Understanding the concept of outlier and its relevance to the assessment of data quality: Probabilistic background theory

Davaadorjin Monhor and Shuzo Takemoto

Department of Geophysics, Graduate School of Science, Kyoto University, Kitashirakawa Oiwake-cho,
Sakyo-ku, Kyoto City, Kyoto 606-8502, Japan

In recent years an increasing interest in the studies on outlier can be observed, however, for the time being there exists no general definition of outlier. In the present paper we introduced a generic descriptive definition of outlier. We observed that the outlier problems had so far been treated in statistical way without paying proper attention to probabilistic-theoretic backgrounds. In view of this gap, we made an attempt to establish a probabilistic background theory. Within this framework, the large deviations are considered as probabilistic-theoretic model of outlier, and the interrelationship of the laws of large numbers, the central limit theorems and the large deviations are clarified. These considerations are specialized for the case of statistical sample, which is important from the point of view of the assessment of data quality. Some methodological and historical aspects of geodesy, geophysics and astronomy are mentioned, too. We revealed that the data analysis carried out by Kepler in the process of discovery of his famous elliptic law of planetary motion has relevance to the outlier problem. This methodologically interesting fact is a new result in the history of geosciences. We established that the accuracy of Chebyshev inequality increases as the deviation of the random variable involved from its expectation, increases. The possibility of application of Chebyshev inequality to the outlier problem is pointed out.
Key words: Assessment of data quality, Berry-Esseen theorem, Chebyshev inequality, large deviations, outliers.

1. Motivation and Introduction

In recent years an increasing interest in the studies on outlier can be observed, however, for the time being there exists no general definition of outlier. To emphasize that the above statement is not our subjective opinion, but—indeed—an objective reality, we proceed to the following passages. Barnett (1984, p. 360), the author of the book "Outliers in Statistical Data" writes: "The outlier problem seems to be arousing more interest today than it has ever done, in spite its long history. The pages of the international statistical journals ever more contributions to this fascinating and useful field of study, as it evidenced by the substantial list of references at the end of the book". In a research paper by Gather (2000) we read: "The problem of outliers in random data sets is a very interesting, important and common one. Nevertheless there is no formal and generally accepted definition of what is meant by an *outlier*. Terms like outlier, spurious observation, contaminant, gross error and others are used with different and overlapping meanings". These statements clearly indicate that the consideration of the nature and background of outlier is a problem of significance both in theoretical and practical terms.

The aim of the present paper is as follows: (i) to give a general definition of outlier which at the same time gives rise to a categorization of outliers; (ii) to introduce a methodologically new way of unifying presentation of the

laws of large numbers and central limit theorem in conjunction with outliers (iii) to emphasize the role of Chebyshev inequality for detection of outliers. These considerations in an aggregated way constitute a probabilistic background theory for studies on outliers.

The paper is organized as follows. Section 2 presents the three instances from the very origin of outliers which help a correct understanding of outliers. In the history of science the way of how Kepler discovered his famous elliptic law is known (Aiton, 1969; Wilson, 1968). However, Kepler's data analysis has so far not considered from the standpoint of outlier problem. We revealed the outlier detection aspect of Kepler's way of handling the measurement data. The other instances also contain useful elements. Section 3 addresses the the question of definition of outlier and introduces a new generic descriptive definition of outlier. In Section 4 we give the basic mathematical techniques of the asymptotic growth rate of functions which are necessary for the presentations in the subsequent sections. In Section 5 we consider the laws of large numbers (LLN) in a specialized for statistical sample form. The content of this section serves several purposes: The LLN is considered as a first step towards the central limit theorem (CLT), at the same time the LLN is presented as being the probabilistic-theoretic background of the sample mean. Bearing in mind the needs of data analysis, we here introduced a novel way of consideration of the LLN. Section 6 considers the transition from the LLN to CLT specialized for the case of statistical sample. This attitude results in the consideration of the CLT as being of further refinement of the LLN. Here

we presented Berry-Esseen theorem and related research results obtained by Zolotarev (1967) and van Beer (1972).

In Section 7 we introduced the theory of large deviations in simplified setting. The theory of large deviations is a new branch of probability theory. Here the large deviations are considered as probabilistic-theoretic model of outlier. This is a new theoretical approach. In Section 8 a numerical study of Chebyshev inequality in comparison with the normal distribution is given. Based on this, the role of Chebyshev inequality for the outlier problems is clarified. Section 9 summarizes the main conclusions of the paper.

2. Three Instances Related to the Origin of Outlier: New Findings and Lessons

The concept of outlier stemmed from the mathematically processing geodetic and astronomical measurements data. To determine the mean ellipticity of the Earth from measurements data, Maire and Boscovich (1755), and Boscovich (1757) removed two data on the basis that they were too much deviated from the remaining data. In modern terminology, this means that Maire and Boscovich used outlier rejection technique. Legendre (1810) also used and proposed outlier rejection technique. It is worthwhile to note that the name of A. M. Legendre (1752–1833) is well-known as being one of the discoverers of the famous least squares method. However, it is not widely known that he conducted significant research in the area of the determination of the figure of the Earth and related geophysical and geodetic problems. The volume II of the book by Todhunter (1873) contains three chapters that consider Legendre's memoirs on the topic.

Among early papers on outlier, the paper titled "The most probable choice between several discrepant observations and the formation therefrom of the most likely induction" by Daniel Bernoulli (1961) seems to be one of the most interesting one. He wrote: "I see no way of drawing a dividing line between those that are utterly to be rejected and those that are to be wholly retained; it may even happen that the rejected observation is the one that would have supplied the best correction to the others. Nevertheless, I do not condemn in every case the principle of rejecting one or other observations, indeed approve it, whenever in the course of observation an accident occurs which in itself raises an immediate scruple in the mind of observer, before he has considered the event and compared it with other observations". In our opinion the most important element in Daniel Bernoulli's remarks is that he emphasizes—in translation of modern terms—the uncertainty deep-rooted in the very concept of the outlier and sees the main difficulty in this uncertainty. If we think merely in terms of probability theory, then it may be said that today the situation concerning the concept of the outlier is not much changed. However, there had existed no highly-developed theory of probability, at that time. Nowadays, we have the modern theory of probability which we should support for understanding the nature of the outlier. In the present paper we made some initial attempts towards this attitude.

In the relevance to outlier problem we would like to mention here Kepler's way of handling the data in his famous elliptic law of motions of planets. From the paper by Aiton (1969) and Wilson (1968) we can see that Kepler first attempted with the circular law for long time and analyzing the fitting of the circular law to the data of motion of Mars, and he found the deviation of the magnitude of 8′. He considered this magnitude as being too large, and finally he decided to change his circular law to the elliptic one. We see that in modern terminology Kepler detected an outlier, and based on this detection, he modified his model. In this historical fact, we get the following important new findings. It is Kepler who first considered such an outlier that is not "gross error". Secondly, it is Kepler who first carried out the model-fitting analysis based on residuals.

3. How to Define the Outlier?

We all know what means the term statistical sample or in short the sample in mathematical statistics. This is mathematically very simple and clearly understandable definition. However, there are at least two reasons to recall here the definition. First, all the theory, methods, precise definitions and statements of modern statistics are based on the definition of the sample directly or indirectly (Wilks, 1962). So, behind of every precise definition should directly or indirectly be the correct understanding of the sample. Secondly, nevertheless, the mathematical definition is so simple, in real life cases, it is not so simple, even sometimes difficult to consider whether the data in question is sample or not. It is not an occasional case that the difficulties associated with the sample from the data of geology are in detailed way considered by Wilks (1963) who is one of the outstanding figures in modern mathematical statistics.

In probabilistic setting, the statistical sample simply is a set of n independent random variables with a common probability distribution. Here n is the sample size. In real life data, we usually have n pieces of numbers which are measurements results and we should consider each of these numbers are the representative of the random variables that are in the probabilistic definition of the sample.

General Basic Definition: *An outlier is or are such a element or elements which displays or display a considerable discrepancy from distributional behavior of remainders that constitute a statistical sample or the presence of this or these elements disturbs or disturb the being the statistical sample of the remainders.*

This basic general definition gives only generic descriptive character of outliers, however, it lacks in being "concrete", because it is difficult to distinguish between the outlier and the remainders. The lack can be corrected by the following categorizing or refining subsidiary definitions.

Refining Definitions.

(a) *The undesirable outliers: This is a kind of outliers that occur in making measurements, computer processing the measurement data (input, output, transmission of data and mathematically and computer processing data) that disturb the correct mathematical processing the remainder data.*

(b) *Outliers arising from model fitting analysis and/or unexpected (exploratory) outlier.*

(c) *Outliers that belong neither (a) nor (b), but by their appearance manifest the intrinsic nature of fluctuation of the randomness.*

We consider some explanatory comments on the above definitions. The outliers that belong to the (a) category is known in geodetic and geophysical literature under the name of gross-errors and/or blunders. The very origin of the concept of outlier is related to these outliers, consequently, this type of outliers constitute the "ancestors" of outliers in general. It is evident that this kind of outliers should be removed from the statistical sample, provided the outlier is detected, although the clear-cut identification of this kind of outliers is not an easy task. This is the reason why the rejection of outliers has historically been the main direction in the treatment of outliers.

Nowadays, we all know the achievements modern measurement technology result in an extremely high precision in the measurement process, i.e., the dangers of the "gross errors" seem to be passing "their time". However, there is "the reverse of the medal", too. The equipment failure, errors hidden in software, keypunch errors—for example—wrong decimal points and similar errors may be termed as "modern" gross errors. In this connection, from Kubik, Weng and Frederiksen (1985) we here cite the following passage: "...there is an increasing danger: more and more data are automatically processed on a computer without being scrutinized by a competent person...". A similar warnings can be found in Detrekoi (1986). These observations clearly indicate that we are still unable to get rid of the gross errors.

The important distinctive character of the category (b) is that this kind of outliers are not directly related to measurement data, but they have relevance to the exploratory analysis of data and/or examinations of appropriateness of models or model-fitting analysis. In recent years, in the model-fitting analysis of residuals of regression models the term outlier are becoming apparent. As far as we know, it is Srikantan (1961) who first observed that the model-fitting analysis in regression can be considered as an outlier detection problem. Following Srikantan, Stefansky (1971, 1972), O'Gorman and Myers (1987) and many others considered the model-fitting analysis in regression in context of outliers.

The outliers of the category (c) can be thought of as a marked manifestation of random fluctuation of intrinsic variability of such random phenomena that belong neither to the category (a) nor (b). This category can be easily understood by the following examples. The life span of an elementary particle can be considered a certain concrete number. However, this attitude is satisfactory only for some simple considerations. The situation changes, when a closer examination is given to the question. In more precise setting, the life span is a random variable. The high precision technology measurements used in modern physics from time to time detect large fluctuations. Another example from physics. Let us consider a certain kind of gas contained in a closed vessel. We all know that the walls of the vessel are affected by the pressure of the gas which is usually considered to be a certain number. On the other hand, the

pressure of the gas is the mean effect determined by the number and velocity of molecules hitting the walls of the vessel per time unit. If we denote the effect of a single molecule identified by index i by ξ_i and suppose that the vessel contains n molecules, then the pressure of the gas is $\frac{\xi_1 + \xi_2 + \cdots + \xi_n}{n}$, where $\xi_1, \xi_2, \ldots, \xi_n$ are independent and identically distributed random variables. The gas pressure is usually "almost" constant. In fact, the precise measurements of modern physics indicate small fluctuations which are usually neglected. But sometimes, there may occur cases where the fluctuations reach relatively big enough amount. In both examples, it is usual practice that physicists speak about fluctuations or large fluctuations, and they are unfamiliar with the mathematical term "outlier", indeed, for them the term large fluctuations are the same as outliers. Here we see that these outliers are neither "gross errors" nor "errors arising from mathematical modelling".

The following example is an everyday life analogue of the first example. It is obvious that human life span is also an random variable like that of elementary particle. In a concrete community it is a usual practice to speak about the life expectancy or average life. However, there may occur a rare case of very high longevity. This is again such outlier that belongs neither to the category (a) nor (b), because—obviously—the such longevity is—again—neither "gross error" nor "error arising from mathematical modelling". A little more detailed consideration of this simple example provides us with some useful insights into the significance of probabilistic background for correctly understanding the essence of outliers. From media we are sometimes informed that somewhere somebody celebrated—say—105th birthday and so on. Here we are soon be faced the question that which numbers can actually present the life span of a person, if we prefer a concrete number for the life span. Is there exist such a number that below which the life is achievable, but beyond which the life is impossible? Both intuition and logic support the non-existence of such precise limit, because otherwise we would have a clear-cut maximal age for a person. However, based on common sense, we all will hesitate to admit that somebody will reach an age of—say—300 years. Consequently, there is a contradiction or paradox. This kind of difficulty is easily solved by giving a probabilistic background to the problem in question. Indeed, the theory and practice of modern actuarial science admit no bounds to human life span. Instead carefully compiled mortality tables which are nothing other than numerical realization of a probability distribution are used.

These three examples have a common feature that the quantities in question are *ab initio* kinds of "natural random variables", and the "fluctuations of considerable size", i.e., outliers are apparent to be an intrinsic feature of the randomness. Through these illustrative examples we clearly see that these outliers are neither "the gross error" nor "the error arising from model fitting".

4. The Asymptotic Growth Rate of Functions and Related Asymptotic Notations

The comparing of the rates at which functions of real or integer variable grow as their argument becomes large, plays a considerable role in various areas of theory and ap-

plications like as advanced calculus, asymptotic analysis, limit theorems of probability and theoretical computer science. The mathematical technique for such comparison first appeared in number theory which is an important branch of pure mathematics. Paul Bachmann German number theorist treated different problems of number theory in his five volume work appeared in the period of 1892–1905. In the second volume titled *Analytische Zahlentheorie* and appeared in 1894 he writes: "...

$$\tau(n) = n \log n + O(n),$$

wenn wir durch das Zeichen $O(n)$ ein Grösse ausdrücken, deren Ordnung in Bezug auf die n Ordnung von n nicht überschreitet. [...where the notation $O(n)$ stands for expressing such a quantity the order of which with respect to n does not exceed the order of n]". This fact clearly indicates that it is Bachmann (1894) who first introduced a well-chosen mathematical notation for handling asymptotic growth rate of functions of integer number, namely, the O-notation (the big-oh notation). This initial step was much developed by Edmund Landau (1877–1938) German mathematician who mainly contributed to number theory and calculus. A seven-page special section titled *Analytische Zahlentheorie: Einleitung* of the second volume of the book: *Vorlesungen über Zahlentheorie* von Edmund Landau, Leipzig, 1927, considers the precise mathematical definition of the O-notation and introduces o-notation (the little-oh notation) for functions of real variable, i.e., Landau further develops the O-notation of Bachmann and introduces the new concept of order of growth rate denoted by o-notation. In addition, he presents the mathematical techniques for the comparison of asymptotic growth rate in an easy-to-understand and consistent way based rigorous definitions and notations. He also reformulated and refined many important asymptotic results of calculus and number theory which were known at that time in terms of asymptotic notations developed by him. Landau established two basic rules for algebraic manipulations that enable considerable simplifications in asymptotic calculations:

$$o(g(x)) = o(g(x)) + o(g(x))$$
$$O(cg(x)) = O(g(x))$$

Using these algebraic rules, he refined the asymptotic formula for distribution of prime numbers. Under the influence of Landau's excellent presentation of the comparison of asymptotic growth rate functions, these techniques gradually became a commonly used standard method in advanced calculus, asymptotic analysis and recently in the studies of the efficiency of algorithms in computer science. Knuth (1976) further developed by introducing Ω-and Θ-notations. These developments are arisen from the need of theoretical computer science.

In what follows, we consider the essential elements of the mathematical techniques of asymptotic comparisons and asymptotic notations simplified, but rigorous mathematical setting.

Definition 4.1. A function $f(\cdot)$ is said to be of smaller order than $g(\cdot)$ as $x \longrightarrow \infty$, if

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = 0. \qquad (4.1)$$

In this case, we write $f(x) = o(g(x))$ as $x \longrightarrow \infty$, and we read as "f of x equals little-oh of g of x, or in short, "f is little-oh of g".

The meaning of the definition is to convey the idea that the function $f(\cdot)$ is more quickly than $g(\cdot)$ tending to zero, i.e., quickly decreases, as x grows. Here the attentive reader can observe that we are concerned with the rate of decrease rather than growth, as being "contradicting" to the section title. This is because that in the asymptotic analysis it is usual practice to speak about "algebraic growth", when the quantity actually decreases. On the other hand, the most of the applications of the asymptotic analysis are in the growth case. Essentially, we are considering the rate of change of functions.

Remark 4.1. In the special case when $g(\cdot) = constant = 1$, the formal application of (4.1) gives us that $\lim_{x \to \infty} \frac{f(x)}{g(x)} = \lim_{x \to \infty} \frac{f(x)}{1} = \lim_{x \to \infty} f(x) = 0$. This leads to the commonly used practice that $f(x) = o(1)$, as $x \longrightarrow \infty$. Although this means that simply function $f(x)$ itself tends to 0, as $x \longrightarrow \infty$, it turns out that the use of this simple symbol is very convenient in many circumstances.

Definition 4.2. Functions $f(\cdot)$ and $g(\cdot)$ are said to be growing at the same rate as $x \longrightarrow \infty$, if

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = C, \qquad (4.2)$$

where C is finite positive real number. We denote this by writing $f(x) = O(g(x))$, and read as "f of x is big-oh of g of x", or in short, simply "f is big-oh of g".

In recent years, it is becoming a usual practice that in the above definition instead of (4.2.) the condition

$$\left| \frac{f(x)}{g(x)} \right| \leq C \qquad (4.3)$$

is often used. The (4.2.) implies the (4.3.), but the implication is not reversible, and this means that the definition given by the condition (4.3.) is more general than Definition (4.2.)

Remark 4.3. The notation $O(1)$ means that the function $f(\cdot)$ is bounded, as $x \longrightarrow \infty$.

One special case of Definition (4.2.) is worth looking at: We say that functions $f(\cdot)$ and $g(\cdot)$ are asymptotically equivalent as $x \longrightarrow \infty$, if in Definition 4.2 $C = 1$, and we indicate this by the notation $f(x) \sim g(x)$, as $x \longrightarrow \infty$. It is evident that $f(x) \sim g(x)$, as $x \longrightarrow \infty$, is the same as $f(x) = g(x)(1 + o(1))$, as $x \longrightarrow \infty$.

The basic properties of the asymptotic notations are expressed as the algebraic manipulation rules that are very often useful in calculating complicated functions involved in asymptotic analysis. The most important ones among these properties are as follows:

1) $o(g(x)) \pm o(g(x)) = o(g(x))$;

2) $O(g(x)) \pm O(g(x)) = O(g(x))$;

3) $o(cg(x)) = o(g(x))$, where $c \neq 0$;

4) $O(cg(x)) = O(g(x))$, where $c > 0$;

5) $o(g(x)) \cdot (q(x)) = o(g(x) \cdot (q(x))$;

6) $O(g(x)) \cdot (q(x)) = O(g(x) \cdot (q(x)))$;

7) $o(cg(x)) = O(g(x))$;

8) $O(O(g(x))) = O(g(x))$;

9) $o(O(g(x))) = o(g(x))$;

10) $O(o(g(x))) = o(g(x))$.

Remark 4.4.   In the applications of the asymptotic notations to advanced calculus, not only the case of the argument x approaching to ∞, but the case when $x \longrightarrow x_0$, where x_0 any real number, is also considered. In the present paper, we do not need the latter case. The readers interested in this aspect are referred to any comprehensive textbooks and monographs on modern advanced calculus, for example, to Finney and Thomas (1990).

5.   The Laws of Large Numbers as Probabilistic-Theoretic Background for Sample Mean

We often encounter the arithmetic mean in various areas ranging from daily life to scientific analyzing data. The use of the arithmetic mean in mathematically processing the results of measurements goes back to the 3th century B.C. In observational astronomy, later in geodesy and geophysics the arithmetic mean had for long period been used as a basic mathematical tool for the improvement of precision of measurement results. As Plackett (1958) remarks, the Babylonian astronomers developed a systematic mathematical theory based on simple arithmetic schemes for the calculation of motions of the sun, moon and planets.

At the end of 16th century, Tycho Brahe, Danish astronomer observed that by using the arithmetic mean of observational data the accuracy of measurement can be improved, and he systematically used to eliminate measurement errors (Dreyer, 1890). After Tycho Brahe's experience, it gradually became commonplace to use the arithmetic mean as an estimate of the true value to be determined.

From these facts we can conclude that (i) the arithmetic mean already gained important applications in astronomy, geodesy and geophysics long before the emergence of modern probability theory; (ii) it is empirically established that the arithmetic mean possesses some kind of stability around a fixed number which can be considered as the true value of the object under measurement.

These empirical findings, however, lacked in a sound mathematical foundation given in terms of probability. The foundation is the law of large numbers (LLN). As Pearson (1925) remarks, the LLN was discovered by James Bernoulli in 1713. After Bernoulli, Poisson (1837) made an important contribution to the developments of the LLN. He introduced the term of the law of large numbers for indicating the universal character of the stabilization of relative frequencies and arithmetic mean when the number of observations related to the same phenomenon and the same circumstances becomes very large, independently of the nature of individual phenomenon. We would like to emphasize that among these diverse phenomena Poisson considered such geophysical observation like the mean sea level. Unfortunately, the modern theory of the LLN is extremely abstract, and the historical roots of its practical applications have been disappeared.

In a simplified setting, the main topic of the LLN is the study of convergence of the average

$$\frac{\xi_1 + \xi_2 + \cdots + \xi_n}{n} \tag{5.1}$$

of random variables $\xi_1, \xi_2, \ldots, \xi_n$.

Depending on the type of convergence, in modern probability theory there are two kinds of law of large numbers: weak law of large numbers (WLLN) and strong law of large numbers (SLLN). The statement of the WLLN is that the arithmetic mean of independently and identically distributed random variables, for large n, approximately is equal to the expectation with probability large enough. The SLLN asserts the same, but with probability one. Therefore, the statement of the strong law of large numbers is more stronger than the weak law.

Theorem 5.1. (Weak law of large numbers (Khinchine (1929))).   *The sequence of random variables*

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots \tag{5.2}$$

is assumed to be independently and identically distributed. If their expectations exist, i.e., $\mathbf{E}(\xi_k) = a < \infty$, $k = 1, 2, \ldots$, then for any small $\varepsilon > 0$, we have

$$\mathbf{P}\left(\left| \frac{\xi_1 + \xi_2 + \cdots + \xi_n}{n} - a \right| \le \varepsilon \right) \longrightarrow 1, \tag{5.3}$$

as $n \longrightarrow \infty$.

Remark 5.1.   In the theory of probability the consideration of independently and identically distributed random variables is very often and this expression is abbreviated as iid. Therefore, the above theorem is in the "language" of probability theory. However, in the language of mathematical statistics the finite set of iid random variables is nothing other than sample. Taking into account this delicate nuance, the above theorem is easily translated into the language of statistics as follows:

Theorem 5.2. (Weak law of large numbers (Khinchine (1929))).   *Let*

$$\xi_1, \xi_2, \ldots, \xi_n \tag{5.4}$$

be a sample. If their expectations exist, i.e., $\mathbf{E}(\xi_k) = a < \infty$, $k = 1, 2, \ldots, n$, then for any small $\varepsilon > 0$, we have

$$\mathbf{P}\left(\left| \frac{\xi_1 + \xi_2 + \cdots + \xi_n}{n} - a \right| \le \varepsilon \right) \longrightarrow 1, \tag{5.5}$$

as sample size $n \longrightarrow \infty$.

Kolmogorov (1930) proved that the version of the SLLN is also valid under the conditions of the Khinchine's WLLN, i.e., the SLLN holds under the conditions of Theorem 5.2.

Theorem 5.3.   (Strong law of large numbers (Kolmogorov (1930))).   *Let*

$$\xi_1, \xi_2, \ldots, \xi_n \tag{5.6}$$

be a sample. If their expectations exist, i.e., $\mathbf{E}(\xi_k) = a < \infty$, $k = 1, 2, \ldots, n$, then , we have

$$\mathbf{P}\left(\lim_{n \to \infty} \frac{\xi_1 + \xi_2 + \cdots + \xi_n}{n} = a \right) = 1. \tag{5.7}$$

The meaning of Khinchine's LLN is that the probability that the arithmetic mean (5.1) will differ from its mathematical expectation by less than any previously chosen $\varepsilon > 0$ tends to one, as the sample size increases. The meaning of Kolmogorov's SLLN is that this statement takes place with probability one.

This means that both laws of large numbers assert that by increasing the number of measurement the precision can be improved at an arbitrary degree, or expressing in terms of an "idealized" case of infinite number of measurements, "the true" value can be achieved. However, we all know that the "infinite number of measurements" can never be achieved. In real-life practical circumstances the number of independent measurements is to be limited because of the cost of running measuring equipments, time, manpower and so on. On the other hand, due to modern developments in measurement technology, the possibility of the large number of repeated measurements is ever increasing process. Therefore, the role of the laws of large numbers will increase.

6. The Central Limit Theorem as Quantitative Probabilistic Refinement of the Laws of Large Numbers

James Bernoulli, the inventor of the LLN, was not concerned with the numerical determination or estimation of probability content, instead his aim was to establish the fundamental law which shows the stabilization process of certain random phenomena like as the frequency of probability and the arithmetic mean of random quantities. In 1733, De Moivre discovered the central limit theorem (CLT) for the special case of binomial random variables and applied it to calculating probabilities related to insurance problems. After De Moivre, Laplace (1810, 1812) formulated and proved a rather general version of the CLT and applied it to the problem of determining the probability that the mean inclination of orbits of any number of comets is within prescribed error limits. With the help of his CLT, Laplace calculated the error limits for at that time known 97 comets.

In the light of the above facts of historical developments we are in position to easily grasp the connection between the LLN numbers and CLT and its meaning. To see this situation clearly, we below consider the following special case of the the central limit theorem.

Let
$$\xi_1, \xi_2, \ldots, \xi_n \tag{6.1}$$
be a sample with probability distribution function $F(\cdot)$ having the mathematical expectation a and variance σ. We introduce the following notations.

$$F_n(x) = \mathbf{P}\left(\frac{\frac{\xi_1 + \xi_2 + \cdots + \xi_n}{n} - a}{\sigma n^{-1/2}} < x\right) \tag{6.2}$$

$$= \mathbf{P}\left(\frac{\xi_1 + \xi_2 + \cdots + \xi_n - na}{\sigma n^{1/2}} < x\right) \tag{6.3}$$

Theorem 6.1 (The Central Limit Theorem for Sample). *Let $\xi_1, \xi_2, \ldots, \xi_n$ be the (6.1) sample. Then, we have*
$$F_n(x) \longrightarrow \Phi(x), \tag{6.4}$$
as sample size $n \longrightarrow \infty$. Here $\Phi(\cdot)$ denotes the probability distribution function of the standard normal distribution.

We see that (6.4) is much stronger statement than (5.7), because the (6.4) supplies us with a quantitative estimate for the probability of the deviation of the sample mean from its mathematical expectation, while the LLN proves only the convergence to the mathematical expectation. But the LLN are valid,—even—when the random variables have no finite variance, if we consider the LLNs in the form of Khinchine and Kolmogorov. In this sense, the LLN are more general than the CLT. Therefore, we should use both the LLN and the CLT for the data analysis. However, the LLN should be used not in the conventional form, but in the form that is presented in the previous section, because the conventional form of the LLN requires the existence of the variance in order merely to supply us with a mathematically elegant proof of the LLN based on Chebyshev inequality.

Comparing conditions of theorems in the previous section with that of the above CLT, we see that the requirements for the probability distribution of the parent population, i.e., for the probability distribution that models measurement data is changed towards a little more stronger condition, because the existence of the higher order moments imply that of the lower order moments and the reverse of this statement does not hold in general. The condition in the case of the LLN was only the existence of first order moment, i.e., mathematical expectation, while in the case of the CLT the existence of second order moment, i.e., variance. If we continue this way of thinking, then the third order moment should be the next step requirement. The most significant and interesting result in this area of mathematical research is the following theorem discovered independently from each other by Berry (1941) and Esseen (1942).

Theorem 6.2. (Berry-Esseen Theorem). *We supply the above CLT with the condition that $\mathbf{E}(|\xi_1^3|)$ exists. Then the inequality*

$$|F_n(x) - \Phi(x)| \leq \frac{C}{\sqrt{n}(1 + (|x|)^3)} \frac{\mathbf{E}(|\xi_1 - a|^3)}{\sigma^3} \tag{6.5}$$

holds for all x, where C is a universal constant.

Remark 6.1. We recall that $\xi_1, \xi_2, \ldots, \xi_n$ can be considered as being different "replicas" of only one random variable, or—in other words—different independent realizations of the random variable that serves as a model for the measurement process, because these random variables constitute the statistical sample. This is simultaneously the condition of both Theorems 6.1 and 6.2. In this sense, we write $\mathbf{E}(|\xi_1^3|)$ instead of

$$\mathbf{E}(|\xi_1^3|) = \mathbf{E}(|\xi_2^3|) = \ldots = \mathbf{E}(|\xi_n^3|).$$

The same remark applies to Inequality 6.5, i.e., the inequality is valid for any of the random variables $\xi_1, \xi_2, \ldots, \xi_n$.

The right-hand side of Inequality 6.5 is the absolute difference between the probability given by 6.2 or 6.3 and the probability distribution function of the standard normal random variable. This difference, i.e., the distance of an approximation of the the first probability by the second one is from above estimated by the quantity $\frac{C}{\sqrt{n}(1+(|x|)^3)} \frac{\mathbf{E}(|\xi_1 - a|^3)}{\sigma^3}$ which depends on an universal constant C, n, x and the quotient of 3rd and 2nd order moments. From (6.5) we see that the quality of the approximation is of the order of

$o(\frac{1}{\sqrt{n}})$ and $o(\frac{1}{(1+(|x|)^3)})$, i.e,. the approximation can be improved either by increasing repetition of measurements or the absolute value of x, or the both of them. The most remarkable finding in this analysis of Theorem 6.2 is that if the 3rd order moment exists, then for sufficiently large x, independently from the probability distribution type of the sample, the probability $F_n(x)$ can be asymptotically evaluated by the standard normal distribution. We will return to this important fact in Section 8 in connection with Chebyshev inequality.

The question of estimating the universal constant has been the subject of intensive research. In their original papers both Berry and Esseen obtained upper bound. Esseen (1956) gave the following lower bound.

$$\frac{3+\sqrt{10}}{6\sqrt{2\pi}} \le C \qquad (6.6)$$

Zolotarev (1967) proved that $C < 0.9051$. The best upper bound was obtained by van Beer (1972) reducing Zolotarev's result to 0.7975. Consequently, Theorem 6.2 can be extended by the following inequality:

$$\frac{3+\sqrt{10}}{6\sqrt{2\pi}} \le C < 0.7975. \qquad (6.7)$$

More information on the CLT can be found in Monhor and Takemoto (2004).

7. The Large Deviations as Probabilistic-Theoretic Models for Outliers

During the past two-three decades, one important theory has emerged within the modern probability theory. This is the theory of large deviations. In recent years, the theory of large deviations has rapidly been developed in different directions finding various applications to such diverse areas like statistics, operations research, information theory, statistical physics and engineering. The theory of large deviations deals with the probabilistic descriptions and asymptotic analysis of random events where a sum of random variables deviates from its mean by more than "normally expected" amount and generalizations of these results to such abstract mathematical spaces like Banach space and topological vector space. Bearing in mind the intention of applying some results and ideas of the large deviation theory to the clarification of the concept of outlier, we reformulate this description—in simplified setting—in more concrete terms as follows: the theory of large deviations deals with asymptotic evaluating the probabilities that $\xi_1 + \xi_2 + \ldots + \xi_n$ exceeds a large threshold value.

If we think that this threshold value as the boundary that determines outlier, then we will be in position to estimate probabilities for occurrences of outliers, and thereby we also put a probabilistic background for outliers. The most important point here is that the threshold value is not a concrete number, but it can flexibly and functionally be controlled by the large deviation probability relationship. This flexibility is in a complete harmony with the nature of the outliers. This is a completely new approach to understanding the nature of outliers in terms of probability theory, and—as far as we know—the results of the large deviation theory have not yet been used in outlier problems, geodesy, geophysics and astronomy.

The main emphasis of LLN presented in Section 5 was on the probabilistic law that describes the stability in sample mean. However, in real-life situations there may sometimes arise considerable fluctuations or deviations from the mean. By incorporating some useful results from the large deviations theory into the LLN, these real-life situations can be taken into consideration. Bahadur and Rao (1960) published a theoretical paper titled "On deviations of sample mean". In the paper by Bahadur and Rao (1960) the asymptotic behavior of the probability

$$P_n = \mathbf{P}\left(\frac{\xi_1 + \xi_2 + \cdots + \xi_n}{n} \ge d\right), \qquad (7.1)$$

where d to a certain extent is arbitrary real number, was studied. One of the main results of this paper can be formulated in the following simplified form.

Theorem 7.1. (Bahadur and Rao (1960)). *There exists a sequence of positive real numbers that*

$$P_n = \frac{\rho^n}{\sqrt{2\pi n}} b_n((1+o(1)), \ln b_n = O(1), \qquad (7.2)$$

where $n \longrightarrow \infty$.

The above theorem is non-constructive, because the constants ρ and b_n are not given in an explicit form. But the existence of large deviations and their dependence of the number of measurements is given in the functional form.

The asymptotic analysis of probabilities for large deviations can be considered as modern developments of the CLT. We here present the results due to Cramer (1938) and Linnik (1961).

Theorem 7.2 *Under the condition that $\mathbf{E}(\exp(\lambda\xi))$ exists for some $\lambda > 0$ and for positive $x = o(\sqrt{n})$, we have the following asymptotic relationship:*

$$\frac{1-F_n(x)}{1-\Phi(x)} = \exp\left\{\frac{x^3}{\sqrt{n}}\right\}\lambda\left(\frac{x}{n}\right)\left[1+O\left(\frac{x+1}{\sqrt{n}}\varphi(x)\right)\right] \qquad (7.3)$$

$$\frac{F_n(-x)}{\Phi(-x)} = \exp\left\{\frac{-x^3}{\sqrt{n}}\right\}\lambda\left(\frac{-x}{n}\right)\left[1+O\left(\frac{x+1}{\sqrt{n}}\varphi(x)\right)\right] \qquad (7.4)$$

Here $\lambda(\cdot)$ is a power series which converges in a certain neighborhood of zero whose coefficients depend only on the moments of sample. The $\varphi(x)$ denotes the probability density function of the standard normal distribution.

Remark 7.1. Multiplying by $1 - \Phi(x)$ the both sides of Eq. (7.3), we get the analytical expression for the probability that a positive outlier occurs. Similarly, Eq. (7.4) gives the probability for the case of negative outliers. Theorem (7.2.) is one of the fundamental results in the large deviations theory. The coefficients of the power series $\lambda(\cdot)$ in the theorem is not given explicitly. However, computationally tractable results can be obtained from the theory of large deviations. Recent results of the large deviations theory can be found in Cramer (1938), Linnik (1961), Nagaev (1979) and their references.

8.   Chebyshev Inequality, Outlier Detection and Data Quality: Probabilistic Background

In the theory of probability the Chebyshev inequality is the most frequently used tool for proving different convergence processes, and it plays fundamental role in proofs of various forms of laws of large numbers. This means that the Chebyshev inequality has pure theoretical "applications" in the theory of probability.

Chebyshev inequality.   *If a random variable ξ has finite variance, then, for any $\lambda > 1$, we have*

$$\mathbf{P}\left(|\xi - \mathbf{E}(\xi)| > \lambda\sigma(\xi)\right) \leq \frac{1}{\lambda^2}, \qquad (8.1)$$

where $var(\xi) = \sigma^2(\xi)$.

Remark 8.1.   If $0 < \lambda \leq 1$, then we have $\frac{1}{\lambda^2} \geq 1$, i.e., the right-hand side of (8.1) becomes to be equal or greater than 1. This means that the case $0 < \lambda \leq 1$ leads to a trivial inequality. Consequently, the condition that $\lambda > 1$ is not a restriction, but avoids the trivial case.

Remark 8.2.   In the theory of probability the following form of Chebyshev inequality is very often used.

$$\mathbf{P}\left(|\xi - \mathbf{E}(\xi)| > \varepsilon\right) \leq \frac{var(\xi)}{\varepsilon^2}, \qquad (8.2)$$

where ε is an arbitrary small positive number. This form is suitable for formulating different conditions of convergence in terms of variance. From the right-hand side of Inequality (8.2), it is evident that the accuracy of the inequality will be spoilt, if the positive real number ε decreases. In addition, the ε should be chosen as small as possible, in order to prove a convergence of a sequence of random variables. In this situation Inequality (8.2) is not used for a single random variable ξ, but for a sequence of random variables— say—for ξ_n, $n = 1, 2, \ldots$. However, in the proving the convergence of a sequence of random variables, the accuracy of the probability for a single random variable is not essential, but the sequence of variances. This means that the convergence to zero of the sequence of numerators in the right-hand side of Inequality (8.2) should be managed. This goal can be achieved in numerous ways. This is an important research area in the probability theory.

Remark 8.3.   In what follows we will see the Chebyshev inequality from a different or a nonconventional angle. By writing the Chebyshev inequality in the form (8.1), we have the probability that the deviation of a random variable ξ from its mathematical expectation exceeds $\lambda\sigma(\xi)$. The consideration of the $\lambda\sigma(\xi)$ supplies us with two kinds of generalizations of the well-known 3-σ rule. The first one

is that instead of the multiplier 3 we have the multiplier λ. The second one is that instead of the normal distribution any such probability distribution that has variance can be considered. Both of them lead to flexibility. On the other hand, the increase of $\lambda\sigma(\xi)$ in Inequality (8.1) will improve the accuracy of the inequality. This can not be seen from Inequality (8.1). However, Inequality (8.2) suggests this observation to us, but it is still not clear. A little later, this aspect will be studied numerically. By increasing the multiplier λ, we get larger deviations, i.e., the left-hand side of (8.1) becomes the probability which is suitable for the consideration of outlier problem. This is the great advantage of the Chebyshev inequality written in the form (8.1).

Remark 8.4.   Chebyshev inequality can be formulated for the case of sample. Let $\xi_1, \xi_2, \ldots, \xi_n$ be a sample. Simple algebraic manipulations give us the following inequality.

$$\mathbf{P}\left(\sqrt{n}\left|\frac{\xi_1 + \xi_2 + \cdots + \xi_n}{n} - \mathbf{E}(\xi_1)\right| > \lambda\sigma(\xi_1)\right) \leq \frac{1}{\lambda^2}$$
$$(8.3)$$

Remark 8.5.   Here we would like to mention that Chebyshev inequality is still subject to mathematical research (Bickel and Krieger, 1992).

We see that the Chebyshev inequality gives an estimation or a bound for the probability of deviation of a random variable from its mathematical expectation in terms of its variance. The viewing Chebyshev inequality as a probability bound is essential for the translation from its conventional theoretical applications to the practical setting, and from this standpoint we consider further insights into Chebyshev inequality. Now, we consider the numerical aspect of Chebyshev inequality in comparison with the standard normal distribution. The random variable ξ in the following table is the standard normal random variable.

The table displays that the accuracy of Chebyshev inequality increases with the multiplier λ. This means that we can use the Chebyshev inequality for an estimation for large enough deviations. This observation has the following two important consequences. (i) Chebyshev inequality gives a numerically applicable estimate for the probability of deviation of a random variable from its mathematical expectation when the deviation is large enough, i.e., it exceeds more than two or three times its variance; (ii) in connection with the above observation, we recall that in geodesy and geophysics, it is commonplace to use the outlier detection rule which rejects the data of a deviation exceeding more than three times the variance in conjunction with normal distribution (for example, in Imanishi, Higashi and Fukuda

	Probabilities, their bounds and errors		
λ	$\mathbf{P}\left(|\xi - \mathbf{E}(\xi)| \leq \lambda\sigma(\xi)\right) = 2\Phi(\lambda) - 1$	$\mathbf{P}\left(|\xi - \mathbf{E}(\xi)| \leq \lambda\sigma(\xi)\right) \geq 1 - \frac{1}{\lambda^2}$	absolute error of the bounds
1	0.683	0	0.683
1.5	0.866	0.556	0.310
2	0.954	0.750	0.204
2.5	0.988	0.840	0.148
3	0.997	0.889	0.108
3.5	0.998	0.918	0.080

(2002) for the calibration of superconducting gravimeter a deviation exceeding four times the variance was used to reject outliers).

In the geodetic literature the statistical testing procedure introduced by Baarda (1968) is often used as an outlier detection method. In rigorous mathematical setting, the attitude of Baarda is based on *a priori* assumed normal distribution. Consequently, the method of Baarda works essentially for the case of the normal distribution.

Chebyshev inequality is valid for all the probability distributions having finite second moment. Consequently, the Chebyshev inequality holds for all the probability distributions having finite third order moment. In this case, Berry-Esseen theorem states that the probability of the deviation of sample mean from its mean can approximately be evaluated by the normal distribution provided the threshold value and/or sample size is large enough. Combining this fact with our finding expressed by the above table, we can conclude that Chebyshev inequality can be used in estimating large deviations probabilities. This finding and the probabilistic aspects presented in previous sections establish an integrated probabilistic background theory for outliers, especially outliers of the category of (a). Finally, we would like to mention that the outliers of the category (a) are essentially connected with the assessment of data quality and the very origin of the 3-σ rule stems from the statistical control of quality in mass-production.

9. Concluding Remarks

The most useful and important points of the results of the present paper can be summarized as follows.

Nevertheless the concept of outlier has a long history, there has been no definition of it. This is a quite conflicting situation, it gives contingency for subjective judgements and misunderstandings. We do not know the reason of this undesirable situation, and the aim of the present paper is not looking at this reason. The new generic and categorizing definition given in Section 3 contributes to the correction of this unfavorable situation. The categorization integrates the diverse and heterogenous appearance of outliers and helps the comprehensive grasp of the concept of outlier, since any kind of outliers falls to one of the three categories.

We observed that the outlier problems had so far been treated only in statistical and/or "semi-empirical" way without paying proper attention to probabilistic-theoretic background. Taking into account this observation, we consider outliers in the light of achievements of the modern theory of probability. This novel attitude is realized through the methodologically new way of the unified presentation of the LLN, CLT and theory of large deviations. Avoiding the "theorem-proof" type technicalities very often used by theoretical mathematicians, the main emphasis is placed on the understanding of these important areas of the modern probability theory and connecting them with the outlier problem.

In terms of the theory of large deviations the new probabilistic-theoretic model of outliers is given. As far as we know, it is the first time that the theory of large deviations is considered in the literature of geodesy, geophysics and astronomy.

We established that Chebyshev inequality gives a numerically applicable estimate for the probability that the discrepancy between the random variable and its mathematical expectation is large enough. On the other hand, we also observed that Berry-Esseen theorem indicates that—for all the probability distributions having the 3rd order moment— the probability of the sample mean can be approximated by normal distribution. These findings are useful and new facts from the standpoint of the consideration of outliers in terms of probability.

The consideration of the way how Kepler arrived to his famous law in the light of outlier problem is a new result from the point of view of history of science, since until now the analysis by Kepler has never been studied within the framework of outlier.

Besides these main results, many other useful subsidiary observations and remarks have been presented. For example, Section 4 is not only useful for understanding the subsequent sections, but it is of an independent interest, too. It may be the first time that here we presented the most important tools of the modern asymptotic analysis in geodetic literature.

References

Aiton, E. J., Kepler's second law of Planetary Motion, *Isis*, **60**, 75–90, 1969.

Baarda, W., A testing procedure for use in geodetic networks, Publications on geodesy, Vol. 2, No. 5. Netherlands Geodetic Commission, Deft, 1968.

Bachmann, P., Zahlentheorie: Die Analytische Zahlentheorie, Zweiter Theil, B. G. Teubner, Leipzig, 1894.

Bahudar, R. and R. R. Rao, On deviations of sample mean, *Annals of Mathematical Statistics*, **31**, 1015–1027, 1960.

Barnett, V. and T. Lewis, *Outliers in Statistical data*, Second Edition, John Wiley, New York, 1984.

van Beek, P., An application of the Fourier methods to the problem of sharpening the Berry-Esseen inequality, Z. *Wahrscheinlichkeitstheorie ver. Geb.*, **23**, 187–196, 1972.

Bernoulli, D., The most probable choice between several discrepant observations and the formations therefrom of the most likely induction, *Reprinted in Biometrika*, **48**, 1–18, 1961.

Bernoulli, J., Wahrscheinlichkeitsrechnung, Whilh. Engelmann, Leipzig, 1899.

Berry, A. C., The accuracy of the Gaussian approximation to the sum of independent variates, *Trans. Amer. Math. Soc.*, **49**, 122–136, 1941.

Bickel, P. J. and A. M. Krieger, Extensions of Chebyshev's inequality with applcations, *Probability and Mathematical Statistics*, **13**, 293–310, 1992.

Boscovich, R. J., De litteraria expeditione per pontificiam ditionem, et synopsis amplioris operis, ac habentur plura ejus ex exemplaria etiam sensorum impessa, *Bononiensi Scientiarum et Artum Instuto Atque Academia Commentarii*, **4**, 353–396, 1757.

Chebyshev, P. L., Des valeurs moyennes, Liouville's, *J. Math. Pures Appl.*,

12, 177–184, 1867.

Cramèr, H., Sur un nouveau théorème-limite de la théorie des probabilités, *Actualités Scientifiques et Industrielles*, **736**, 5–23, 1938.

Detrekoi, A., On the taking of gross errors into consideration in processing measurement data, *Geodezia es Kartgrafia*, No. 3, 155–160, 1986 (in Hungarian).

Dreyer, J. L. E., *Tycho Brahe: a Picture of Scientific Life and Work in the XVIth Century*, Black, Edinburgh, 1890.

Esseen, C. G., On the Liapunov limit error in the theory of probability, *Ark. Mat. Astr. Fys.*, **28**, 1–19, 1942.

Finney, R. L. and G. B. Thomas, *Calculus*, Addison-Wesley, New York, 1990.

Gather, U., Outlier models and some related inferential issues, in *The Exponential Distribution*, edited by N. Balakrishnan and A. P. Basu, pp. 221–239, University of Missouri-Columbia, Gordon and Breach Publishers, 2000.

Imanishi, Y., T. Higashi, and Y. Fukuda, Calibration of the superconducting gravimeter T011 by parallel observation with the absolute gravimeter FG5#210—a Bayesian approach, *Geophys. J. Int.*, **151**, 867–878, 2002.

Khinchine, A. I., Sur la loi des grands nombres, *Comptes rendus de l'Académie des Sciences*, **189**, 477–479, 1929.

Knuth, D. E., "Big Omicron and big Omega and big Theta", SIGACT News, *Special Interest Group on Algorithms and Computation Theory*, **8**, 18–14, 1976.

Kolmogorov, A. N., Sur la loi forte des grands nombres, *Comptes rendus de l'Académie des Sciences*, **191**, 910–912, 1930.

Kolmogorov, A. N., *Foundations of the Theory of Probability*, Chelsea, New York, 1950.

Kubik, K., W. Weng, and P. Frederiksen, Oh, Gross Errors!, *Australian Journal of Geodesy, Photogrammetry and Surveying*, **42**, 1–18, 1985.

Landau, E., Vorlesungen über Zahlentheorie: Aus der Analytischen und geometrischen Zahlentheorie, Zweiter Band, Hirzel, Leipzig, 1927.

Laplace, P. S., Mémoire sur les approximations des formules qui sont fonctions de très grands nombres et sur leur applications aux probabilités, *Mémoires de l'Académie des Sciences de Paris*, 353–415, Supplement 559–569, 1810.

Laplace, P. S., *Théorie Analytique des Probabilités*, Gauthier-Villars, Paris 1st ed., 1812., 2nd ed., 1814 and 3rd ed., 1820.

Legendre, A. M., Méthods des moindres carrés, pour trouver le milien le plus probable entre les résultats de differéntes observations, *Mem. Inst. de France*, 149–154, 1810.

Linnik, Y. V., On the probability of large deviations for the sums of independent variables, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume II, 289–306, 1961.

Maire, C. and R. J. Boscovich, De litteraria expeditione per pontificiam ditionem ad dimetiendas duas Meridiani gradus, et corrigendum mappam geographicam, jussu, et auspiciis Benedicti XIV pont. Max. Suscepta. Ramae, or its French translation, 1755.

Maire, C. and R. J. Boscovich, Voyage Astronomique et Géographique dans l'Etal de l'Eglise, entrepis par l'Ordre et sous les Auspices du Pope Benoit XIV, pour mesurer deux degrés du méridien, et corriger la Carte de l'État ecclesiastique, Paris, 1770.

Monhor, D. and S. Takemoto, Geodetic and astronomical contributions to the invention of the normal distribution: some refinements and new evidences, *J. Geod. Soc. Japan*, 2004 (submitted).

Nagaev, S. V., Large deviations of sums of independent random variables, *The Annals of Probability*, **7**, 745–789, 1979.

O'Gorman, M. A. and R. H. Myers, Measures of errors with outlier in regression, *Comm. Statist. Simula.*, **16**, 771–789, 1987.

Pearson, K., James Brenoulli's theorem, *Biometrika*, **17**, 202–211, 1925.

Plackett, R. L., The principle of the arithmetic mean, *Biometrika*, **45**, 130–135, 1958.

Poisson, S. D., Recherches sur la Probabilité des Jugements en Matière Criminalle et en Matière Civile, précedées des Règles Générales du Calcul des Probabilités, Bachelier, Paris, 1837. Translated into German by C. H. Schnuse under the title: Lehrbuch der Wahrscheinlichkeitsrechung und deren wichtigen Anwendungen, Braunschweig, 1841.

Srikantan, K. S., Testing the outlier in a regression model, *Sankhya, A*, **23**, 251–260, 1961.

Stefansky, W., Rejecting outliers by maximum normal residual, *The Annals of Statistics*, **42**, 35–45, 1971.

Stefansky, W., Rejecting outliers in factorial designs, *Tecnnometics*, **14**, 469–479, 1972.

Todhunter, I., *A History of the Mathematical Theories of Attraction and the Figure of the Earth*, in two volumes, Macmillan and Co., London, 1873.

Wilks, S. S., *Mathematical Statistics*, Wiley, New York, 1962.

Wilks, S. S., Statistical inference in geology, *The Earth Sciences: Problems and Progress in Current Research*, edited by T. W. Donnelly, Rice University Semicentenial Publications, pp. 105–136, 1963.

Wilson, C., Kepler's derivation of the elliptic path, *Isis*, **59**, 4–25, 1968.

Zolotarev, V. M., A sharpening of the inequality of Berry-Esseen, *Z. Wahrscheinlichkeitstheorie ver. Geb.*, **8**, 332–342, 1967.

D. Monhor (e-mail: monhor@ella.hu) and S. Takemoto