LIMITS AND ACCURACY IN MEASUREMENTS

G.A. Ososkov
Joint Institute for Nuclear Research, Dubna 141980, Russia

E.L. Kosarev
P.L. Kapitza Institute for Physical Problems, Moscow 119334, Russia

Keywords: Measurement, Sample, Accuracy, Statistics, Error, Bias, Estimation, Maximum likelihood, Least squares, Parametric method, Non-parametric, Resolution, Superresolution, Robust statistics

Contents

Summary

The natural way to find the limits of the accuracy in measurements is based on the probability theory and the mathematical statistics. In this general review distribution-free methods and non-parametric methods, point and interval estimations of the unknown parameters are discussed. The connection between maximum likelihood method and least squares method is shown. The special sections are devoted to robust approach and to resolution of digital signals. Cramér-Rao lower-bound for accuracy is demonstrated.

1. Introduction

The word **ACCURACY** (from Lat. accuratus – made with taking care of) has several definitions:
- the freedom from errors;

- property of a human statement to be close to truth;
- degree of conformity of a measurement to a true value, i.e. to a standard or to a model;

which are reflected by its synonyms: exactness, correctness, precision. Staying more in the EOLSS context we would rather consider the case of measurements distinguishing between **direct** and **indirect** measurements. Direct measurements are accomplished either by counting the number of some events within a given time interval (as for instance, for Geiger counter), or by comparing a measured object with a standard, i.e. its accuracy can be evaluated quantitatively in units of a minimal scale factor. However in contemporary sciences and technologies direct measurements are inherent in the lowest level of a procedure of more sophisticated indirect measuring of an observable phenomenon. Such phenomena are described, as a rule, by theoretical models with given quantitative characteristics of parameters. Thus indirect measurements suppose to be a subject of calculations that leads to the problem of the accuracy estimation from the set of measured values. This problem is caused not only by the complexity of a functional dependence connecting a chosen model, its parameters and measurements, but mainly, due to errors of the latter. These errors are inherent in any measurements, direct or indirect, regardless of the thoroughness, with which the measurements have been done. The accuracy of parameters in question is inversely proportional to those errors. Therefore they have to be classified according to their sources and analyzed in order to be decreased as much as possible.

There are several types of errors distinguished depending either on their sources – such as **instrumental** and **model errors**, or on their statistical behaviour – such as **bias** and **random errors**.

Instrumental errors appear due to inevitable distortions introduced into the measurements by various maladjustments of a measuring device while its construction or by misalignments of its parts. Such errors can be observed in the process of a special calibration procedure, when an especially designed standard object is measured. Results of these calibration measurements are then handled to be compared with well known features of the standard. Such calibration data handling has twofold goals: (1) to evaluate and approximate distortions of the measuring device in order to compensate them mathematically; (2) to determine a functional transformation from the scales of measuring device to the standard coordinate system. From a mathematical point of view calibration problems belong to the more general class of unfolding problems described below.

Model errors are specific for hierarchical, indirect measurements and can often result in more serious errors in interpretation of experimental data. As soon as one tries to describe a certain phenomenon by a functional dependence on measured data and some parameters, then the choice of the type of function and values of its parameter can appear critical in verifying of such a description. We include here also errors of the method implementation, such as errors of approximation, round-off and discarding of expansion members which are of higher order of smallness. A typical example of such model errors appears when one tries to approximate observations of an unknown dependence by a polynomial. A wrong choice of this polynomial degree leads to an unavoidable approximation error. It is just the error of the wrong model and results usually in a significant accuracy loss.

However, all the errors listed above are developed in statistics of observations and, therefore, each of them can be classified statistically either as a bias (systematic error) or as a random error.

The systematic errors are caused by factors acting identically during the whole measuring process. The simplest example is weighing with wrong weights. It would always give you a wrong result unless you weigh a well-known standard weight, i.e. you make a calibration of your balance by calculating the difference between the previous biased measurement and *a priori* known weight of your standard. Then you can weigh any object and obtain its correct weight by adding that difference to the result of this biased weighing, which is, in fact, an example of the alignment transformation.

The random errors vary even for completely identical conditions of measurements depending on many occasional reasons which influences can not be taken into account in advance. We do not consider here gross errors of measurements that usually can be avoided by a careful experiment design or be eliminated later by a corresponding cut-off procedure.

Thus depending on the measurement process of any experiment some of errors listed above must be taken into account in order to improve the accuracy of measurements by a correct choice of statistical procedures embodying data handling algorithms.

2. Mathematical Formalism

In mathematical formulation we have a set of measurements (a sample)

$$x_1, x_2, \ldots, x_n \qquad (1)$$

to be processed statistically to extract the maximum of useful information related to the explored phenomenon with an acceptable level of accuracy. If our sample consists of equally distributed, independent random variables, then the first problem is usually to estimate their mean value and variance. More sophisticated problem is to estimate either the cumulative distribution function of our sample or its probability density function (p.d.f.). Depending on the nature of data and our *a priori* knowledge it can be done in several ways.

2.1. Distribution-free Methods

If the type of the sample distribution is unknown, one of **distribution-free methods** can be applied to estimate the sample mean value and even the distribution law of our sample. Theses methods are usually based on the **order statistics** $x_{(i)}$ obtained from (1) by reordering the sample in ascending order, so that $x_{(1)} \le x_{(2)} \le \ldots, \le x_{(n)}$ and the ordered measurements $x_{(i)}$ are called the order statistics. In particular, one of those statistics, namely $x_{(n/2)}$ named **the median** is a good estimation of the distribution mean value.

The *empirical probability distribution function* of the order statistics defined as

$$F_n(x) = \begin{cases} 0 & x < x_{(1)} \\ i/n & x_{(i)} \le x < x_{(i+1)} \\ 1 & x_{(n)} \le x \end{cases} \qquad (2)$$

can serve as a good estimation of the sample distribution law, whose accuracy is increased asymptotically with increasing n. That allows us to determine the type of the sample distribution by some of distribution-free **goodness-of-fit tests** in order to apply afterwards one of the parametric methods described in the next subsection.

2.2. Parameter Estimation

In a parametric case the type of the sample distribution is known and the problem is to estimate its parameters. Given the sample (1), estimation consists in determining either a value (so-called point estimation) or an interval most likely including the unknown parameter value in question (interval estimation).

2.2.1. Point Estimators

Both terms: *estimation* and *estimator* are often used. There is a minor difference between them: the first one often denotes the process or the procedure of the parameter estimation whereas the second one more often denotes the specific function of the sample data which is used for parameter estimation. We shall use both terms.

Estimators are constructed as functions of our sample data and, therefore, are random values, whose accuracy related properties can be expressed in probability terms only: its mean value, variance and probability of a big deviation from the estimated parameter. Thus having chosen an estimator, one can consider its goodness in terms of following basic properties:

- consistency,
- unbiasedness,
- efficiency,
- robustness.

An estimator is called *consistent* if its estimates converge toward the true value θ of the unknown parameter as the number n of measurements increases. The convergence is understood *in probability*, i.e. given any ε and any η, $\hat{\theta}_n$ is a consistent estimator of θ if an N exists such that

$$P\left(\left|\hat{\theta}_n - \theta\right| > \varepsilon\right) < \eta$$

for all $n > N$. One of the most widely known estimators of the center of the sampling distribution is the *arithmetic mean* of the sample data (we call it further *the sample mean*)

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i . \qquad (3)$$

Its consistency follows from the famous *law of large numbers* for the majority of distributions. Although as a function of random measurements the sample mean is a random variable, it is more precise than any of these measurements, since its variance is $\sqrt{n}$ times smaller.

It should be borne in mind that there are distributions for which the law of large numbers is not valid and the arithmetic mean is an inconsistent estimator for these distributions. As an example, consider the probability density function of the Cauchy distribution

$$p(x) = \frac{1}{\pi\left(1 + x^2\right)}, \quad -\infty < x < \infty. \qquad (4)$$

Both the mean value and the variance do not exist for the Cauchy distribution.

Denoting by E the *mathematical expectation* of a random variable we define the *bias* b of the estimator $\hat{\theta}_n$ as the deviation of its expectation from the true value θ_0,

$$b_n\left(\hat{\theta}_n\right) = E\left(\hat{\theta}_n - \theta_0\right).$$

Thus, an estimator is *unbiased* if for all n and θ_0

$$b_n\left(\hat{\theta}_n\right) = 0$$

or

$$E\left(\hat{\theta}_n\right) = \theta_0 .$$

Let us take as an example such an important characteristic of any distribution as its variance, i.e. the expectation of squared deviations of a random variable from its mean

$$\sigma_x^2 = E\left(x - E\left(x\right)\right)^2 .$$

Calculating the *sample variance* one should replace the unknown value of the distribution mean by its statistical analog, i.e. the sample mean (3) that gives

$$S_x^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 , \qquad (5)$$

which is, in fact, an estimation of the variance of our sampling distribution. However the replacement we made introduces the bias in this estimator. As it is easy to calculate, its expectation is equal to

$$E\left(S_x^2\right) = \frac{n-1}{n}\sigma_x^2 = \sigma_x^2 - \frac{1}{n}\sigma_x^2,$$

which means that we have the bias term σ_x^2/n. The estimator (5) is consistent and the presence of the small bias is not important when the sample size n is very large, but for small n we have to correct our estimator to make it unbiased

$$\tilde{S}_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2.$$

The above two properties are important, but not enough to describe the goodness of an estimator. Since it is a random variable, its precision can be evaluated in terms of its variance. For instance, from the above mentioned consistent estimators of the distribution center, namely, the median and the arithmetic mean of the sample data, the second one has its variance smaller than for the median (in the majority cases, when the variance of the sampling distribution does exist). The arithmetic mean can be clearly considered as a more efficient estimator than the median.

Thus, in general, the *efficiency* of an estimator is determined by its variance: the smaller it is the more efficient is this estimator.

The estimator *robustness* means that it should be independent of the distribution, or insensitive to departure from the assumed distribution. In such a sense the median is more robust estimation for the sample distribution center than the sample mean (3), especially for distributions like Cauchy distribution (4), one having no mean value at all.

More detailed consideration of robust estimates is given in Section 3 below.

In the choice of a good estimator one can note a conflict between efficiency and robustness requirements. It is a typical situation when one wants to choose an estimator which must meet all requirements stated above and even some more needed to satisfy such realistic demands as minimum computer time or a simplicity in understanding and, in general, minimum loss of scientists' time. To find a compromise one must establish an order of importance between these requirements taking into account statistical and other merits, like cost or time (urgency of completing a research). In frames of our present considerations we focus ourselves further on statistical merits.

From this standpoint one of the most powerful statistical methods for estimating parameters is the *maximum likelihood method* (MLM) invented by R.Fisher (1912). Suppose for the sample (1) we know the probability density function $f(x,\Theta)$ common

for each x_i with unknown parameter vector $\Theta = (\theta_1, \ldots, \theta_m)$. Then so-called **likelihood function**

$$L(\Theta) = \prod_i^n p(x_i). \tag{6}$$

is the density function for obtaining this sample if Θ is fixed. The MLM consists in finding an estimate of parameters $\hat{\Theta}$, which maximizes $L(\hat{\Theta})$. Since the maximum of L is also the maximum of $\ln L$, it is easier to maximize the latter function by solving *the likelihood equations*

$$\frac{\partial \ln(L(\hat{\Theta}))}{\partial \theta_k} = 0, \quad k = 1, 2, \ldots, m \tag{7}$$

in order to obtain *the maximum likelihood estimation* $\hat{\Theta}$. Its remarkable properties such as *asymptotic consistency*, *efficiency* and *normality* are proven.

Now one can easily find MLM-estimates for a known sample distribution.

Example 1. The Gaussian distribution is given by

$$f(x; a, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2\sigma^2} \right) \tag{8}$$

For the sample (1) taken from a normal population one obtains

$$\ln\left(L(a, \sigma)\right) = -n \ln(\sigma) + \ln\left(2\pi^{n/2}\right) - \frac{1}{2\sigma^2} \sum_i (x_i - a)^2 \tag{9}$$

The solution of the likelihood equations gives two MLM-estimates

$$\hat{a} = \frac{1}{n} \sum_{i=1}^n x_i, \qquad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{a})^2, \tag{10}$$

which we already had before.

Example 2. The Poisson distribution. It is a discrete distribution of the random variable taking values equal to whole positive numbers k with the probability

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}. \tag{11}$$

For the sample $k_1,\ldots,k_n$ one has

$$\ln\left(L\left(\lambda\right)\right) = -\lambda n + \ln(\lambda)\sum_{i=1}^{n} k_i + \ln\left(\prod_{i=1}^{n}\frac{1}{\left(k_i\right)!}\right). \qquad (12)$$

The solution of the likelihood equation gives the estimate of the parameter λ

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} k_i.$$

2.2.2. Interval Estimators

Each of the point estimators discussed above gives us a value intended to estimate an unknown parameter. As it was pointed out, these estimators are random by their nature, but obtaining a value we do not feel that its randomness is concealed and it could deceive about its accuracy or probability of being close enough to an unknown parameter. Therefore an experimenter prefers to use the estimators that include explicitly the range

$$\theta_a \le \theta \le \theta_b,$$

which contains the true value θ_0 with probability β. Given a measurement x from a p.d.f. $f\left(x|\theta\right)$ with a known parameter θ, the probability content β can be calculated as

$$\beta = P\left(a \le x \le b\right) = \int_a^b f\left(x|\theta\right)dx. \qquad (13)$$

However in our case we have an unknown parameter and too large arbitrariness in choosing the interval borders a and b. It would be better to choose an interval which has minimal length among all intervals $\left[\theta_a,\theta_b\right]$ with the same probability β. Such intervals are called *confidence intervals* for θ *with probability* β. Since the parameter θ is unknown, one has to take a different variable $z = z(x,\theta)$, a function of the measurement x and the parameter θ, but such that its p.d.f. is independent of the unknown θ. If it can be found, we can re-express Eq.(13) as a problem of interval estimation: given β, find the optimal range $\left[\theta_a,\theta_b\right]$ in θ-space such that

$$P\left(\theta_a \le \theta_0 \le \theta_b\right) = \beta. \qquad (14)$$

It is better to explain this scheme with a particular example of the confidence interval for the mean of normally distributed sample data with the cumulative distribution function

$$\Phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2\sigma^2}\right) dt. \qquad (15)$$

When both distribution parameters μ and σ are known one can calculate β from (13):

$$\beta = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

However when μ is unknown (but σ is known), one can instead calculate the probability β that some of functions of our measurements, say the sample mean $\bar{x}$, lie in an interval that includes its unknown mean. Let us take a symmetrical interval $[\mu - c, \mu + c]$. Then

$$\begin{aligned} \beta &= P(\mu - c \le \bar{x} \le \mu + c) \\ &= \frac{1}{\sigma^*\sqrt{2\pi}} \int_{\mu-c}^{\mu+c} \exp\left(-\frac{(t-\mu)^2}{2\sigma^{*2}}\right) dt = \Phi\left(-\frac{c}{\sigma^*}\right) - \Phi\left(\frac{c}{\sigma^*}\right), \end{aligned} \qquad (16)$$

where for the sample mean $\sigma^* = \sigma/\sqrt{n}$. We can now invert the probability statement in (16) in order to take the form of the statement (14):

$$\beta = P(\bar{x} - c \le \mu \le \bar{x} + c).$$

As it is known in the case of the normal distribution, one can obtain $\beta = 0.95$ if the constant c is chosen as

$$c = 1.96\,\sigma^* = 1.96\,\sigma/\sqrt{n}.$$

-
-
-

Bibliography

Brandt S. (1999) *Data Analysis. Statistical and Computational Methods for Scientists and Engineers*. 3-rd Edition, 652 p., Springer Verlag, .,. [It is a well known textbook where a good collection of recipes of statistical data analysis is illustrated by many Fortran programs]

Cramer H. (1999) *Mathematical Methods of Statistics*, Princeton Univ. Press, 575p., Princeton, N-Y. [This book presents all essential aspects and approaches to mathematical statistics in a general and comprehensive way]

Huber P. (1981) *Robust Statistics*, J. Willey & Sons, 308p., NY. [This book presents main approaches to robust methods of statistics]

Kolganova E., Kosarev E., Ososkov G. (2000) *Superresolution algorithms for data analysis of discrete detectors in nuclear physics*, Nucl.Instr.Meth. **A443** No.2-3, pp.464-477. [In this survey parametric and non-parametric approaches are presented for the recovery of particle coordinates from distorted and noisy data with limiting superresolution efficiency]

Kosarev E.L. et al (1993) *Programs for signal recovery from noisy data using the maximum likelihood principle. Part 1: General description*, Computer Physics Communications, Vol.74, No.3, pp.335-348, *Part 2: Program implementation*. ibid, pp.349-357. [Theory of a robust method for the signal recovery and description of the RECOVERY program, implementing this method]

Mosteller F., Tukey W. (1977). *Data analysis and regression: a second course in statistics*, 588p., Addison-Wesley. [The book of one of the most experienced specialists in exploratory data analysis with many practical recommendations and robust algorithms]

Ososkov G.A. (1987). *Robust regression for the heavy contaminated sample*, Proc. 2nd International Tampere Conference in Statistics (Tampere, Finland) [At this paper optimal weight are proposed for the robust estimates of regression parameters, joint estimates of location and scale parameters are also studied]

Shannon C.E. (1948) *A mathematical theory of communication*, Bell Sys. Techn. Journal, **27**, pp.379-423, 623-656. See also in the URL: http://www.cs.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf. [The introduction to the communication theory, where the famous theorem about limitations of signal transfer via a noisy channel is proven]

Shannon C.E. (1949) *Communication in the presence of noise*, Proc. of the IRE, **37**, No.1, pp.10-21. [Another classical C.Shannon paper, where his theorem is presented using geometric approach]
Van der Waerden B.L. (1969) *Mathematical Statistics*, 367p., Springer Verlag. [This is a widely used textbook that presents essential principles of mathematical statistics]

Biographical Sketches

Ososkov, Gennadii Alexeevich, graduated from the Moscow State University. in 1953, becoming candidate of sciences in 1957, and doctor of physico-mathematical sciences in 1987. From 1961 until the present time he has worked at Joint Institute for Nuclear research, Dubna Russia, now are being the leading scientist of this Institute. In addition he is a professor of computational statistics of the Dubna University and Ivanovo State University. His scientific interests include: mathematics, theory of probability, mathematical and computational statistics, pattern recognition, artificial intelligence, Monte-Carlo simulations. As an experienced expert in robust data handling in experimental physics, G.A.O. has been many times invited for scientific consultings to famous physical centers (to the Max Planck Institute for Nuclear Physics, Heidelberg, Germany in 1995, 1996, 1997, to the Yale University, New Haven, and Alabama University, Birmingham, USA in 1998, to DESY, Hamburg in 2000). In 1997 he was awarded by the first JINR prize, in 2000 the presidium RAS has awarded him by the monthly state grant in the field of information science for three years. G.A.O. has published more than 200 papers in scientific journals and proceedings of international conferences. He is the author of the monograph: "Pattern Recognition Application in High Energy Physics ", Serie "Lectures for young scientists", No.26, JINR publ. P10-83-187, Dubna, 1983, and coauthor of the second one: "Automation of measurements and data processing of physical experiment data ", Moscow State Univ. Publ., Moscow, 1986 (both in Russian)

Kosarev, Evgenij Leonidovich, graduated from the Moscow Institute for Physics and Technology (MIPT) in 1962, becoming candidate of sciences in 1971, and doctor of physics and mathematical

sciences in 1995. From 1962 until the present time he has worked at P.L.Kapitza Institute for Physical Problems, Moscow, Russia, now are being the head scientist of this Institute and Head for computing there. He is also an Associate Professor in MIPT. His scientific interests include: experimental physics inverse problems for applications in X-ray detectors, any kind of spectroscopy, high energy physics particle detectors, muon-spin-rotation experiments, etc…He was invited for research visits to world physics laboratories: Daresbury Laboratory, UK in 1980, 1988, CERN, Switzerland in 1982, Photon Factory, KEK, Japan in 1990. E.L.Kosarev has published more than 50 papers in scientific journals and proceedings of international conferences.