

Statistical and other errors

Volker Blobel – University of Hamburg
March 2005

1. Statistical errors
2. Other errors
3. Correlations
4. Systematic errors

Statistical errors

Statistical errors are due to statistical uncertainties:

- arise from stochastic fluctuations (random quantum processes),
- are uncorrelated with previous measurements,
- follow well-developed theory;
- examples are finite statistics (Poisson distribution) and measurement resolution.

A statistical error is usually given by the

standard deviation σ ,

equal to the square root of the variance (expectation value of the square of the difference to the mean value):

$$V[x] = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx = \sigma^2$$

Limits and Confidence intervals

The statistical error or uncertainty of a measurement can be described by a [confidence interval](#). In this concept, first developed by NEYMAN, an interval $[a_{\text{lower}}, a_{\text{upper}}]$ is defined for a measured variable \hat{x} for a certain confidence level $1 - \gamma$, such that

$$\text{probability} \quad P(a_{\text{lower}} \leq x \leq a_{\text{upper}}) = 1 - \gamma$$

holds regardless of the true value of the variable x .

The quantity $1 - \gamma$ is also called [coverage probability](#). The interval $[a_{\text{lower}}, a_{\text{upper}}]$ would

- include this true value of the variable x in a single measurement with probability $1 - \gamma$ or,
- in repeated measurements, would include the true value in a fraction of $1 - \gamma$ of the experiments.

The result of a measurement is often represented by

$$\hat{x} \pm .$$

or, in a graphical representation, by negative and positive error bars of length $\hat{x} - a_{\text{lower}}$ and $a_{\text{upper}} - \hat{x}$.

In certain cases one-sided confidence intervals of limits are appropriate. The value a_{lower} represents a lower limit on the parameter x such that $a_{\text{lower}} \leq x$ with probability α :

$$P(a_{\text{lower}} \leq x) = 1 - \alpha .$$

The value a_{upper} represents an upper limit on the parameter x such that $a_{\text{upper}} \geq x$ with probability β :

$$P(a_{\text{upper}} \geq x) = 1 - \beta .$$

The calculation of confidence intervals and limits requires the knowledge of the distribution of the measured value \hat{x} for a given true value x_{true} of the parameter, i.e. the p.d.f. of the measured value \hat{x} , denoted by

$$g(\hat{x}; x_{\text{true}})$$

has to be known. Using this density the lower and upper limits are defined in the classical definition of limits by

$$\begin{aligned} \alpha &= \int_{x_{\text{obs}}}^{\infty} g(x; a_{\text{lower}}) dx && \text{lower limit} \\ \beta &= \int_{-\infty}^{x_{\text{obs}}} g(x; a_{\text{upper}}) dx && \text{upper limit} \end{aligned}$$

Two-sided intervals determined for the confidence interval $1 - \gamma$ are not uniquely defined. The so-called **central confidence interval** is defined by lower and upper limits with $\alpha = \beta = 1 - \gamma$; the probabilities α and β are equal, but in general the lower and upper limits a_{lower} and a_{upper} are not equidistant.

P-Values

The *p-value* is defined as the probability, that a random variable could assume a value greater than or equal to the observed (measured) value x . Its calculation requires the assumption of a hypothesis H with a probability density function $g(x; H)$. In application values of the random variable greater than the observed value x usually represent values more extreme than that observed.

The *p-value* is defined by

$$p = \int_{x_{\text{obs}}}^{\infty} g(x; H) dx .$$

and thus the *p-value* is itself a random variable. The calculation of p can be considered as a transformation of the measured value x_{obs} .

The concept of *p-values* is used in the procedure for rejecting the hypothesis H underlying the density $g(x; H)$. The procedure is:

- Decide on the significance α *before* the experiment is performed;
- perform the experiment;
- the hypothesis H is rejected for $p < \alpha$.

Example: test of goodness-of-fit with theoretical χ^2 -distribution.

An *equivalent* procedure is to calculate a limiting value x_{limit} from the selected significance α by

$$\alpha = \int_{x_{\text{limit}}}^{\infty} g(x'; H) dx'$$

and to reject hypothesis H if the observed value $x_{\text{obs}} > x_{\text{limit}}$.

Coverage

A parameter $\hat{\mu}$ is estimated from a data sample \mathbf{x} , perhaps by a fitting procedure, together with errors σ_1 and σ_2 .

The result

$\hat{\mu} \begin{matrix} +\sigma_2 \\ -\sigma_1 \end{matrix}$ determines an interval $[\hat{\mu} - \sigma_1, \hat{\mu} + \sigma_2]$ (the region between error bars) .

What is the coverage probability (or short: coverage) of the interval?

The data follow a probability density $g(\mathbf{x}, \mu)$ of the data \mathbf{x} for a given value of parameter μ .

Coverage $C(\mu)$ is a function of the (unknown) parameter μ , defined as the probability that, with $\hat{\mu} \equiv \hat{\mu}(\mathbf{x})$, $\sigma_1 \equiv \sigma_1(\mathbf{x})$ and $\sigma_2 \equiv \sigma_2(\mathbf{x})$

$$\hat{\mu} - \sigma_1 \leq \mu \leq \hat{\mu} + \sigma_2 ,$$

where μ is regarded as the fixed (true) value of the parameter and the probability statement applies to the random variables $\hat{\mu}$, σ_1 and σ_2 , which depend on the data sample \mathbf{x} .

In other words: C is for any given value of μ the probability that an experiment will obtain an interval, that includes, or “covers” the true value μ .

Frequentist require $C(\mu) = C_0$ or at least $C(\mu) \geq C_0$ with $C_0 = 0.6827$ or $C_0 = 0.95$, corresponding to 1 sigma or ≈ 2 sigma in case of a Gaussian distribution.

Poisson distributed data

The origin of a statistical error in particle physics is most often the observation of Poisson distributed data.

The probability of observing k events, if the mean value is μ , is given by

$$P(k; \mu) = \frac{e^{-\mu} \mu^k}{k!}$$

(k is a variable and μ is fixed).

There are several different methods for the error-bar scheme.

Joel G. Heinrich, Coverage of error bars for Poisson data, CDF/MEMO/STATISTICS/PUBLIC/6438, 2003

Pearson's χ^2 intervals

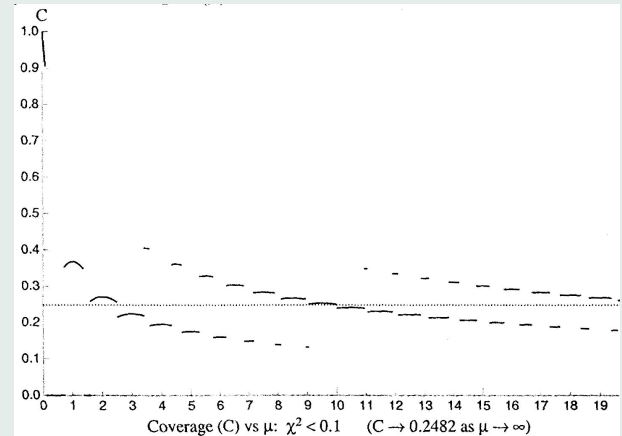
Pearson's χ^2 is given by

$$\chi^2(\mu; k) = \frac{(k - \mu)^2}{\mu}$$

After observation of k events the central estimate is determined by minimizing the χ^2 w.r.t. μ : $\hat{\mu} = k$ (where $\chi^2 = 0$).

Error bars are given by the interval such that $\chi^2(\mu; k) \leq \Delta$ with the result

$$\sigma_1 = \sqrt{k\Delta + \Delta^2/4} - \Delta/2 \quad \sigma_2 = \sqrt{k\Delta + \Delta^2/4} + \Delta/2$$



The figure shows the coverage $C(\mu)$ for the case $\Delta = 0.1$. There are many discontinuities and the values are between 0 (zero coverage regions) and 1 !

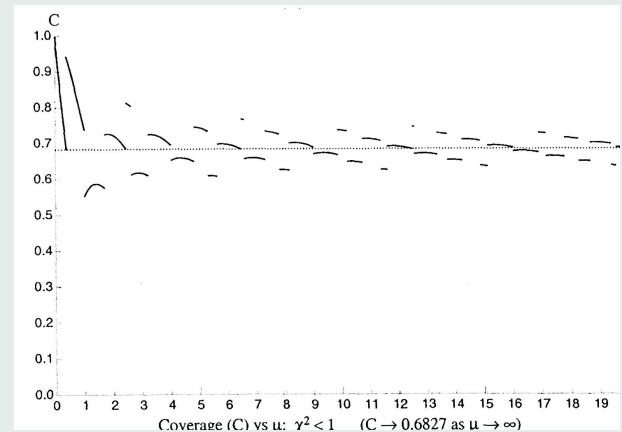
The plot shows the coverage with physicist' standard choice $\Delta = 1$.

The minimum value for $C(\mu)$ is $1.5^{-1} = 0.5518$, attained in the limit as μ approaches 1 from above. Example: 6 events observed. According to the scheme

the result would be

$$\mu = 6 \begin{matrix} +3 \\ -2 \end{matrix}$$

with coverage greater than 55.18 %.



Neymans's modified χ^2 intervals

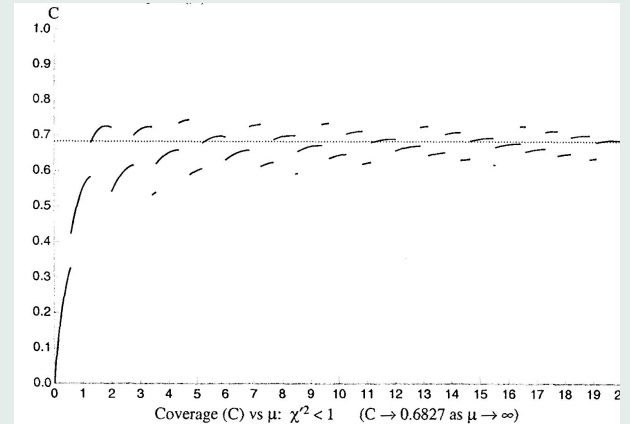
Instead of Pearson's χ^2 often Neymans's modified χ^2 is used:

$$\chi^2(\mu; k) = \frac{(k - \mu)^2}{k}$$

(number of observed events k replaces μ in the denominator). Error bars given by interval such that

$\chi^2(\mu; k) \leq \Delta$ with the result

$$\sigma_1 = \min\left(k, \sqrt{k\Delta}\right) \quad \sigma_2 = \sqrt{k\Delta}$$



The coverage is shown for the standard case $\Delta = 1$. Minimum coverage is zero!

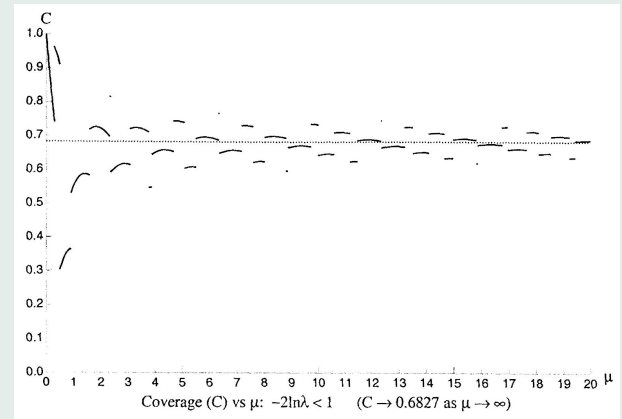
Likelihood intervals

Error intervals based on the value of the likelihood (which is minimized when one does a maximum-likelihood fit to the Poisson distribution).

Having observed k events the expression is

$$-2 \log \mathcal{L}(\mu, k) = 2 [(\mu - k) + k \log(k/\mu)]$$

The plot is for the standard choice $\Delta = 1$.



The minimum coverage is 0.3033, which occurs in the neighborhood of $\mu = 0.5$.

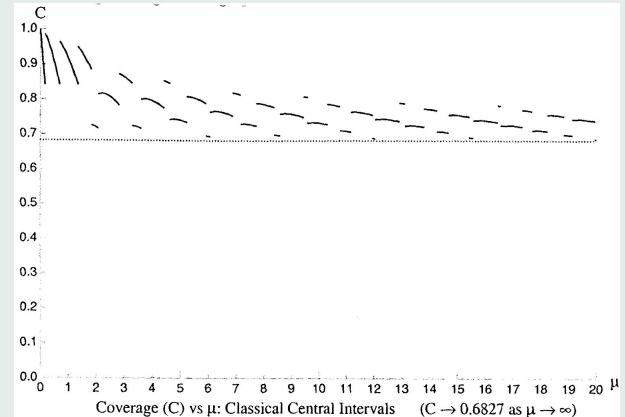
Classical-Frequentists central intervals

Coverage achieved by the “68.27 %” central intervals of the classical frequentists approach.

For k observed events the central interval is given by the set of all μ such that

$$\mu \text{ is upper limit } \sum_{n=0}^k \frac{e^{-\mu} \mu^n}{n!} \geq \frac{1 - C_0}{2}$$

$$\mu \text{ is lower limit } \sum_{n=k}^{\infty} \frac{e^{-\mu} \mu^n}{n!} \geq \frac{1 - C_0}{2}$$



The minimum coverage is guaranteed to be $\geq C_0$, but the average coverage is much larger.

$$\text{Special case } k = 0 : \quad \mu \leq \ln \frac{2}{1 - C_0} \quad (\text{upper limit})$$

The concept of the p -value is often used with the Poisson distribution as underlying distribution. Consider the case where k events are observed in a certain fixed condition, e.g. in a certain kinematic range, which are due to a certain rare process and/or background. The expected background is n_{back} , assumed to be known without error. Then the p -value is given by

$$p = \sum_{n=k}^{\infty} \frac{n_{\text{back}}^n e^{-n_{\text{back}}}}{n!} = 1 - \sum_{n=0}^{k-1} \frac{n_{\text{back}}^n e^{-n_{\text{back}}}}{n!}$$

In the case of background with an uncertainty or other more complicated cases a Monte Carlo simulation may be used. The estimated background is used as the mean of a simulated Poisson distributed number; the uncertainty of the background can be taken into account by e.g. a normal distribution with the estimated mean and standard deviation of the background.

An example for such a calculation is the following:

Observed events: $k = 7$ expected background: 3.2 ± 1.1 p -value = 0.072

How to calculate P-values

The incomplete gamma and beta functions are useful for calculating limits:

$$\begin{aligned} \text{Incomplete gamma function} \quad P(a, x) &= \frac{1}{\Gamma(a)} \int_0^x e^{-t} t^{a-1} dt \quad (a > 0) \\ \text{Complement} \quad Q(a, x) &= 1 - P(a, x) \\ \text{Beta function} \quad B(a, b) &= \int_0^1 t^{a-1} (1-t)^{b-1} dt \\ \text{Incomplete beta function} \quad I_x(a, b) &= \frac{1}{B(a, b)} \int_0^x t^{a-1} (1-t)^{b-1} dt \quad (a, b > 0) \end{aligned}$$

Limit calculation requires to solve an equation with given integral (or sum):

$$\text{Poisson limits: } \sum_{j=k}^{\infty} P_{\mu}(j) = P(k, \mu)$$

$$\text{Gaussian limits: } \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt = P(1/2, x^2) \quad (x \geq 0)$$

$$\chi^2 \text{ limits: } \int_{\chi^2}^{\infty} f(\chi^2, n) d\chi^2 = 1 - P(n/2, \chi^2/2) = Q(n/2, \chi^2/2)$$

$$\text{Binomial limits: } \sum_{j=k}^n \binom{n}{j} p^j (1-p)^{n-j} = I_p(k, n-k+1)$$

Other Errors

”There is another class of errors which are not so easy to detect and for which statistical analysis is not generally useful.”

R. Bevington: Data Reduction and Analysis for the Physical Sciences

This statement is only half true.

Systematic uncertainties:

- due to uncertainties in the apparatus or model
- usually correlated with previous measurements
- limited theoretical framework
- examples are calibrations uncertainties, detector acceptance and poorly-known theoretical parameters.

... using talks by R. Barlow (SLUO and Durham 2002)

Other errors and correlated errors

Other errors:

systematic errors, normalization errors ...

The determination or estimation of systematic errors requires an expert knowledge of the measurement process and the physics behind.

”It requires ingenuity, experience, cunning and a fair degree of paranoia. These are well provided in particle physics. It also requires honesty and bravery, which are scarcer.”

R. Barlow

...and there are ”correlated” errors:

Correlations are **always** present, if parameters are determined in a common fit or analysis, and are described by the covariance matrix. Often correlations are neglected (by mistake).

How to write a thesis/paper?

- 1 Devise cuts, get results
- 2 Do analysis for random errors (using e.g. Poisson statistics)
- 3 Make big table, alter cuts by arbitrary amounts, put in table, alter cuts, put in table, alter cuts ... until time/money/supervisors patience is exhausted.
- 4 Add all the variations in quadrature.
- 5 Quote result as "systematic error, and call it "conservative"

This give the pusillanimous researcher a comfortably large error estimate to hide behind.

This practice combines evaluation of errors with checks for mistakes, in an inappropriate way.

There is no cause to include it as contribution to the systematic error, if there is no reason to suspect a systematic effect from a particular source, and if there is no numerical evidence for a significant variation.

Example for "systematic" error determination

trigger	$\pm 5 \%$
hadronic energy scale	$\pm 4 \%$
electron energy scale	$\pm 1 \%$
electron scattering angle ± 2 mrad	$\pm 2 \%$
electron energy cut	$\pm 6 \%$
lepton recognition	$\pm 12 \%$
track efficiency of 98 %	$\pm 3 \%$
z -vertex uncertainty	$\pm 2 \%$
fiducial cut	$\pm 4 \%$
structure function	$\pm 3 \%$
radiative corrections	$\pm 4 \%$
non-resonant background	$\pm 4 \%$
branching ratio	$\pm 3.2 \%$
fitting procedure	$\pm 3 \%$
bin-center correction	$\pm 2 \%$
Luminosity	$\pm 1.5 \%$
total systematic error	18%

Why is an uncertainty called "error"? – A proposal

The result of a measurement has to be accompanied by a quantitative statement of its **uncertainty**. Each component of uncertainty is represented by an estimated standard deviation, termed **standard uncertainty** ($= +\sqrt{V}$ = square-root of the variance V).

The **uncertainty** of the result of a measurement generally consists of several components which may be grouped into two categories according to the **method** used to estimate their numerical values:

- A. those which are evaluated by **statistical methods**,
- B. those which are evaluated by **other means**.

A Type B evaluation is usually based on scientific judgment using all the relevant information available, which may include previous measurement data, experience with the behavior and property of relevant materials and instruments, calibration data.

Terminology: Error (of measurement) = result of a measurement minus the value of the measurand; in general this is unknown.

Guidelines for Evaluating and Expressing the Uncertainty of Measurement Results, Technical Note 1297, by B.N. Taylor and C.E. Kuyatt, National Institute of Standards and Technology (NIST), USA

Correlations

Consider the formula for the transverse momentum (in GeV/c)

$$p_T = 0.3 \cdot B \cdot R$$

where

0.3 is a constant (speed of light $\times 10^{-9}$);

R is radius (in meter) of curvature from a track; for each track measured with some accuracy;

B is magnetic field (in Tesla); measured *once* with finite accuracy, but influencing every track.

No problem for one track: e.g. transverse momentum p_T has error 5 %, if we know R to 4 % and B to 3 %. For two tracks the transverse momenta are correlated and this has to be taken into account, if anything is done with the two tracks (e.g. forming a “mass”).

General law of approximate ”error” propagation for $\mathbf{x} \rightarrow \mathbf{y}$:

$$\mathbf{V}_y = \mathbf{A} \mathbf{V}_x \mathbf{A}^T$$

Transformation of errors – one track

Measured (2-dim) and transformed quantity (1-dim):

$$\begin{aligned} \text{Measured: } \mathbf{x} &= \begin{pmatrix} R_1 \\ B \end{pmatrix} & \mathbf{V}_x &= \begin{pmatrix} \sigma_{R_1}^2 & 0 \\ 0 & \sigma_B^2 \end{pmatrix} \\ \text{Transformed quantity: } y &= p_{T1} = 0.3BR_1 \end{aligned}$$

Resulting covariance "matrix"

$$\begin{aligned} \mathbf{V}_y &= \mathbf{A}\mathbf{V}_x\mathbf{A}^T \quad \text{with } \mathbf{A} = (0.3B \quad 0.3R_1) \\ &= (0.3B\sigma_{R_1})^2 + (0.3R_1\sigma_B)^2 \end{aligned}$$

Transformation of "errors" – two tracks

$$\text{Measured: } \mathbf{x} = \begin{pmatrix} R_1 \\ R_2 \\ B \end{pmatrix} \quad \mathbf{V}_x = \begin{pmatrix} \sigma_{R_1}^2 & 0 & 0 \\ 0 & \sigma_{R_2}^2 & 0 \\ 0 & 0 & \sigma_B^2 \end{pmatrix}$$

$$\text{Transformed quantity: } \mathbf{y} = \begin{pmatrix} p_{T1} \\ p_{T2} \end{pmatrix} = \begin{pmatrix} 0.3BR_1 \\ 0.3BR_2 \end{pmatrix}$$

Resulting covariance matrix

$$\begin{aligned} \mathbf{V}_y &= \mathbf{A}\mathbf{V}_x\mathbf{A}^T \quad \text{with } \mathbf{A} = \begin{pmatrix} 0.3B & 0 & 0.3R_1 \\ 0 & 0.3B & 0.3R_2 \end{pmatrix} \\ &= \begin{pmatrix} (0.3B\sigma_{R_1})^2 & 0 \\ 0 & (0.3B\sigma_{R_2})^2 \end{pmatrix} + \begin{pmatrix} (0.3R_1\sigma_B)^2 & (0.3\sigma_B)^2 R_1 R_2 \\ (0.3\sigma_B)^2 R_1 R_2 & (0.3R_2\sigma_B)^2 \end{pmatrix} \\ &= \begin{pmatrix} (0.3B\sigma_{R_1})^2 + (0.3R_1\sigma_B)^2 & (0.3\sigma_B)^2 R_1 R_2 \\ (0.3\sigma_B)^2 R_1 R_2 & (0.3B\sigma_{R_2})^2 + (0.3R_2\sigma_B)^2 \end{pmatrix} \end{aligned}$$

Combining random and systematic errors

Statistical and systematic errors are assumed to be independent (by definition): add in quadrature.

Consider two measurements x_1 and x_2 with

- individual (independent) random errors σ_1 and σ_2 , and
- common systematic error S .

Joint covariance matrix:

$$\mathbf{V} = \begin{pmatrix} \sigma_1^2 + S^2 & S^2 \\ S^2 & \sigma_2^2 + S^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} + \begin{pmatrix} S^2 & S^2 \\ S^2 & S^2 \end{pmatrix}$$

The "estimation" of a combined covariance matrix may be dangerous: a mistake will produce *unexpected* (and wrong) results.

Three measurements

Consider three measurements x_1 , x_2 and x_3 with

- individual independent random errors σ_1 , σ_2 and σ_3 , and
- common systematic error S .
- common systematic error T , shared by x_2 and x_3 , but not x_1 .

$$\begin{aligned} \text{cov. matrix } \mathbf{V} &= \begin{pmatrix} \sigma_1^2 + S^2 & S^2 & S^2 \\ S^2 & \sigma_2^2 + S^2 + T^2 & S^2 + T^2 \\ S^2 & S^2 + T^2 & \sigma_3^2 + S^2 + T^2 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix} + \begin{pmatrix} S^2 & S^2 & S^2 \\ S^2 & S^2 & S^2 \\ S^2 & S^2 & S^2 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & T^2 & T^2 \\ 0 & T^2 & T^2 \end{pmatrix} \end{aligned}$$

Example: effect of multiple scattering in material on space points measured in a track detector.

What is the "nature" of the error (or deviation)?

- B is uniform and constant, and was measured with a flux meter good to 1%.
- B is uniform and constant. It was measured with a highly accurate flux meter. However changes in geometry since then (e.g. inserting a drift chamber, opening and closing doors) will have changed that field by an amount we estimate at 1%.
- B is uniform and well measured, however it varies with time (temperature effects on power supplies) at the 1 % level.
- B is constant but not uniform: different tracks see different fields, but still obey $p_T = 0.3BR$.
- B is constant but not uniform: the same track sees different fields along it path, and $p_T = 0.3BR$ is inadequate.
- B is constant, uniform and measured accurately, but a wrong number has been typed in
- tracks loose momentum in the beam pipe so all the measurements are low
- tracks loose momentum in the drift chamber so they are not pure circular helices
- the drift chamber is badly aligned
- ...

Systematic errors

Distinguish between

- errors from manifest sources (e.g. uncertain calibration applied to all energy measurements)
→ determine error contribution
- unsuspected errors (e.g. change in detector performance with external conditions)
→ check for systematic effects, but usually no error contribution.

”Statistical” methods can often be used to determine ”systematic errors”.

For errors added in quadrature the final error is usually dominated by one or two contributions → try to reduce the large ones.

Analysis is most effective at the planning stage of an experiment.

Manifest systematic errors

These are factors with known errors, used explicitly in the data analysis, e.g.

- conversion factors and associated errors from calibration run,
- predicted background contribution depending on parameters used in the MC simulation, ...

Standard procedure is to repeat the analysis with e.g. modified constants, establishing the dependence of the final result on the constant.

Special case: no number quantifies the assumption, eg. different background predicted by different models (HERWIG, JETSET, ...).

Estimation method 1: take several assumptions and take standard deviation of correction factor as standard deviation of systematic error (for two values: $\sigma \approx \Delta/\sqrt{2}$; $\sigma = 0.707\Delta$).

Estimation method 2: take values under **extreme** assumptions, and assume $\sigma = \Delta$, divided by $\sqrt{12}$ (uniform dist.); $\sigma = 0.289\Delta$.

Unsuspected systematic errors

e.g. failure of an electronic component, new source of noise, program bug, ...

Make checks for absence of systematic effects, i.e. repeat analysis with minor differences and calculate shifts of parameters.

This is **not** the measurement of an effect, but a **check** whether an effect exists, with the hope to find no effect. Do nothing if shifts of parameters are "small" (see later).

A large shift after a minor modification would indicate the presence of an unsuspected systematic effect, which has to be treated seriously (e.g. influence of TGV on LEP energy). If it can not be *explained* it is folded into the systematic error as a last resort.

Check for systematic effects

Systematic errors which are not anticipated

Standard and sensible practice is to repeat the analysis in different forms, for example

- varying the range of the data used to extract the result,
- varying cuts applied to ensure the purity and quality of the data,
- including/excluding subsets of data taken under different experimental conditions,
- using histograms with different bin sizes,
- determining quantities by simple counting and by fits of a parametrized curve,

i.e. repeat the analysis with minor differences.

Values obtained by different methods will disagree, even in the absence of systematic effects. If the size of such deviations may occur due to statistical fluctuations → do nothing.

Comparison of two methods of estimation - significance

Suppose the parameter a is estimated from **one** data sample by two different methods:

- estimate \hat{a}_1 with statistical error σ_1 (method I)
- estimate \hat{a}_2 with statistical error σ_2 (method II)

The difference Δ between the estimates and its variance is

$$\Delta = \hat{a}_1 - \hat{a}_2 \quad \sigma_{\Delta}^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$$

$$\rho = \text{correlation coefficient with } -1 \leq \rho \leq +1$$

In general the correlation ρ between the two estimates will be positive (estimates share some of the same data). For $\rho = +1$

$$\sigma_{\Delta}^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2 = (\sigma_1 - \sigma_2)^2$$

Is the difference Δ large? Determine correlation coefficient ρ and test value of Δ/σ_{Δ} .

Example: mean of subset of the data

Estimate a_1 is the mean value of the total number N_T of data:

$$a_1 = \frac{1}{N_T} \sum_{N_T} x_i \quad \sigma_1 = \frac{\sigma}{\sqrt{N_T}}$$

Check: using a subset of the data has the result $a_2 \pm \sigma_2$

$$a_2 = \frac{1}{N_S} \sum_{N_S} x_i \quad \sigma_2 = \frac{\sigma}{\sqrt{N_S}}$$

Is the difference $\Delta = a_1 - a_2$ small?

The relations

$$\text{cov}(a_1, a_2) = N_S \frac{\sigma}{N_T} \frac{\sigma}{N_S} \quad \rightarrow \quad \rho = \sigma_1 / \sigma_2$$

have the result

$$\sigma_{\Delta}^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2 = \sigma_2^2 - \sigma_1^2 \quad (\text{i. e. subtract in quadrature})$$

General limits on the correlation coefficient

Calculate weighted mean of two *correlated* estimates:

$$\hat{a} = w\hat{a}_1 + (1 - w)\hat{a}_2 \quad \sigma^2 = w^2\sigma_1^2 + (1 - w)^2\sigma_2^2 + w(1 - w)\rho\sigma_1\sigma_2$$

Minimum error σ is obtained from

$$\partial\sigma^2/\partial w = 0 \rightarrow w = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

$$\sigma^2 = \frac{\sigma_1^2\sigma_2^2(1 - \rho^2)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

From statistical theory there is a **Minimum Variance Bound (MVB)** σ_0^2 (see example).

For $\sigma^2 \geq \sigma_0^2$ there is an expression for ρ which is satisfied only between the two roots of the quadratic equation:

$$\rho = \frac{\sigma_0^2 \pm \sqrt{(\sigma_1^2 - \sigma_0^2)(\sigma_2^2 - \sigma_0^2)}}{\sigma_1\sigma_2}$$

Allowed range: $\sigma_{\Delta}^{\min} \leq \sigma_{\Delta} \leq \sigma_{\Delta}^{\max}$

$$\begin{aligned}\sigma_{\Delta}^{\max} &= \sqrt{(\sigma_1^2 - \sigma_0^2)} + \sqrt{(\sigma_2^2 - \sigma_0^2)} && \text{(from minimum value of } \rho) \\ \sigma_{\Delta}^{\min} &= \left| \sqrt{(\sigma_1^2 - \sigma_0^2)} - \sqrt{(\sigma_2^2 - \sigma_0^2)} \right| && \text{(from maximum value of } \rho)\end{aligned}$$

Range of ρ often narrow; if one estimator highly efficient, then MVB saturated and

$$\sigma_{\Delta} = \sqrt{|\sigma_1^2 - \sigma_2^2|}$$

Example: angular distribution of μ pairs in Z decays

The (normalized) probability distribution is assumed to be

$$P(x) = \frac{3}{8} (1 + x^2) + ax \quad \text{with } x \simeq \cos \vartheta$$

where a is the asymmetry between "forward" and "backward" events. With $N = 33\,464$ events the asymmetry a is determined by two methods.

- $a_1 = 0.0123 \pm 0.0055$ by counting of forward and backward events,
- $a_2 = 0.0084 \pm 0.0051$ by fitting the curve.

Does the difference of $\Delta = 0.0039$ indicate a systematic effect?

From the Minimum Variance Bound (**MVB**):

$$\frac{1}{\sigma_0^2} = N \int_{-1}^{+1} P(x) \left(\frac{d \ln P}{da} \right)^2 dx = \frac{8N}{3} \int_{-1}^{+1} \frac{x^2}{1 + x^2 + 8Ax/3} dx \approx \frac{8N}{3} \frac{4 - \pi}{2}$$

with result $\sigma_0 = 0.0051$ (equal to value σ_2 from fit).

$$\sigma_\Delta = \sqrt{\sigma_1^2 - \sigma_2^2} = 0.0021 \quad |\Delta/\sigma_\Delta| = 0.0039/0.0021 < 2$$

Statistical and other errors

Statistical errors	2
Limits and Confidence intervals	3
P-Values	5
Coverage	6
Poisson distributed data	7
Pearson's χ^2 intervals	8
Neymans's modified χ^2 intervals	10
Likelihood intervals	11
Classical-Frequentists central intervals	12
P-Value and the Poisson distribution	13
How to calculate P-values	14
Other Errors	15
Other errors and correlated errors	16
Common practice	17
Example for "systematic" error determination	18
Why is an uncertainty called "error"? – A proposal	19
Correlations	20
Transformation of errors – one track	21
Transformation of "errors" – two tracks	22
Combining random and systematic errors	23
Three measurements	24
What is the "nature" of the error (or deviation)?	25
Systematic errors	26
Manifest systematic errors	27
Unsuspected systematic errors	28
Check for systematic effects	29
Comparison of two methods of estimation - significance	30
Example mean of subset of the data	31
General limits on the correlation coefficient	32
Example: angular distribution of μ pairs in Z decays	34