Terascale School on data combination and limit setting

04 October to 07 October 2011 – DESY

Data Combination in Particle Physics

Correlated and non-Gaussian data with systematic uncertainties

Volker Blobel – Universität Hamburg

Constrained least squares as a natural method is a more general alternative to χ^2-function minimization, especially for data combination.

Part 1. Combining correlated data

Part 2. Constrained least squares

Part 3. Combining non-Gaussian data

Part 1. Combining correlated data

1. Combining data
2. Averaging by linear least squares
3. Mean values, variances and covariances, correlations
4. Combining correlated data of a single quantity
5. Weights by Lagrange multiplier method
6. Least squares: Gauss, Legendre and Lagrange
7. Charm particle lifetime
8. Common additive systematic error I
9. Common multiplicative systematic error I
10. Average of two correlated data
11. Two-by-two covariance matrix from maximum likelihood
12. The two-dimensional normal distribution
13. Dzero result
14. Dzero: how big is the correlation coefficient?
15. Two alternative, but equivalent data combination methods
16. The PDG strategy

1. Combining data

For many physics analyses, several channels are combined into one result. In a similar way results from different experiments are merged. The aim of this procedure is to increase the precision of the results . . . **no bias, as accurate as possible** . . .

Lifetime of charmed particles: four different values extracted with different methods from the same data – hence correlated

Like-sign dimuon charge asymmetry: two values determined from different events samples – average has significant deviation from the SM

Mass of the top quark: several measured values with several systematic errors

Non-linearity: e.g. normalization uncertainty: straightforward methods produce biased results

Non-linearity: combining over-determined measurements of triangle parameters

Non-Gaussian data: Poisson-distributed data, lognormal factors

"Error" propagation: quantity depends on several measured values

From linear problems with Gaussian variables . . . to non-linear problems with non-Gaussian variables and several sources of systematic errors.

Essential: Understanding of physics, detector behaviour and data analysis.

2. Averaging by linear least squares

Weighted average: The average value $\boldsymbol{x}_{\text{ave}}$ of single values $\boldsymbol{x}_i$, $i = 1, 2, \ldots n$, with covariance matrix $\boldsymbol{V}_x$, is the weighted sum:

$$x_{\text{ave}} = \sum_i w_i\, x_i \qquad \text{with} \quad \sum_i w_i = 1 \qquad (\rightsquigarrow x_{\text{ave}} \text{ unbiased, if } x_i \text{ unbiased})$$

where the weights w_i are usually positive, but can be negative in certain cases.
The weighted average according to the equation is unbiased, if the single values are unbiased:

$$\mathrm{E}\left[\boldsymbol{x}_{\text{ave}}\right] = \sum_i w_i \mathrm{E}\left[\boldsymbol{x}_i\right] = \sum_i w_i \boldsymbol{\mu} = \boldsymbol{\mu} \qquad \text{for } \sum_i w_i = 1$$

The variance σ^2_{ave} follows from the law of (linear) propagation of uncertainties.

Uncorrelated data $x_i \pm \sigma_i$: ($\boldsymbol{V}_x$ diagonal)

$$w_i = \left(\sum_i \frac{1}{\sigma_i^2}\right)^{-1} \cdot \frac{1}{\sigma_i^2} \qquad\qquad \sigma^2_{\text{ave}} = \left(\sum_i \frac{1}{\sigma_i^2}\right)^{-1}$$

3. Mean values, variances and covariances, correlations

1-dim. random variable: mean value $\mu = \mathrm{E}[x] = \int x \cdot p(x)\,\mathrm{d}x \qquad p(x) = \text{ pdf}$

variance $\sigma^2 = \mathrm{E}\left[(x-\mu)^2\right] = \mathrm{V}[x] = \int (x-\mu)^2 \cdot p(x)\,\mathrm{d}x\ .$

n-dim. random variable: mean value $\boldsymbol{\mu} = \mathrm{E}[\boldsymbol{x}] = \int \ldots \int \boldsymbol{x} \cdot p(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$

covariance matrix $\boldsymbol{V}_x = \mathrm{E}\left[(\boldsymbol{x}-\boldsymbol{\mu})\,(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\right] = \mathrm{V}[\boldsymbol{x}]\ .$

Random vector $\boldsymbol{x}$, covariance matrix $\boldsymbol{V}_x$ and correlation matrix $\boldsymbol{C}_x$:

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ \ldots \\ x_n \end{pmatrix} \qquad \boldsymbol{V}_x = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2n} \\ \ldots & \ldots & & \ldots \\ \ldots & \ldots & & \ldots \\ \sigma_{n1} & \sigma_{n2} & \ldots & \sigma_{nn} \end{pmatrix} \qquad \boldsymbol{C}_x = \begin{pmatrix} 1 & \rho_{12} & \ldots & \rho_{1n} \\ \rho_{21} & 1 & \ldots & \rho_{2n} \\ \ldots & \ldots & & \ldots \\ \ldots & \ldots & & \ldots \\ \rho_{n1} & \rho_{n2} & \ldots & 1 \end{pmatrix}$$

The elements $(\boldsymbol{V})_{jk}$ are the covariances $(\boldsymbol{V})_{jk} = \sigma_{jk} = \rho_{jk}\sigma_j\sigma_k$, and ρ_{jk} are the correlation coefficients with $-1 \le \rho_{jk} \le +1$ (values of the correlation coefficients are often printed in %).

4. Combining correlated data of a single quantity

Correlated data x_i with (non-diagonal) covariance matrix $\boldsymbol{V}_x$: needs inverse $\boldsymbol{V}_x^{-1}$

$$w_i = \left(\sum_{j,k} \left(\boldsymbol{V}_x^{-1} \right)_{jk} \right)^{-1} \sum_j \left(\boldsymbol{V}_x^{-1} \right)_{ij} \qquad \sigma^2_{\text{ave}} = \boldsymbol{w} \boldsymbol{V}_x \boldsymbol{w}^{\mathrm{T}} = \sum_{i,j=1}^{n} w_i w_j \left(\boldsymbol{V}_x \right)_{ij}$$

Optimal values of the weights are obtained by χ^2-minimization (linear least squares). The inverse of $\boldsymbol{V}_x$ is used as weight matrix in the weighted sum of squares: $(\boldsymbol{x} = (x_1,\, x_2, \ldots x_n)^{\mathrm{T}})$

$$\chi^2\text{-function} \quad S(x_{\text{ave}}) = (x_{\text{ave}} \cdot \mathbf{1} - \boldsymbol{x})^{\mathrm{T}} \, \boldsymbol{V}_x^{-1} \, (x_{\text{ave}} \cdot \mathbf{1} - \boldsymbol{x}) = \sum_{i,j} (x_{\text{ave}} - x_i) \left(\boldsymbol{V}_x^{-1} \right)_{ij} (x_{\text{ave}} - x_j)$$

(all components of the vector $\mathbf{1}$ are one). The χ^2-function $S(x_{\text{ave}})$ is minimized with respect to the one parameter x_{ave}.

Generalization to different problems with multiple quantities possible, but constrained least squares is simpler.

See NIMA A 500 (2003) 391–405); NIMA A 270 (1988) 110–117

Correlated data x_i with (non-diagonal) covariance matrix $\boldsymbol{V}_x$:

The derivative of the weighted sum of squares $S(x_{\text{ave}})$ with respect to the parameter x_{ave}

$$\frac{1}{2}\frac{\partial S}{\partial x_{\text{ave}}} = \mathbf{1}^{\mathrm{T}}\boldsymbol{V}_x^{-1}\mathbf{1}\cdot x_{\text{ave}} - \mathbf{1}^{\mathrm{T}}\boldsymbol{V}_x^{-1}\boldsymbol{x}$$

is set to *zero* to obtain the solution

$$x_{\text{ave}} = \left[\left(\mathbf{1}^{\mathrm{T}}\boldsymbol{V}_x^{-1}\mathbf{1}\right)^{-1}\mathbf{1}^{\mathrm{T}}\boldsymbol{V}_x^{-1}\right]\boldsymbol{x} = \boldsymbol{w}\,\boldsymbol{x}$$

The expression in []-parentheses is the weight $\boldsymbol{w}$

$$\boldsymbol{w} = \underbrace{\left(\mathbf{1}^{\mathrm{T}}\boldsymbol{V}_x^{-1}\mathbf{1}\right)^{-1}}_{\text{scalar}}\underbrace{\mathbf{1}^{\mathrm{T}}\boldsymbol{V}_x^{-1}}_{\text{row vector}} \qquad\qquad w_i = \left(\sum_{j,k}\left(\boldsymbol{V}_x^{-1}\right)_{jk}\right)^{-1}\sum_{j}\left(\boldsymbol{V}_x^{-1}\right)_{ij}$$

The weight $\boldsymbol{w}$ is a 1-by-n matrix $\boldsymbol{w} = (w_1,\, w_2 \ldots w_n)$ (or a row vector) and average value x_{ave} and its variance σ^2_{ave} are given by

$$x_{\text{ave}} = \boldsymbol{w}\,\boldsymbol{x} = \sum_{i=1}^{n} w_i x_i \qquad\qquad \sigma^2_{\text{ave}} = \boldsymbol{w}\boldsymbol{V}_x\boldsymbol{w}^{\mathrm{T}} = \sum_{i,j=1}^{n} w_i w_j \left(\boldsymbol{V}_x\right)_{ij}$$

5. Weights by Lagrange multiplier method

Alternative method for the determination of weights: Minimization of the variance σ^2_{ave} of the average

$$\sigma^2_{\text{ave}} = \boldsymbol{w}\boldsymbol{V}_x\boldsymbol{w}^{\text{T}} = \sum_{i,j=1}^{n} w_i w_j \left(\boldsymbol{V}_x\right)_{ij}$$

subject to the equality constraint

$$\sum_i w_i = 1$$

Method of Lagrange multipliers with Lagrange function:

$$\mathcal{L}(\boldsymbol{w},\, \lambda) = \sum_{i,j=1}^{n} w_i w_j \left(\boldsymbol{V}_x\right)_{ij} + \lambda \left(\sum_i w_i - 1 \right)$$

with a single Lagrange multiplier λ.

The constrained problem is solved by setting the derivative of $\mathcal{L}(\boldsymbol{w},\, \lambda)$ with respect to the weights w_i and to the Lagrange multiplier λ to zero. The result is *identical* to the previous result.

6. Least squares: Gauss, Legendre and Lagrange

Carl Friedrich Gauß (1777 –1855) Used the least squares method already around 1794, but did not publish it at that time, but later in 1809. "Our principle, which we have made use of since 1795, has lately been published by Legendre ...." Proved in 1821 and 1823 the optimality of the least squares estimate without any assumptions that the random varibales follow a particular distribution (rediscovered by Markoff in 1912).

Adrien-Marie Legendre (1752 – 1833) Was the first to publish the method in 1805.

Joseph-Louis Lagrange (1736 –1813) Method of Lagrange multipliers (optimization of functions of several variables subject to equality constraints) formulated in "Leçons sur le calcul des fonctions" (1804). Used the principle of minimizing the sum of the absolute residuals $\sum_i |r_i|$, with $\sum_i r_i = 0$ in 1799.

Gauß-Markoff-Theorem: The vector $\boldsymbol{x} \in \mathbb{R}^n$ of measurements with a vector $\boldsymbol{\epsilon}$ of random errors is assumed to be related to an unknown parameter (or parameters) by a fixed linear model relation. All elements ϵ_i of $\boldsymbol{\epsilon}$ have zero means (no bias), are uncorrelated and have the same variance. The residuals r_i are the differences between the measured values x_i and the values, given by the parametrization (linear model). The **best** (i.e. minimum variance) **linear unbiased estimator** (BLUE) for the parameter(s) is the least squares estimator, minimizing the sum of squared residuals $||\boldsymbol{r}||^2$. No assumptions are required that the random errors follow a particular distribution; i.e. the normal (Gaussian) distribution is not required.

7. Charm particle lifetime

Lifetime (in units of 10^{-13} sec) of charmed particles in a CERN experiment determined using four different methods from the same data (hence correlated):

$$\boldsymbol{\tau} = \begin{pmatrix} 9.5\,^{+1.7}_{-1.2} \\ 11.9\,^{+1.5}_{-1.3} \\ 11.1\,^{+1.8}_{-1.2} \\ 8.9\,^{+1.6}_{-1.2} \end{pmatrix} \qquad \boldsymbol{V}_\tau = \begin{pmatrix} 2.74 & 1.15 & 0.86 & 1.31 \\ 1.15 & 1.67 & 0.82 & 1.32 \\ 0.86 & 0.82 & 2.12 & 1.05 \\ 1.31 & 1.32 & 1.05 & 2.93 \end{pmatrix}$$

$$\boldsymbol{C}_\tau = \left(\begin{array}{c|cccc} & \tau_1 & \tau_2 & \tau_3 & \tau_4 \\ \hline \tau_1 & - & & & \\ \tau_2 & 0.54 & - & & \\ \tau_3 & 0.36 & 0.44 & - & \\ \tau_4 & 0.46 & 0.60 & 0.42 & - \end{array} \right)$$

Upper and lower errors transformed to a symmetric interval by the geometric mean, for example to replace $^{+1.7}_{-1.2}$ by $\sqrt{1.7 \cdot 1.2} = 1.43$.
Full error matrix estimated by MC simulation of large number of "experiments", analysed in exactly the same way as real data (plus some corrections).

weight factors $\qquad \boldsymbol{w} = (0.14,\ 0.47,\ 0.35,\ 0.04)$

Final average for the charm lifetime:

$$\tau_{\text{ave}} = (11.16 \pm 1.13) \times 10^{-13}\ \text{s}$$
$$[\ \tau_{\text{ave}} = (10.62 \pm 0.75) \times 10^{-13}\ \text{s} \qquad \text{if correlations ignored}\ ]$$

Bias of ≈ 1 standard deviation and underestimation of accuracy, if correlations are ignored.

L. Lyons et al., Combining correlated estimates of a single physical quantity, NIMA **A270** (1988) 110–117

8. Common additive systematic error I

Common additive uncertainty : $x_i \pm \sigma_i \pm \Delta$ (identical *systematic* error Δ)

- (PDG:) first average $x_i \pm \sigma_i$, then combine error with Δ^2;
- (PDG:) apply factor $(1 + \Delta^2 \left(\sum_i 1/\sigma_i^2\right))^{1/2}$ to all errors, and treat as uncorrelated;
- syst. uncertainty Δ as additional measured value

... as correlated data ...

$$x_1 = x_1' + a$$
$$x_2 = x_2' + a$$

$$a = \begin{array}{lll} x_1' & \pm\sigma_1 \\ x_2' & \pm\sigma_2 \\ 0 & \pm\Delta \end{array} \qquad \boldsymbol{V}\begin{pmatrix} x_1' \\ x_2' \\ a \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \Delta^2 \end{pmatrix}$$

- define non-diagonal covariance matrix by law of (linear) propagation of uncertainties, with $(\boldsymbol{V}_x)_{ii} = \sigma_i^2 + \Delta^2 \qquad (\boldsymbol{V}_x)_{ij} = \Delta^2 \quad i \neq j.$

$$\boldsymbol{V}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \boldsymbol{V} \begin{pmatrix} x_1' \\ x_2' \\ a \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 + \Delta^2 & \Delta^2 \\ \Delta^2 & \sigma_2^2 + \Delta^2 \end{pmatrix}$$

all methods are equivalent.

9. Common multiplicative systematic error I

Common multiplicative uncertainty: i.e. normalization or calibration uncertainty

$$\text{e.g.} \quad (x_i \pm \sigma_i)\,(1 \pm \Delta) \quad \rightsquigarrow \text{non-linear!}$$

Data with common multiplicative systematic uncertainty

... as correlated data ...

$$x_1 = x_1' \times a$$
$$x_2 = x_2' \times a$$

$$a = \begin{array}{ccc} x_1' & \pm\,\sigma_1 \\ x_2' & \pm\,\sigma_2 \\ 1 & \pm\,\Delta \end{array} \qquad \boldsymbol{V}\begin{pmatrix} x_1' \\ x_2' \\ a \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \Delta^2 \end{pmatrix}$$

Non-diagonal covariance matrix by law of propagation of uncertainties: non-linear transformation

$$\boldsymbol{V}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a & 0 & x_1 \\ 0 & a & x_2 \end{pmatrix} \boldsymbol{V}\begin{pmatrix} x_1' \\ x_2' \\ a \end{pmatrix}\begin{pmatrix} a & 0 \\ 0 & a \\ x_1 & x_2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 + x_1^2\Delta^2 & x_1 x_2 \Delta^2 \\ x_1 x_2 \Delta^2 & \sigma_2^2 + x_2^2 \Delta^2 \end{pmatrix} \quad \text{with } a = 1$$

Dangerous: elements of the transformation matrix are not constant; the two representations are **not equivalent**.

... discussed later.

10. Average of two correlated data

Data vector $\boldsymbol{x}$ and covariance matrix $\boldsymbol{V}_x$:

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad \boldsymbol{V}_x = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

$$\text{inverse covariance matrix } = \boldsymbol{V}_x^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1/\sigma_1^2 & -\rho/(\sigma_1\sigma_2) \\ -\rho/(\sigma_1\sigma_2) & 1/\sigma_2^2 \end{pmatrix}$$

Weight factors and variance of average determined by σ_1, σ_2 and ρ:

$$w_1 = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \qquad w_2 = 1 - w_1 = \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

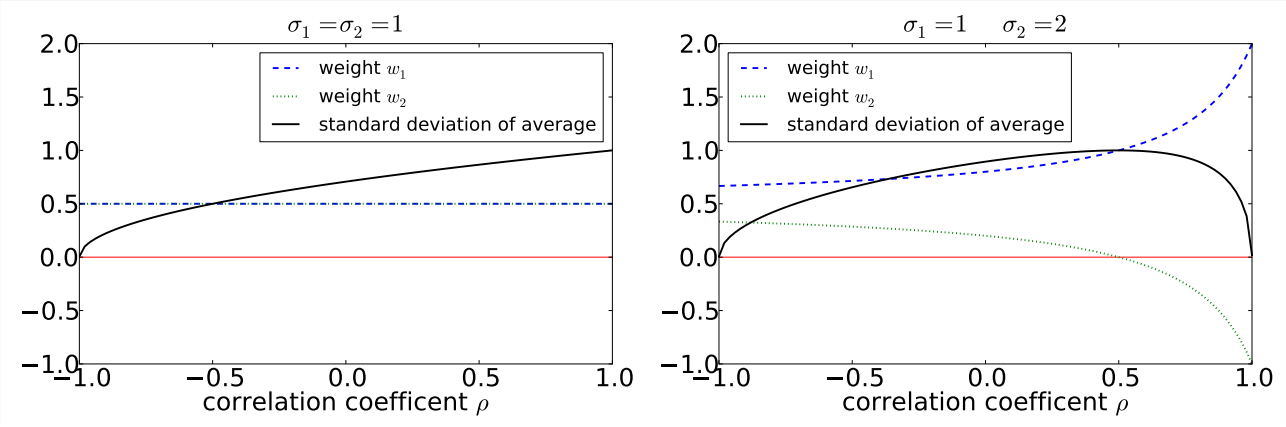
$$\sigma^2_{\text{ave}} = \left(1-\rho^2\right) \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} = \qquad \text{variance of average}$$

No correlation: $\boldsymbol{\rho = 0}$

$$w_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \qquad w_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \qquad \sigma^2_{\text{ave}} = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-1}$$

negative correlation $\boldsymbol{\rho < 0}$: both weights are always positive and x_{ave} is between x_1 and x_2.

positive correlation $\boldsymbol{\rho > 0}$: weight of the less accurate value may become negative and then the average x_{ave} is outside the range $x_1 \ldots x_2$.



- Weight $w_2 < 0$ for large positive correlation $\rho > +\frac{\sigma_1}{\sigma_2}$
  $\rightsquigarrow$ **not meaningful?**
- negative weights impossible for correlations by overlapping data samples,
- $x_{\text{ave}} \equiv x_1$ and $\sigma_{\text{ave}} \equiv \sigma_1$, **no improvement** for $\rho = \frac{\sigma_1}{\sigma_2}$,
- expected difference^2 $E\left[(x_1 - x_2)^2\right] = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$
- smaller value of σ_{ave} for negative correlation.

11. Two-by-two covariance matrix from maximum likelihood

Maximum likelihood estimate $\widehat{\boldsymbol{x}}$: $\quad \mathcal{L}(\widehat{\boldsymbol{x}}) = \max_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x})$

likelihood equation: $\quad \partial \log \mathcal{L} / \partial x_j = 0 \qquad j = 1,\, 2, \ldots,\, n$

Hessian (second order derivative matrix of the negative log likelihood function) with elements

$$(\boldsymbol{H}_x)_{jk} = -\frac{\partial^2 \log \mathcal{L}}{\partial x_j \partial x_k} \,.$$

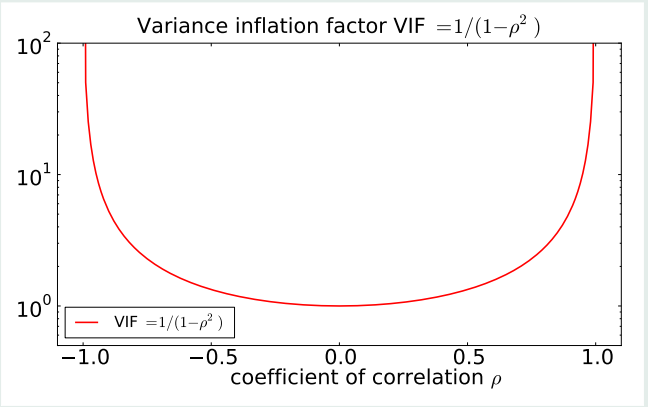
Cramér-Rao inequality $\quad (\boldsymbol{V}_x)_{jk} \geq (\boldsymbol{H}_x)_{jk} \qquad$ ... in practice: $\boldsymbol{V}_x = \boldsymbol{H}^{-1}$

The 2-dimensional case:

$$\boldsymbol{H} = \begin{pmatrix} H_{11} & H_{12} \\ H_{12} & H_{22} \end{pmatrix} \qquad \boldsymbol{V}_x = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

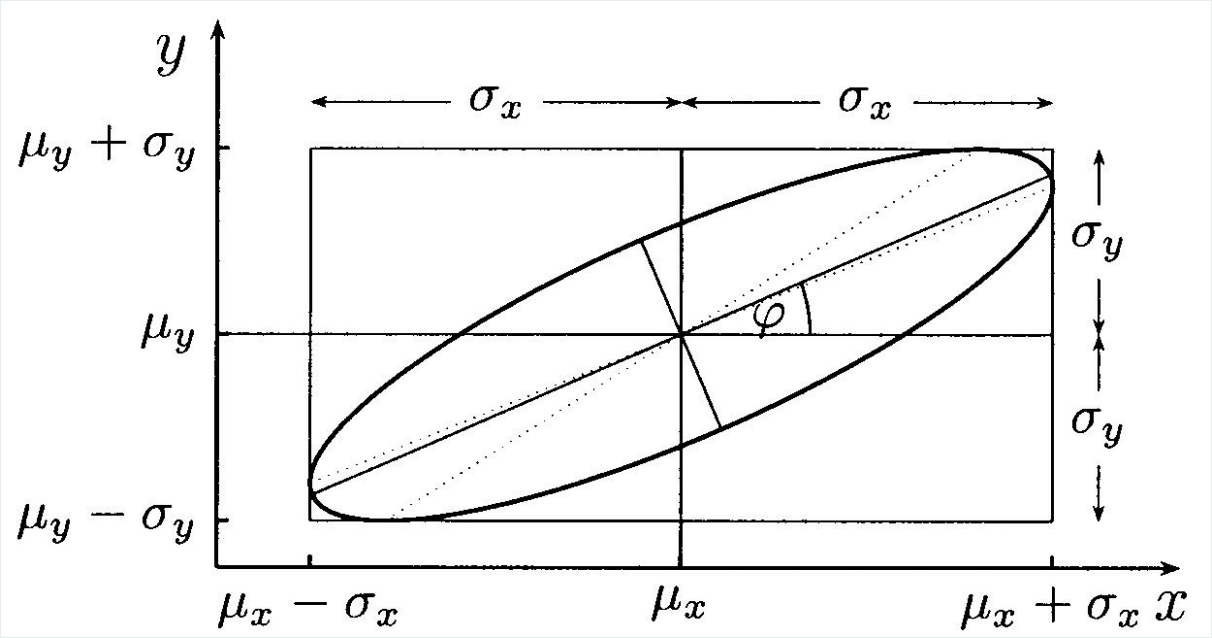
$$\rho = -\frac{H_{12}}{\sqrt{H_{11} H_{22}}} \qquad \sigma_1^2 = \left(\frac{1}{1-\rho^2}\right) \frac{1}{H_{11}}$$

$$\sigma_2^2 = \left(\frac{1}{1-\rho^2}\right) \frac{1}{H_{22}}$$



For non-zero correlation $\rho \neq 0$ the variances are larger by the factor $1/(1-\rho^2)$, the **variance inflation factor**, compared to zero correlation.

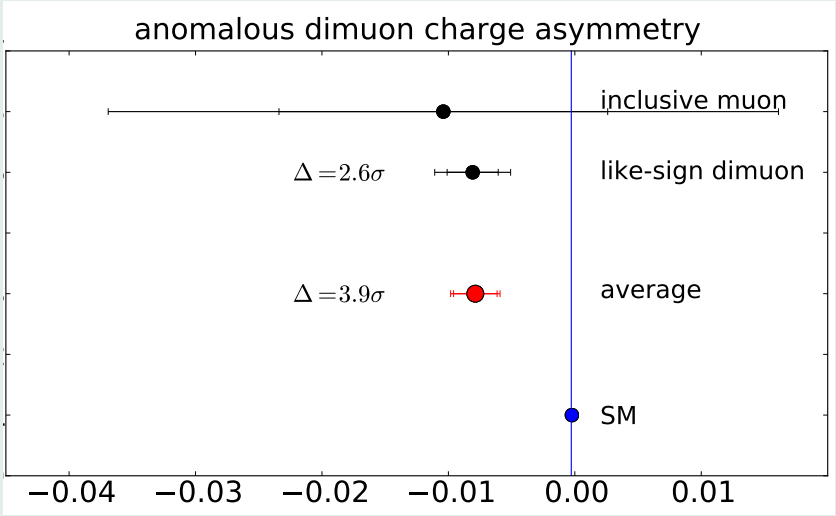
12. The two-dimensional normal distribution



13. Dzero result

Measurement of the anomalous like-sign dimuon charge asymmetry with 9 fb^{-1} of pp collisions arXiv:1106.6308[hep:ex]; Fermilab-Pub-11/307-E

inclusive muon sample	$A_{sl}^b = (-1.04 \pm 1.30\,(\text{stat}) \pm 2.31\,(\text{syst}))\,\%$
like-sign dimuon sample	$A_{sl}^b = (-0.808 \pm 0.202\,(\text{stat}) \pm 0.222\,(\text{syst}))\,\%$
combined result	$A_{sl}^b = (-0.787 \pm 0.172\,(\text{stat}) \pm 0.093\,(\text{syst}))\,\%$

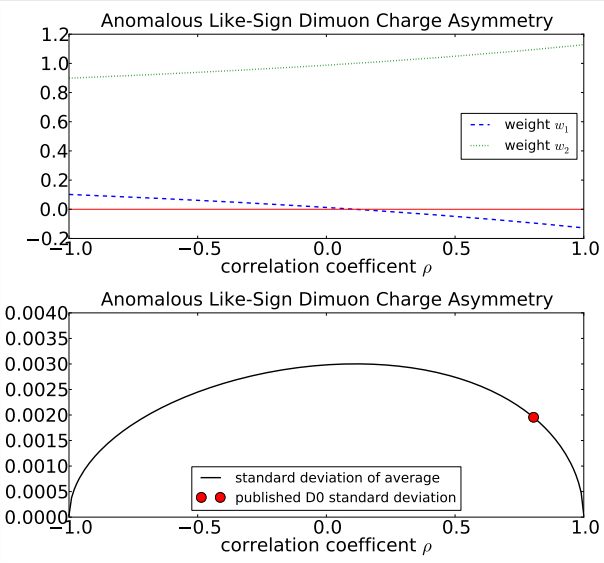


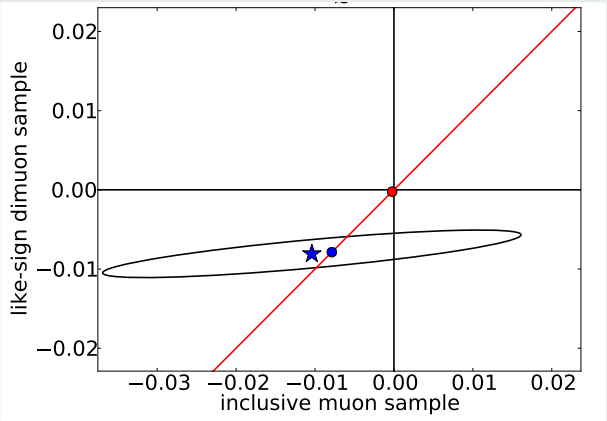
14. Dzero: how big is the correlation coefficient?

Statistical and systematic error quadratically added:

$$\begin{aligned} \text{inclusive muon sample} \quad & A_{sl}^b = (-1.04 \pm 2.65\,(\text{tot}))\,\% \\ \text{like-sign dimuon sample} \quad & A_{sl}^b = (-0.808 \pm 0.300\,(\text{tot}))\,\% \end{aligned}$$

Weights w_1, w_2, and st.dev. of average, as a function of ρ:





measured values (star) with error ellipse

15. Two alternative, but equivalent data combination methods

Assuming the two measured values with uncorrelated statistical and fully ($\rho = +1$) correlated systematic uncertainties:

Three uncorrelated data and one constraint:

$$x'_1 = (-1.04 \pm 1.3)\,\%$$
$$x'_2 = (-0.808 \pm 0.202)\,\%$$
$$\Delta = (0 \pm 1)\,\%$$
$$f_1 = (x'_1 + 2.31 \cdot \Delta) - (x'_2 + 0.222 \cdot \Delta)$$

i.e. extra variable Δ for systematic uncertainty.

Two correlated data and one constraint:

$$x_1 = (-1.04 \pm 1.3)\,\%$$
$$x_2 = (-0.808 \pm 0.202)\,\%$$
$$\boldsymbol{V} = \begin{pmatrix} 1.3^2 + 2.31^2 & 2.31 \cdot 0.222 \\ 2.31 \cdot 0.222 & 0.202^2 + 0.222^2 \end{pmatrix}$$
$$f_1 = x_1 - x_2$$

i.e. covariance matrix with statistical and systematic uncertainties.

The two results are identical:

$$A^b_{sl} = (-0.792 \pm 0.246)\,\% \qquad \chi^2_1 = 0.009 \qquad p\text{-value} = 92.5\%$$

published $\qquad A^b_{sl} = (-0.787 \pm 0.196)\,\%$

16. The PDG strategy

Linear least squares, minimizing value of χ^2 expression, using standard averaging: value $\overline{x}$ with error $\delta\overline{x}$.
Selection of reliable data from the literature:

- add statistical and systematic errors in quadrature, handle "asymmetric errors";
- e.g. do not use data from preprints;
- ... check consistency of the data, using χ^2, compared to expected value $N-1$, for N data.
- $\boldsymbol{\chi^2 \leq N-1}$ **:** accept result $\overline{x}$ as average and $\delta\overline{x}$ as error.

PDG:

$\boldsymbol{\chi^2 \gg N-1}$ **:** bad χ^2 – data incompatible – reject result for average

$\boldsymbol{? > \chi^2 > N-1}$ **:** χ^2 greater than expected value, but not greatly so: errors may be underestimated – average $\overline{x}$ remains unchanged, but error $\delta\overline{x}$ increased by scale factor

$$S = \left(\frac{\chi^2}{N-1}\right)^{1/2}$$

... undetected correlations?

$\boldsymbol{\chi^2}$ **very large:** data inconsistent or errors underestimated or undetected negative correlations;

$\boldsymbol{\chi^2}$ **very small:** either errors overestimated (could decrease uncertainty $\delta\overline{x}$ of average), or undetected positive correlations (could increase uncertainty $\delta\overline{x}$ of average).

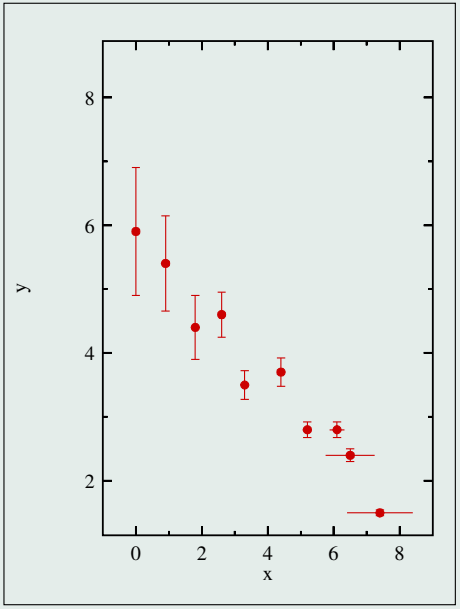
Part 2. Constrained Least Squares

1. x-y-data with uncertainties in both coordinates
2. Constrained Least Squares
3. Alternative least squares methods for fitting/averaging
4. Comparison
5. Constrained least squares fit program APLCON
6. Averaging correlated scattering lengths
7. Straight line with uncertainties in both coordinates
8. Straight line and correlated data
9. Uncertainties of fit parameters
10. Mathematics: solution of constrained least squares

1. x-y-data with uncertainties in both coordinates

The subject is discussed by Press et. al. (Numerical Recipes) with the remarks:

> "*If experimental data are subject to measurement error not only in the y_i's, but also in the x_i's, then the task of fitting a straight-line model*
>
> $$y(x) = a + bx$$
>
> *is considerably harder ... Be aware that the literature on the seemingly straightforward subject of this section is generally confusing and sometimes plain wrong.*"



(data in figure from C.A. Cantrell)

What is the uncertainty of residual $r_i = y_i - (a + b\,x_i)$?

C.A. Cantrell [*Atmos. Chem. Phys.*, **8**, 5477-5487, 2008] lists > 30 publications for methods (including methods giving wrong results), only for straight-line fits, almost all for uncorrelated data only.

$\rightsquigarrow$ try "Deming regression" or "error-in-variables-model (EIV)" or "total least squares (TLS)" in Google.

Vector of variables $\boldsymbol{x}$, including

- measured variables $\boldsymbol{x}_m$ (with covariance matrix $\boldsymbol{V}_x$) and
- unmeasured variables $\boldsymbol{x}_u$ (e.g. model parameters).

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_m \\ \boldsymbol{x}_u \end{pmatrix} \qquad \text{minimize} \quad \Delta\boldsymbol{x}_m^{\mathrm{T}}\, \boldsymbol{V}_x^{-1}\, \Delta\boldsymbol{x}_m$$

$$\text{subject to equality constraints} \quad f_j\left(\boldsymbol{x} + \Delta\boldsymbol{x}\right) = 0 \quad j = 1,\, 2 \ldots m$$

Best estimate for variables is $\widehat{\boldsymbol{x}} = \boldsymbol{x} + \Delta\boldsymbol{x}$.

Rudolf Boeck, Application of a generalized method of least squares for kinematical analysis of tracks in bubble chambers, CERN 60–30

From publications: "In practice, the added technical complexity of a constrained fit with extra free parameters is not justified . . . "

"The application of Lagrange multipliers is unnecessarily complicated and the linear approximation requires additional assumptions and iterations."

3. Alternative least squares methods for fitting/averaging

$\boldsymbol{x}_m$ = measured variables, with covariance matrix $\boldsymbol{V}_x$

$\boldsymbol{x}_u$ = unmeasured variables, "parameters"

$\boldsymbol{x} = (\boldsymbol{x}_m, \boldsymbol{x}_u)$ = measured *and* unmeasured variables $\qquad$ $\boldsymbol{t}$ = independent coordinates

χ^2-function minimization

$$S(\boldsymbol{x}_u) = \sum_i \frac{((\boldsymbol{x}_m)_i - f(t_i, \boldsymbol{x}_u))^2}{\sigma_i^2} = \min$$
$$\rightarrow \boldsymbol{r}^{\mathrm{T}} \boldsymbol{V}_x^{-1} \boldsymbol{r} = \min$$

Residuals $\boldsymbol{r}$: χ^2-function to be minimized is sum of squares of residuals; problems, if residuals

- depend on > 1 measurement, and/or depend on > 1 error contribution, especially contributions changing the normalization.

Constrained Least Squares

$$S(\Delta\boldsymbol{x}_m) = \Delta\boldsymbol{x}_m^{\mathrm{T}} \boldsymbol{V}_x^{-1} \Delta\boldsymbol{x}_m = \min$$
$$f_j\,(\boldsymbol{x}_m + \Delta\boldsymbol{x}_m,\, \boldsymbol{x}_u + \Delta\boldsymbol{x}_u,\, \boldsymbol{t}) = 0 \quad j = 1,\, 2 \ldots m$$

Individuals corrections $\Delta\boldsymbol{x}_m$ for measured variables: expression to be minimized is sum of squares of corrections.

- Constraints $f_j(\boldsymbol{x}) = 0$ may be implicit expressions;
- bias reduced or avoided.

Both alternatives are equivalent, with identical results, for simple problems. In both alternatives the data may be correlated and the functions/constraints may be non-linear.

4. Comparison

χ^2-function minimization, e.g. using MINUIT

- User has to provide the function $S(\boldsymbol{x})$, which is "seen" by MINUIT. The user function includes all data, uncertainties, the physical and statistical model.
- MINUIT calculates by finite differences the first derivative of $S(\boldsymbol{x})$, and approximates, using the VM method, the full Hessian in $\geq n$ iterations for linear and non-linear problems.
- Variables are the parameters (= unmeasured variables).

Constrained least squares, e.g. using APLCON 2.0

- User describes set of variables incl. covariance matrix, and individual model functions $f_j(\boldsymbol{x})$.
- APLCON calculates by finite differences the first derivative of all individual model functions $f_j(\boldsymbol{x})$, which allows to calculate the full Hessian during each iteration (Gauss-Newton matrix).
- Many variables: measured and unmeasured variables plus Lagrange multipliers.
- Principle used in HEP for $>$ 50 years, mainly with kinematical constraints for particle reactions and decays; APLCON 1.0 in use for 35 years.

5. Constrained least squares fit program APLCON

$$\begin{array}{ll} \text{minimize} & \Delta\boldsymbol{x}^{\mathrm{T}}\boldsymbol{V}_x^{-1}\Delta\boldsymbol{x} \\ \text{subject to} & f_j\left(\boldsymbol{x},\,\boldsymbol{t}\right)=0 \quad j=1,\,2\ldots m \end{array}$$

Properties:

- Extreme form of constrained least squares, with separation into a quadratic expression, and a set of constraints $f_j(\boldsymbol{x})$ with all nonlinearities; solved using Lagrange multipliers;
- simple to use: derivatives calculated by numerical methods, no step definition necessary, no principle distinction between measured ($\mathbf{X}_m$) and unmeasured variables ($\mathbf{X}_u$); full initial and final covariance matrix, and pulls;
- Extension to **non-Gaussian** variables: selected variables can be treated e.g as Poisson- or log-normal-distributed;
- Extension to advanced analysis of uncertainties: profile likelihood
- APLCON is a method for difficult problems to follow accurately the assumed physical and statistical model of the measurement process, and to avoid a **bias** in the result;
- APLCON (Fortran) available from `www.desy.de/~blobel`

6. Averaging correlated scattering lengths

50 years old data on the isospin 1/2 and 3/2 scattering lengths in πp-scattering in the s-state:

Experiment (1): $a_1 = 0.170 \pm 0.0240$; $a_3 = -0.107 \pm 0.0197$; corr. coefficient $\rho = -39.1\%$.

Experiment (2): $a_3' = -0.104 \pm 0.006$.

Input to the APLCON fit to average the two a_3-values and, at the same time, improve the correlated a_1:

$$\boldsymbol{x}_m = \begin{pmatrix} a_1 \\ a_3 \\ a_3' \end{pmatrix} = \begin{pmatrix} 0.170 \pm 0.0240 \\ -0.107 \pm 0.0197 \\ -0.104 \pm 0.0060 \end{pmatrix} \qquad \boldsymbol{V}_x = \begin{pmatrix} 0.580 & -0.185 & 0 \\ -0.185 & 0.388 & 0 \\ 0 & 0 & 0.036 \end{pmatrix} \times 10^{-3}$$

and after the code $f_1 = a_3 - a_3'$ the result by APLCON is

$$\boldsymbol{x} = \begin{pmatrix} a_1 \\ a_3 \\ a_3' \end{pmatrix} = \begin{pmatrix} 0.169 \pm 0.0220 \\ -0.1043 \pm 0.0057 \\ -0.1043 \pm 0.0057 \end{pmatrix} \qquad \boldsymbol{V}_x = \left(\begin{array}{cc|c} 0.499 & -0.0157 & -0.0157 \\ -0.0157 & 0.0329 & 0.0329 \\ \hline -0.0157 & 0.0329 & 0.0329 \end{array}\right) \times 10^{-3} \,.$$

7. Straight line with uncertainties in both coordinates

$\mathbf{X} := \ldots$ *(variable array)*
$\mathbf{V}_x := \ldots$ *(matrix array)*

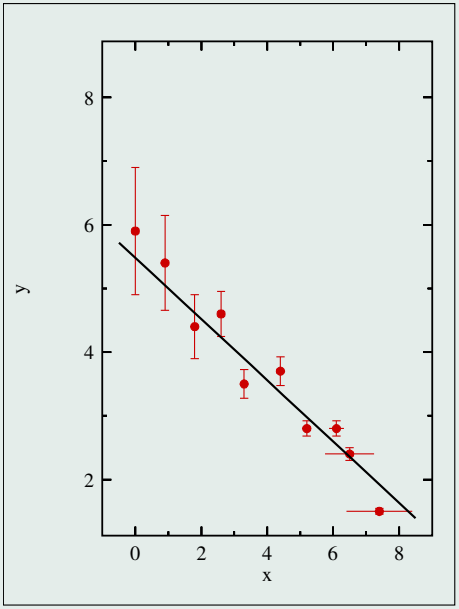
for $j = 1$ to N

`f(j)` $= a + b \cdot x_j - y_j$

final result in $\mathbf{X}$ *and* $\mathbf{V}_x$

variable array

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_m \\ \hline \mathbf{X}_u \end{pmatrix} = \begin{pmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ \vdots \\ x_N \\ y_N \\ \hline a \\ b \end{pmatrix}$$



Note: order of measured and unmeasured variable irrelevant – distinguished by **zero** elements in input covariance matrix $\mathbf{V}_x$.

If measurement of slope b exists before: add variance of b to $\boldsymbol{V}_x$, with no change in the program code

8. Straight line and correlated data

Now: correlation between x and y in data $\neq 0$, and fit of straight line required

$\mathbf{X} := \ldots$ *(variable array)*
$\mathbf{V}_x := \ldots$ *(matrix array)*

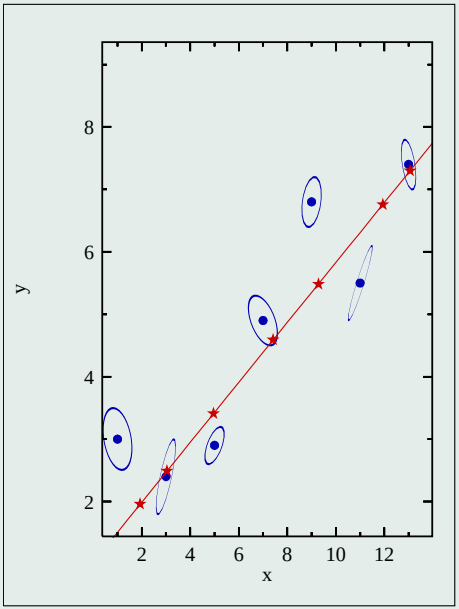
for $j = 1$ to N

`f(j)` $= a + b \cdot x_j - y_j$

final result in $\mathbf{X}$ *and* $\mathbf{V}_x$

variable array

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_m \\ \hline \mathbf{X}_u \end{pmatrix} = \begin{pmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ \vdots \\ x_N \\ y_N \\ \hline a \\ b \end{pmatrix}$$



add off-diagonal elements to $\boldsymbol{V}_x$
no change of code

red star is fitted xy-value

Extension to parabola simple: include c in $\mathbf{X}$ and add $+ c \cdot {x_j}^2$ to `f(j)`

9. Uncertainties of fit parameters

Aplcon provides

full covariance matrix $\boldsymbol{V}_x$ for combined variables: fitted values of measured variables and of unmeasured variables ("parameters"), from the inverse of the Hessian (by the law of propagation of uncertainties);

pulls for all measured variables: should follow $N(0,1)$ distributions;

- Covariance matrix is accurate in simple cases: measured data Gaussian and constraints linear, or asymptotically in the limit of ∞ data;
- Matrix may be inaccurate (and non-Gaussian) for non-Gaussian data, constraints from non-linear models and low statistic $\rightsquigarrow$ statistically improved information is required on confidence intervals for important parameters.

confidence intervals on selected parameters by **profile analysis** (optional):
realized by repeated fits with one additional internal constraint;

contours for selected parameters pairs by **profile analysis** (optional):
realized by repeated fits with two additional internal constraints.

10. Mathematics: solution of constrained least squares

Special case: no unmeasured variables (parameters), $\boldsymbol{x} \equiv \boldsymbol{x}_m$, and linear equality constraints:

minimize $\Delta\boldsymbol{x}^{\mathrm{T}}\boldsymbol{V}_x^{-1}\Delta\boldsymbol{x}$ subject to equality constraints $\boldsymbol{f} + \boldsymbol{A}\Delta\boldsymbol{x} = 0$

Solved by stationary point of Lagrange function

$$\mathcal{L}(\widehat{\boldsymbol{x}}) = \Delta\boldsymbol{x}^{\mathrm{T}}\boldsymbol{V}_x^{-1}\Delta\boldsymbol{x} + 2\boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{f})$$

Derivatives w.r.t $\Delta\boldsymbol{x}$ and $\boldsymbol{\lambda}$ set to zero:

$$\begin{array}{rcl} \boldsymbol{V}_x^{-1}\Delta\boldsymbol{x} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} &=& 0 \\ \boldsymbol{A}\Delta\boldsymbol{x} &=& -\boldsymbol{f} \end{array} \qquad \begin{pmatrix} \boldsymbol{V}_x^{-1} & \boldsymbol{A}^{\mathrm{T}} \\ \boldsymbol{A} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \Delta\boldsymbol{x} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ -\boldsymbol{f} \end{pmatrix}$$

$$\text{solution:} \quad \begin{pmatrix} \Delta\boldsymbol{x} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \boldsymbol{V}_x^{-1} & \boldsymbol{A}^{\mathrm{T}} \\ \boldsymbol{A} & \boldsymbol{0} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{0} \\ -\boldsymbol{f} \end{pmatrix}$$

Solution by determination of partitioned matrix:

$$\boldsymbol{W}_A = (\boldsymbol{A}\boldsymbol{V}_x\boldsymbol{A}^{\mathrm{T}})^{-1} \qquad \begin{pmatrix} \boldsymbol{V}_x^{-1} & \boldsymbol{A}^{\mathrm{T}} \\ \boldsymbol{A} & \boldsymbol{0} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{V}_x - \boldsymbol{V}_x\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}_A\boldsymbol{A}\boldsymbol{V}_x & \boldsymbol{V}_x\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}_A \\ \boldsymbol{W}_A\boldsymbol{A}\boldsymbol{V}_x & \boldsymbol{W}_A \end{pmatrix}$$

Note: inverse matrix $\boldsymbol{V}_x^{-1}$ not used in algorithm (may be singular).

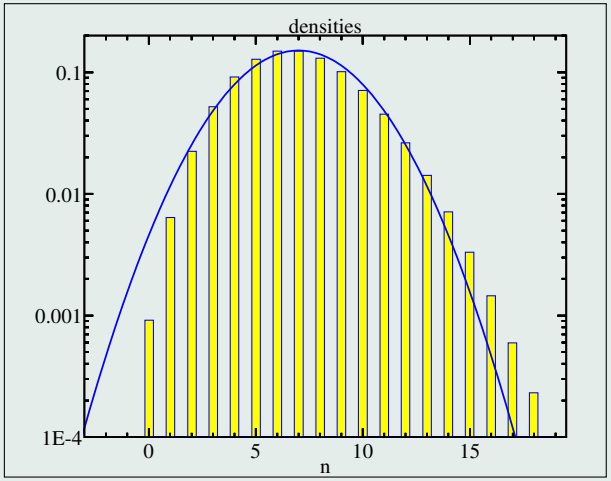
Part 3. Non-Gaussian data and nonlinearities

1. Poisson distributed data – counts
2. Cross section measurement
3. Log-normal distribution
4. Averaging with "normalization" uncertainty
5. Covariance matrix plot
6. Solution with constraints
7. Triangle parameters
8. Correlated additive systematic errors II
9. Correlated multiplicative systematic errors II
10. Combination of FNAL results on the mass of the top quark
11. Branching ratios

1. Poisson distributed data – counts

Poisson data, having a variance equal to the mean, have the problem of non-uniform variance. The Poisson distribution for a small mean value is asymmetric. The normal distribution is a bad approximation for small mean values and in the tails.

Poisson and normal density for $\mu = \sigma^2 = 7$ $\longrightarrow$



Least squares requires data with constant variance, independent of fit result.
What happens, if x_i are not normal distributed or do not have constant variance? $\rightsquigarrow$ **Bias**

Example: Average of data following (or proportional to) Poisson distribution

$x_1 = \quad 9 \pm 3$
$x_2 = \quad 16 \pm 4$

Weighted mean (LS) $x_{\text{ave}} = 11.52 \pm 2.40$
Using Poisson statistic (ML) $x_{\text{ave}} = 12.5 \pm 2.5$

2. Cross section measurement

Least squares popular in particle physics for cross section fits and averaging, using data from ≥ 1 experiment.
Cross sections x_i are measured via counted numbers $n_i = S \cdot x_i$ of events:

$$\text{cross section} \quad x_i = S^{-1} \cdot n_i, \quad i = 1, \ldots$$

where the sensitivity factor S is a product

$$S = A_1 \cdot A_2 \cdots A_a \cdot \int \mathcal{L}\, \mathrm{d}t \cdot \Delta x$$

of many factors (trigger, detection, reconstruction ... probabilities, luminosity, bin width).

number n_i: follows Poisson distribution,

sensitivity S: will follow a log-normal distribution (log of S normal distributed) – the inverse S^{-1} will follow a log-normal distribution too:

$$\text{cross section } x_i \; = S^{-1} \text{ (log-normal)} \times n_i \text{ (Poisson)} \qquad \text{– possible with APLCON}$$

Normalization factors will approximately follow the log-normal distribution, as a consequence of the Central Limit-Theorem: product of many factors with small uncertainty.

cross section: log-normal × Poisson, but in practice often assumed to follow the normal distribution with x_i assumed to be independent (with diagonal covariance matrix); even resolution-corrected ("unfolded") cross sections usually assumed to be independent (!).

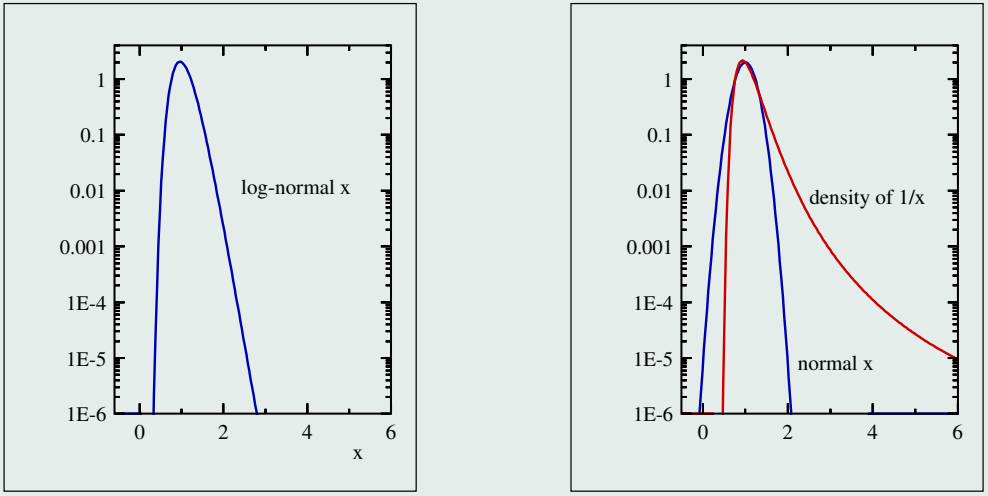
3. Log-normal distribution

Log-normal distribution for e.g normalization factors (and other variables, which by definition are *positive*):

log-normal variable (with uncertainty $\propto$ value): external $\alpha \quad \Rightarrow \quad \exp[\ \alpha'\ ]$

with new internal Gaussian variable $\alpha' \quad \equiv \quad \ln \alpha$

Example: $\alpha = 1 \pm 0.2$



Data in HEP are often given with uncertainty in %, i.e relative uncertainty.
This indicates the log-normal (instead of the normal) distribution with constant *relative* uncertainty.

In a **publication** (NIM A) the following measurement for two data points x_1, x_2 and a common normalization factor α with uncertainty ϵ is given:

$$x_1 = 8.0 \pm 2\% \qquad x_2 = 8.5 \pm 2\% \qquad \alpha = 1 \pm \epsilon \qquad \text{with} \quad \epsilon = 0.1$$

"Assuming that the two measurements refer to the same physical quantity, the best estimate of its true value can be obtained by fitting the points to a constant" (from the publication).

Result of straightforward unweighted average: $x_{\text{ave}} = (x_1 + x_2)/2 = 8.25$, but ...

Publication: average x_{ave} by "χ^2-function minimization", the covariance matrix $\boldsymbol{V}$ is defined to include the normalization uncertainty:

$$\chi^2 = \boldsymbol{\Delta}^{\mathrm{T}} \boldsymbol{V}^{-1} \boldsymbol{\Delta} = \text{minimum} \qquad \text{with} \quad \boldsymbol{V} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} + \epsilon^2 \cdot \begin{pmatrix} x_1^2 & x_1 x_2 \\ x_1 x_2 & x_2^2 \end{pmatrix}$$

($\boldsymbol{\Delta}$ is "the vector of the differences" between x_i and average x_{ave}).

Resulting average is $x_{\text{ave}} = 7.87 \pm 0.81$, outside (!) the range of the two input values

... apparently wrong $\rightsquigarrow$ large bias with constructed non-diagonal covariance matrix.

Note: weights $w_1 = +1.25$ and $w_2 = -0.25$ because $\sigma_1 < \sigma_2$, and large positive correlation.

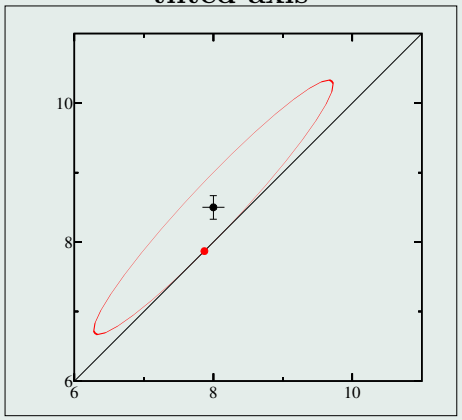
$\rightsquigarrow$ try 'Peelles puzzle" in Google (or NIMA A 346 (1994) 306–311)

Axis of covariance ellipse is slightly **tilted** (left) because input values x_1 and x_2 (and σ_1, σ_2) are not equal; this causes the "strange" value of the average.

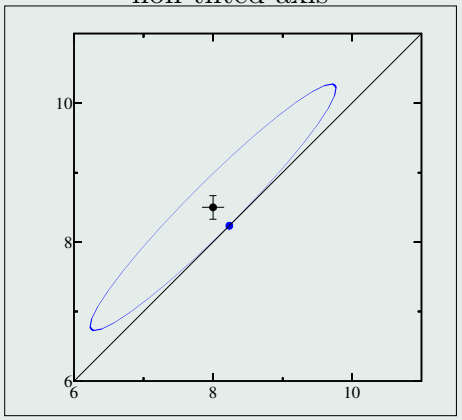
$$\chi^2 = \boldsymbol{\Delta}^{\mathrm{T}} \boldsymbol{V}^{-1} \boldsymbol{\Delta} = \text{minimum} \qquad \text{with} \quad \boldsymbol{V} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} + \epsilon^2 \cdot \begin{pmatrix} x_1^2 & x_1 x_2 \\ x_1 x_2 & x_2^2 \end{pmatrix}$$

($\boldsymbol{\Delta}$ is "the vector of the differences" between x_i and average x_{ave}).

tilted axis



non-tilted axis



Axis of covariance ellipse is not tilted for $\sigma_1 = \sigma_2$ (right).

With two constraints the average x_{ave}, multiplied by the normalization factor α, is forced to agree with the two measurements.

$\mathbf{X} := \ldots$ *(variable array)*
$\mathbf{V}_x := \ldots$ *(matrix array)*

$\texttt{f(1)} = \alpha \cdot x_{\text{ave}} - x_1$
$\texttt{f(2)} = \alpha \cdot x_{\text{ave}} - x_2$

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_m \\ \hline \mathbf{X}_u \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \alpha \\ \hline x_{\text{ave}} \end{pmatrix}$$

variable	measured		fit result		pull
x_1	8.0	$\pm 2\%$	8.235	± 0.116	2.14
x_2	8.5	$\pm 2\%$	8.235	± 0.116	-2.14
α	1	$\pm 10\%$	1.000	± 0.100	-2.14
x_{ave}			8.235	± 0.832	

$\chi^2 = 4.6$
$n_{\text{df}} = 1$
p-value $= 3.2\%$

$\rightsquigarrow$ no problem with normalization uncertainty with constrained least squares.
Data contain **no** information about the normalization $\rightsquigarrow$ normalization factor unchanged!

7. Triangle parameters

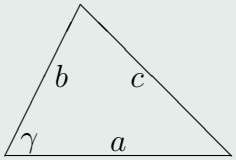
Determination of triangle parameters:

- three sides a, b and c are measured,
- one angle γ is measured.

Three values are sufficient for a complete definition of a triangle. Thus the least squares method can be used to *improve* the measured values. The parameter of interest is assumed to be the

- triangle area A with uncertainty ("error" propagation):

$$A = \sqrt{p(p-a)(p-b)(p-c)} = p\,(p-c)\tan\gamma/2 \qquad \text{with} \quad p = (a+b+c)/2$$



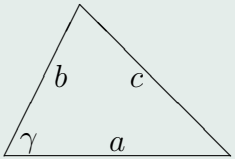
The least squares problem with 4 measured value, evtl. with non-diagonal covariance matrix, and 1 (unmeasured) parameter can be solved easily by `aplcon`, including propagation of uncertainties.

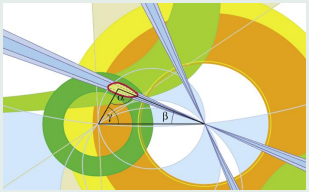
$\mathbf{X} := \ldots$ *(variable array)*
$\mathbf{V}_x := \ldots$ *(matrix array)*

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_m \\ \hline \mathbf{X}_u \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \\ \gamma \\ \hline A \end{pmatrix}$$

$p = (a+b+c)/2$! circumference/2
$A = \sqrt{p(p-a)(p-b)(p-c)}$! area of triangle
`f(1)` $= \tan(\gamma/2) - A/(p(p-c))$! angle constraint

variable	measured		fit result		pull
a	10	± 0.05	10.01	± 0.05	1.75
b	7	± 0.2	7.06	± 0.20	1.75
c	9	± 0.2	8.72	± 0.12	-1.75
γ	1	± 0.02	1.019	± 0.017	1.75
A			30.10	± 0.87	





$\rightsquigarrow$ e.g. unitarity triangle, representing interactions between quarks

8. Correlated additive systematic errors II

$x_1 \pm \sigma_1 \pm \Delta_{a1} \pm \Delta_{b1}$
$x_2 \pm \sigma_2 \pm \Delta_{a2} \pm \Delta_{b2}$

statistical errors σ_1 and σ_2 uncorrelated;
but sources a and b of systematic errors assumed to be fully (positive) correlated for Δ_{a1}, Δ_{a2} and for Δ_{b1}, Δ_{b2}

- Either define non-diagonal covariance matrix by law of (linear) propagation of uncertainties:

$$\boldsymbol{V}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 + \Delta_{a1}^2 + \Delta_{b1}^2 & \Delta_{a1}\Delta_{a2} + \Delta_{b1}\Delta_{b2} \\ \Delta_{a1}\Delta_{a2} + \Delta_{b1}\Delta_{b2} & \sigma_2^2 + \Delta_{a2}^2 + \Delta_{b2}^2 \end{pmatrix}$$

  i.e. insert all systematic uncertainty contributions into covariance matrix, which has to be inverted for the fit.

- or introduce one additional measured variable for each source of uncertainty:

... as correlated data ...

$$x_1 = x_1' + a \cdot \Delta_{1a} + b \cdot \Delta_{1b}$$
$$x_2 = x_2' + a \cdot \Delta_{2a} + b \cdot \Delta_{2b}$$

$$\begin{array}{rrr} x_1' & \pm\,\sigma_1 \\ x_2' & \pm\,\sigma_2 \\ a = & 0 & \pm\,1 \\ b = & 0 & \pm\,1 \end{array} \qquad \boldsymbol{V}\begin{pmatrix} x_1' \\ x_2' \\ a \\ b \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

... using `aplcon` with constraints: e.g. pull for each source of systematic uncertainty avaialable.

... both methods are equivalent (for additive uncertainties).

9. Correlated multiplicative systematic errors II

Multiplicative systematic errors (e.g. normalization) are expressed by *relative* errors $\Delta_{\rm rel}$, sometimes expressed as a factor $(1 + \Delta_{\rm rel})$ applied to measured or fitted values x.

Now errors Δ_{a1}, Δ_{b1}, Δ_{a2}, Δ_{b2} are defined as relative errors!

1. factor $(1 \pm \Delta_{a1})\,(1 \pm \Delta_{b1}) \cdots$
2. factor $(1 \pm \Delta_{a2})\,(1 \pm \Delta_{b2}) \cdots$

statistical errors σ_1 and σ_2 uncorrelated;
but systematic multiplicative relative errors are assumed to be fully (positive) correlated for Δ_{a1}, Δ_{a2} and for Δ_{b1}, Δ_{b2}

The factors can be assumed to follow the log-normal distribution:

$$\exp\left(\Delta_{\rm rel}\right) \approx 1 + \Delta_{\rm rel} + \frac{1}{2!}\Delta_{\rm rel}^2 + \ldots$$

Factors, applied to fitted values x:

1. factor $(1 + a \cdot \Delta_{a1})\,(1 + b \cdot \Delta_{b1}) \cdots \quad \rightarrow \quad \exp\left(a \cdot \Delta_{a1}\right) \exp\left(b \cdot \Delta_{b1}\right) \cdots$
2. factor $(1 + a \cdot \Delta_{a2})\,(1 + b \cdot \Delta_{b2}) \cdots \quad \rightarrow \quad \exp\left(a \cdot \Delta_{a2}\right) \exp\left(b \cdot \Delta_{b2}\right) \cdots$

$$\begin{array}{rrl} & x_1' & \pm\,\sigma_1 \\ & x_2' & \pm\,\sigma_2 \\ a = & 1 & \pm\,1 \\ b = & 1 & \pm\,1 \end{array} \qquad \boldsymbol{V} \begin{pmatrix} x_1' \\ x_2' \\ a \\ b \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

10. Combination of FNAL results on the mass of the top quark

2010 data: 11 measurement and 14 systematic uncertainty categories

Experiment	Data	mass ± stat	weight w_i
CDF Run I	l+j	176.1 ± 5.1	−0.025
	di-l	167.4 ± 10.3	−0.005
	all-j	186.0 ± 10.0	−0.007
D0 Run I	l+j	180.1 ± 3.6	+0.013
	di-l	168.4 ± 12.3	+0.002
CDF Run II (publ)	all-j	174.80 ± 1.70	+0.105
	trk	175.30 ± 6.20	−0.005
DO Run II (prel)	l+j*	173.75 ± 0.83	+0.262
	di-l	174.66 ± 2.92	−0.021
CDF RUN II (prel)	l+j*	173.00 ± 0.65	+0.700
	d1-l	170.56 ± 2.19	−0.018

Systematic uncertainty categories
top mass uncertainty
Combined values (GeV/c^2)

cat.	Δm	cat.	Δm
iJES	0.46	Signal	0.19
aJES	0.21	Backgr.	0.23
bJES	0.20	Fit	0.11
cJES	0.13	MC	0.40
dJES	0.19	UN/MI	0.02
rJES	0.15	CR	0.39
LepPt	0.15	MHI	0.09

FERMILAB-TM-2466-E (arXiv:1007.3178v1 [hep-ex]): Combination of CDF and D0 results on the mass of the top quark using up to . . . ,

based on BLUE (best linear unbiased estimate), all systematic uncertainties treated as additive uncertainties, uncorrelated (Fit, iJES), or with full (positive) correlation either for Runs, for experiments, for channels.

From the paper: "The weights of some of the measurements are negative . . . a negative weight means that it affects the resulting m_{top} central value and helps reduce the total uncertainty."

Averaging with `aplcon`, treating all uncertainties (except statistical error and multiple hadron interactions) as multiplicative, assuming log-normal factors (see before).

Comparison

2009 $m_{\text{top}} = 173.12 \pm 1.26$ GeV/c^2

2010 $m_{\text{top}} = 173.32 \pm 1.06$ GeV/c^2

`aplcon` $m_{\text{top}} = 173.09 \pm 1.13$ GeV/c^2

2011 $m_{\text{top}} = 173.18 \pm 0.94$ GeV/c^2

Note: average mass value dominated by two experimental values.

In `aplcon` for each systematic error category the corresponding factor (a, b ...) is fitted and the pull is calculated, but no weights (no weights w_i with $\sum_i w_i = 1$) exist.
The calculated uncertainty in `aplcon` is slightly larger.

11. Branching ratios

In total 16 values $x_1 \ldots x_{16}$ measured for five different decay channels of the $f_1(1285)$:

$$\frac{\Gamma_4}{\Gamma_1} \qquad \frac{\Gamma_4}{(\Gamma_2+\Gamma_3)} \qquad \frac{\Gamma_2}{(\Gamma_2+\Gamma_3)} \qquad \frac{\Gamma_1}{(\Gamma_2+\Gamma_3)} \qquad \frac{\Gamma_5}{1/3\Gamma_1} \qquad \frac{\Gamma_4}{\Gamma} \qquad \frac{(\Gamma_2+\Gamma_3)}{\Gamma_5}$$

Branching rations $B_1 \ldots B_5$ determined by constrained fit using all data with constraints e.g.

$$f(.) = B_4 - X_1 \cdot B_1 \qquad \ldots \qquad f(.) = 3 \cdot B_5 - B_1 \cdot X_{13} \qquad \ldots \qquad f(.) = \sum_j B_j - 1$$

Symbol	Mode	PDG-Fraction Γ_i/Γ	aplcon: Γ_i/Γ	$(\Gamma_i/\Gamma)^*$
B_1	$f_1(1285) \to 4\pi$	$(33.1^{+2.1}_{-1.9})\%$	$(29.7 \pm 1.7)\%$	$(31.9 \pm 1.6)\%$
B_2	$f_1(1285) \to a_0(980)\pi$	$(36 \pm 7)\%$	$(39 \pm 8)\%$	$(37 \pm 8)\%$
B_3	$f_1(1285) \to \eta\pi\pi$	$(16 \pm 7)\%$	$(16 \pm 8)\%$	$(16 \pm 7)\%$
B_4	$f_1(1285) \to K\bar{K}\pi$	$(9.0 \pm 0.4)\%$	$(7.7 \pm 0.4)\%$	$(8.6 \pm 0.4)\%$
B_5	$f_1(1285) \to \gamma\rho_0$	$(5.5 \pm 1.3)\%$	$(6.8 \pm 0.7)\%$	$(6.4 \pm 0.6)\%$

PDG: 10 scale factors up to $S = 2.8$ are applied to original data;

aplcon 1.result: result with unmodifeed original data;

aplcon 2.result*: scale factor $S = 2.8$ applied to the value with largest pull.

Summary

Data combination

- Increase of precision by data combination, and test for data compatibility;
- Requires understanding of physics, detector behaviour and data analysis, ...
- ... and estimation of all relevant systematic uncertainties, ...
- ... and of concept of statistical correlations and distributions,

Constrained least squares fits allow solution of complicated combination problems:

- non-linearities in the model;
- model relations with > 1 measured values and with correlated uncertainties;
- non-Gaussian variables, e.g. following the Poisson and the log-normal distribution

with simple code, reduced to the essentials.

Contents