Statistical Inference Problems in High Energy Physics (06w5054)
Banff International Research Station July 15 – July 20, 2006

Dealing with systematics for chi-square and for log likelihood goodness of fit statistics

Volker Blobel – Universität Hamburg

---

1. Global analysis of data from HEP experiments
2. Normalization errors (muliplicative)
3. Additive systematic errors
4. Correlated statistical errors
5. Parameter uncertainties: Profile likelihods and χ^2
6. Outliers and their influence on the fit

---

1. Global analysis of data from HEP experiments

Systematic errors are at the origin of the unsatisfactory situation, when data from many experiments are used by theoreticians in a global analysis and parameter estimation, and when attempts are made to determine e.g. uncertainties of predictions from parton distribution functions.

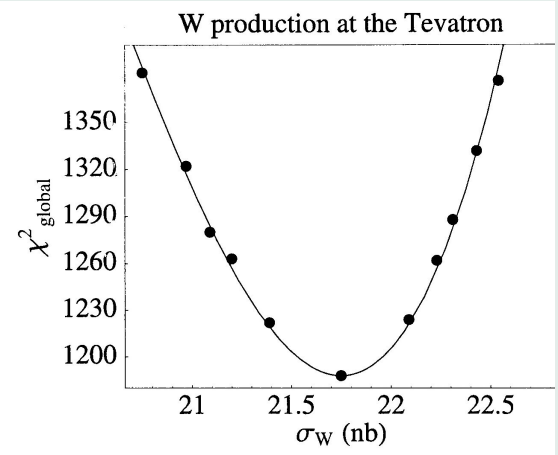
- It is difficult to estimate systematic errors correctly, and the estimated errors are only as good as the model used for systematic errors.
- It is difficult to construct a correct chisquare expression for the estimation of parameters, taking into account the systematic contributions.
- The construction of the chisquare expression is often done incorrectly and, if published in a respected journal, it is often the model for the next generation of scientists.
- There is a tendency to allow extremely large $\Delta\chi^2$ in an error analysis.

The situation is e.g. described in the paper: D. Stump et al., Phys. Rev. **D65**, 014012

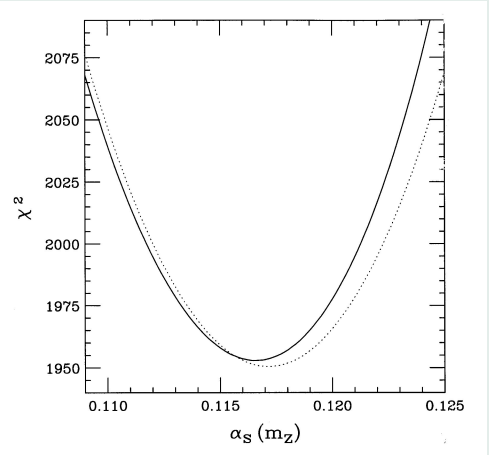
Examples with large $\Delta\chi^2$

The large, artificial and arbitrary magnification of errors is hardly acceptable – the procedure points to a deep problem in the whole data analysis. Two examples from parton distribution fits:

W production at the Tevatron



$\alpha_S(M_Z^2)$



Both curves are parabolas to a very good approximation over a range of $\Delta\chi^2 > 100$ ...

... while usually one would consider only a range of $\Delta\chi^2 \approx 4$, corresponding to two standard deviations.

" ... determine the increase in χ^2_{global} that corresponds to our estimated uncertainty $\Delta\sigma^W$ in the σ^W prediction. ... corresponds to $\Delta\chi^2_{\text{global}} \approx 180$."

Data uncertainties in HEP experiments

Statistical errors: The statistical error are given by the Poisson statistics of event numbers ($n \pm \sqrt{n}$), without correlations between bins ...
... but numbers corrected for finite resolution are correlated $\Longrightarrow$ Correlated statistical errors.

Normalization error: There is an uncertainty in the factor used to convert *event numbers* to *cross sections*. This so-called normalization error applies to cross sections and to statistical errors of cross sections as well. Because of its origin – product of many factors, each with some uncertainty – the factor perhaps follows a log-normal distribution due the multiplicative central limit theorem.

Systematic errors: There are uncertainties in the detector behaviour, e.g. energy measurements by a calorimeter may have a general relative error of a few %. Often the experiments analysis is repeated with $\pm$ few % relative change of e.g. calorimeter data; from the change in the result, error contributions for *all* data are related to the single systematic uncertainty. A single error contribution is a rank-1 contribution to the covariance matrix of the data.

Correlated statistical errors, unfolding error: The smearing effect of the detector is usually corrected for in a way, which, due to the *smoothing* aspect, introduces at least *positive* correlations between neightboring points.

Several different error contributions are reported by the experiments, but some may be missing; the *unfolding error* (with positive correlations) usually remains unpublished.

2. Normalization errors (multiplicative)

Normalization error: There is an uncertainty in the factor used to convert *event numbers* to *cross sections*. This so-called normalization error applies to cross sections and to statistical errors of cross sections as well. Because of its origin – product of many factors, each with some uncertainty – the factor perhaps follows a log-normal distribution due the multiplicative central limit theorem.

From a paper: "... Then, in addition, the fully correlated normalization error of the experiment is usually specified separately. For this reason, it is naturally to adopt the following definition for the effective χ^2 (as done in previous ... analyses)

$$\chi^2_{\text{global}} = \sum_n w_n \chi^2_n(a) \qquad (n \quad \text{labels the different experiments}),$$

$$\chi^2_n(a) = \left(\frac{1-\mathcal{N}_n}{\sigma^N_n}\right)^2 + \sum_\ell \left(\frac{\boxed{\mathcal{N}_n D_{n\ell}} - T_{n\ell}(a)}{\sigma^D_{n\ell}}\right)^2$$

For the nth experiment, $D_{n\ell}$, $\sigma^D_{n\ell}$, and $T_{n\ell}(a)$ denote the data value, measurement uncertainty (*statistical and systematic combined*) and theoretical value (dependent on $\{a\}$) for the ℓth data point, σ^N_n is the experimental normalization uncertainty, and $\mathcal{N}_n$ is an overall normalization factor (with default value 1) for the data of experiment n. ..."

The nuisance parameter $\mathcal{N}_n$ should be applied as a factor to $T_{n\ell}(a)$, instead of $D_{n\ell}$, or alternatively to both $D_{n\ell}$ and $\sigma^D_{n\ell}$; otherwise a normalization bias is introduced.

"Data are $y_1 = 8.0 \pm 2\%$ and $y_2 = 8.5 \pm 2\%$, with a common (relative) normalisation error of $\varepsilon = 10\%$. The mean value (constraint $\widehat{y}_1 = \widehat{y}_2$) resulting from a χ^2 minimisation of $\chi^2 = \boldsymbol{\Delta}^T \boldsymbol{V}^{-1} \boldsymbol{\Delta}$ is:

$$\overline{y} = 7.87 \pm 0.81 \qquad \text{i.e. } < y_1 \text{ and } < y_2 \qquad \boldsymbol{\Delta} = \begin{pmatrix} y_1 - \overline{y} \\ y_2 - \overline{y} \end{pmatrix}$$

– this is apparently wrong.
The method used was to define a full covariance matrix for the correlated data by

$$\boldsymbol{V}_a = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} + \varepsilon^2 \cdot \begin{pmatrix} y_1^2 & y_1 y_2 \\ y_1 y_2 & y_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 + \varepsilon^2 y_1^2 & \varepsilon^2 y_1 y_2 \\ \varepsilon^2 y_1 y_2 & \sigma_2^2 + \varepsilon^2 y_2^2 \end{pmatrix}$$

Conclusion in the paper:

> "... that including normalisation errors in the correlation matrix will produce a fit which is biased towards smaller values ... the effect is a direct consequence of the hypothesis to estimate the empirical covariance matrix, namely the linearisation on which the usual error propagation relies."

But the matrix $\boldsymbol{V}_a$ is wrong! Correct model: the normalisation errors $\varepsilon \cdot$ value are identical

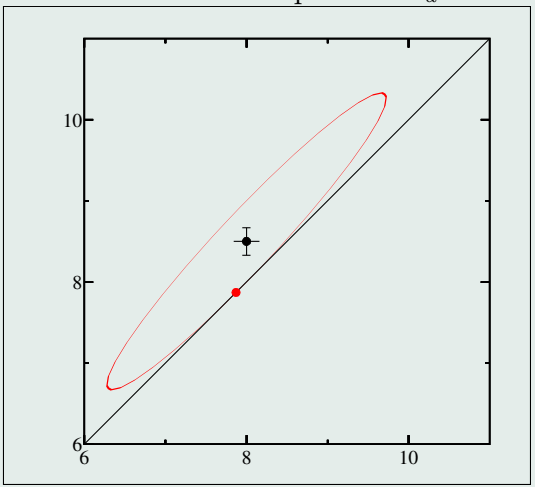
$$\boldsymbol{V}_b = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} + \varepsilon^2 \cdot \begin{pmatrix} \overline{y}^2 & \overline{y}^2 \\ \overline{y}^2 & \overline{y}^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 + \varepsilon^2 \overline{y}^2 & \varepsilon^2 \overline{y}^2 \\ \varepsilon^2 \overline{y} & \sigma_2^2 + \varepsilon^2 \overline{y}^2 \end{pmatrix}$$

will give the correct result with $y_1 < \overline{y} < y_2$.

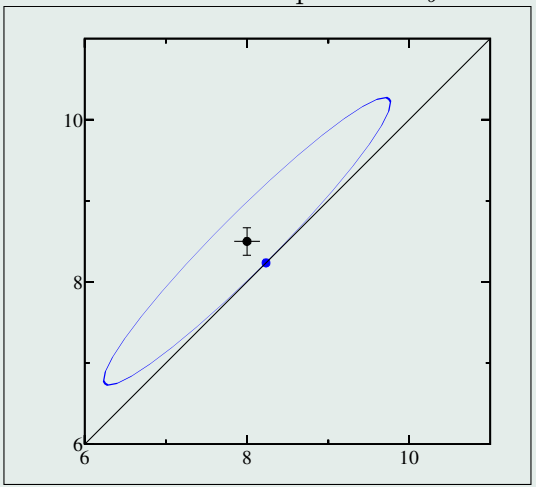
Covariance matrix comparison

Plot of one measured value vs. the other measured value, with the assumed covariance ellipse; the mean value is on the diagonal.

Covariance ellipse for $\boldsymbol{V}_a$



Covariance ellipse for $\boldsymbol{V}_b$



Axis of ellipse is tilted w.r.t. the diagonal and ellipse touches the diagonal at a biased point.

Axis of the ellipse is $\approx 45°$ and ellipse touches the diagonal at the correct point.

The result of χ^2 minimisation may depend critically on details of the model implementation!

The method with a nuisance parameter ...

Another method often used is to define

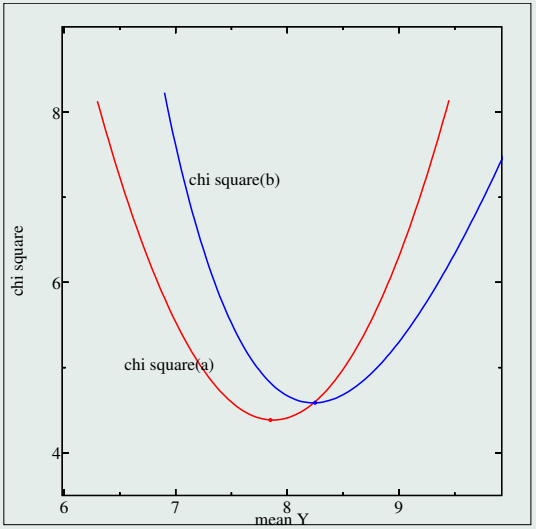
$$\chi^2_a = \sum_k \frac{(f \cdot y_k - \overline{y})^2}{\sigma_k^2} + \frac{(f-1)^2}{\varepsilon^2} \,,$$

which will again produce a biased result.

The χ^2 definition for this problem

$$\chi^2_b = \sum_k \frac{(y_k - f \cdot \overline{y})^2}{\sigma_k^2} + \frac{(f-1)^2}{\varepsilon^2}$$

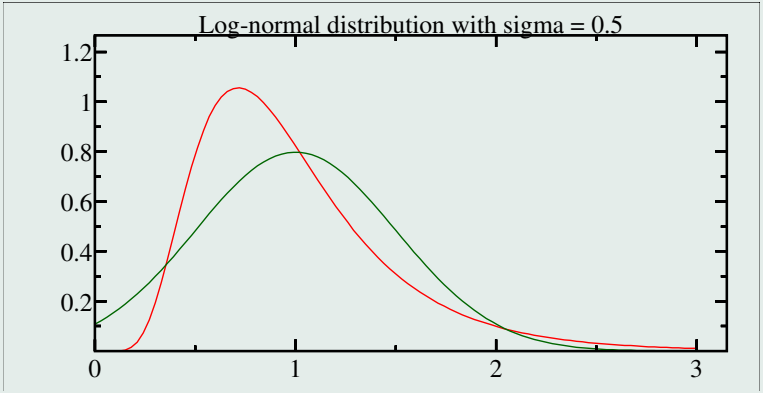
will give the correct result (data unchanged and fitted value according to the model), as seen by blue curve.



The normalisation factor determined in an experiment is more the product (luminosity, detector acceptance, efficiency) than the sum of random variables. According to the *multiplicative* central limit theorem the product of positive random variables follows the log-normal distribution, i.e. the logarithm of the normalisation factor follows the normal distribution.

For a log-normal distribution of a random variable α with $E[\alpha] = 1$ and standard deviation of ε the contribution to the χ^2-function $S(\boldsymbol{a}, \alpha)$ is

$$\begin{aligned} \Delta S^{\text{norm}} &= \ln\alpha \left(3 + \frac{\ln\alpha}{\ln\left(1+\varepsilon^2\right)}\right) \\ &\rightarrow \frac{(\alpha-1)^2}{\varepsilon^2} \quad \text{for small } \varepsilon \end{aligned}$$



The normal and the log-normal distribution, both with mean 1 and standard deviation $\varepsilon = 0.5$.

$$\text{log-normal density}(\alpha) = \exp\left[-\tfrac{1}{2}\ln\alpha\left(3 + \tfrac{\ln\alpha}{\ln(1+\varepsilon^2)}\right) + \tfrac{\varepsilon^2}{8} - \ln\varepsilon - \ln\sqrt{2\pi}\right]$$

Aymmetric normalisation errors

Proposed method, to take the normalisation error ε into account, if data from > 1 experiment are combined:

Introduce one additional factor α for each experiment as nuisance parameter, which has been measured to be $\alpha = 1 \pm \varepsilon$, modify the expectation according to

$$f_i = \alpha \cdot f(x_i, \boldsymbol{a})$$

and make fit with

$$S(\boldsymbol{a}) = \sum_i \frac{(y_i - \alpha \cdot f(x_i, \boldsymbol{a}))^2}{\sigma_i^2} + \Delta S^{\text{norm}} \qquad \text{with} \quad \Delta S^{\text{norm}} = \frac{(\alpha - 1)^2}{\varepsilon^2}$$

$$\text{or} \quad \Delta S^{\text{norm}} = \ln \alpha \left(3 + \frac{\ln \alpha}{\ln (1 + \varepsilon^2)} \right) \qquad \text{lognormal distribution}$$

3. Additive systematic errors

Systematic errors: There are uncertainties in the detector behaviour, e.g. energy measurements by a calorimeter may have a general relative error of a few %. Often the experiments analysis is repeated with ± few % relative change of e.g. calorimeter data; from the change in the result, error contributions for *all* data are related to the single systematic uncertainty. A single error contribution is a rank-1 contribution to the covariance matrix of the data.

Experimental method: e.g. run MC for each systematic (UNISIM) with constant varied by 1 σ and redetermine result – determine signed **shifts** s_i of data values y_i.

1. Method: Modify covariance matrix to include contribution(s) due to systematic errors

$$\boldsymbol{V}_a = \boldsymbol{V}_{\text{stat}} + \boldsymbol{V}_{\text{syst}} \qquad \text{with} \quad \boldsymbol{V}_{\text{syst}} = \boldsymbol{s}\boldsymbol{s}^T \quad \text{(rank-1 matrix)} \qquad \text{or} \sum_k \boldsymbol{s}_k \boldsymbol{s}_k^T$$

and use inverse matrix $\boldsymbol{V}^{-1}$ as weight matrix in the χ^2 function. (→ simplified calculation of inverse)

2. Method (recommended): Introduce one nuisance parameter β, assumed to be measured as 0 ± 1, for each systematic error source, and make fit with

$$S(\boldsymbol{a}) = \sum_i \frac{\left(y_i + \beta s_i - \alpha \cdot f(x_i, \boldsymbol{a})\right)^2}{\sigma_i^2} \; + \; \beta^2$$

Other method: MULTISIM: vary all systematic parameters randomly using their assumed probability distribution and redetermine result.

Simplified calculation of inverse

Assume that the inverse $\boldsymbol{A}^{-1}$ of the n-by-n matrix $\boldsymbol{A}$ is known. If a small change in $\boldsymbol{A}$ is done in one of the two forms below, the corresponding change in $\boldsymbol{A}^{-1}$ is calculated faster by the formulas below.

Sherman-Morrison Formula: $\boldsymbol{u}$ and $\boldsymbol{v}$ are n-vectors and the change $\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ of $\boldsymbol{A}$ is of rank 1.

$$\left(\boldsymbol{A}+\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\right)^{-1}=\boldsymbol{A}^{-1}-\frac{1}{1+\lambda}\left(\boldsymbol{A}^{-1}\boldsymbol{u}\right)\left(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\right)\qquad\text{with}\quad\lambda=\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}$$

Woodbury Formula: $\boldsymbol{U}$ and $\boldsymbol{V}$ are n-by-k matrices with $k<n$ and usually $k\ll n$.

$$\left(\boldsymbol{A}+\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\right)^{-1}=\boldsymbol{A}^{-1}-\left[\boldsymbol{A}^{-1}\boldsymbol{U}\left(\mathbf{1}+\boldsymbol{V}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{U}\right)^{-1}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{A}^{-1}\right]$$

Only a k-by-k matrix has to be inverted.

Inversion of a 100-by-100 matrix takes a few 10^{-3} seconds.

W.H.Press et al., Numerical Recipes, The Art of Scientific Computing, Cambridge University Press

"For larger k the direct methods may be faster and more accurate because of the stabilizing advantages of pivoting."

Example from an experimental paper

Systematic uncertainties:

- "Correlated systematic uncertainties:" e.g. $-0.2\%\ -0.7\%\ +1.0\%\ +0.9\%\ -0.6\% \ldots \sum = 1.6\%$
  - electron energy
  - electron angle
  - hadronic calibration
  - calorimeter noise contribution
  - photoproduction background
- "Uncorrelated systematic uncertainties:" e.g. $0.4\%\ \ 0.8\%\ \ldots \sum = 2.1\%$
  - Monte carlo statistic
  - trigger efficiency
  - detector efficiency
  - radiative corrections
- "Total cross section uncertainty" e.g. $2 - 3\%$

Shown are typical (small) values for error contributions.

Example from analysis papers

"All the experiment included in our analysis provide full correlated systematics, as well as normalization errors. The covariance matrix can be computed from these as

$$\mathrm{cov}_{ij} = \left(\sum_{k=1}^{N_{\mathrm{sys}}} \sigma_{i,k}\sigma_{j,k} + F_i F_j \sigma_N^2 \right) + \delta_{ij}\sigma_{i,t}^2 \,;$$

where F_i, F_j are central experimental values, $\sigma_{i,k}$ are the N_{sys} correlated systematics, σ_N is the total normalization uncertainty, and the uncorrelated uncertainty $\sigma_{i,t}$ is the sum of the statistical uncertainty $\sigma_{i,s}$ and the N_u uncorrelated systematic uncertainties (when present) . . . "

The inverse of the covariance matrix cov above is used as a weight matrix for a χ^2 calculation.

" . . . However . . . correlations between measurement errors, and correlated theoretical errors, are not included in its definition."

" . . . Instead, the evaluation of likelihoods and estimation of global uncertainty will be carried out . . . after sets of optimal sample PDF's for the physical variable of interest have been obtained."

Comments: The quadratic combination of statistical and systematic measurement uncertainty neglects the known correlation, inherent in the systematic effect. Neither unbiased optimal parameters values nor a usable χ^2 or parameter errors can be expected from the χ^2 function.

4. Correlated statistical errors

Statistical deviations per bin of measured quantities from event numbers $(n \pm \sqrt{n})$ are independent. The covariance matrix is diagonal.

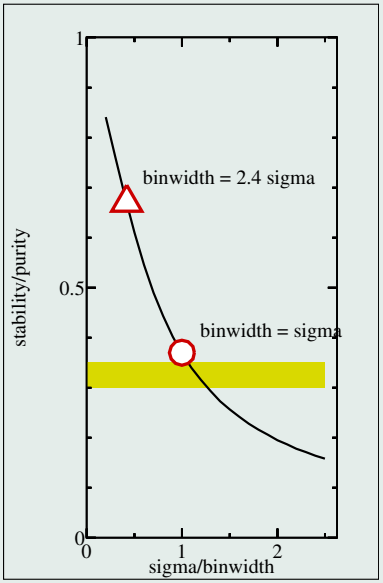
but

> "... the selected event samples are corrected for detector acceptance and migration using the simulation and are converted to bin-centred cross sections. ... The bins used in the measurement are required to have *stability* and *purity* larger than 30 %. ... The stability (purity) is defined as the number of simulated events which originate from a bin and which are reconstructed in it, divided by the number of generated (reconstructed) events in that bin ..."

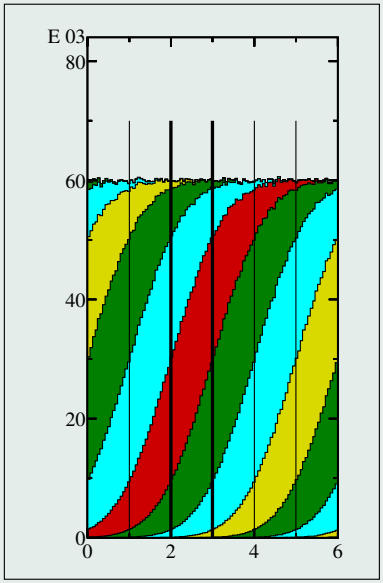
- A stability/purity of 30 % corresponds to a bin size narrower than 1 standard deviation of the measurement accuracy (for a flat distribution)!
- Up to 70 % of the events in a bin come from the $2 + 2$ neighbour bins.
- There must be a strong correlation of the corrected data in neighbour bins.
- The covariance matrix for corrected data is non-diagonal and the variances are magnified. This is visible in the eigenvalue spectrum of the migration matrix.

... the published covariance matrices of statistical errors are diagonal.

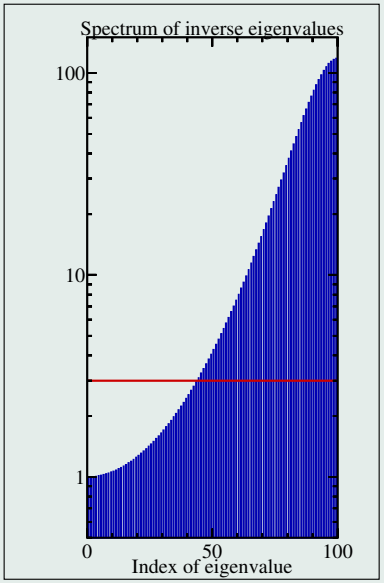
Why the cov matrix is non-diagonal ...



Assuming a uniform distribution, the stability/purity is plotted as a function of the ratio of σ to binwidth.



Consider bin $[2,3]$ with $\sigma =$ binwidth: the red contribution originated from this bin. $2+2$ neighbour bins contribute significantly to the bin $[2,3]$.



Assuming $\sigma =$ binwidth the eigenvalue spectrum is shown for a 100-bin histogram. The eigenvalue is a magnification factor for the variance of linear combinations.

5. Parameter uncertainties: profile likelihoods and χ^2

Covariance matrix: parameter uncertainties and correlations are given by the covariance matrix $\boldsymbol{V}$ of the fit, obtained by inversion of the matrix of second derivatives (Hessian) of the log-likelihood function (Fishers Information $\boldsymbol{I} = \boldsymbol{V}^{-1}$).
The covariance matrix is usually assumed to be sufficient to describe the parameter uncertainties.

χ^2 contour: the surface of the error ellipsoid corresponds to the area of $\chi^2_{\text{mininum}} + 1$.

Single-parameter uncertainty: the information is supplemented by the *profile likelihood*, obtained by minimizing, for many fixed values in the range $\pm$ several σ of one of the parameters, with respect to all other parameters. This is used e.g. in MINUIT in the MINOS option to check the uncertainties and evtl. define asymmetric errors.

Function of parameters: the uncertainty of functions $g(\boldsymbol{a})$ of the parameters is determined by the error propagation formula (derivatives of $g(\boldsymbol{a})$ w.r.t. the parameters and covariance matrix $\boldsymbol{V}$.

The information is supplemented by the *profile likelihood* for functions $g(\boldsymbol{a})$.

Profile likelihood for a function

The profile likelihood for a function $g(\boldsymbol{a})$ of parameters, e.g. the W production cross section σ^W at the Tevatron, or $\alpha_S(M_Z^2)$ can be calculated by the use of the Lagrange multiplier method (Lagrange 1736-1813):

For many fixed values $g_{\rm fix}$ of the function $g(\boldsymbol{a})$ the likelihood function is minimized. The standard method is to define a **Lagrange function**

$$\mathcal{L}(\boldsymbol{a}) = S(\boldsymbol{a}) + \lambda \cdot (g(\boldsymbol{a}) - g_{\rm fix})$$

and to find the stationary point w.r.t the parameters $\boldsymbol{a}$ and the Lagrange multiplier λ, given $g_{\rm fix}$. The constraint defines a set of parameter values for each value of $g_{\rm fix}$, e.g. σ_W.

An alternative method, used in a recent paper, is

" ...to assume, by trial-and-error, fixed values of the Lagrange multiplier λ and to minimize

$$S(\boldsymbol{a}) + \lambda \cdot g(\boldsymbol{a})$$

and after minimization, to calculate the corresponding fixed $g(\boldsymbol{a})$ (allows to use MINUIT)."

Comments to a recent paper

Global fit: The overall ("global") fit is done, taking into account the normalization errors, but neglecting certain systematic error contributions (non-diagonal) in the experimental data.
Why are the non-diagonal error contributions neglected, but later used?

Profile χ^2-function: In the determination of the profile χ^2-function however the normalization is fixed at the previously fitted value.
Fixing certain parameters will not result in a correct profile.

Single experiment analysis: Afterwards a χ^2-analysis is done, separately for each experiment, now taking into account the systematic error contributions (non-diagonal) in the experimental data.
For each experiment the "profile" χ^2-function is evaluated, however with the parameters from the global fit.
Each experiment has a different parameter covariance matrix, depending on the kinematical region and the accuracy. Evaluating the χ^2-function using parameter sets from the global fit will not result in a correct profile.

From papers ...

"Notice that the covariance matrix

$$V_{ij}^{p} = \langle \Delta_i \Delta_j \rangle = \Delta\chi^2 \cdot H_{ij}^{-1}$$

depends on the choice of $\Delta\chi^2$ which usually, but not always, is taken to be $\Delta\chi^2 = 1$. This choice ... corresponds to the definition of the width of a Gaussian distribution."

"In full global fit art in choosing correct" $\Delta\chi^2$ given complication of errors. Ideally $\Delta\chi^2 = 1$, but unrealistic."

" ... and $\Delta\chi^2$ is the allowed variation in χ^2. ... and a suitable choice of $\Delta\chi^2$ ... and $\Delta\chi^2$ is the allowed deterioration in fit quality for the error determination."

"... Our standard PDF set S_0 is a parametrized fit to 1295 data points with 16 fitting parameters. The minimum of χ^2_{global} is approximately 1200. Naively, it seems that an increase of χ^2_{global} by merely 1, say from 1200 to 1201, could not possibly represent a standard deviation of the fit. Naively one might suppose that a standard deviation would have $\Delta\chi^2 \sim \sqrt{1295}$ rather than 1. However this is an misconception. If the errors are uncorrelated (or if the correlations are incorporated into χ^2) then indeed $\Delta\chi^2 = 1$ would represent a standard deviation. But this theorem is irrelevant to our problem, because the large correlations of systematic errors are not taken into account in χ^2_{global}. ..."
(Phys. Rev.)

Use of redundant parameters

An example from parton density fits: the gluon parametrization is

$$xg(x, Q_0^2) = \quad \ldots \quad - A_- \, (1-x)^{\eta_-} \, x^{-\delta_-}$$

where $A_- \sim 0.2$, $\delta_- \sim 0.3$ and η_- fixed at ~ 10. A change of δ_- changes both shape *and* normalisation.

> "...we notice that a certain amount of redundancy in parameters leads to potentially disatrous departures ...For example, in the negative term in the gluon parameterization very small changes in the value of δ_- can be compensated almost exactly by a change in A_- and in the other gluon parameters ..."

> "We found our input parameterization was sufficiently flexible to accomodate data, and indeed there is a certain redundancy evident."

In the case of highly correlated, redundant parameters the Hessian will be (almost) singular, inversion may be impossible and the convergence of the fit is doubtful. Redundant parameters have to be avoided!

6. Outliers and their influence on the fit

> "Everyone believes in the normal law of errors, the experimenters because they think it is a mathematical theorem, the mathematicians because they think it is an experimental fact." [Poincaré]

Outliers – single unusual large or small values among a sample – are dangerous and will usually, because of their large influence, introduce a bias in the result:

- in the final χ^2 value,
- in the values of the fitted parameters, and
- in the parameter uncertainties.

A method for outlier treatment: **M-estimation**, closely related to the maximum-likelihood method. For data with a probability density pdf(z) the method of maximum-likelihood requires to minimize

$$S(\boldsymbol{a}) = -\sum_{i=1}^{n} \ln \mathrm{pdf}(z_i) = \sum_{i=1}^{n} \rho(z_i)$$

with $\rho(z) = \ln \mathrm{pdf}(z)$. For a Gaussian distribution $\rho(z) = \frac{1}{2}z^2$. The function $\rho(z)$ is modified in M-estimation by down-weighting.

M-estimates

Generalization of least-squares, following from Maximum Likelihood arguments.

$$\text{Abbreviation} \quad z_i = \frac{y_i - f(x_i;a)}{\sigma_i} \qquad (\sim N(0,1) \quad \text{for Gaussian measurement})$$

$$\text{Least-squares:} \quad \text{minimize} \sum_i \frac{1}{2} z_i^2 \quad \text{solve} \sum_i \frac{y_i - f(x_i;a)}{\sigma_i^2} \frac{\partial f}{\partial a_j} = 0 \quad j = 1, 2 \ldots p$$

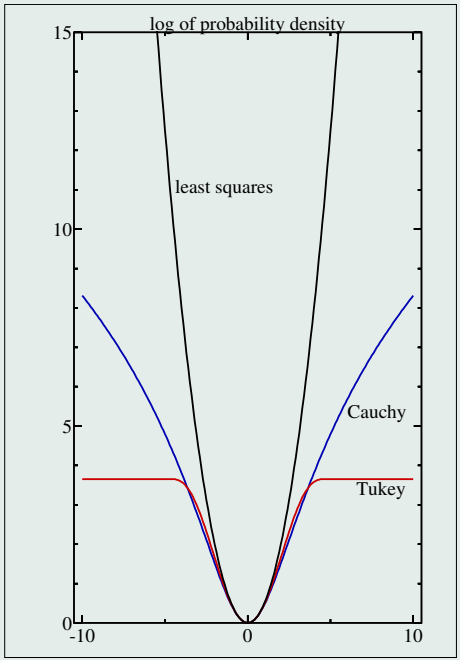
$$\text{M-estimates:} \quad \text{minimize} \sum_i \rho(z_i) \quad \text{solve} \sum_i \frac{y_i - f(x_i;a)}{\sigma_i^2} w(z_i) \frac{\partial f}{\partial a_j} = 0 \quad j = 1, 2 \ldots p$$

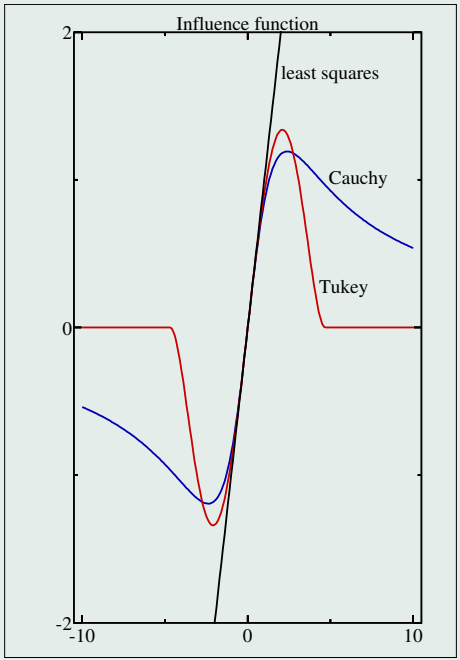
with influence function $\psi(z) = \dfrac{\mathrm{d}\rho}{\mathrm{d}z}$ and with additional weight $\boxed{w(z) = \psi(z)/z}$

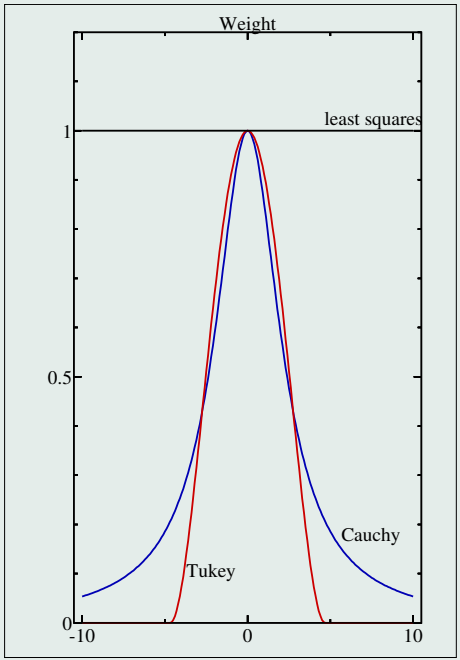
$$\rho(z) = \frac{1}{2} z^2 \qquad \psi(z) = z \qquad w(z) = 1 \qquad \text{Case of least-squares}$$

Requires iteration (non-linearity(!)) e.g with weight $w(z)$, calculated from previous values.

$$z \equiv \left(\frac{y - f(x)}{\sigma} \right)^2$$







$\rho(z)$ $\psi(z) = \mathrm{d}\rho(z)/\mathrm{d}z$ $w(z) = \psi(z)/z$

Commonly used M-estimators

	$\rho(z) = \ln \mathrm{pdf}(z)$	influence function $\psi(z) = \mathrm{d}\rho(z)/\mathrm{d}z$	weight $w(z) = \psi(z)/z$
Least squares	$= \frac{1}{2} z^2$	$= z$	$= 1$
Cauchy	$= \frac{c^2}{2} \ln\left(a + (z/c)^2\right)$	$= \frac{z}{1 + (z/c)^2}$	$= \frac{1}{1 + (z/c)^2}$
Tukey $\begin{cases} \text{if } \lvert z\rvert \le c \\ \text{if } \lvert z\rvert > c \end{cases}$	$= \begin{cases} c^2/6 \left(1 - [1 - (z/c)^2]^3\right) \\ c^2/6 \end{cases}$	$= \begin{cases} z\,[1 - (z/c)^2]^2 \\ 0 \end{cases}$	$= \begin{cases} [1 - (z/c)^2]^2 \\ 0 \end{cases}$
Huber $\begin{cases} \text{if } \lvert z\rvert \le c \\ \text{if } \lvert z\rvert > c \end{cases}$	$= \begin{cases} z^2/2 \\ c\,(\lvert z\rvert - c/2) \end{cases}$	$= \begin{cases} z \\ c \cdot \mathrm{sign}\,(z) \end{cases}$	$= \begin{cases} 1 \\ c/\lvert z\rvert \end{cases}$

Asymptotic efficiency of 95 % on the normal distribution obtained with $c = 2.3849$ (Cauchy), $c = 4.6851$ (Tukey) and $c = 1.345$ (Huber).

M-estimation: reduces the influence of outliers and improves fitted parameter values and uncertainties ... but the final χ^2-function values does not follow the standard χ^2 distribution.

Conclusion

Systematic errors of experimental data: Recent experiments provide a lot of information on systematic error contributions – this information should be used in a global analysis.

Statistical errors of experimental data: The published statistical errors are often too optimistic, because correlations (especially between neighbour bins) are neglected.

Construction of χ^2-function: Each error contribution should be taken into account with the correct underlying model of the error contribution.

Parametrization: Redundant parameters have to be avoided in a global fit.

Contents

1. Global analysis of data from HEP experiments **2**
Examples with large $\Delta\chi^2$ . . . 3
Data uncertainties in HEP experiments . . . 4

2. Normalization errors (multiplicative) **5**
Common normalisation errors . . . 6
Covariance matrix comparison . . . 7
The method with a nuisance parameter 8
The log-normal distribution . . . 9
Aymmetric normalisation errors . . . 10

3. Additive systematic errors **11**
Simplified calculation of inverse . . . 12
Example from an experimental paper . . . 13
Example from analysis papers . . . 14

4. Correlated statistical errors **15**
Why the cov matrix is non-diagonal 16

5. Parameter uncertainties: profile likelihoods and χ^2 **17**
Profile likelihood for a function . . . 18
Comments to a recent paper . . . 19
From papers 20
Use of redundant parameters . . . 21

6. Outliers and their influence on the fit **22**
M-estimates . . . 23
Influence functions and weights . . . 24
Commonly used M-estimators . . . 25

Conclusion **26**

Table of contents **27**