieves 95% efficiency at normal distribution* (relative to the mean) *and 33% efficiency at the Cauchy* (relative to the maximum likelihood estimate) *simultaneously* for all sample sizes 5, 10, 20, 40' [ANDREWS et al. 1972, p. 253]. Remembering the behaviour of the most frequent values at small sample sizes (showing approximately the asymptotic behaviour) and looking at the efficiencies of 96% and 77% of M''' in Table VI, the conclusion can be drawn

that the application of the general most frequent value calculations can definitely be proposed for geophysicists and geologists (and also for experts in other branches of science, too,) because of their robustness and other properties, see Table III. The flexibility must be separately enhanced: the optimum procedure is defined always by only a single parameter (k).

Estimates	Defining factor (k)	Efficiencies of the estimates	
		for Gaussian distribution	for Cauchy distribution
M	1.0	73.7%	100 %
M'	1.9	90.2%	90.2%
M''	2.4	94.1%	83.2%
M'''	2.8	96.0%	77.4%
E	∞	100 %	0 %

Table VI. Efficiencies of various generalized most frequent value calculations (the calculation of arithmetic mean denoted by E as the limiting case is included), for two types of probability distributions

VI. táblázat. Hatásfokok kétféle eloszlásra és különböző általános leggyakoribb értékekre (beleértve határesetként az E-vel jelölt számtani átlagképzést is)

Таблица VI. Эффективности двух распределений и обобщенного наиболее частого значения, включая — в качестве предельного случая — среднеарифметическое E.

8.2.3 M_k-fittings

It is obvious that fitting problems with more than only one single unknown occur in practice much more frequently (as a part of our geophysical interpretation algorithms) than the simplest case. This short introductory monograph has not the possibility to deal with these cases in detail but after a slightly more detailed discussion of the generalization of the simplest case and remembering Sections 2.4 and 3.3, fitting problems can be treated easily.

Adjustment according to the general most frequent value can be generalized in the same way as the simplest case: the weights

$$\frac{(k\varepsilon)^2}{(k\varepsilon)^2 + [x_i - T(\mathbf{p}; \mathbf{y}_i)]^2} \tag{117}$$

are to be used (in every fitting step of the twofold iteration) instead of

$$\frac{\varepsilon^2}{\varepsilon^2 + [x_i - T(\mathbf{p}; \mathbf{y}_i)]^2},$$

supposing that the errors of x are distributed according to $f_a(x)$, $(k = k(a)$, see Eq. (111)).

In some cases there are connections to be exactly fulfilled among the components $p_1, p_2, ..., p_j, ..., p_J$ of the parameter vector $\mathbf{p}$. (A classical example is that the sum of the angles of a triangle must be 180°.)

Introducing the symbol

$$d_i = x_i - T(\mathbf{p}; \mathbf{y}_i),\qquad(118)$$

the *l*-th constraint can be written as

$$C_l(\mathbf{p}) = 0.\qquad(119)$$

These constraints can be taken into consideration most easily for least squares calculations by using the so-called Lagrange multipliers and fulfilling

$$\sum_{l=1}^{L} \lambda_l C_l(\mathbf{p}) + \sum_{i=1}^{n} d_i^2 = \min.\qquad(120)$$

Since, clearly,

$$\prod_{i=1}^{n} [(k\varepsilon)^2 + d_i^2] = \min.\qquad(121)$$

is the generalization of Eq. (38a), in the case of conditions to be fulfilled, analogously to Eq. (120),

$$\exp\left[\sum_{l=1}^{L} \lambda_l C_l(\bar{p})\right] \cdot \prod_{i=1}^{n} [(k\varepsilon)^2 + d_i^2] = \min.\qquad(122)$$

is to be solved (after logarithmization the analogy is obvious). An important application is that of fitting with spline functions. (The program for spline fitting according to the general most frequent values—also containing some results from the following section—is available in the Geophysical Department of the University of Miskolc.)

8.2.4 Calculation of errors

In close connection with Eq. (121), it should be mentioned that a possible measure of the error is

$$U_k = \varepsilon \left\{ \prod_{i=1}^{n} \left[1 + \left(\frac{d_i}{k\varepsilon}\right)^2 \right] \right\}^{\frac{1}{2n}}\qquad(123)$$

[FERENCZY et al. 1988], if *k* was used during the fitting. If nothing is known about the type of distribution, and/or only the error itself is of interest, fitting is to be made using $k = 2$, and the following formula is applied for uncertainty:

$$U = \varepsilon \left\{ \prod_{i=1}^{n} \left[1 + \left(\frac{d_i}{2\varepsilon}\right)^2 \right] \right\}^{\frac{1}{2n}}\qquad(123a)$$

New error definitions are unavoidable as the sample variance is sensitive to outliers to an extreme degree, and—on the other hand—its asymptotic variance (the 'error of the error') is infinite already at $f_5(x)$ (and naturally for all $f_a(x)$ distributions for which $a < 5$ holds [HAJAGOS and STEINER 1988]). These new error definitions give for the whole $f_a(x)$ supermodel with good approximation the value of $(1/2) \cdot [F^{-1}(1 - 1/6) - F^{-1}(1/6)]$ [CSERNYÁK et. al. 1988], i.e. it is twice as likely that for the actual deviation $|d_i| < U$ holds than the fulfilment of the relation $d_i > U$. (N. B. the same is approximately valid for σ in the case of Gaussian distribution.)

U and U_k obviously characterize the error of a single measurement. As for the errors of the general most frequent values, however, for them naturally the asymptotic variance or asymptotic scatter is characteristic. As HAJAGOS [1985a] has shown, in this case the formula

$$A(M_k) = \frac{k\varepsilon\sqrt{n(k\varepsilon) - n_2}}{2n_2 - n(k\varepsilon)} \tag{124}$$

holds, where $n(.)$ is defined by Eq. (22) (with similar heuristic meaning: this is the effective number of data divided by n); n_2 is defined by

$$n_2 = \int_{-\infty}^{\infty} \left(\frac{(k\varepsilon)^2}{(k\varepsilon)^2 + d_i^2} \right)^2 f(x)\,dx$$

(i.e. n_2 is the mean of the squared weights). Equation (81) is a special case of Eq. (124) with $k = 1$; namely, in this case $n_2 = 3/4 \cdot n(k\varepsilon)$ holds. For the supermodel $f_a(x)$, $A(M_k)$ differs from U only by some per cent, $1.06U = A(M_k)$ is approximately valid, and on the basis of this fact it is understandable that small samples also behave approximately according to the asymptotic rules.

8.3 Variants with increased resistance: M_{kb}- and M_{kc}-fittings

The IC-function of the M_k-estimates has the form

$$IC_k(x) = \text{const}\,\frac{x(k\varepsilon)^2}{(k\varepsilon)^2 + x^2} \tag{125}$$

(the factor depending upon $f(x)$). For great values of $|x|$, $|IC|$ is proportional to $1/|x|$—but there are cases for which this degree of diminishing is not enough [ZILAHI–SEBESS 1987]. In the following, two possibilities are shown for increasing the resistance (in paragraphs 8.3.1 and 8.3.2).

8.3.1 Cutting at $\varphi_{\min}$ and its influence on the efficiencies

Using the weight function

$$
\varphi_{kb}(x) = \begin{cases} \varphi_k(x) = \dfrac{(k\varepsilon)^2}{(k\varepsilon)^2 + x^2}, & \text{if} \quad \varphi_k(x) \geqq \varphi_{\min}, \\[4mm] 0, & \text{if} \quad \varphi_k(x) < \varphi_{\min}. \end{cases} \tag{126}
$$

$IC_{kb}(x) = 0$ holds for $|x| > bk\varepsilon$, where

$$
b = \sqrt{\frac{1 - \varphi_{\min}}{\varphi_{\min}}}. \tag{126a}
$$

If $f(x)$ is symmetric to the origin, the corresponding asymptotic variance of the M_{kb} values obtained is given [HAJAGOS and STEINER 1988] by the formula

$$
A_{kb}^2 = \frac{\displaystyle\int_{-bk\varepsilon}^{bk\varepsilon} \frac{x^2(k\varepsilon)^4}{[(k\varepsilon)^2 + x^2]^2} f(x)\, dx}{\left[\displaystyle\int_{-bk\varepsilon}^{+bk\varepsilon} \frac{(k\varepsilon)^4 - x^2(k\varepsilon)^2}{[(k\varepsilon)^2 + x^2]^2} f(x)\, dx - \frac{2b(k\varepsilon)^3}{(k\varepsilon)^2 + (bk\varepsilon)^2} f(bk\varepsilon)\right]^2}. \tag{127}
$$

Figure 54 shows the efficiency curves for $\varphi_{\min} = 0.1$ and 0.2; in both cases $k = 1.9$. The efficiency curve is also drawn for $\varphi_{\min} = 0$ (i.e. also for M') for comparison. It is clearly seen that $\varphi_{\min} = 0.1$ is by all means permitted and also at $\varphi_{\min} = 0.2$ the decrease of efficiency is nowhere greater than 3%.

The increase in the resistance unavoidably causes the decrease of efficiencies as 'good' data are also cut. The analogy of ANSCOMBE [1960] for such cases is very appropriate, therefore the paragraph relating to this is cited from HUBER [1981 p. 73]: 'ANSCOMBE's [1960] comparison of the situation with an insurance problem is very helpful. Typically a so-called classical procedure is the optimal procedure for some ideal (usually normal) model. If ... we want to insure against accidents caused by deviations from the model, we clearly will have to pay for it by sacrificing some efficiency at the model. The questions are, of course, how much efficiency we are willing to sacrifice, and against how bad a deviation we would like to insure.'
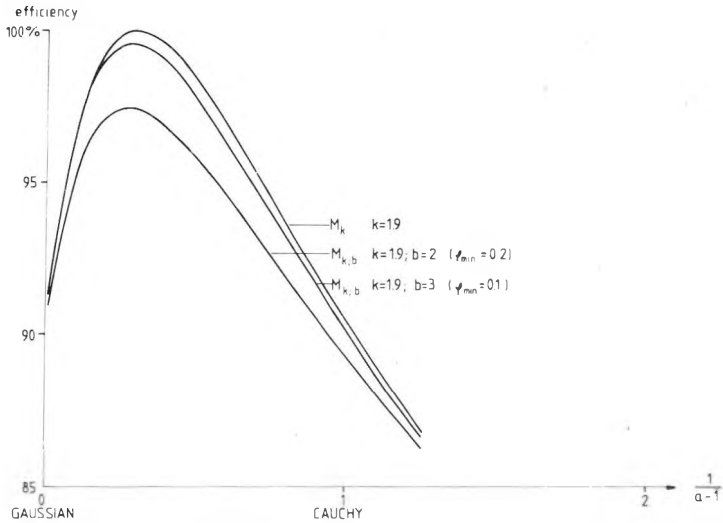
Fig. 54. The efficiencies show not too great decrease if data with weights less than 0.1 and 0.2, are neglected, in order to increase the resistance

54. ábra. A hatásfokok nem csökkennek túlságosan nagy mértékben, ha a rezisztencia növelése céljából elhagyjuk a 0,1-nél vagy 0,2-nél kisebb súlyú adatokat

Рис. 54. Не слишком значительное убывание эффективностей при исключении данных весами менее 0,1 или 0,2 с целью увеличения устойчивости.

8.3.2 Increasing the resistance by M_{kc}-fitting

M_{kc}-fitting is defined using the following weight function [HAJAGOS and STEINER 1988]:

$$\varphi_{kc}(x) = \begin{cases} \varphi_k(x) = \dfrac{(k\varepsilon)^2}{(k\varepsilon)^2 + x^2}, & \text{if} \quad |x| \leq ck\varepsilon; \\[2ex] \dfrac{1}{1+c^2} \exp\left\{ \dfrac{c^2}{1+c^2}\left[1 - \dfrac{x^2}{(ck\varepsilon)^2} \right] \right\}, & \text{if} \quad |x| > ck\varepsilon. \end{cases} \quad (128)$$

This is obviously a smooth weight function but it guarantees excellent resistance.

A detailed discussion of this case (e.g. the formula for the corresponding asymptotic variance) is given by HAJAGOS and STEINER [1988]. *Figure 55* shows efficiency curves for $c = 1.5$ and $c = 1.0$, using $k = 1.9$ in both cases, and for comparison the efficiency curve of M' (i.e. of M_k with $k = 1.9$) is also given.

In Fig. 55 it can clearly be seen that this way of increasing the resistance is always applicable for $c = 1.5$ and is often applicable for $c = 1.0$ the maximum decrease of the efficiency being even in the latter case only 2%.
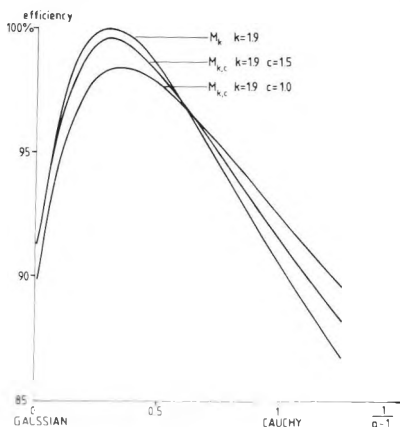
Fig. 55. Efficiency curves versus $1/(a-1)$ of the more resistant variants M_{kc} of M'

55. ábra. Hatásfok-görbék $1/(a-1)$ függvényében az M' fokozott rezisztenciájú M_{kc} változataira

Puc. 55. Кривые эффективностей как функций $1/(a-1)$ для вариантов M' с повышенной устойчивостью (M_{kc}).

If we know nothing about the type of actual distribution and increased resistance against outliers seems to be necessary, our standard procedure of fitting will be defined by $k = 1.9$ or 2, and $c = 1.5$.

* * *

The problems treated in the Appendices are instructive and useful both for a deeper understanding of concept dealt with in the present short monograph, and for applying these concept in geophysics and in geology (or in other branches of science and in practical applications).

APPENDIX I
Comparisons with other robust methods

CLAERBOUT and MUIR [1973] have proposed the use of the median (or more generally the use of the L_1 norm) for geophysical purposes. The median is perhaps the oldest robust estimate. But what about its efficiency? The answer is given for the $f_a(x)$ supermodel in *Fig. 56* comparing the asymptotic scatter of the median (m) with the optimum one. The curve for E (i.e. for the L_2 norm) is also given. In *Fig. 57* the surplus of data are shown (in per cent) needed for the same accuracy which characterizes the optimum algorithm. The curves for the median (m) and for M' (M_k with $k = 1.9$) considerably differ from each other: we find the minimum surplus of 20% (and occasionally a surplus of greater than 50%) too much from the economical point of view in the case of m (to say nothing about the less favourable calculating procedures based on the L_1 norm).
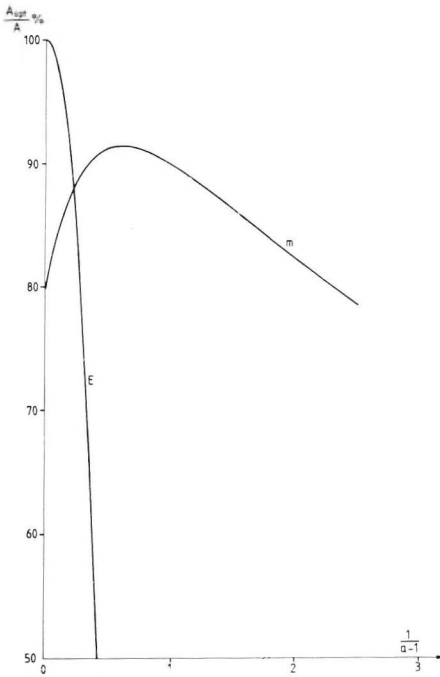
Fig. 56. A_{opt}/A curves (square root of efficiency) of sample medians (*m*) and of arithmetic means (*E*) versus $1/(a-1)$

56. ábra. A hatásfokok gyökei $1/(a-1)$ függvényében az *m* mintamediánokra és az *E* számtani átlagokra

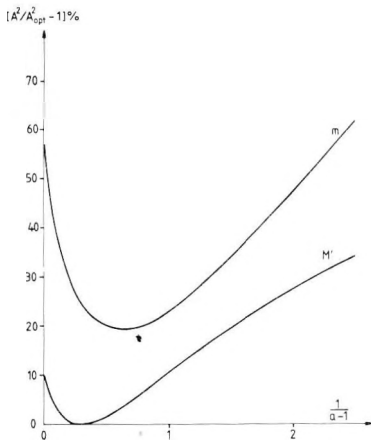*Рис. 56.* Кривые квадратного корня эффективностей как функций $1/(a-1)$ для медиан *m* и для среднеарифметических *E*.

Fig. 57. The curves show the percentage of extra data needed to achieve the same accuracy as by always using the optimal estimation procedure

57. ábra. A görbék azt mutatják, hogy hány százalék adattöbblet szükséges ahhoz viszonyítva, mintha minden eloszlástípushoz az optimális becslési eljárást alkalmaznánk

Рис. 57. Относительное количество дополнительных данных (в %) по сравнению с тем, что было бы необходимо при применении способа оптимальных оценок.

More sophisticated robust methods than the median are the α-trimmed mean (say, with $\alpha = 0.1$) and the Hodges–Lehmann estimate, both having a maximum efficiency in the Jeffreys interval *(Fig. 58)*. The basic formulae for the calculations are given in Chapter 4 and we get in this way e.g. for the efficiencies of the Hodges–Lehmann estimate, the following expression in the case of the $f_a(x)$ supermodel:

$$e_{H.L} = \frac{12(a+2)c^2(2a)}{a(a-1)c^4(a)}$$

where $c(a)$ is to be calculated according to Eq. (88a) (Table V). If we choose a most frequent value characterized by a similar place of maximum ($k = 2.8$ and $c = 1.0$), the much greater robustness and the economic advantages of the most frequent value calculations are obvious.

The generalization of the Hodges–Lehmann procedure into a fitting method itself is also problematic. Problems of this nature cause no difficulties when using the so-called Danish method [cited, for example by DETREKŐI 1986]. One version of this is defined by the weight function:

$$\varphi_D(d_i) = \begin{cases} 1, & \text{if} \quad d_i^2 < B^2 \\ \exp[1 - d_i^2/B^2], & \text{if} \quad d_i^2 \geq B^2; \end{cases} \tag{129}$$

d_i is given in Eq. (118) and B is defined by

$$B = \frac{D \cdot \text{med} |d_i|}{0.6745}. \tag{129a}$$
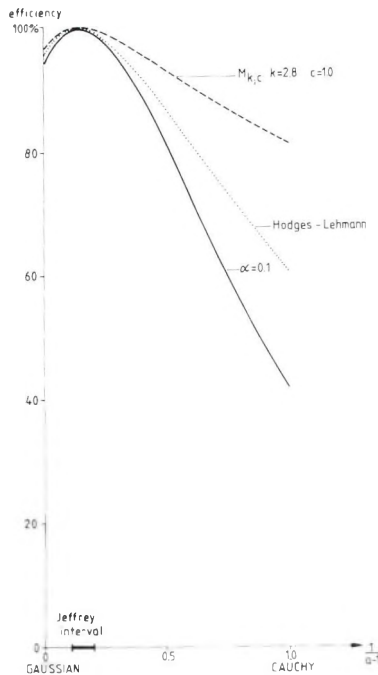
Fig. 58. Efficiency curves for three types of estimation and for supermodel $f_a(x)$

58. ábra. Hatásfok-görbék háromféle becslésre és az $f_a(x)$ szupermodellre

Рис. 58. Кривые эффективностей для трех различных оценок и для супермодели $f_a(x)$.

The corresponding efficiency curve for $D = 1.5$ is shown in *Fig. 59* (*D*-curve), having a maximum efficiency of about 90% [STEINER 1987]. The efficiencies of m, M' and E are also shown to compare them with the efficiency of the Danish method. The M' procedure seems to be even more advantageous than this new type of robust procedures.
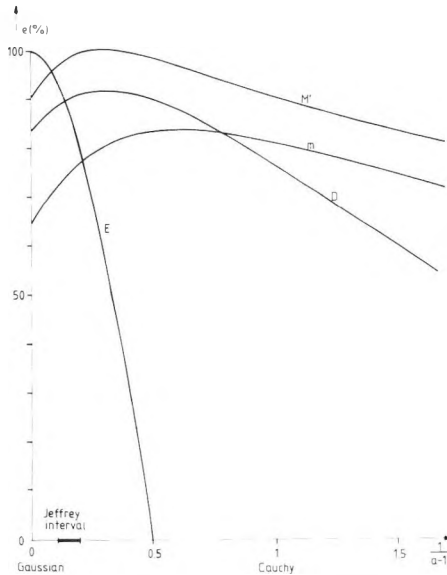
Fig. 59. Efficiency curves for the generalized most frequent value M', for the sample median m, for the arithmetic mean E, and for an alternative version (D) of the so-called Danish method

59. ábra. Hatásfok-görbék az M' általános leggyakoribb értékre, az m mintamediánra, az E számtani átlagra, valamint az ún. dán módszer egy változatára (D)

Рис. 59. Кривые эффективностей для обобщенного наиболее частого значения M', для медианы m, для среднеарифметического E, а также для варианта т.н. датского способа (D).

APPENDIX II
M^*-estimates

If we use the square of the weight given in Eq. (117), i.e. applying the weight function

$$\varphi^*(x_i) = \frac{(k\varepsilon)^4}{[(k\varepsilon)^2 + d_i^2]^2} \tag{130}$$

with d_i-s given in Eq. (118), a new procedure is defined. Similarly to M, M', M'' and M''' the symbols M^*, $M^{*\prime}$, $M^{*\prime\prime}$ and $M^{*\prime\prime\prime}$ are defined by $k = 2; 3.27; 4$ and 5 (with M^* having maximum efficiency at $a = 2$, i.e. at the Cauchy distribution, M^* at $a = 6$ and $M^{*\prime\prime\prime}$ at $a \approx 10$; $M^{*\prime}$ has the same efficiency for the Gaussian and the Cauchy distribution).

It is well known that the middle of the Jeffreys interval belongs to $a = 8$. For comparison, this density function is shown in *Figs. 60* and *61*, in which primarily the four variants of weight functions are shown both for $\varphi(.)$ according to Eq. (117) (Fig. 60) and for $\varphi^*(.)$ according to Eq. (130) (Fig. 61). The much better resistance of the latter is obvious but for adequate investigation of this question the IC-curves are necessary.
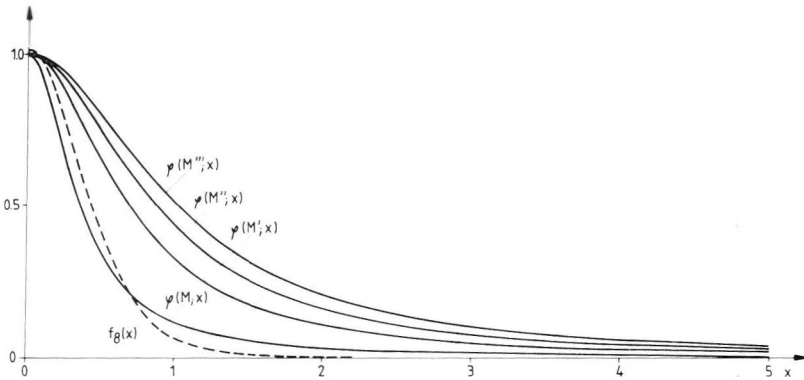
Fig. 60. Weight functions belonging to standard $f_8(x)$ distribution and to various M-estimates. Dotted line: density function $f_8(x)$

60. ábra. Standard $f_8(x)$-eloszláshoz és különböző M-becslésekhez tartozó súlyfüggvények. Az $f_8(x)$ sürüségfüggvényt a szaggatott görbe mutatja

Рис. 60. Весовые функции для стандартного распределения $f_8(x)$ и разных оценок M. Пунктирная линия обозначает плотностную функцию $f_8(x)$.
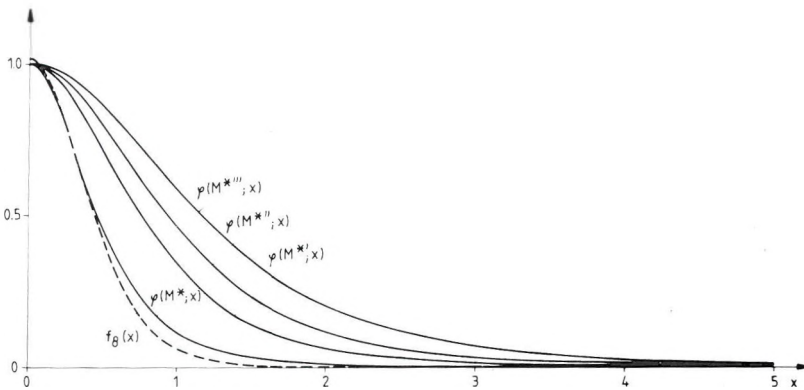
Fig. 61. Weight functions belonging to standard $f_8(x)$ distribution and to various M^*-estimates. Dotted line: density function $f_8(x)$

61. ábra. Standard $f_8(x)$-eloszláshoz és különböző M^*-becslésekhez tartozó súlyfüggvények. Az $f_8(x)$ sürüségfüggvényt a szaggatott görbe mutatja

Рис. 61. Весовые функции для стандартного распределения $f_8(x)$ и разных оценок M^*. Пунктирная линия обозначает плотностную функцию $f_8(x)$.

To calculate IC-values, first the formula of the corresponding $\varphi^*(.)$ must be given:

$$\varphi^*(x) = \frac{x}{(1+x^2)^2} \; ; \tag{131}$$

then, according to Eq. 64 (with $S = k\varepsilon$), the IC-functions are the following for $f_8(x)$ (to be calculated by $\varepsilon = 0.359$):

$$\text{IC}(x, F_8, M_k) = \frac{1}{\displaystyle\int_{-\infty}^{\infty} \frac{(k\varepsilon)^2 - y^2}{[(k\varepsilon)^2 + y^2]^2} \, f_8(y) \, dy} \cdot \frac{x}{(k\varepsilon)^2 + x^2} \, , \tag{132}$$

and

$$\text{IC}(x, F_8, M_k^*) = \frac{1}{\displaystyle\int_{-\infty}^{\infty} \frac{(k\varepsilon)^2 - y^2}{[(k\varepsilon)^2 + y^2]^3} \, f_8(y) \, dy} \cdot \frac{x}{[(k\varepsilon)^2 + x^2]^2} \, . \tag{133}$$

The corresponding IC curves are shown in *Figs. 62* and *63*, and for very large values of $|x|$ in *Fig. 64*. The resistance of M^*-methods is really much better than that of the original versions of the M_k-fitting (original means here that M_k-methods were compared and not M_{kb}- or M_{kc}-methods of increased resistance, see 8.3.1 and 8.3.2). But what about the efficiencies?

On the grounds of Eq. (71) and of Eq. (133) the asymptotic variance for M^*-estimates is given by

$$A^2 = \frac{\displaystyle\int_{-\infty}^{\infty} \frac{x^2}{[(k\varepsilon)^2 + x^2]^4} \, f_a(x) \, dx}{\left[\displaystyle\int_{-\infty}^{\infty} \frac{(k\varepsilon)^2 - 3x^2}{[(k\varepsilon)^2 + x^2]^3} \, f_a(x) \, dx \right]^2} \; ; \tag{134}$$
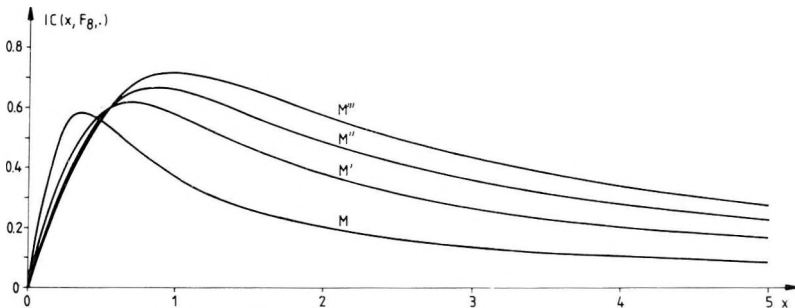

Fig. 62. Influence curves for various M-estimates of the distribution $f_8(x)$

62. ábra. Hatásfüggvények (IC-görbék) az $f_8(x)$ eloszláshoz tartozó különböző M-becslésekhez

Рис. 62. Кривые эффекта для различных оценок M распределения $f_x(x)$.

also taking Eq. (114) into consideration, the efficiency is easily calculated as A_{opt}^2/A^2.—Although on the one hand, the maximum efficiencies cannot reach 100%, they are very near it and, on the other hand (the efficiency curves being slightly more flat than those of the M_k-estimates), the demand of robustness is somewhat better fulfilled in the case of M_k^*-estimates than for M_k-procedures (see *Fig. 65*).
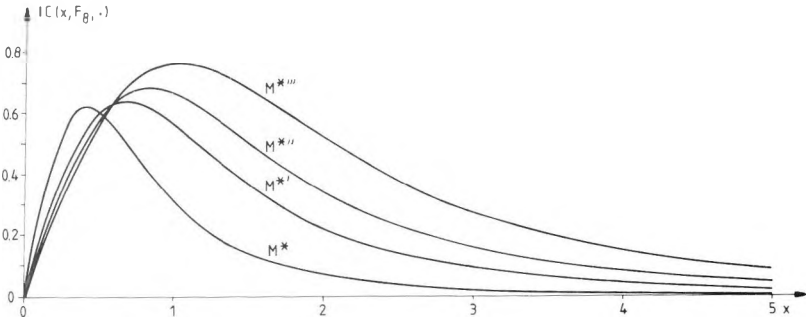

Fig. 63. Influence curves for various M^*-estimates of the distribution $f_8(x)$

63. ábra. Hatásfüggvények (IC-görbék) az $f_8(x)$ eloszláshoz tartozó különböző M^*-becslésekhez

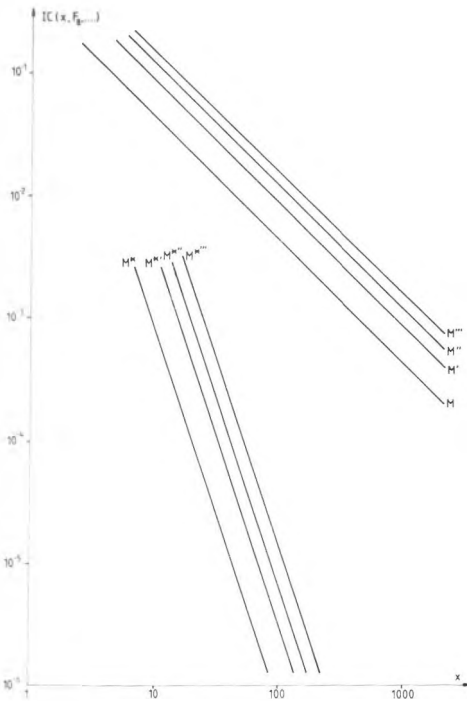*Рис. 63.* Кривые эффекта для различных оценок M^* распределения $f_8(x)$.

Fig. 64. Differing resistance of the M- and M^*-estimates shown with IC-curves for large values of x

64. ábra. M ill. M^*-becslésekhez tartozó hatásgörbék nagy x-ekre, a rezisztenciában mutatkozó jelentős különbségek szemléltetésére

Рис. 64. Кривые эффекта для различных оценок M и M^* при больших x-ах. Четко выявляются существенные различия в устойчивостях.
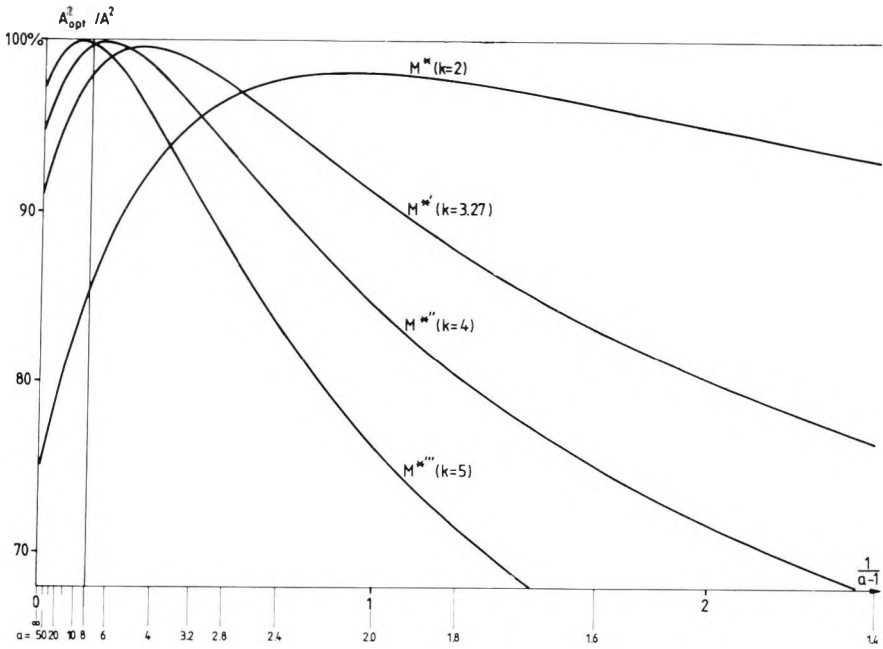
Fig. 65. Efficiency curves of M^*-estimates for the supermodel $f_a(x)$ (the distribution type is characterized by $1/(a-1)$ and by a)

65. ábra. Hatásfok-görbék különböző M^*-becslésekre és az $f_a(x)$ szupermodellre (az eloszlástipust $1/(a-1)$ és a egyaránt jelzi)

Рис. 65. Кривые эффективностей для различным оценок M^* и для супермодели $f_a(x)$ (тип распределения равным образом характеризуется как $1/(a-1)$, так и a).

As a curiosity it is mentioned finally that if

$$\sum_{i=1}^{n} \frac{S^2}{S^2 + (x_i - M^*)^2} (x_i - M^*)^2 = \text{minimum} \tag{135}$$

were demanded primarily (fulfilled only step-wise in M_k iterations), after differentiation and rearranging we get

$$M^* = \frac{\displaystyle\sum_{i=1}^{n} \frac{S^4 x_i}{[S^2 + (x_i - M^*)^2]^2}}{\displaystyle\sum_{i=1}^{n} \frac{S^4}{[S^2 + (x_i - M^*)^2]^2}}, \tag{136}$$

which is really the iteration formula for M^*-estimates (see the defining expression of weights in Eq. (130)).

APPENDIX III
Dangers of the χ^2-test

The most commonly applied method for investigating "normality" is the well known χ^2-test. Not only is the method itself given but also the table needed to perform such tests is given (for the commonly used significance levels) in all statistics handbooks [VINCZE 1968]. The question arises as to which values of a is the $f_a(x)$ distribution with a not negligible probability declared to be a Gaussian distribution, according to the χ^2-test. According to the Monte Carlo investigations of HAJAGOS [1984] 'the normality was proved' on the grounds of all samples investigated coming from the $f_a(x)$ distribution at $a = 10$, and the overwhelming majority of samples gave the same result at $a = 8$.

There are both economical and theoretical consequences. Concerning the first, if we use the least squares techniques (as "the normality is proved"), in the case of $f_8(x)$ we have—according to Eq. (116)—an efficiency of only 89%, and even in the case of $a = 10$ the efficiency is equal to 93%. Geophysical and geological data are too expensive for us to have to systematically throw out about 10% of our data.

If we regard the situation more generally, the main danger is that those experts can be misled on questions of normality who are cautious enough to make such tests, and this leads to secondary effects that are very much more dangerous for the great masses of appliers. In this sense we can even speak about the trap of the χ^2-test.

APPENDIX IV
Sterility of distributions

The notion 'sterility of distributions' can have various but rather trivial statistical meanings, too: 1) there are no contaminations, 2) $f(x) = 0$ holds outside a finite interval ('absolute sterility'), etc.

For the applier a distribution can be called sterile if the flanks are too small to characterize real situations in the given discipline. As $f(x) > 0$ for all x values for the commonly used models it is rather difficult to give an exact definition for the sterility. The definition in question cannot be a given analytical classification of the asymptotic behaviour of the density function for large $|x|$-values. To avoid subjectivity, this notion must be defined on the grounds of probability theory, regarding at the same time the specific properties of the discipline studied. Consequently, the acceptable definition of sterility may very much depend on the discipline, or even on the concrete task within the given discipline.

In the opinion of the author both demands are fulfilled concerning geophysics, if we analyse (for symmetrical distributions) the following estimate of a symmetry point:

$$T = \frac{\max (x_i) + \min (x_i)}{2},$$ (137)

stating that for this estimate the law of large numbers cannot be fulfilled in the overwhelming majority of practical cases of geophysics and geology. In other words estimate in Eq. (137) is extremely sensitive even to a single outlier; because not all effects can be taken into consideration in the theoretical treatment of geophysical problems, the probability increases with n that just such a neglected effect occurs—and this is equivalent with the statement that the reverse of the law of large numbers is fulfilled.

CSERNYÁK [1984] investigated this criterion of sterility for the $f_a(x)$ supermodel. He proved that the density function of the estimates defined by Eq. (137) is

$$g(z) = 2n(n-1) \int_{-\infty}^{z} F(2z-x) - F(x)^{n-2} f(2z-x) f(x) \, dx.$$ (138)

On the basis of this formula CSERNYÁK proved in various ways that the law of large numbers is inversely fulfilled in a definite manner up to a-values of 5 or 6; after an a-interval which is approximately independent of n, the law of large numbers is fulfilled and, in the case of $a \to \infty$ the increase in the accuracy of the estimate given by Eq. (137) is already considerable. Consequently, the Gaussian distribution and distributions near it can really be regarded as 'sterile' from the geophysical point of view.

APPENDIX V
Permitted percentage of outliers for fittings executed according to the concept of the generalized most frequent value

The breakdown point was defined in Chapter 4 somewhat too arbitrarily: outliers were very far from each other in this model. Some actual practical cases need other models, too, as outliers can also occur in gathered form. CLAERBOUT and MUIR [1973] give a geophysical example: If line defects occur in the global seismological network, the value zero becomes very probable. Defects in the equipment however may also often cause the maximum value to occur. Consequently, the asymmetric U-distribution consisting of two Dirac functions at points $x = +1$ and -1 as an adequate model for many such cases seems to be useful *(Fig. 66)*: K is the ratio of the outliers.

The resulting general most frequent value estimate of the location parameter is $+1$ from $K=0$ to a given maximum value of K; that is, this value is
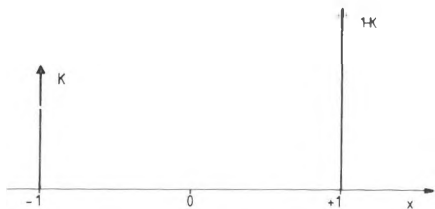
fully independent of outliers in this range of K values. This maximum value naturally depends on k or a (i.e. on the defining factors of the general most frequent value procedure); it also depends, to a limited degree on the choice of variant of the general most frequent value used. For the variant proposed in the present monograph this is the K_{max} curve in *Fig. 67,* but if twofold iteration is carried out using

$$M_k = \frac{\sum_{i=1}^{n} \dfrac{S^2 x_i}{S^2 + (x_i - M_k)^2}}{\sum_{i=1}^{n} \dfrac{S^2}{S^2 + (x_i - M_k)^2}} \tag{139}$$

and Eq. (109), the results are shown by the $K_{max}^{(a)}$ curve [HAJAGOS 1985b]. Let us remember that these two variants are fully equal for $f_a(x)$ distributions, but our present model (Fig. 66) is clearly very far from being such a distribution. In the
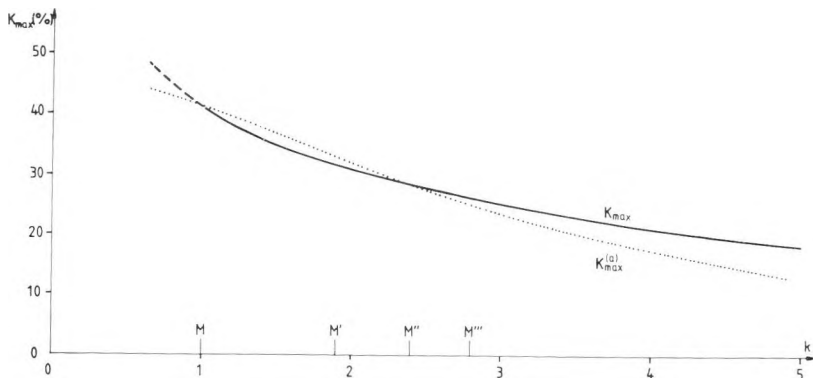

Fig. 67. Maximum percentage of concentrated outliers for various most frequent values M_k (full line). Dotted line belongs to a variant of the generalized most frequent value calculations where a scale parameter other than dihesion is primarily calculated

67. ábra. Koncentráltan jelentkező kieső adatok maximális aránya különböző M_k leggyakoribb értékekhez (folytonos vonal). A pontozott görbe az általános leggyakoribb értékszámítás azon variánsához tartozik, amelynél nem a dihézió-számítás a procedúra egyik alapeleme

Рис. 67. Максимальное отношение выпадающих данных, появляющихся концентрированно, к различным наиболее частым значениям M_k (сплошная линия). Пунктирная кривая принадлежит к тому варианту обобщенного расчета наиболее частого значения, при котором расчет дигезии не является одним из основных элементов процедуры.

interval $1 \leqq k \leqq 2.8$—which is also used in practice—differences are fortunately small thereby showing that when treating such problems the two variants can be regarded as being approximately the same.

In Eq. (109) a role is played by a therefore the meaning of the abscissa in *Fig. 68* is the type of general most frequent value procedure used (defined by the type of $f_a(x)$ distribution for which the procedure in question is an optimum one). The known curve $K_{max}^{(a)}$ is once more shown in Fig. 68; no $K_{max}^{(a)}$ value is greater than 45.68% on the one hand and, on the other, $K_{max}^{(a)}$ tends to zero if $a \to \infty$ (i.e. if we approximate the least squares technique).
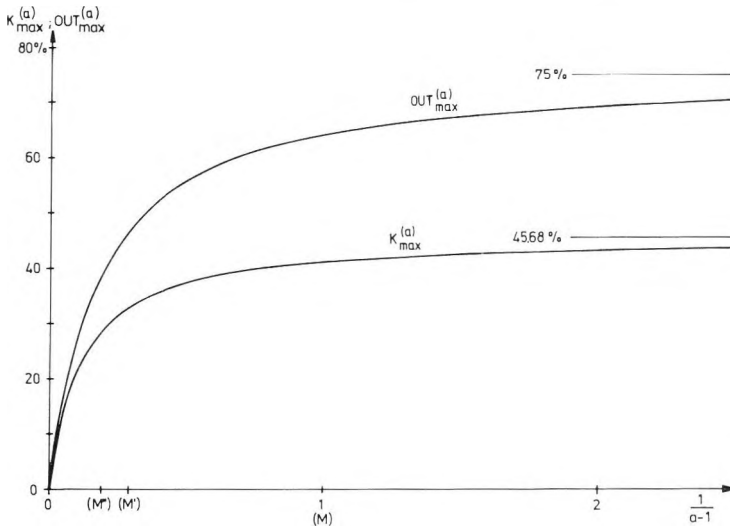

Fig. 68. Maximum percentage of concentrated outliers (K_{max}^a), and of dispersed outliers (OUT_{max}^a). *a* in the abscissa, is the parameter used in the second variant of the generalized most frequent value calculations, having maximum efficiency for $f_a(x)$

68. ábra. Koncentráltan jelentkező kieső adatok maximális aránya (K_{max}^a) és ugyanez, ha a kieső adatok nem tömörödnek (OUT_{max}^a). Az abszcisszában *a* az általános leggyakoribb értékszámítás második variánsában használt paramétert jelenti, amely az $f_a(x)$ eloszlásra szolgáltat maximális hatásfokú eljárást

Рис. 68. Верхний предел процентного отношения ($K_{max}^{(a)}$) и то же при отсутсвии тенденции к концентрированию данны с грубыми ошибками ($OUT_{max}^{(a)}$). Параметр *a* в значении $1/(a-1)$ абсцисс обозначает параметр, используемый во втором варианте расчета обобщенного наиболее частого значения обладающий максимальной эффективностью для распределения $f_a(x)$.

Outliers, however, may occur in a form other than asymmetrically. If the dihesion ε characterizes the single gathering of the data and other values occur arbitrarily (but without any expressed gathering), then

$$OUT^{(a)} = 1 - \frac{(a+1)^2}{(a+1)^2 + 4(a+2)} \tag{140}$$

gives the maximum permitted ratio of outliers in the sense that if the ratio of outliers of the investigated type is smaller than $OUT^{(a)}$, the procedure finds this to be a single gathering [STEINER 1985b]; a obviously refers here to the procedure used. The corresponding curve in Fig. 68 shows that this ratio may be greater than 50% (for $a < 3.8$; that is, for $1/(a-1) > 0.35$). Perhaps this is surprising but in special tasks (e.g. in astronomy) such a property can be advantageous. (The upper limit of $OUT^{(a)}$ is 75%). If, for the geophysical problem investigated it seems unreal to permit a greater ratio than, say, 50%, then M' ($k = 1.9$) can be proposed, at 40% the procedure M'' ($k = 2.4$) may be appropriate, and so on.

Fitting problems (with more than one single unknown) must be judged with an increased cautiousness from this aspect as the fitting procedure itself may produce a gathering of the d_i deviations characterized even by $\varepsilon = 0$ if n is small, resulting in a simple interpolating hypersurface based on only a few points and completely neglecting reliable data. Therefore, if J is the number of unknowns to be determined, then for the number of data relation (141) must hold with minimum $c = 5$ or 10:

$$n \geqq c \frac{J}{1 - OUT^{(a)}} \tag{141}$$

(In a very simple way we can say that we have c items of data for each of the unknowns.)

APPENDIX VI
Supermodel $f_m(x)$ and the possibility of increasing the efficiency

The first step in general most frequent value calculations is to fix the value of k to be used. If we know something about a, the k value is given by Eq. (111), if not (but maximum efficiency is required), primarily the type of distribution should be estimated. A possible solution of the problem was briefly mentioned in paragraph 8.2.1: the sum counterpart of Eq. (113), i.e.

$$A_k^2 = \frac{\sum_{i=1}^{n} \dfrac{(x_i - M_k)^2}{[(k\varepsilon)^2 + (x_i - M_k)^2]^2}}{\left[\sum_{i=1}^{n} \dfrac{(k\varepsilon)^2 - (x_i - M_k)^2}{[(k\varepsilon)^2 + (x_i - M_k)^2]^2} \right]^2}, \tag{142}$$

is to be minimized according to k, and then Eq. (111) defines a. But determinations of this sort are not so easy to carry out in practice and we have no possibility here to deal in detail with such problems. It is obvious that n is required to be sufficiently large, as a *sine qua non* of whatever solution of the problem.

We are now interested, however, just in the k-value to be used, i.e. we need

that k value which minimizes the expression in Eq. (142). The sample size n must also be large in this case. The following details relate to a supermodel differing from $f_a(x)$ as for the latter distribution family this method obviously always results in efficiencies of 100%. The probability distributions of the $f_a(x)$ supermodel, however, are very near the distributions occurring in practice, therefore, from the majority of the distributions of the new supermodel the only constraint is that they should not be too far from the actual cases occurring in geophysics.

For the above mentioned purposes, the following supermodel is defined:

$$f_m(x) = \frac{m^{1-1/m}}{2\Gamma\left(\dfrac{1}{m}\right)} e^{-\frac{x^m}{m}} \tag{143}$$

The behaviour of the flanks causes no problem for whatever sort of estimates since all moments are finite for all distributions of the supermodel. The expression for calculating the j-th moment is:

$$\frac{m^{j/m}\Gamma\left(\dfrac{j+1}{m}\right)}{\Gamma\left(\dfrac{1}{m}\right)}. \tag{144}$$

If m (the type parameter) equals 1, $f_m(x)$ gives the Laplace distribution; for $m = 2$, $f_m(x)$ is clearly the Gaussian distribution, and if $m \to \infty$ we get uniform distribution, i.e. the $f_m(x)$ supermodel contains three distributions from the most well-known ones. For small values of m we get distributions being even more peaked than the Laplace distribution (see in *Fig. 69* the density function for $m = 0.5$—like the Eiffel tower). And for large values of parameter m we get somehow rounded off uniform distributions (see, e.g., in the same figure the $f_m(x)$ curve for $m = 10$). *Figure 70* shows the curve of the dihesions ε, versus $1/m$.

When discussing absolute effectivities, the formula for the Cramér-Rao bound (i.e. for the minimum possible asymptotic variance, see 4.4.2) must also be given:

$$A_{\min}^2 = \frac{\Gamma\left(\dfrac{1}{m}\right)}{m^{2-2/m}\Gamma\left(2-\dfrac{1}{m}\right)}. \tag{145}$$

If we carry out most frequent value calculations in the original form, i.e. with $k = 1$, the asymptotic variance must be calculated according to Eq. (81) as $\varepsilon^2/n(\varepsilon)$. The formula for the efficiency is consequently

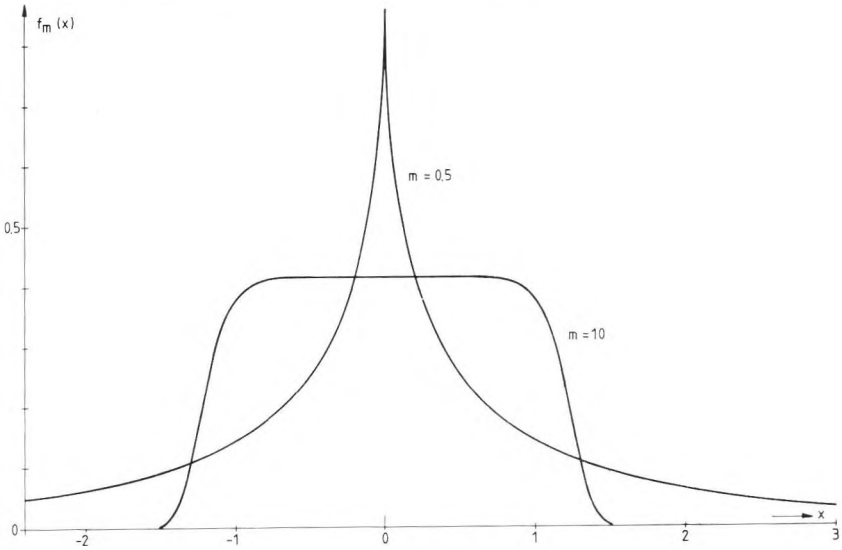$$e_M = \frac{A_{\min}^2 n(\varepsilon)}{\varepsilon^2}. \tag{146}$$

Fig. 69. Two distribution types shown by their density functions of the supermodel $f_m(x)$

69. *ábra.* Két $f_m(x)$ szupermodellbeli eloszlástípus sürüségfüggvénye

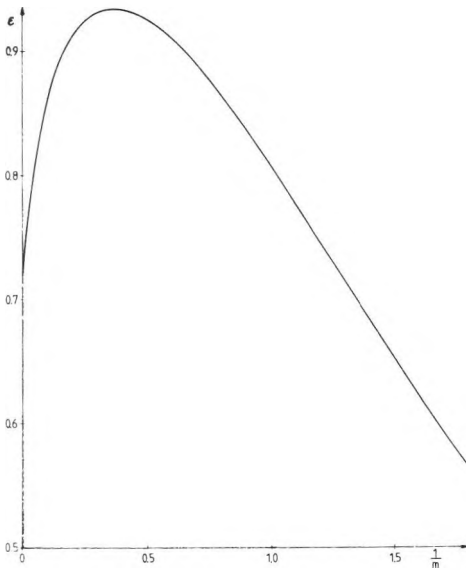*Рис.* 69. Плотностные функции двух типов распределения в супермодели $f_m(x)$.

Fig. 70. Values of dihesion ε for supermodel $f_m(x)$

70. *ábra.* Az ε dihézió értéke az $f_m(x)$ szupermodellre

Рис. 70. Значения дигезии ε для супермодели $f_m(x)$.

In Fig. 72 the thick continuous line (denoted by M) gives the e_M values (versus $1/m$). Although the most frequent value calculation is the optimum algorithm for the Cauchy distribution—which is extremely unusual for the $f_m(x)$ supermodel,—we find the maximum efficiency value of 92% to be satisfactory. Since the arithmetic means calculation is the optimum algorithm for the Gaussian distribution (i.e. for a member of the supermodel $f_m(x)$), it is evident that the least squares efficiency curve reaches the value of 100% at $m = 2$ (thin continuous line in Fig. 72 denoted by E). These efficiencies are to be calculated according to

$$e_{\text{mean}} = \frac{\Gamma^2\left(\dfrac{1}{m}\right)}{m^2 \Gamma\left(2 - \dfrac{1}{m}\right) \Gamma\left(\dfrac{3}{m}\right)} \tag{147}$$

(see Eqs. (144) and (145)).

The efficiency curve of the medians (thin dotted line in Fig. 72 denoted by m) also reaches the maximum possible value as the Laplace distribution is also a member of the supermodel $f_m(x)$ (for $m = 1$), and the calculation of sample medians is the well known optimum algorithm for the Laplace distribution. In that the asymptotic variance for the $f_m(x)$ supermodel in the case of sample median calculations is given by

$$A^2_{\text{median}} = \frac{\Gamma^2\left(\dfrac{1}{m}\right)}{m^{2 - 2/m}}, \tag{148}$$

the formula for efficiencies is clearly

$$e_{\text{median}} = \frac{1}{\Gamma\left(\dfrac{1}{m}\right) \Gamma\left(2 - \dfrac{1}{m}\right)}. \tag{149}$$

Our question is: How can the efficiency be increased if we use—within the concept of the most frequent value calculations—the k value which minimizes the asymptotic variance?

In our theoretical investigation, based on the $f_m(x)$ supermodel, in the first step those k values are to be determined which minimize the expression

$$A^2_{M_k} = \frac{\Gamma\left(\dfrac{1}{m}\right)}{m^{1 - 1/m}} \cdot \frac{\displaystyle\int_0^\infty \frac{x^2}{[(k\varepsilon)^2 + x^2]^2} e^{-\frac{x^m}{x}}\, dx}{\left[\displaystyle\int_0^\infty \frac{(k\varepsilon)^2 - x^2}{[(k\varepsilon)^2 + x^2]^2} e^{-\frac{x^m}{m}}\, dx\right]^2}. \tag{150}$$

The results are shown as k and $1/k$ curves versus $1/m$ in *Fig. 71*. The interval $m > 2$ is characterized by $k = \infty$, i.e. the least squares method is the best generalized most frequent value technique for these distributions, these distributions seldom occur in geophysical practice. Denoting by $e_{M,\text{max}}$ the maximum possible efficiency within the concept of most frequent value calculations, we have to determine for all m values the minimum $A^2_{M_k}$ value based on Eq. (150), $A^2_{\min}$ based on Eq. (145), and $e_{M,\text{max}}$ according to

$$e_{M,\text{max}} = \frac{A^2_{\min}}{A^2_{M_k}}. \tag{151}$$

The $e_{M,\text{max}}$ curve is given in *Fig. 72* (denoted by $\bar{M}$). The relation

$$e_{M,\text{max}} \geqq e_M$$

trivially holds. The fact, however, that sometimes

$$e_{M,\text{max}} > 2e_M$$

can be valid, calls our attention to the fact that $\bar{M}$-calculations are to be regarded as a reserve possibility within the concept of the most frequent value calculations, to further increase the efficiency.
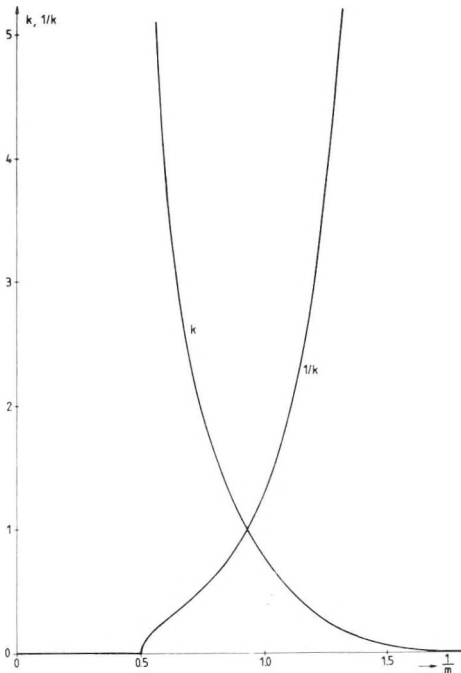
*　　*　　*

Fig. 71. Values of k (and $1/k$, respectively) to obtain minimal asymptotic variance of the generalized most frequent values for the supermodel $f_m(x)$

71. ábra. Azoknak a k (ill. $1/k$) értékeknek a görbéi, amelyekkel minimális az általánosított leggyakoribb érték-számítás aszimptotikus szórása az $f_m(x)$ szupermodell eloszlástípusaira

Рис. 71. Кривые значений k или $1/k$, при которых асимптотическая дисперсия обобщенного наиболее частого значения минимальна для типов распределения супермодели $f_m(x)$.
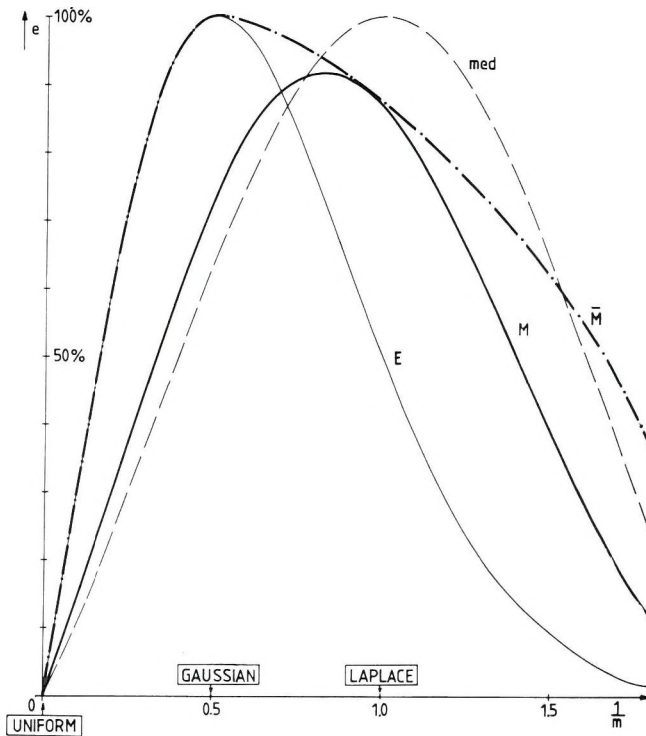
Fig. 72. Efficiency curves for supermodel $f_m(x)$ and for the following estimation procedures: arithmetic mean (E), sample median (*med*), most frequent value (M), generalized most frequent value minimizing the asymptotic variance ($\bar{M}$)

72. ábra. Hatásfok-görbék az $f_m(x)$ szupermodellre és a következő becslési eljárásokra: számtani átlagképzés (E), mintamedián (*med*) számítása, leggyakoribb érték-számítás (M) és az aszimptotikus szórást minimalizáló általános leggyakoribb érték-képzés ($\bar{M}$)

Рис. 72. Кривые эффективности для супермодели $f_m(x)$ и для следующих оценок: среднеарифметическое E, медианы m, наиболее частое значение M и наиболее частое значение $\bar{M}$, полученное путем минимализации асимптотической дисперсии.

Epilogue

It is very much hoped that this short monograph has given a general picture of the principal features of the method of most frequent value procedures and related topics. Although formulations were more concise than detailed, sufficient detail was provided to enable such statistical procedures to be creatively applied in the development of geophysical algorithms. Some theoretical results (concerning, for example, unicity and existence), and a number of complicated applications, were deliberately not mentioned.

Acknowledgement

The author is indebted to Prof. L. Csernyák and to B. Hajagos for the very many discussions; their outstanding role in the development of the theory can clearly be seen from the list of references. In conclusion, I should appreciate it, if readers would be so kind as to send me their comments, suggestions, etc, in order that they can be utilized in our further work.

REFERENCES

ACZÉL E., STOMFAI R. 1968: Measurement of the Hungarian geomagnetic basic network in 1964–65 (in Hungarian). Geofizikai Közlemények, **17**, 3, pp. 5–36
ANDREWS D. F., BICKEL P. J., HAMPEL F. R., HUBER P. J., ROGERS W., TUKEY J. W. 1972: Robust Estimates of Location. Princeton University Press, Princeton N. J.
ANSCOMBE F. J. 1960: Rejection of Outliers. Technometrics, **2**, 1, pp. 123–147
BAULE B. 1963: Ausgleichs- und Näherungsrechnung. S. Hirzel Verlag, Leipzig, 101 p.
BOX G. E. P., HUNTER G. W., HUNTER J. S. 1978: Statistics for Experimenters. Wiley, New York. 653 p.
CLAERBOUT J. F., MUIR F. 1973: Robust modelling with erratic data. Geophysics, **38**, 5, pp. 826–844
CRAMÉR H. 1958: Mathematical Methods of Statistics. Princeton University Press, Princeton, 575 p.
CSERNYÁK L. 1973: On the most frequent value and cohesion of probability distributions. Acta Geodaet., Geophys. et Montanist. Acad. Sci. Hung. **8**, 3–4, pp. 397–401
CSERNYÁK L. 1982: Bemerkung zum Artikel 'Der häufigste Wert, als eine Abschätzung von minimalem Informationsverlust etc.' von B. Hajagos. Publications of the Technical University for Heavy Industry, Series A, Mining, **37**, 1–2, pp. 115–118
CSERNYÁK L. 1984: Flanken-Diagnostik. (Manuscript)
CSERNYÁK L., HAJAGOS B., LANDY I., LANTOS M., STEINER F. 1982: Die Abhängigkeit der geschätzten Werte von der Gliederzahl n der Stichprobe. Publications of the Technical University for Heavy Industry, Series A, Mining, **37**, 1–2, pp. 65–67
CSERNYÁK L., HAJAGOS B., STEINER F. 1981: General validity of the law of large numbers in case of adjustments according to the most frequent value. Acta Geodaet., Geophys. et Montanist. Acad. Sci. Hung. **16**, 1, pp. 73–90
CSERNYÁK L., STEINER F. 1980: Practical computation of the most frequent value of data systems. Acta Geodaet., Geophys. et Montanist. Acad. Sci. Hung. **15**, 1, pp. 59–73
CSERNYÁK L., STEINER F. 1982: Untersuchungen über das Erfüllungstempo des Gesetzes der großen Zahlen. Publications of the Technical University for Heavy Industry, Series A, Mining, **37**, 1–2, pp. 47–64
CSERNYÁK L., STEINER F. 1983a: Relation between two estimates of location: the most frequent value and the arithmetic mean. Acta Geodaet., Geophys. et. Montanist. Hung. **18**, 1–2, pp. 141–156
CSERNYÁK L., STEINER F. 1983b: Limit distribution of the most frequent values of samples from symmetrical distributions. Acta Geodaet., Geophys. et Montanist. Hung. **18**, 1–2, pp. 135–140
CSERNYÁK L., STEINER F. 1985a: Bemerkungen zur CML-Abschätzung. Publications of the Technical University for Heavy Industry, Series A, Mining, **38**, 3–4, pp. 203–209
CSERNYÁK L., STEINER F. 1985b: Bemerkungen zu den häufigsten Werten von asymmetrischen U-Verteilungen. Publications of the Technical University for Heavy Industry, Series A, Mining, **39**, 1–4, pp. 159–172
CSERNYÁK L., STEINER F. 1985c: Die Suche nach einer geeigneten Abschätzungsmethode für die Geophysik. Publications of the Technical University for Heavy Industry, Series A, Mining, **40**, 1–4, pp. 225–238

CSERNYÁK L., HAJAGOS B., STEINER F. 1988: Uncertainties of the error determination (in Hungarian). Magyar Geofizika, **29**, 3, pp. 113–127

CSÓKÁS J., GYULAI Á., MOLNÁR D. 1979: Geoelectrical bed-soundings for constructing prognose maps of tectonic distrurbances (in Hungarian). Magyar Geofizika, **20**, 6, pp. 219–227

DETREKÖI Á. 1986: Considerations of robust errors in the processing of survey data (in Hungarian) Geodézia és Kartográfia, **38**, 3, pp. 155–160

FERENCZY L. 1980: M-fitting in the processing of magnetotelluric data. Acte Geodaet., Geophys. et. Montanist. Acad. Sci. Hung. **15**, 1, pp. 45–58

FERENCZY L., HAJAGOS B., STEINER F. 1988: Shortcomings of the conventional definition of error. Proposal for using a new definition of error (in Hungarian) Magyar Geofizika, **29**, 3, pp. 104–112

FLETCHER R. (Ed.) 1969: Optimization. Academic Press, London and New York. 354 p.

GNEDENKO B. V., KOLMOGOROV A. 1949: Limiting distribution for the sum of random variables. (in Russian) Gosztehizdat, Moszkva

HAJAGOS B. 1980: Method for the rapid calculation of the most frequent value and M-fitting problems. Acta Geodaet., Geophys. et Montanist. Acad. Sci. Hung. **15**, 1, pp. 75–85

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. 1984: Normalitätsuntersuchungen mit Hilfe der χ^2-Probe an Stichproben, die aus Studentschen Mutterverteilungen stammen. (Manuscript)

HAJAGOS B. 1985a: Die verallgemeinerten Student-schen 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. 1985b: Ausreißer-Insensitivität für asymmetrisch liegende Datengruppen. (Manuscript)

HAJAGOS B., STEINER F. (in press): Methods to increase the resistance of the computation of general most frequent values. Acta Geodaet., Geophys. et Montanist. Hung.

HAMPEL F. R. 1968: Contribution to the theory of robust estimation. Ph. D. Thesis, Univ. of Calif., Berkeley

HAZAY I. 1968: Adjusting calculations in surveying. (in Hungarian) (II. ed.) Tankönyvkiadó, Budapest, 598 p.

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

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

JEFFREYS H. 1932: An alternative to the rejection of observations. Proceedings of the Royal Soc. of London. Ser A., 137, pp. 78–87

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

KERÉKFY P. 1978: On robust estimates (in Hungarian) Alkalmazott Matematikai Lapok 4 (1978) pp. 327–357

KORN G. A., KORN TH. M. 1961: Mathematical Handbook for Scientists and Engineers. McGraw-Hill, New York–Toronto–London, 1130 p.

KRARUP T., KUBIK K., JUHL I. 1980: Götterdämmerung over least squares adjustment. Proceedings of the 14. congress of ISP. Hamburg

KRARUP T., KUBIK K. 1983: The Danish method, experience and phylosophy. – Seminar Mathematical Models of Geodetic Photogrammetric Point Determination. DGK Reihe A Nr 98, München, pp. 131–134

KULLBACK S. 1959: Information Theory and Statistics. John Wiley and Sons, New York. 393 p.

LANDY I., LANTOS M. 1982: Ausgleich tellurischer Geraden. Publications of the Technical University for Heavy Industry Series A, Mining, **37**, 1–2, pp. 7–16

LINNIK J. W. 1961: Die Methode der kleinsten Quadrate in moderner Darstellung. VEB Deutscher Verlag der Wissenschaften, Berlin, 314 p.

MATHAI A. M., RATHIE P. M. 1975: Basic Concepts in Information Theory and Statistics. Wiley Eastern Limited, New Delhi, 137 p.

NAAS J., SCHMID H. L. 1965: Mathematisches Wörterbuch. Akademie-Verlag, Berlin, Teubner, Leipzig, 1043 p.

NEMESI L. 1963: On interpretation of telluric measurements (in Hungarian). Magyar Geofizika, **4**, 3–4, pp. 167–176

NETTLETON L. L. 1939: Determination of density for reduction of gravimeter observations. Geophysics, **4**, 3, pp. 176–183

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

PEREZ A. 1967: Information theoretical risk estimates in statistical decisions. Kybernetika, **3**, pp. 1–21

PIRKLE L. F., et al. 1982: Application of the Huber M-estimator to aerial radiometric data. Report. Los Alamos National Laboratory and Bendix Field Engineering Corporation

POINCARÉ H. 1912: Calcul des Probabilités. Gauthier-Villars, Paris, 274 p.

PRÉKOPA A. 1962: Probability theory with technical applications (in Hungarian). Műszaki Könyvkiadó, Budapest, 440 p.

RÉNYI A. 1961: On measures of entropy and information. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley-Los Angeles, pp. 547–561

RÉNYI A. 1962: Wahrscheinlichkeitsrechnung. Deutscher Verlag der Wissenschaften, Berlin, 547 p.

RÉNYI A. 1967: Statistics and information theory. Studia Sci. Math. Acad. Sci. Hung. **2**, 1–2, pp. 249–256

RIZA F. M. 1961: An Introduction to Information Theory. McGraw-Hill, New York, 496 p.

SHORT J. 1763: Second paper concerning the parallax of the sun etc. Philos. Trans. Roy. Soc., London, **53**, pp. 300–343

STEINER F. 1959: Zur Ermittlung des Koeffizienten der gravimetrischen Höhenreduktion. Gerlands Beiträge zur Geophysik, **68**, 1, pp. 15–20

STEINER F. 1973: Most frequent value and cohesion of probability distributions. Acta Geodaet., Geophys. et Montanist. Acad. Sci. Hung. **8**, 3–4, pp. 381–395

STEINER F. 1980: M-fitting (fitting according to the most frequent value) and its comparison with the method of least squares. Acta Geodaet., Geophys. et Montanist. Acad. Sci. Hung. **15**, 1, pp. 37–44

STEINER F. 1982: Gewogener M-Ausgleich. Publications of the Technical University for Heavy Industry, Series A, Mining, **37**, 1–2, pp. 17–28

STEINER F. 1984a: Einiges über verschiedene Ausgleichsprinzipien. (Manuscript)

STEINER F. 1984b: Genauigkeitsabweichungen verschiedener häufigster Werte von dem jeweiligen Optimum. (Manuscript)

STEINER F. 1985a: Programmierung von Algorithmen, die mit der Berechnung von häufigsten Werten zusammenhängen. (Manuscript)

STEINER F. 1985b: Konzentrationsempfindlichkeit der verallgemeinerten häufigsten Werte. (Manuscript)

STEINER F. 1987: Possibilities to realize higher effiency in geophysical interpretation. Geophysical Transactions, **33**, 1, pp. 3–9

STEINER F., ZILAHI-SEBESS L. 1988: Interpretation of Filtered Gravity Maps. Akadémiai Kiadó, Budapest

TÁRCZY-HORNOCH A. 1956: Compensating computation (in Hungarian) (III. ed.). Felsőoktatási Jegyzetellátó Vállalat, Budapest, 326 p.

TUKEY J. W. 1960: A survey of sampling from contaminated distribution. Contrib. to Prob. and Statist. (Ed.: Olkin), Stanford University Press, Stanford, Calif. pp. 448–485

TUKEY J. W. 1977: Exploratory Data Analysis. Addison-Wesley, Reading, Mass, 499 p.

VINCZE I. 1953: On an interpretation of a concept of information theory (in Hungarian). Matematikai Lapok, **10**, 3–4, pp. 255–266

VINCZE I. 1968: Statistics with technical applications (in Hungarian). Műszaki Könyvkiadó, Budapest, 352 p.

ZILAHI-SEBESS L. 1987: Die Bestimmung des gravimetrischen regionalen Feldes mit Hilfe des Ausgleiches den häufigsten Werten entsprechend. Publications of the Technical University for Heavy Industry, Series A, Mining **43**, 1–4, pp. 113–131

LEGGYAKORIBB ÉRTÉKEK

STEINER Ferenc

A jelen monográfia a leggyakoribb érték koncepcióján alapuló modern statisztikai algoritmusokkal foglalkozik (eredeti és általánosított alakjában egyaránt értve a leggyakoribb értéket). Modern statisztikai eljárások azért szükségesek, mivel a legkisebb négyzetek elvén alapuló, általánosan használt eljárások nem elég nagy hatásfokúak a földtudományokban általában előforduló hibaeloszlások esetén; a Gauss-eloszlás előfordulása inkább számít kivételnek, semmint szabálynak. Jelen monográfia az elméleti alapokat is tartalmazza, hogy minél több speciális alkalmazást tegyen lehetővé, azaz hogy elősegítse a leggyakoribb értékek koncepcióján alapuló új geofizikai értelmezési algoritmusok kifejlesztését.

НАИБОЛЕЕ ЧАСТЫЕ ЗНАЧЕНИЯ

Ференц ШТЕЙНЕР

В настоящей монографии рассматриваются современные статистические алгоритмы, основанные на представлении о наиболее частом значении (понимаемом как в первичной, так и в обобщенной форме). Необходимость в современных статистических способах возникает в связи с тем, что общепринятые способы, основанные на принципе наименьших квадратов, не достаточно эффективны в случае распределений ошибок, обычных в геологических науках, а распределение Гаусса встречается скорее в качестве исключения, нежели правила. В настоящей монографии приводятся также и теоретические основы, чтобы обеспечить возможность как можно более обширного применения в специальных целях, то-есть, чтобы способствовать разработке новых алгоритмов геофизической интерпретации, основывающихся на представлении о наиболее частых значениях.