

Systematic uncertainties in experiments at LHC

Sergei I. Bityukov

Institute for High Energy Physics, Protvino, Russia

Nikolai V. Krasnikov

Institute for Nuclear Research RAS, Moscow, Russia

Plan

- **Introduction**
- **A recommendation of International Bureau of Weights and Measures**
- **Classification of systematic uncertainties**
- **Typical sources of systematic uncertainties**
- **Treatment of systematic uncertainties**
- **Systematic uncertainties in real experiment**
- **Conclusion**
- **References**

Introduction

Most measurements of physical quantities involve both a statistical uncertainty and a systematic uncertainty. In high energy physics systematic uncertainties play a key role in measurement of physical quantities, as they are often of comparable scale to statistical uncertainties.

Typically, systematic uncertainties are not well defined and are not straightforward to determine. They are also usually not Gaussian distributed, and combining systematic uncertainties from different sources is problematic. Since even the meaning and definition of systematic uncertainties are difficult to quantify.

The consideration of issues of uncertainty propagation and combining is a very important part of planning of experiments at the LHC (see, for example, PTDR CMS v.II [1]).

Such type studies [2, 3, 4, 5] often include proposals to combine statistical and systematic uncertainties, techniques to estimate the magnitude of systematic uncertainties, and the use of standard statistical techniques to take into account systematic uncertainties.

Recommendation INC-1 by the Working Group on the Statement of Uncertainties convened by the BIPM (Bureau International des Poids et Mesures) [6]

1. The uncertainty in the result of a measurement generally consists of several components which may be grouped into two categories according to the way in which their numerical value is estimated.
 - Type A. Those which are evaluated by statistical methods
 - Type B. Those which are evaluated by other means

There is not always a simple correspondence between the classification into categories A or B and the previously used classification into "random" and "systematic" uncertainties. The term "systematic uncertainty" can be misleading and should be avoided.

Any detailed report of uncertainty should consist of a complete list of the components, specifying for each the method used to obtain its numerical value.

2. The components in category A are characterized by the estimated variances s_i^2 (or the estimated "standard deviations" s_i) and the number of degrees of freedom ν_i . Where appropriate the covariances should be given.

3. The components in category B should be characterized by quantities u_j^2 , which may be considered approximations to the corresponding variances, the existence of which is assumed. The quantities u_j^2 may be treated like variances and the quantities u_j like standard deviations. Where appropriate, the covariances should be treated in a similar way.
4. The combined uncertainty should be characterized by the numerical value obtained by applying the usual method for the combination of variances. The combined uncertainty and its components should be expressed in the form of "standard deviations."
5. If for particular applications, it is necessary to multiply the combined uncertainty by an overall uncertainty, the multiplying factor must always be stated.

One can see the contradiction between definition of uncertainty Type B and content in items 3 and 4. Physicists include in notion "systematic uncertainties" both a part of Type A uncertainties and a part of Type B uncertainties. These uncertainties are considered in framework of probabilistic approach.

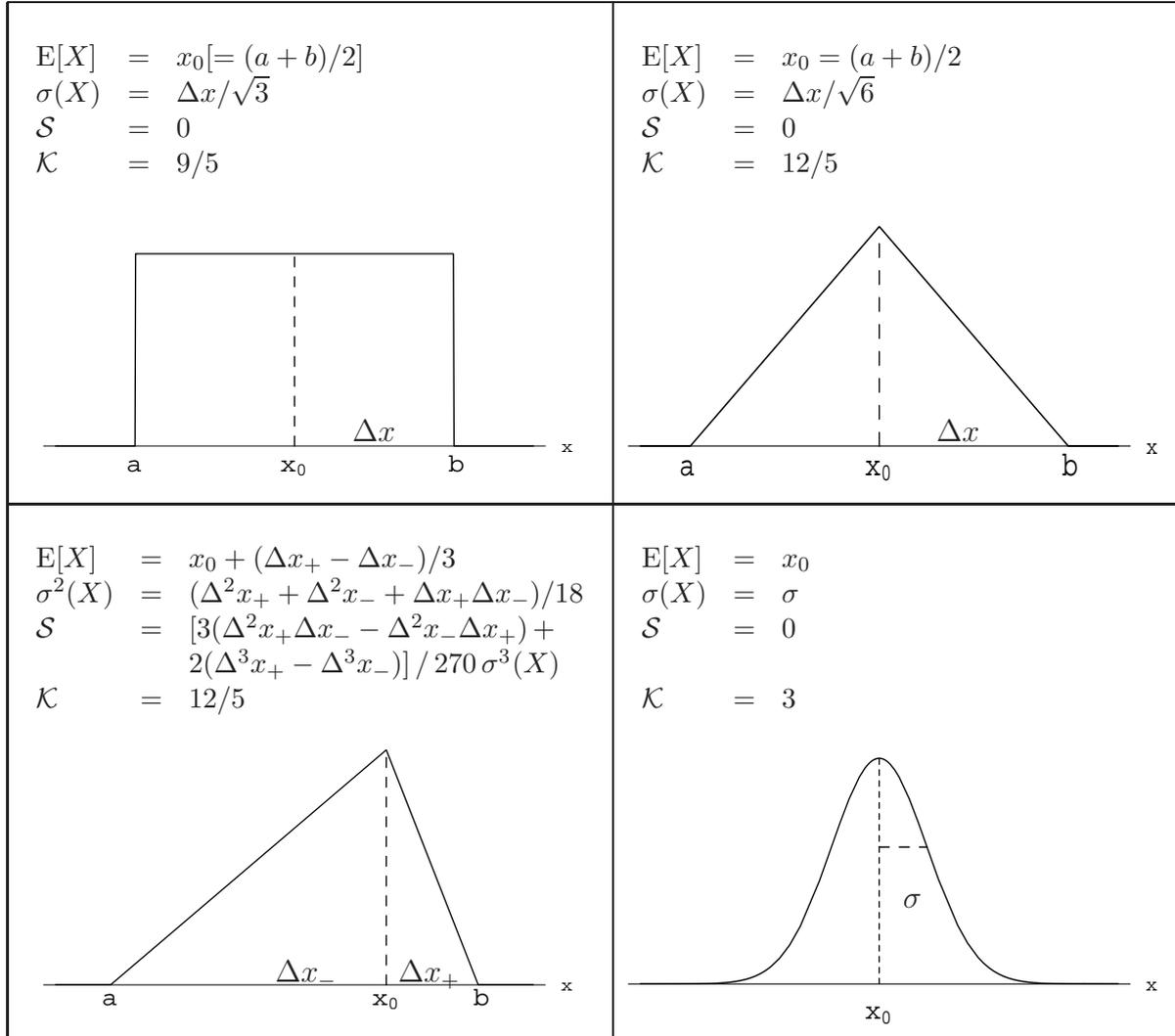


Figure 1: Typical models to assess systematic uncertainties: uniform distribution, symmetric triangular distribution, asymmetric triangular distribution, and Gaussian distribution. The expressions of the most relevant statistical parameters are reported (\mathcal{S} stands for skewness, \mathcal{K} for kurtosis) (figure is taken from [2]).

These pictures describe the possible presentation of systematic uncertainty with mean value x_0 . If systematic uncertainty is a result of direct measurement then we see probability density of this uncertainty. Often systematic uncertainties are used as parameters for correction of observed value. In this case we can say about confidence density in figures above.

Classification of systematic uncertainties

Statistical uncertainties are the result of stochastic fluctuations arising from the fact that a measurement is based on a finite set of observations.

Examples of statistical uncertainties include the finite resolution of an instrument, the Poisson fluctuations associated with measurements involving finite sample sizes and random variations in the system one is examining.

Systematic uncertainties arise from uncertainties associated with the nature of the measurement apparatus, assumptions made by the experimenter, or the model used to make inferences based on the observed data.

Examples of systematic uncertainties include uncertainties that arise from the calibration of the measurement device, the probability of detection of a given type of interaction, and parameters of model used to make inferences that themselves are not precisely known.

Let us consider systematic uncertainties which are usually taken into account in experiment.

In ref.[4] is proposed the classification of systematic uncertainties.

- Class 1 systematics are uncertainties that can be constrained by ancillary measurements and can therefore be treated as statistical uncertainties.
- Class 2 systematics arise from model assumptions in the measurement or from poorly understood features of the data or analysis technique that introduce a potential bias in the experimental outcome.
- Class 3 systematics from uncertainties in the underlying theoretical paradigm used to make inferences using the data.

Class 1 systematics are statistical in nature and will therefore naturally scale with the sample (special or regular) size.

A common technique for estimating the magnitude of systematic uncertainty of Class 2 is to determine the maximum variation in the measurement, associated with the given source of systematic uncertainty.

Class 3 systematics do not depend on how well we understand the measurement per se, but are fundamentally tied to the theoretical model or hypothesis being tested.

The sources of systematic errors can be divided into two main categories: theoretical and experimental.

Typical sources of systematic uncertainties

- **Luminosity from detector:** That part of the luminosity uncertainty that comes from the uncertainty on the luminosity detector acceptance and efficiency.
- **Luminosity from cross section:** That part of the luminosity uncertainty that comes from the uncertainty of the inelastic and diffractive cross sections.
- **Signal modeling:** The systematic uncertainty arising from uncertainties in the modeling of the signal. This includes uncertainties from variations in the ISR, FSR, and PDF descriptions. It also includes difference in the hadronization models.
- **Detector modeling:** The systematic uncertainty arising from the uncertainty on the event detection efficiencies for object identification and MC mismodeling of data.
- **Background from MC:** The systematic uncertainty arising from uncertainties in modeling of the different background sources. It includes normalizations of uncertainties obtained from theoretical calculations.

- **Background from data:** The systematic uncertainty arising from uncertainties in modeling of the different background sources that are obtained using data-driven methods. It also includes the uncertainty on the normalizations of uncertainties, scale factor and shape as well as the uncertainty on the multijet modeling and normalization. It also includes the uncertainty due to MC statistics.
- **Particle identification and mistagging:** The systematic uncertainty coming from the uncertainty on the identification of particles and mistag rate and shape modeling.
- **JES and E_T^{miss} scale:** The JES uncertainty which originates from limitations in the calibration data samples used.

The part of systematic uncertainties depends on the integrated luminosity of experiment. The another part depends on statistics in ancillary measurements during experiment and, correspondingly, has internal limitations. Third part has not dependence from any experimental data.

Of course, systematic uncertainties depend on real study. We present here several typical values.

- **Luminosity:** the uncertainty coming directly from LHC machine estimation is about $\pm 20\%$. The current estimation in detector from the understanding of the online luminosity monitors is that the absolute luminosity calibration for the detector will be around 5% for integrated luminosity $\sim 1 \text{ fb}^{-1}$.
- **Lepton identification and trigger efficiencies** for lepton candidates can be measured in data by using the 'tag and probe' strategy. A systematic uncertainty of a few percent is expected after a few hundred pb^{-1} of data taking.
- The sensitivity of an analysis to the miscalibration and misalignment effects has been estimated of a few percent to an integrated luminosity of 10 pb^{-1} .
- A 5% precision or better is expected in a stable mode of operation of detectors for E_T^{miss} resolution.
- Overall, the total error coming from jet quantities (jet reconstruction + JES) is about 10%.
- PDF uncertainties and NLO corrections are obtained in several analyses close to 5%.
- **Normalization for background:** a 5% uncertainty is considered for the Drell-Yan background and 10% for the di-boson production (Higgs to two W study).
- The contribution on the error from the finite MC statistics for the signal is about 5%, while it is about 10% for the background.

Treatment of systematic uncertainties

We measure the random variable to estimate the parameter of model which describe the behavior of this random variable. Let consider a simple example [7].

Random variable x with parameters μ and $\sigma = \text{const}$

$$x \sim \mathcal{N}(\mu, \sigma). \quad (1)$$

Probability density function here is

$$\varphi(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (2)$$

We can write

$$x = \mu + \epsilon, \quad (3)$$

where $\epsilon \sim \mathcal{N}(0, \sigma)$ and μ is a constant.

Let we have got \hat{x} realization of x . It is an unbiased estimator of parameter μ , then

$$\mu = \hat{x} - \epsilon. \quad (4)$$

As known $(-\epsilon) \sim \mathcal{N}(0, \sigma)$, because of the symmetry of the bell-shaped curve about its central point, i.e.

$$\mu|\hat{x} \sim \mathcal{N}(\hat{x}, \sigma). \quad (5)$$

It means that we construct **the confidence density** of the parameter

$$\tilde{\varphi}(\mu|\hat{x}, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\hat{x}-\mu)^2}{2\sigma^2}}. \quad (6)$$

We see the uncertainty in determination of parameter is associated with probability. Clear, in this case the notation $\mu = \hat{\mu} \pm \sigma$ means $\tilde{\varphi}(\mu|\hat{\mu}, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\hat{\mu}-\mu)^2}{2\sigma^2}}$, i.e. we can describe the systematic uncertainties via **confidence densities** [8,9].

It allows to combine the error estimates of several uncertainties into the overall uncertainty using a root sum-of-the-squares approach. Application of root-sum-square combination of systematic uncertainties is based on the assumption of independent and Gaussian errors.

As mentioned above the systematic uncertainties usually not Gaussian distributed. For example, distributions of efficiencies are asymmetric. In this case the combination of different systematic uncertainties by standard methods is problematic.

Of course, if we know the probability density of uncertainties during planning of experiment we can use MC experiments to imitate the possible result and to construct the confidence density for this result. It is usual practice in preparing of experiments and it is done for experiments at LHC.

Systematic uncertainties in real experiment

The same issues take place for result of real experiment. We have observed value with experimental errors and a set of estimators for systematic uncertainties. How to present result of the experiment ?

The problem with numerical expression and presentation of experimental data has a long story (for example, ref.[10]).

By the using of notion confidence distributions (and, correspondingly, confidence densities) all sources of uncertainties can be included in a logically consistent way to description of overall uncertainty.

The principle of such way is shown in Fig.2.

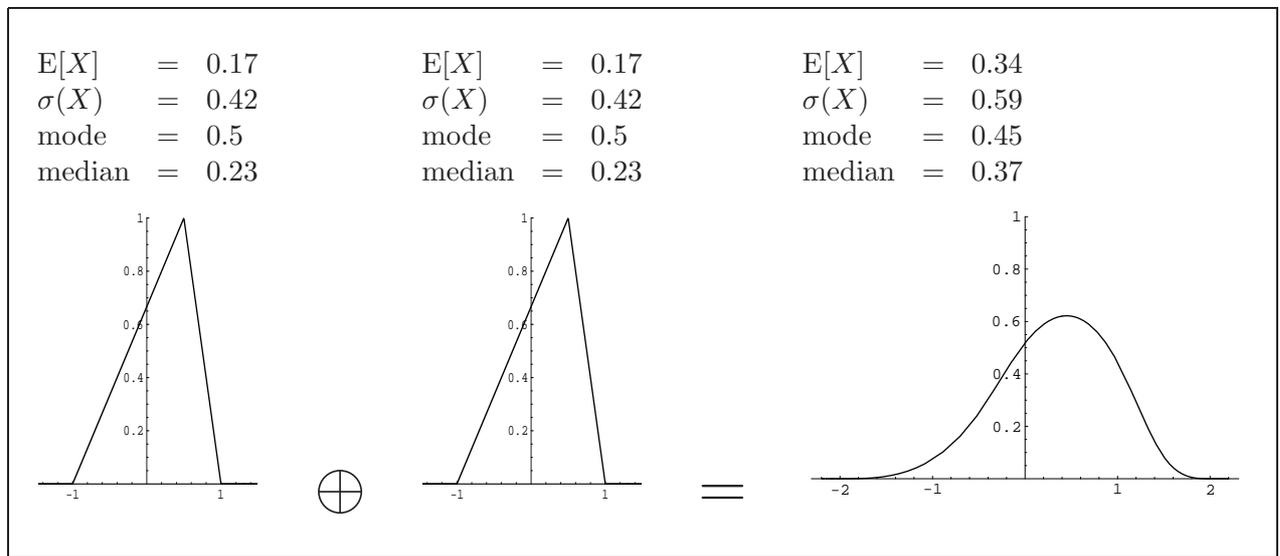


Figure 2: Confidence density function resulting from the sum of two quantities, each described by an asymmetric triangular p.d.f. with $x_0 = 0.5$, $\Delta x_+ = 0.5$ and $\Delta x_- = 1.5$ (figure is taken from [2]).

Conclusions

- We discuss the issues related to systematics in experimental data.
- We consider the possibilities both for classifying and estimating of them and for implementing the estimators of systematic uncertainties in the data analysis of experiments at LHC.
- The confidence distribution is a very useful tool in statistical reporting, and should be a competitive frequentist analogue of the Bayesian posterior distribution.
- We use this notion in statistical studies during preparing experiments at LHC [11,12].

References

- [1] CMS PTDR v.2 (Appendix B), CERN/LHCC 2006-021, 2006
- [2] G. D'Agostini and M. Raso (2000) , e-Print: hep-ex/0002056, 2000
- [3] R. Barlow, e-Print: hep-ex/0207026, 2002
- [4] P.K. Sinervo, PHYSTAT'2003, pp.122-129, 2003; eConf: C030908
- [5] J. Heinrich and L. Lyons, Annu.Rev.Nucl.Part.Sci. 2007.57:145-169
- [6] <http://physics.nist.gov/cuu/Uncertainty/international1.html>
- [7] B. Efron, The American Mathematical Monthly, Controversies in the Foundations of Statistics, 85(4) (1978) 231-246.
- [8] K. Singh, M. Xie, W.E. Strawderman, Combining information from independent sources through confidence distributions, The Annals of Statistics 33 (2005) 159-183.
- [9] S. Bityukov, N. Krasnikov, The estimation of distributions parameters via the using of confidence distributions, IHEP Preprint 2008-10, Protvino, 2008 (in Russian).
- [10] V. Ezhela, report in NSC-08 of RAS (in Russian), http://www.ihep.ru/ihep/doc_seminar/NSC-08/081225m2/Ezhela.pdf
- [11] S.I.Bityukov, Signal Significance in the Presence of Systematic and Statistical Uncertainties, JHEP 09 (2002) 060; e-Print: hep-ph/0207130; S.I. Bityukov, N.V. Krasnikov, NIM A502 (2003) 795-798.
- [12] S. Bityukov, N. Krasnikov, A. Nikitenko, V. Smirnova, Two approaches to Combining Significances, Proc.of Sci. (ACAT08) 118, 2009.