Best fit mehods

Least squares

Maximum likelihood

When the data are gaussian, the ML method gives

$$\text{Max } L = \prod_i \text{Max} \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left[-\frac{1}{2} \frac{(y_i - \mu_i)^2}{\sigma_i^2} \right] \rightarrow \text{ Min } \sum_i \frac{(y_i - \mu_i)^2}{\sigma_i^2}$$

If $y_1, y_2, \ldots, y_n$ are the observed values of $Y_1, Y_2, \ldots, Y_n$ random variables such as

$$\langle Y_i \rangle = \mu_i(\boldsymbol{x}_i, \boldsymbol{\theta}) \equiv \mu_i(\boldsymbol{\theta})$$

where $\boldsymbol{\theta}$ is a set of p parameters and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are the observed values af the predictor and the variances $\mathrm{Var}[Y_i] = \sigma_i^2$ are known. The Least Squares stimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ minimizes the quantity

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{[y_i - \mu_i(\boldsymbol{\theta})]^2}{\sigma_i^2} \ .$$

Here the prior knowledge of the distribution function is not requested.

The method of the Least Squares

χ^2 Degrees of Freedom

From probability calculus

If

$$HZ = X - \mu$$

the Q variable in general is

$$Q \equiv (X - \mu)^\dagger V^{-1} (X - \mu) = Z^\dagger H^\dagger V^{-1} H Z$$
$$= Z^\dagger Z = \sum_{i=1}^{n} Z_i^2$$

Therefore, $Q \sim \chi^2(n)$.

When $|V| = 0$, the general theorem on the quadratic forms says that it is possible to find $n - p$ new independent variables such as

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n-p} T_i^2$$

In conclusion, when

$$Q = (X - \mu)^\dagger W (X - \mu) \ ,$$

one must verify if $|W| = 0$. In this case one has to find new variable such as

$$(X - \mu)^\dagger W (X - \mu) = \sum_{i=1}^{n-p} T_i^2$$

The important principle is that these new variables have not to be found, to reduce the DoF is enough! $\nu = n - p$ (points - equations)

Probability Intervals

- In general: $\int_D p(x,y)\,\mathrm{d}x\,\mathrm{d}y$,
  often D is the ellipse:

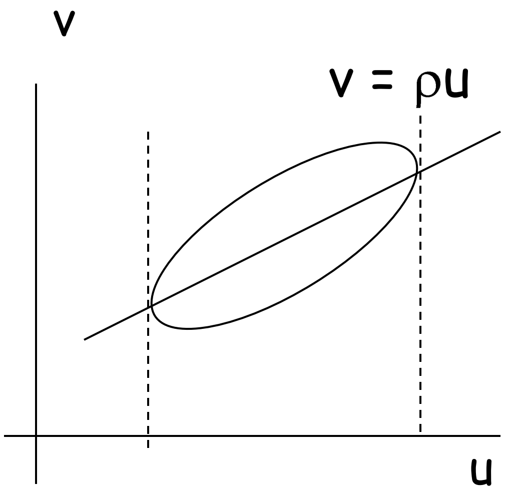
$$Q = (\boldsymbol{X}-\boldsymbol{\mu})^\dagger V^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) \;\rightarrow\; \sum_{i=1}^{n} \frac{(X_i-\mu_i)^2}{\sigma_i^2} \equiv \sum_i Q_i$$

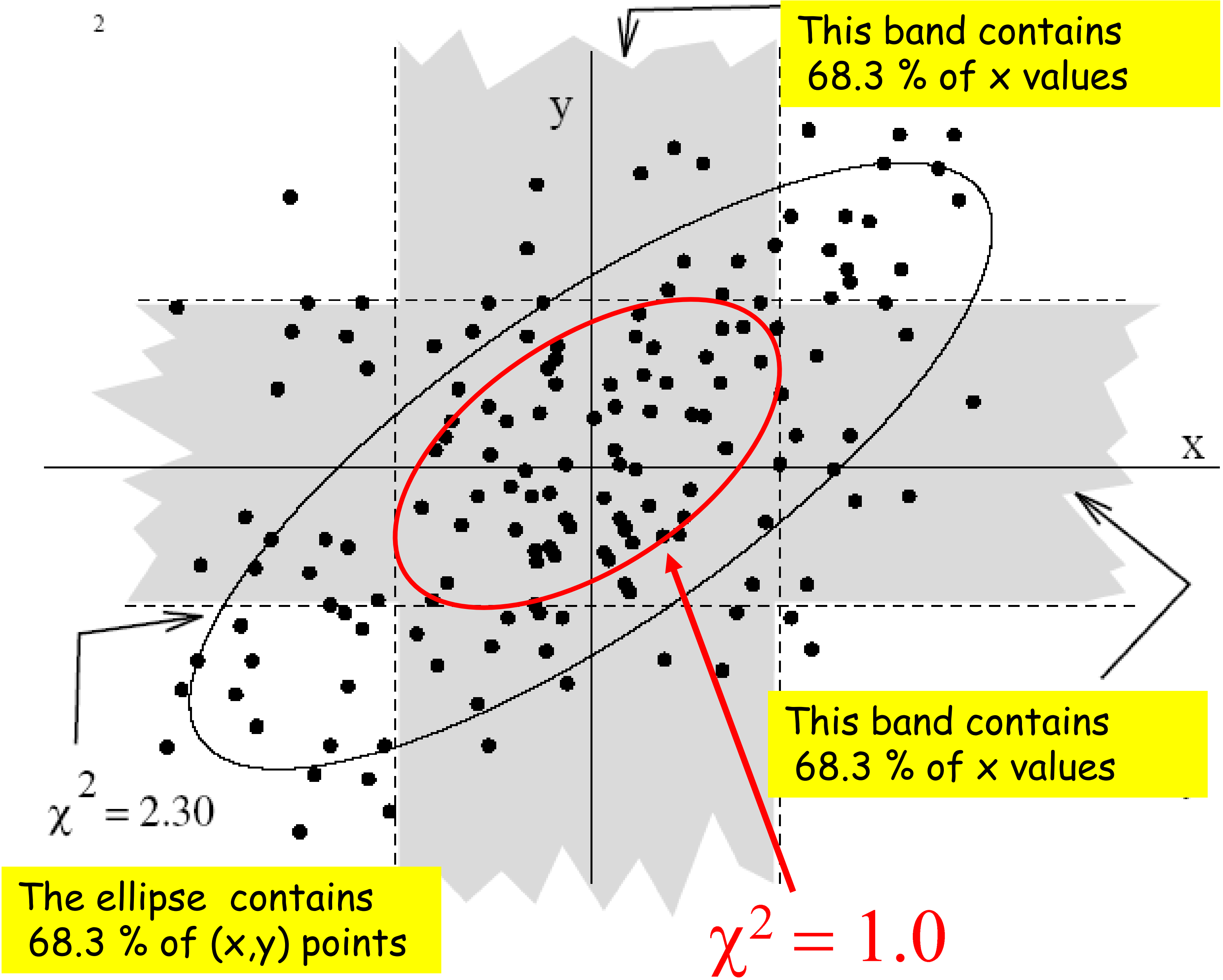
- $(x \pm \sigma_x)$ for each y
- $(y \pm \sigma_y)$ for each x
- $(\langle Y|x\rangle \pm \sqrt{\mathrm{Var}[Y|x]})$
- $(\langle X|y\rangle \pm \sqrt{\mathrm{Var}[X|Y]})$

Q_i follows the $\chi^2(1)$ density, that is $\{Q_i = 1, 4, 9\}$ corresponds to probability levels of 68.3, 95.4 and99.7% In general $\chi^2 = 0$ defines the center and $\chi^2 = 1$ defines 1-σ intervals of the marginal distributions. Example for standard variables in 2D:

$$Q = \gamma(X,Y) = \frac{1}{1-\rho^2}(u^2 - 2\rho uv + v^2) = 1$$

and the intersection with $v = \rho u$ is $u = \pm 1$ and corresponds to $(\mu_x \pm \sigma_x)$.





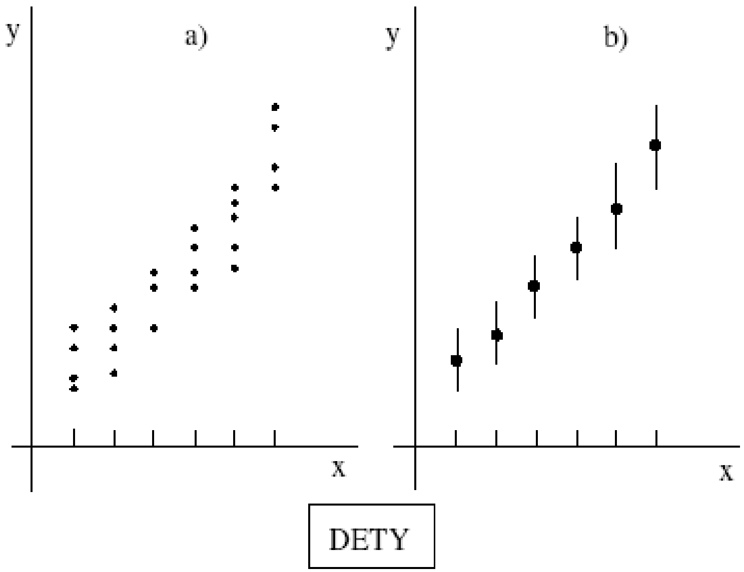
2
This band contains
68.3 % of x values
y
x
This band contains
68.3 % of x values
$\chi^2 = 2.30$
The ellipse contains
68.3 % of (x,y) points
$\chi^2 = 1.0$

Least Squares Case I

The predictors are known:

$$Y = f(x) + Y_R, \quad f(x) \equiv f(x, \boldsymbol{\theta}) \ ,$$

One can always assume $\langle Y \rangle = f(x)$ and $\langle Y_R \rangle = 0$, because if $\langle Y_R \rangle = y_0$ one can set $f(x) + y_0$.



When the model is right the minimum χ^2

$$\chi^2(\boldsymbol{\theta}) = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma_i^2} \ ,$$

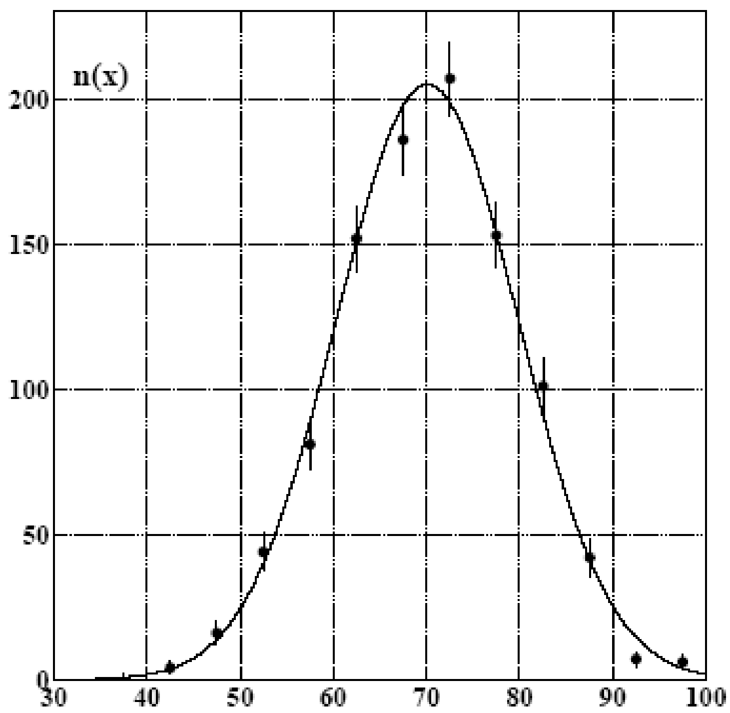
gives the curve $f(x_i, \boldsymbol{\theta})$ with respect to which the point fluctuations are random.

How to decide this? See the χ^2 test below.

χ^2 Test

From Statistics

- 9 points (69 %) in $\pm\sigma_i$;
- 12 points (92 %) in $\pm 2\sigma_i$;
- 13 points (100 %)in $\pm 3\sigma_i$.



More quantitatiely: χ^2 test

$$\chi^2 = \sum_{i=1}^{K} \frac{(n_i - \mu_i)^2}{\mu_i} = \sum_{i=1}^{K} \frac{(n_i - Np_i)^2}{Np_i} .$$

When N is fixed the DoF are K-1

χ^2 Test

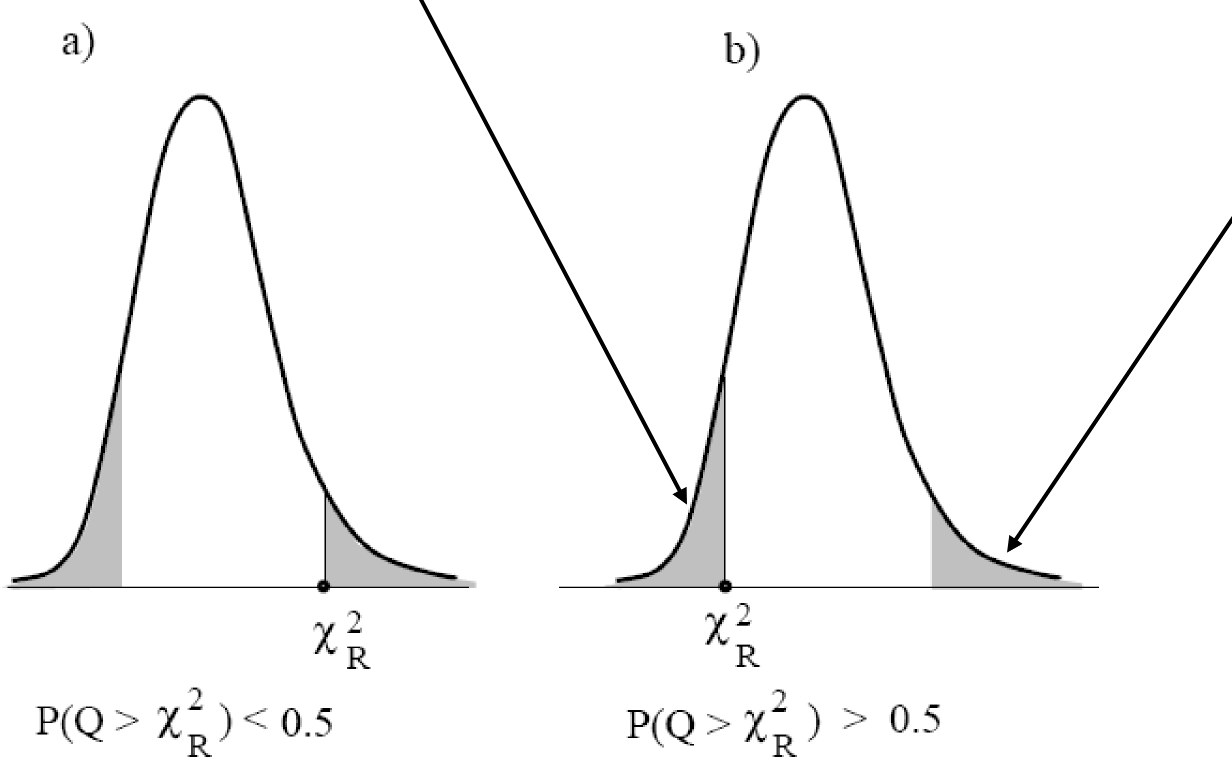
- **one tail test:** SL (o *p-value*) is:

$$SL = P\{Q_R \geq \chi^2_R(\nu)\} \ .$$

- **two tail test:**

$$SL = 2P\left\{Q_R(\nu) > \chi^2_R(\nu)\right\} \ \text{se} \ P\left\{Q_R(\nu) > \chi^2_R(\nu)\right\} < 0.5$$

$$SL = 2P\left\{Q_R(\nu) < \chi^2_R(\nu)\right\} \ \text{se} \ P\left\{Q_R(\nu) > \chi^2_R(\nu)\right\} > 0.5$$



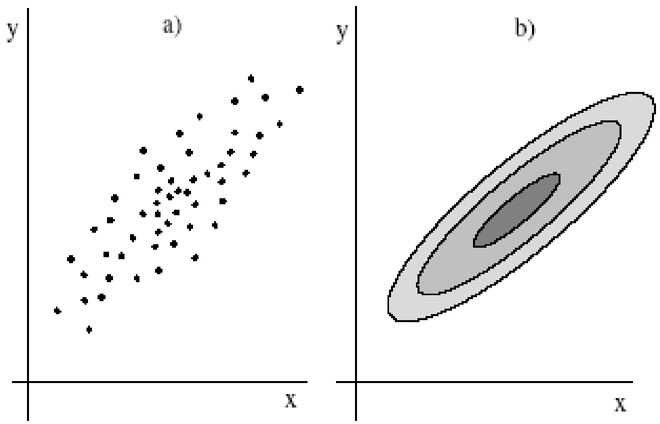
Shaded areas: probability to be wrong if we reject the model (type I error)

Least Squares Case II

Both the variables are random

$$X = x_0 + X_R\ , \quad Y = f(X) + Y_R, \quad f(X) \equiv f(X, \boldsymbol{\theta})\ ,$$

where x_0 is a constant and X_R and Y_R are random ariables with zero mean



STATXY

The minimum χ^2 finds in this case the correlation function $f(X)$:

$$\chi^2(\boldsymbol{\theta}) = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\mathrm{Var}[Y_i|x_i]}\ ,$$

The denominator is the variance of Y for any fixed x_i.

Fortunately, in many cases it is constant,

$$\mathrm{Var}[Y_i|x_i] = \sigma^2$$

Case II

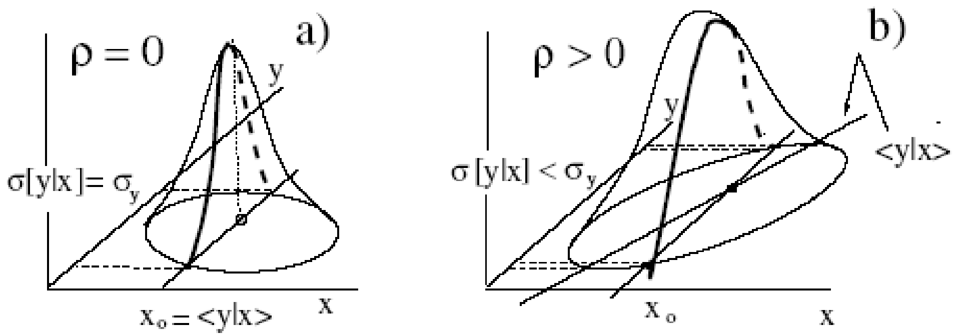
Since $\sigma_y^2 \equiv \mathrm{Var}[Y] = \mathrm{Var}[f(X)] + \mathrm{Var}[Y_R|x]$ and $\mathrm{Var}[Y|x] = \mathrm{Var}[Y_R|x]$, we have

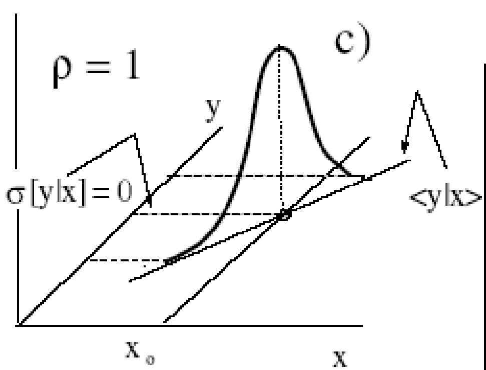
$$\mathrm{Var}[Y|x] = \mathrm{Var}[Y_R|x] = \sigma_y^2\left(1 - \frac{\mathrm{Var}[f(X)]}{\sigma_y^2}\right) .$$

which define the correlation coefficient between X and Y:

$$\rho = \pm\frac{\sigma[f(X)]}{\sigma[Y]} ,$$

Remember that in the gaussian case



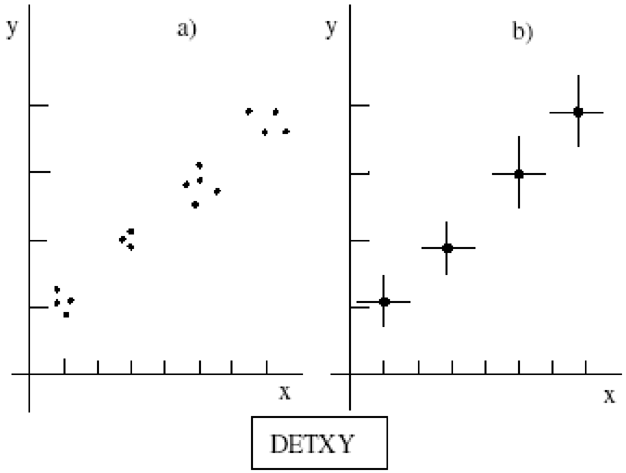


$$\langle y|x\rangle = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x)$$

$$\mathrm{Var}[y|x] = \sigma_y^2(1-\rho^2)$$

In this last case the predictors are measured with error as in many lab measurements:

$$X = x_0 + X_R \ , \quad Y = f(x_0) + Y_R \ , \quad f(x_0) \equiv f(x_0, \boldsymbol{\theta}) \ ,$$



We should minimize in this case:

$$\chi^2(\boldsymbol{\theta}, \underline{x_0}) = \sum_i \frac{(x_i - x_{0i})^2}{\sigma_{x_i}^2} + \sum_i \frac{[y_i - f(x_{0i}, \boldsymbol{\theta})]^2}{\sigma_{y_i}^2} \ .$$

with $n_{x_0} + n_\theta$ free parameters!

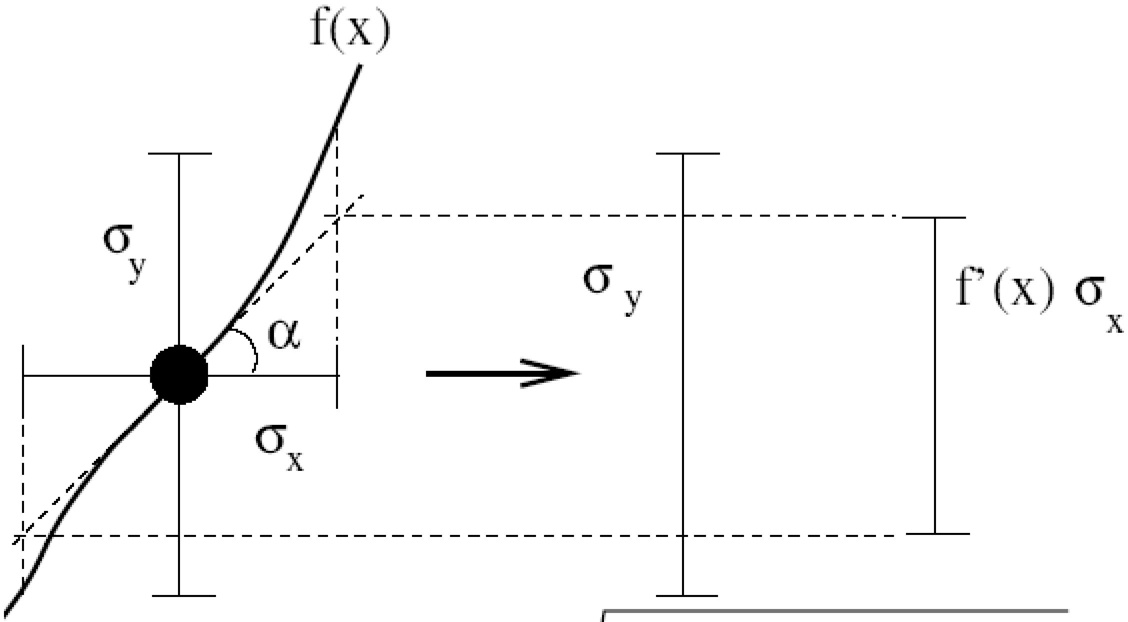
The DoF are $n_{x_0} + n_\theta - n_y = p$ as usual. One should use the effective variance formula:

$$\chi^2(\boldsymbol{\theta}) = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma_{Ei}^2} = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma_{y_i}^2 + f'^2(x_i, \boldsymbol{\theta})\sigma_{x_i}^2} \ ,$$

Effective variance (cont'd)

The geometrical interpretation of the method:

Case III



$$\sqrt{\sigma_y^2 + [f'(x)]^2 \sigma_x^2}$$

When there is a covariance:

$$\chi^2(\boldsymbol{\theta}) = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma_{y_i}^2 + f'^2(x_i, \boldsymbol{\theta})\sigma_{x_i}^2 - 2f'(x_i, \boldsymbol{\theta})\, \sigma_{x_i y_i}} \, .$$

Least Squares formulae

	x	y	χ^2
Standard case	$x = x_0$	$Y = f(x_0) + Y_R$	$\sum \frac{[y - f(x)]^2}{\mathrm{Var}[Y_R]}$
Common in lab measurements	$X = x_0 + X_R$	$Y = f(x_0) + Y_R$	$\sum \frac{[y - f(x)]^2}{\mathrm{Var}[Y_R] + f'^2\ \mathrm{Var}[X_R]}$
$f(x)$ is the correlation funtion	$X = x_0 + X_R$	$Y = f(X) + Y_R$	$\sum \frac{[y - f(x)]^2}{\mathrm{Var}[Y_R \| x]}$

Unknown errors

Once one has minimised, the χ^2 test will say if the chosen model is correct. Errors are crucial

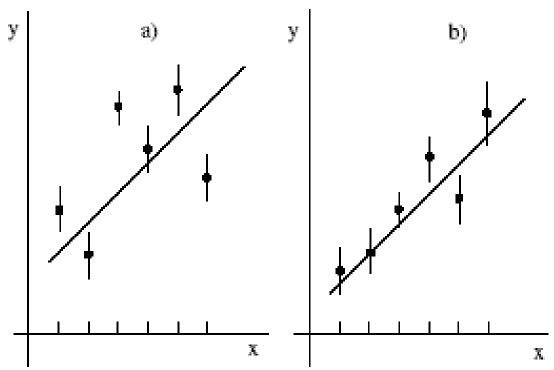


Figure 2: bad fit a) and good fit b); in the second case about the 68% of the points touch the regression curve within an error bar.

If the errors are unknown but all equal,

$$\sigma(y_i) \equiv \sigma \ ,$$

one can use the rescaling technique:

- perform a first fit with $\sigma = 1$: the value of the parameters is correct, but the parameter errors are wrong. Then calculate

$$\sigma^2_{yR} \simeq s^2_y = \frac{1}{n-p} \sum_{i=1}^{n} \left[y_i - \mu(x_i, \hat{\boldsymbol{\theta}})\right]^2 = \frac{\chi^2(\sigma_{yR} = 1, \hat{\boldsymbol{\theta}})}{n-p} \ .$$

- use the previous estimation of σ to refit the data.

Some warnings

First fault: use the error rescaling wihout knowing a priori the functional form:

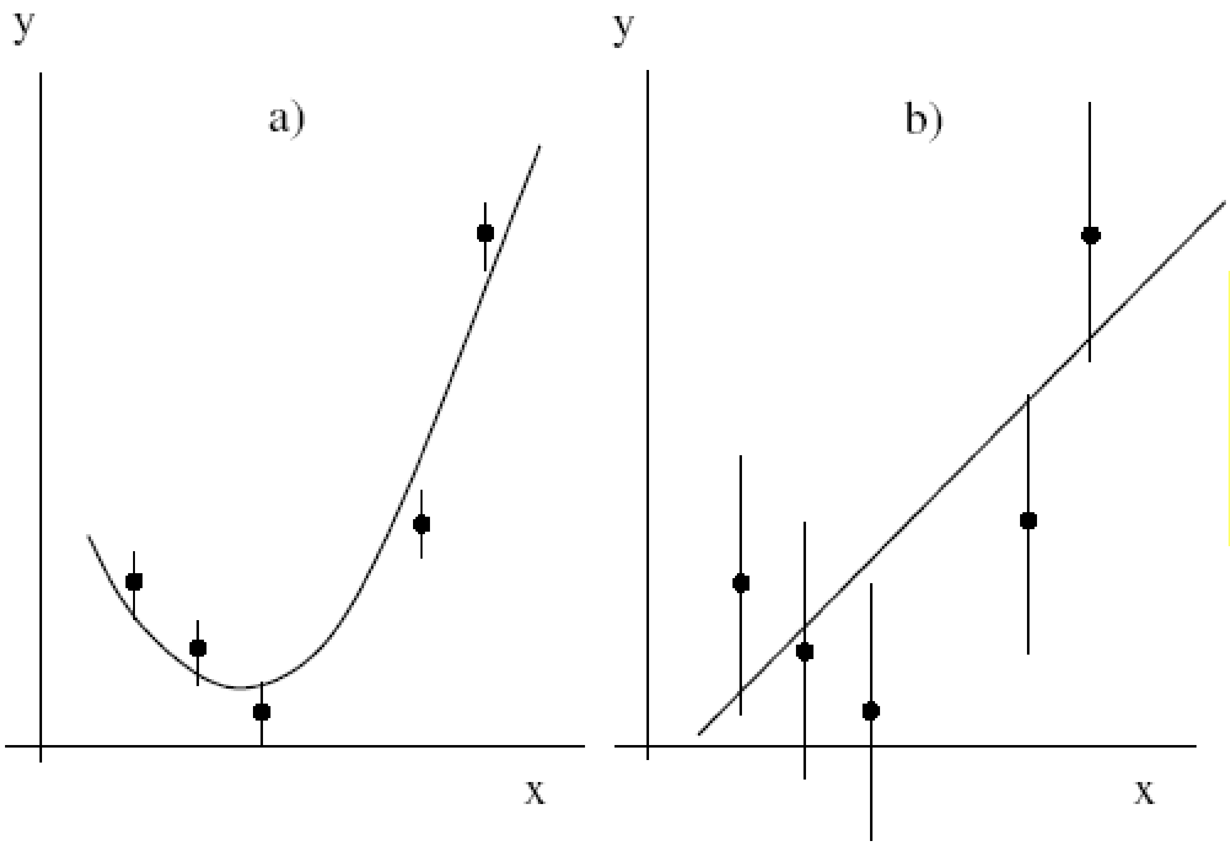


Figure 3: If the data come from a parabola with constant errors as in a), a fit asuming the line as a model and operating the error rescaling will find a good χ^2 overestimating uncorrectly the errors, as in b)

Some warnings

Second Fault: do not give the errors to the program when the errors are varying from point to point

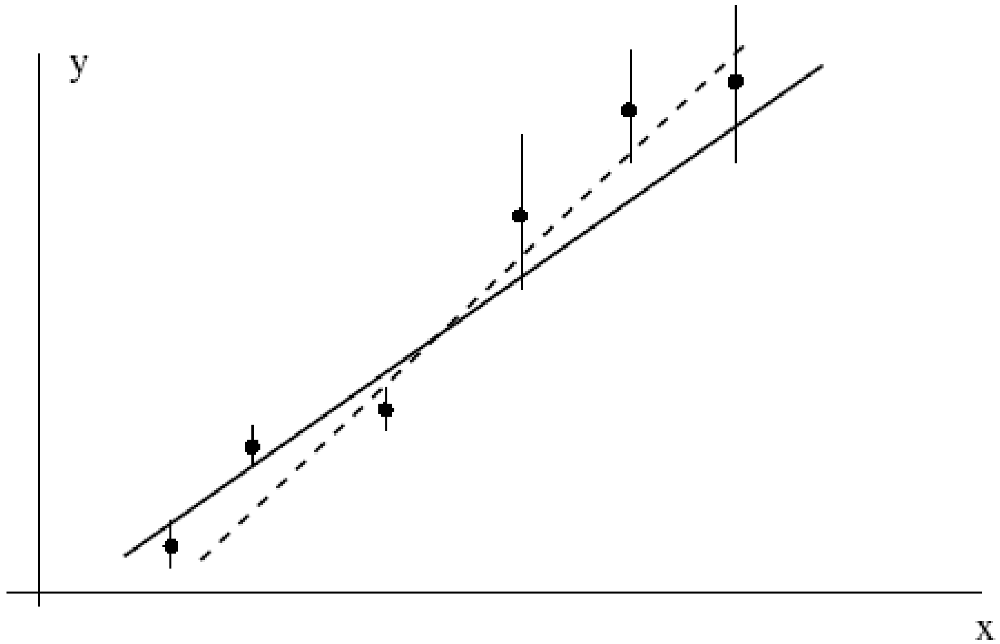


Figure 4: A weighted linear fit takes care of the points with a small error and gives the correct result (full line); the unweighted fit, that assumes equal errors, gives the same inportance to all the points and gives the wrong result (dotted line).

All the existing programs (MINUIT) do not warn you on these errors!!!

Some warnings

In general a linear model is

$$\langle Y|\boldsymbol{x}\rangle \equiv \mu(x,\boldsymbol{\theta}) = f(x,\boldsymbol{\theta}) = \sum_{k=1}^{p} \theta_k f_k(x) \ ,$$

and the minimzation of

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{[y_i - \Sigma_k \theta_k f_k(x_i)]^2}{\sigma_i^2} \ ,$$

leads to the normal equations:

$$\sum_{i=1}^{n} \left[\frac{1}{\sigma_i^2} \left(y_i - \sum_{j=1}^{p} \theta_j f_j(x_i) \right) f_k(x_i) \right] = 0 \ , \quad k = 1, 2, \ldots p \ ,$$

The standard matrices are the derivative, co-variance and the weight matrices

$$F = \begin{pmatrix} f_1(x_1) & f_2(x_1) & \ldots & f_p(x_1) \\ f_1(x_2) & f_2(x_2) & \ldots & f_p(x_2) \\ \ldots & \ldots & \ldots & \ldots \\ f_1(x_n) & f_2(x_n) & \ldots & f_p(x_n) \end{pmatrix} .$$

$$V = \begin{pmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \sigma_n^2 \end{pmatrix} \quad W = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 & \ldots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \frac{1}{\sigma_n^2} \end{pmatrix}$$

Linear Least Squares

$$\mu(x,\theta) = F\ \theta$$

Linear Least Squares

In matricial form we have:

$$\chi^2 = (F\boldsymbol{\theta} - \boldsymbol{y})^\dagger W\,(F\boldsymbol{\theta} - \boldsymbol{y})$$

sometimes expressed in the form:

$$\chi^2 = ||\boldsymbol{y} - F * \boldsymbol{\theta}||^2$$

The normal equations ($\frac{\partial \chi^2}{\partial \theta} = 0$) are:

$$F^\dagger W\,\boldsymbol{y} = (F^\dagger W F)\,\boldsymbol{\theta}$$

The more impotant object is the error matrix

$$F^\dagger W F = \begin{pmatrix} \sum_i \frac{f_1^2(x_i)}{\sigma_i^2} & \sum_i \frac{f_1(x_i) f_2(x_i)}{\sigma_i^2} & \dots & \sum_i \frac{f_1(x_i) f_p(x_i)}{\sigma_i^2} \\ \sum_i \frac{f_2(x_i) f_1(x_i)}{\sigma_i^2} & \sum_i \frac{f_2^2(x_i)}{\sigma_i^2} & \dots & \sum_i \frac{f_2(x_i) f_p(x_i)}{\sigma_i^2} \\ \dots & \dots & \dots & \dots \\ \sum_i \frac{f_p(x_i) f_1(x_i)}{\sigma_i^2} & \sum_i \frac{f_p(x_i) f_2(x_i)}{\sigma_i^2} & \dots & \sum_i \frac{f_p^2(x_i)}{\sigma_i^2} \end{pmatrix},$$

and also

$$F^\dagger\,W = \begin{pmatrix} \frac{f_1(x_1)}{\sigma_1^2} & \frac{f_1(x_2)}{\sigma_2^2} & \dots & \frac{f_1(x_n)}{\sigma_n^2} \\ \frac{f_2(x_1)}{\sigma_1^2} & \frac{f_2(x_2)}{\sigma_2^2} & \dots & \frac{f_2(x_n)}{\sigma_n^2} \\ \dots & \dots & \dots & \dots \\ \frac{f_p(x_1)}{\sigma_1^2} & \frac{f_p(x_2)}{\sigma_2^2} & \dots & \frac{f_p(x_n)}{\sigma_n^2} \end{pmatrix}.$$

Theorems on Least Squares

- when the dependence on the parameters is linear, Ls gives unbiased estimations

$$\langle \hat{\boldsymbol{\theta}} \rangle = (F^{\dagger} W F)^{-1} (F^{\dagger} W) \langle \boldsymbol{y} \rangle = (F^{\dagger} W F)^{-1} F^{\dagger} W F \boldsymbol{\theta} = \boldsymbol{\theta} ,$$

- Gauss Markov theorem
  when the dependence on the parameters is linear, the LS estimator is the best one, that is that with the minimum variance

The Gauss Markov theorem extends the properties of the LS estimator from the gaussian case (where it coincides with the ML one), to a more general case, valid also for non gaussian variables in the case of linear models

The properties of the LS estimator for non gaussian variables and non linear models are not known in general.

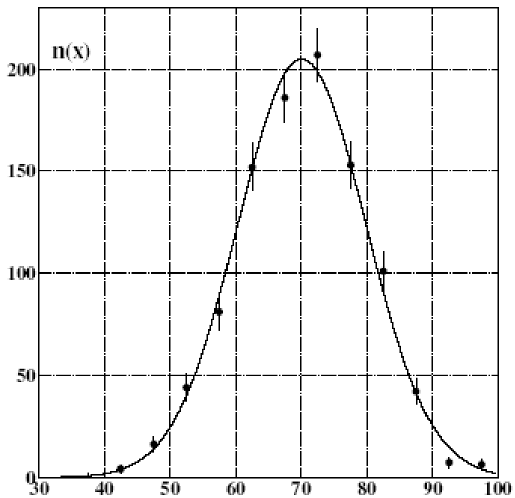
They can be studied case by case with Monte Carlo methods

Binned and unbinned likelihood

Binned likelihood

$$L(\theta) = \prod_{i=1}^{k} p_i(\theta)^{n_i} \quad \Rightarrow$$

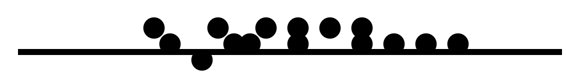
$$-\ln L(\theta, n) = -\sum_{i=1}^{k} n_i \ln[p_i(\theta)]$$



k bins

Unbinned likelihood

$$L(\theta) = \prod_{i=1}^{N} p(x_i, \theta)$$

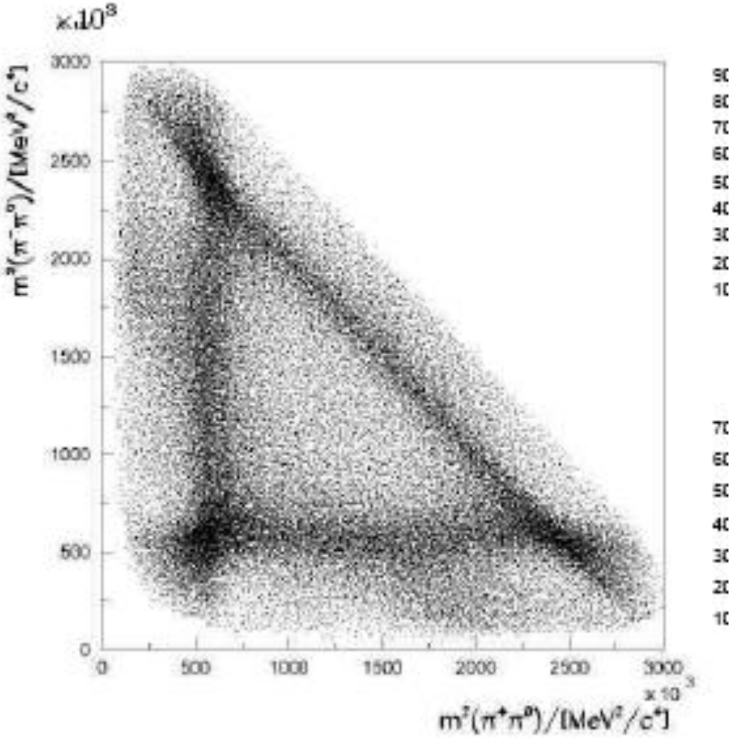


N points

Unbinned likelihood

$$W = T_{i \to f} \; Lips(p_n)$$

First method: generate MC events following phase space, weight them with $T = |<f|T|i>|^2$ and compare with binned data



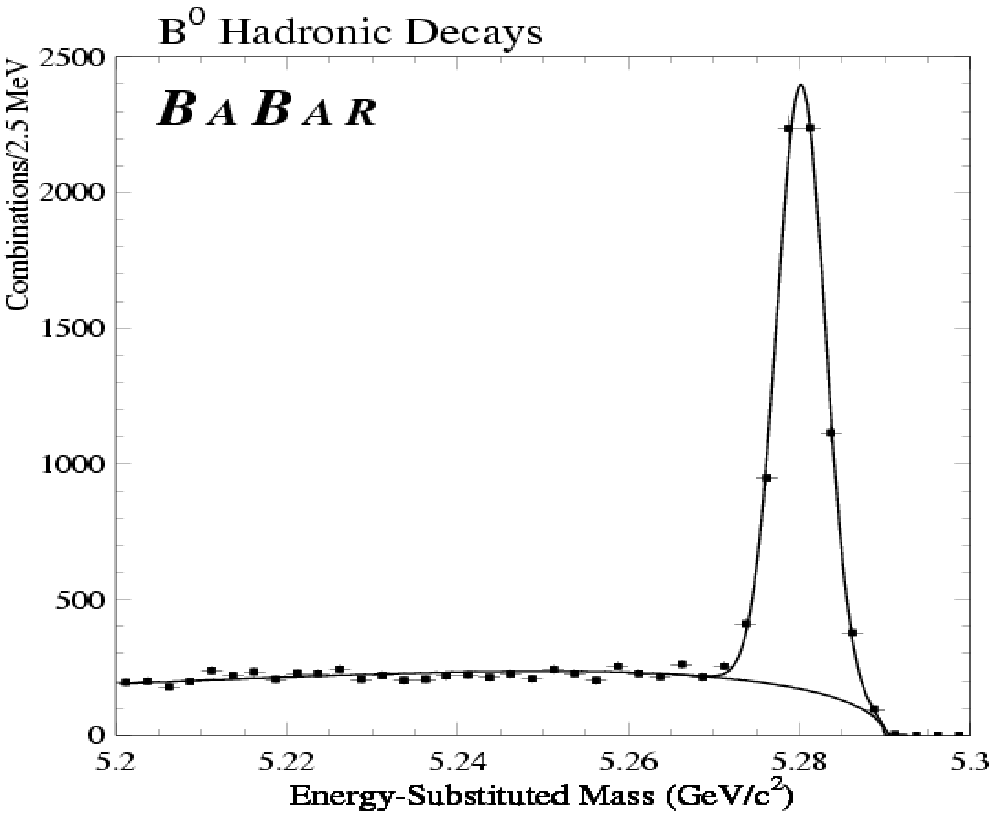
Second method, unbinned likelihood: fill the transition matrix with the measured momenta and maximize

if p_n are the **MEASURED** $4-$momenta

$$L(M,\Gamma) = \prod_{n=1}^{N} T_{i \to f}(p_n, M, \Gamma)$$

Unbinned likelihood

$$L(m, w, A, C, D) = \prod_{n=1}^{N} A\,\text{Breit}(x_n, m, w) + Cx_n + D$$



The extended likelihood

$$L(\theta,\underline{n}) = \prod_i \frac{\mu_i^{n_i}}{n_i!} e^{-\mu_i}$$

$$-\ln L(\theta,n) = -\sum_{i=1}^{k} n_i \ln[\mu_i(\theta)] + \sum_{i=1}^{k} \mu_i(\theta)$$

Since $\mu_i = N\, p_i(\theta), \ \sum \mu_i = N(\theta) \sum p_i(\theta) = N(\theta)$

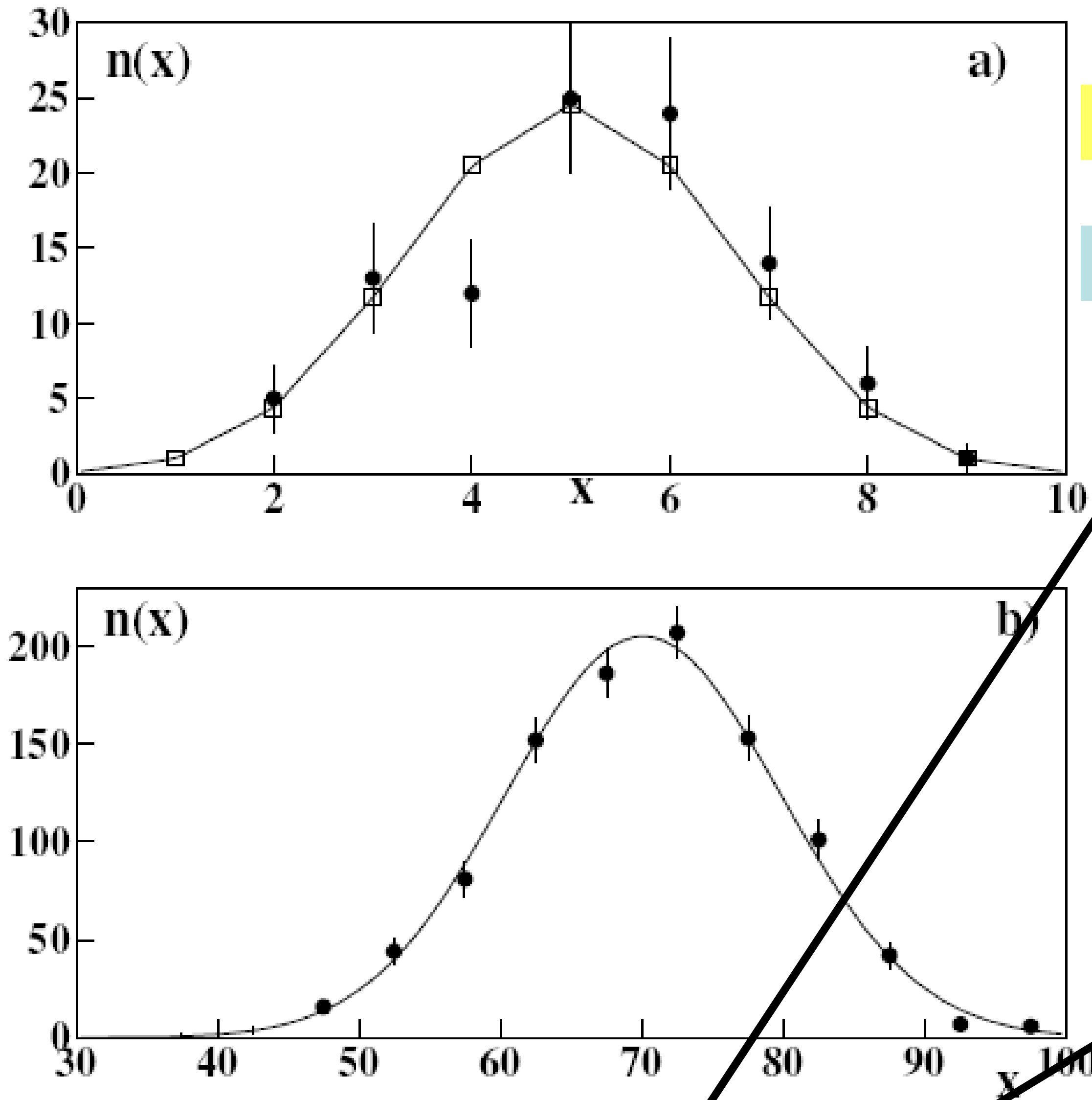
when N is a function of θ as in the case of a detector efficiency,

If there is no functional relation between N and θ

the result is the same as for the non extended likelihood

$$L(\theta) = \prod_{i=1}^{k} p_i(\theta)^{n_i} \qquad \Rightarrow \qquad -\ln L(\theta,n) = -\sum_{i=1}^{k} n_i \ln[p_i(\theta)]$$

Binomial
p=0.5



p= 0.522 0.015

p= 0.528 0.017

Gaussian
μ=70
σ=10

μ =70.09 0.31
σ =9.73 0.22

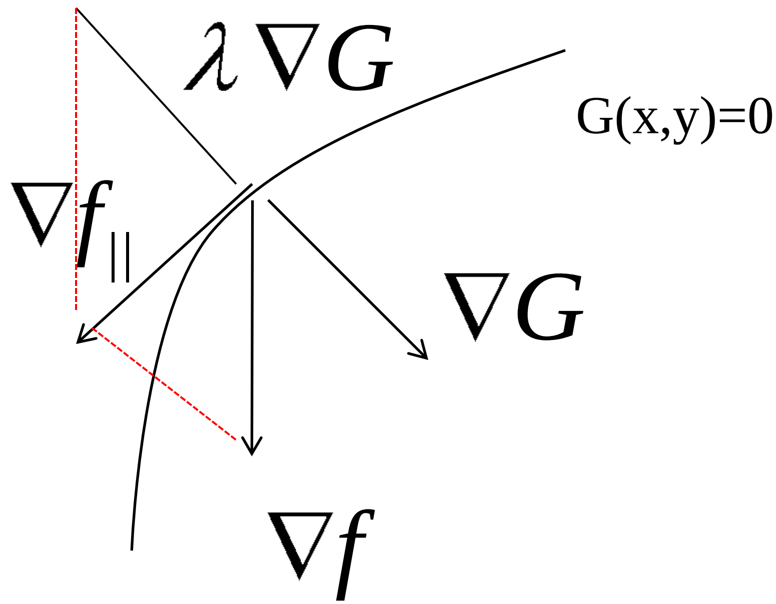
μ =69.97 0.31
σ =9.59 0.22

$\mathcal{L}$

χ^2

Lagrange multipliers

A function f(x,y) to be minimized with the constraint G(x,y)=0



The constrained minimum condition is $\nabla f_{||} = 0$. Hence:

$$\nabla f_{||} = \nabla f + \lambda \nabla G = 0 \text{ with the constraint } G(x,y) = 0 \text{ implies to minimize}$$

$$S = f + \lambda G$$

w.r.t. all free parameters and λ

Degrees of Freedom

Unconstrained χ^2 does not work:

$$\chi^2 = \sum_i \frac{(y_i - a_i)^2}{\sigma_i^2} \quad \Rightarrow \quad a_i = y_i \qquad \text{DOF} = 0$$

Constraint with an internal function:

$$\chi^2 = \sum_i \frac{(y_i - f(a, x_i))^2}{\sigma_i^2} \quad \Rightarrow \qquad \text{DOF} = n(y) - n(a)$$

Constraint with with an external function:

$$\chi^2 = \sum_i \frac{(y_i - a_i)^2}{\sigma_i^2} + \sum_k \lambda_k \, \phi_k(a, z) \quad \Rightarrow \quad n(y) = n(a)$$

$$\text{DOF} = \underbrace{n(y) - n(a)}_{=0} + [n(\phi) - n(z)] = n(\phi) - n(z)$$

Then, we have:

$$\mathrm{d}(\chi^2 + \lambda_k \Phi_k) = \sum_{j=1}^{p} \frac{\partial}{\partial \mu_j}[\chi^2 + \lambda_k \Phi_k]\, \mathrm{d}\mu_j + \sum_{n=1}^{q} \lambda_k \frac{\partial \Phi_k}{\partial \varphi_n}\, \mathrm{d}\varphi_n = 0 \ ,$$
$$k = 1, 2 \ldots, M \ .$$

with independnt $\mathrm{d}\mu_j$.

The parameter independence implies all null derivatives!

The constraint equations are equivalent to the derivatives w.r.t. λ: hence the objective function to be minimized w.r.t. $\mu_j, \varphi_n, \lambda_k$, with $j = 1, 2, \ldots, p,\ \ n = 1, 2, \ldots, q,\ \ k = 1, 2, \ldots, M$ is

$$S(\boldsymbol{\mu}, \boldsymbol{\varphi}, \boldsymbol{\lambda}) = \left[\chi^2(\boldsymbol{\mu}) + 2 \sum_{k=1}^{M} \lambda_k\, \Phi_k(\boldsymbol{\mu}, \boldsymbol{\varphi})\right] \qquad (6)$$

- μ, λ and φ give a number of equations $= p + M + q$
- degrees of fredom: $\nu = n - (p - M + q) = \nu = n - p + M - q$
  number of points - free parameters + constraints - non-measured parameters

Minimization with constraints

Minimization with Constraints: example

The measurement of the angles of two points on a line is

$$y_1 = 20^o \pm 2^o$$
$$y_2 = 195^o \pm 2^o \; .$$

Optmize the measurement taking into account the constraint

$$\langle Y_2 \rangle - \langle Y_1 \rangle - \pi = 0 \; ,$$

where $\langle Y_1 \rangle$ e $\langle Y_2 \rangle$ are the parameter to optimize.
The objective unction is

$$S(\langle Y_1 \rangle, \langle Y_2 \rangle, \lambda) = \frac{(y_1 - \langle Y_1 \rangle)^2}{s^2} + \frac{(y_2 - \langle Y_2 \rangle)^2}{s^2} + 2\,\lambda(\langle Y_2 \rangle - \langle Y_1 \rangle - \pi) \; ,$$

where $s = 2^o$.

$$\frac{\partial S}{\partial \langle Y_1 \rangle} = \frac{y_1 - \langle Y_1 \rangle}{s^2} + \lambda = 0 \; ,$$
$$\frac{\partial S}{\partial \langle Y_2 \rangle} = \frac{y_2 - \langle Y_2 \rangle}{s^2} - \lambda = 0 \; ,$$
$$\frac{\partial S}{\partial \lambda} = \langle Y_2 \rangle - \langle Y_1 \rangle - \pi = 0 \; ,$$

Minimization with Constraints: example

and the solutions is

$$\langle \hat{Y}_1 \rangle = \frac{1}{2}(y_1 + y_2 - \pi) \ ,$$
$$\langle \hat{Y}_2 \rangle = \frac{1}{2}(y_1 + y_2 + \pi) \ , \qquad (7)$$
$$\lambda = \frac{1}{2}\frac{1}{s^2}(y_2 - y_1 - \pi) \ .$$

where λ has not physical significance.

Error prpagation gives:

$$\sigma\left[\langle \hat{Y}_1 \rangle\right] = \sigma\left[\langle \hat{Y}_2 \rangle\right] = \sqrt{\frac{1}{4}s^2(y_1) + s^2(y_2)} = \frac{s}{\sqrt{2}} \ .$$

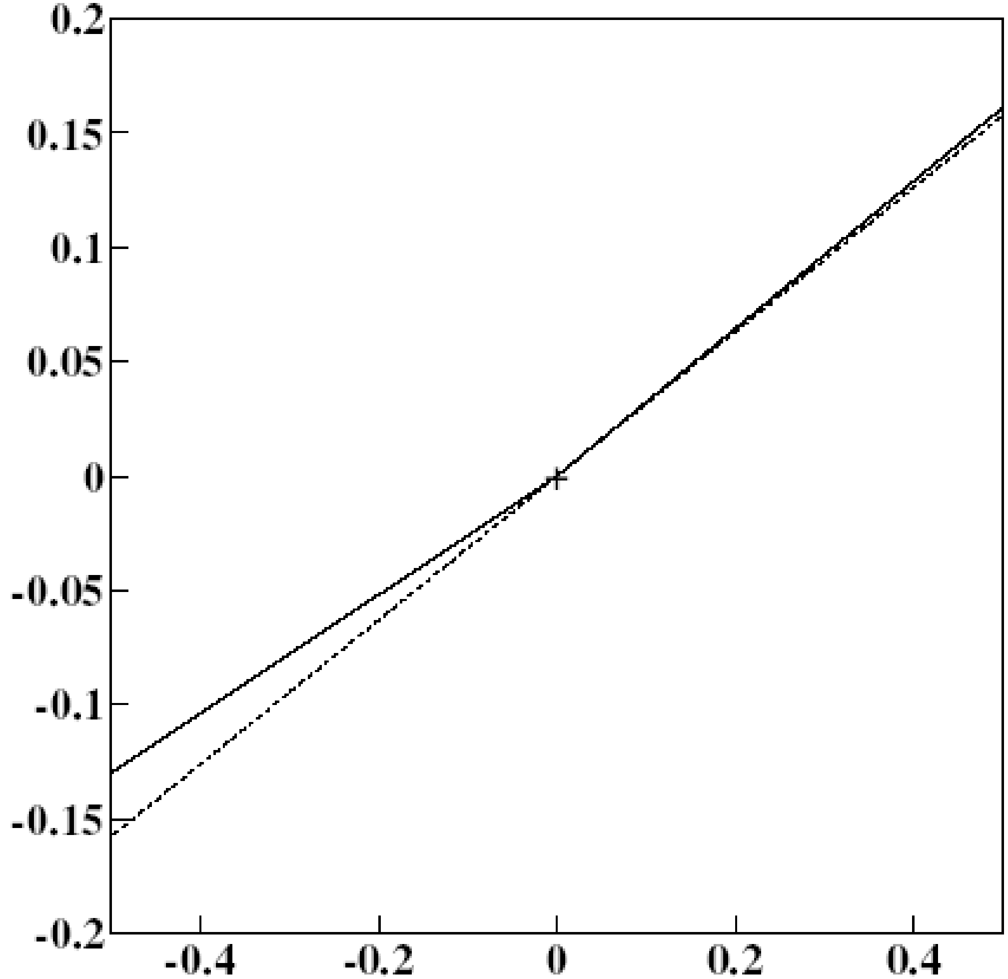
and nymerically

$$\langle \hat{Y}_1 \rangle = 17.5^o \pm 1.4^o$$
$$\langle \hat{Y}_2 \rangle = 197.5^o \pm 1.4^o \ .$$

Some remarks

- the solution obeys to the constraint and has a reduced error
- there is one degree of freedom
- the χ^2 test

$$\hat{\chi}^2 = \frac{(20 - 17.5)^2}{4} + \frac{(195 - 197.5)^2}{4} = 1.62 \ .$$


0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
-0.4
-0.2
0
0.2
0.4

Kinematic fit: degrees of freedom

$$\chi^2 = \sum_{j=1}^{3N} \sum_{i=1}^{3N} (p_i^{meas} - p_i^{fit}) W_{ij} (p_j^{meas} - p_j^{fit}) + 2\lambda\phi(\vec{p}, \vec{p}_u)$$

$$\phi(\vec{p}, \vec{p}_u) = 0, \qquad D\ \delta\alpha + d = 0$$

N(p) means number of the p variables

Degrees of freedom: $N(constr.eq.) - N(p_u)$

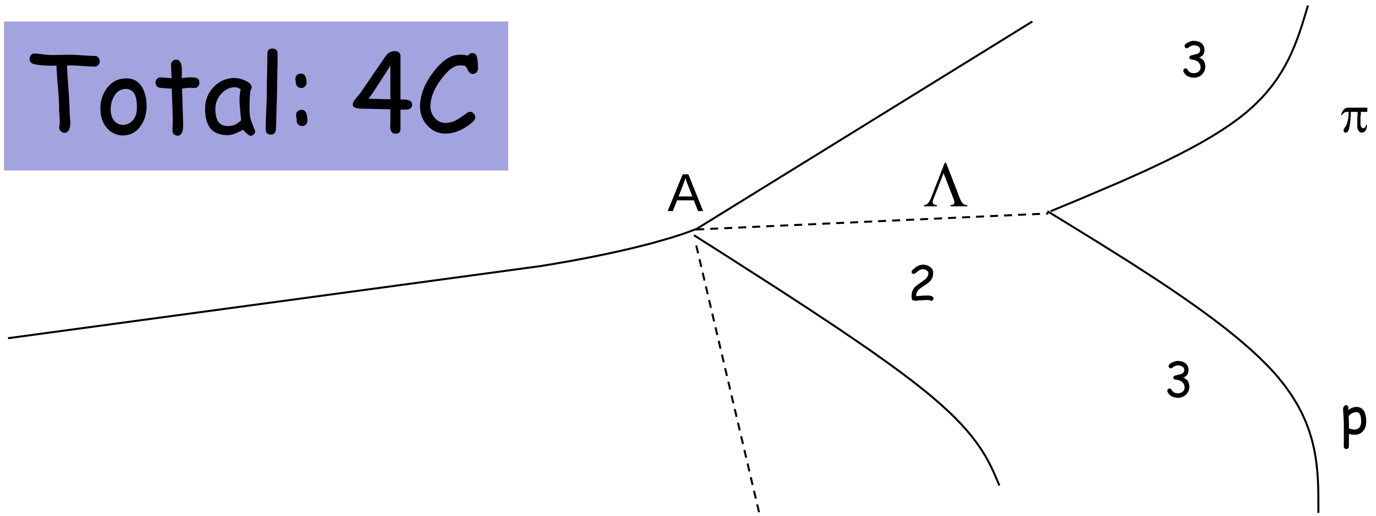
$N(p) + N(p_u) > N(constr.)$ the fit works

$N(p) + N(p_u) = N(constr.)$ constraints are fulfilled without fit

$N(p) + N(p_u) < N(constr.)$ constraints are not satisfied

Kinematic fit: examples

Total: 4C

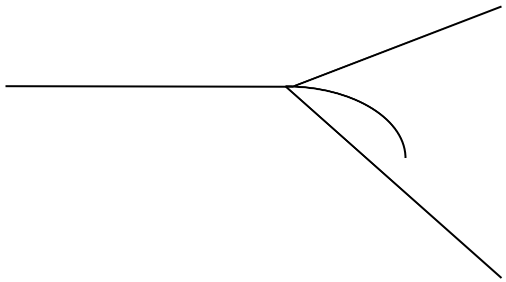


$$\Lambda \to p\pi^-$$

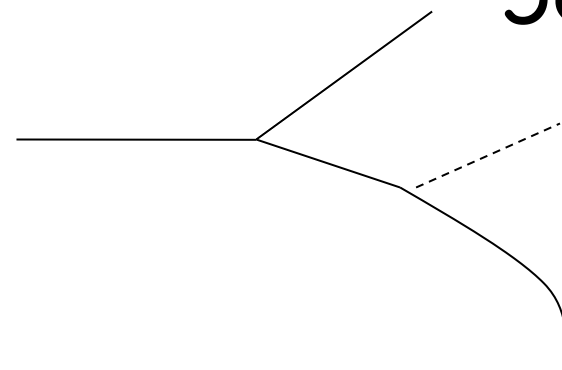
Unmeasured: 1
Constraints: 4

3C

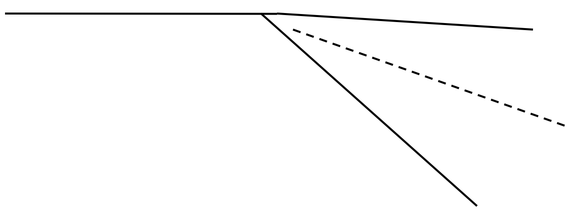
Kinematic fit: examples



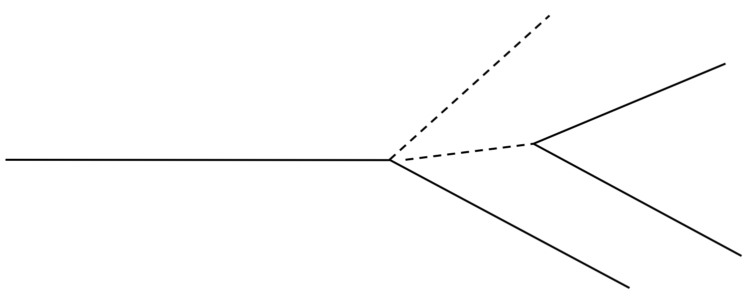
4C



5C



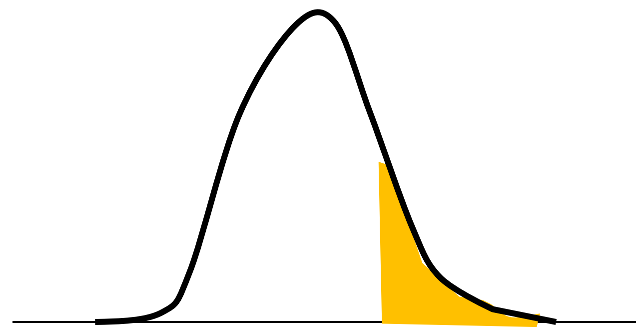
1C

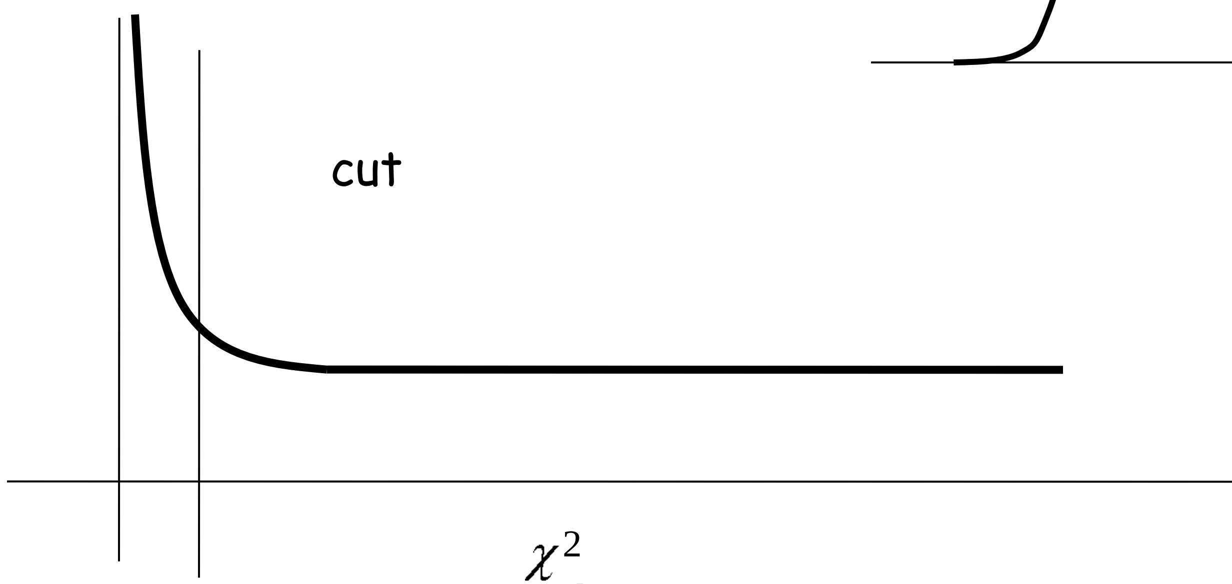


4C

Kinematic fit: goodness of fit

$$C = \int_0^X p(x)\,\mathrm{d}x \approx U(0,1)$$





$$1 - \int_0^{\chi^2} p(\chi^2, \nu)\,\mathrm{d}\chi^2$$

Non linear fits (MINUIT)

The function to minimize is (C is an arbitrary constant)

$$\chi^2(\boldsymbol{\theta}) \simeq -2 \ln L(\boldsymbol{\theta}) + C \ ,$$

All the non linear minimizations use the parabolic approximation

$$\chi^2(\theta) = \chi_0^2 + b\,\theta + \frac{1}{2} c\,\theta^2 \ ,$$

If $c > 0$,the function has a minimum in $\theta = -b/c$:

$$\frac{d\chi^2}{d\theta} = b + c\,\theta = 0 \quad \Longrightarrow \quad \theta = -\frac{b}{c} \ .$$

A necessary condition to be near a minimum is $c > 0$.
When $c < 0$ one moves along $-b$ until $c > 0$
In many dimensions this is negative gradient method

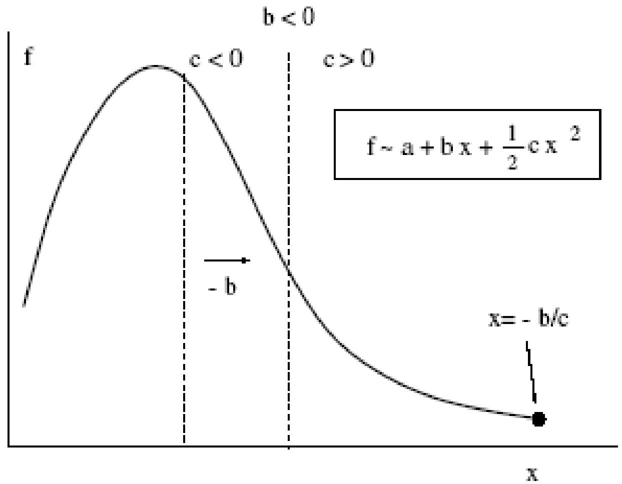


Figure 5: The negative gradient method

$$\chi^2(\boldsymbol{\theta}) = \chi_0^2 + [\boldsymbol{\nabla}\chi^2(\boldsymbol{\theta}_0)]^\dagger \cdot \boldsymbol{\theta} + \boldsymbol{\theta}^\dagger G \boldsymbol{\theta} \ ,$$

where $\boldsymbol{\nabla}$ is the gradient and G is the matrix of the second derivatives. The minimum condition is:

$$\boldsymbol{\theta} = -G^{-1}\boldsymbol{\nabla}\chi^2(\boldsymbol{\theta}_0)$$

and the condition on the 2-nd derivatives becomes:

$$\boldsymbol{\theta}^\dagger G \boldsymbol{\theta} > 0 \ ,$$

that is G must be a positive definite matrix

The search proceeds using the Marquardt algorithm:

$$G^{-1} \quad \rightarrow \quad (G + \lambda I)^{-1} \ , \tag{8}$$

where λ is greather of the more negative G eigenvalue when G is not positive definite (p.d.). When G is p.d. λ decreases and one uses the 2-order approximation:

$$\chi^2 \simeq \chi_0^2 + \sum_j \frac{\partial \chi^2}{\partial \theta_j}\theta_j + \frac{1}{2}\sum_{jk}\frac{\partial^2\chi^2}{\partial\theta_j\partial\theta_k}\theta_j\theta_k \ . \tag{9}$$

Usually one write:

$$\chi^2(\boldsymbol{\theta}) = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma_i^2} \equiv \sum_i H_i^2 \ ,$$

where i refers to the measured points.

The 2-nd derivatives are:

$$\begin{aligned} \frac{\partial^2\chi^2}{\partial\theta_j\partial\theta_k} &= \frac{\partial}{\partial\theta_j}\frac{\partial}{\partial\theta_k}\sum_i H_i^2 = \frac{\partial}{\partial\theta_j}\sum_i 2\,H_i\frac{\partial H_i}{\partial\theta_k} \\ &= 2\sum_i \frac{\partial H_i}{\partial\theta_j}\frac{\partial H_i}{\partial\theta_k} + 2\sum_i H_i \frac{\partial^2 H_i}{\partial\theta_j\partial\theta_k} \ . \end{aligned} \tag{10}$$

Non linear fits (MINUIT)

MINUIT warnings

A very common warning is
CURVATURE MATRIX NOT POSITIVE DEFINITE
In this case the matrix is forced d.p. with the Marquardt formula, but the result could be not reliable.
The cure of this disease usually is:

a) check the correlation matrix and remove too correlated parameters ($\rho > 0.98$) by changing the prametrization

b) assignement of the parameter initial values too far from the true minimum. For instance, you are moving along the valley of a folded foil...

c) use the "double precision" to avoid round-off errors

Another Suggestion
Verify if the minimum is stable by changung the initial alues of the parameters, because the pogram finds
the nearest relative minimum

Minuit function MIGRAD

- Purpose: find minimum

```
**********
**   13 **MIGRAD        1000          1
**********
(some output omitted)
MIGRAD MINIMIZATION HAS CONVERGED.
MIGRAD WILL VERIFY CONVERGENCE AND ERROR MATRIX.
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=257.304 FROM MIGRAD    STATUS=CONVERGED      31 CALLS         32 TOTAL
                     EDM=2.36773e-06    STRATEGY= 1      ERROR MATRIX ACCURATE
 EXT PARAMETER                                   STEP         FIRST
 NO.   NAME      VALUE            ERROR          SIZE       DERIVATIVE
  1   mean        8.84225e-02   3.23862e-01   3.58344e-04  -2.24755e-02
  2   sigma       3.20763e+00   2.39540e-01   2.78628e-04  -5.34724e-02
                            ERR DEF= 0.5
EXTERNAL ERROR MATRIX.    NDIM=  25    NPAR=  2    ERR DEF=0.5
 1.049e-01  3.338e-04
 3.338e-04  5.739e-02
PARAMETER  CORRELATION COEFFICIENTS
      NO.  GLOBAL      1       2
       1  0.00430   1.000   0.004
       2  0.00430   0.004   1.000
```

Progress information, watch for errors here

Parameter values and approximate errors reported by MINUIT

Error definition (in this case 0.5 for a likelihood fit)

Minuit function MIGRAD

- Purpose: find minimum

```
**********
**    13 **MIGR
**********
(some output o
MIGRAD MINIMIZ
MIGRAD WILL VERI                       ERROR MATRIX.
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=257.304 FROM MIGRAD    STATUS=CONVERGED       31 CALLS           32 TOTAL
                         EDM=2.36773e-06     STRATEGY= 1      ERROR MATRIX ACCURATE
 EXT PARAMETER                                        STEP         FIRST
 NO.   NAME       VALUE            ERROR              SIZE      DERIVATIVE
  1  mean         8.84225e-02   3.23862e-01   3.58344e-04  -2.24755e-02
  2  sigma        3.20763e+00   2.39540e-01   2.78628e-04  -5.34724e-02
                                   ERR DEF= 0.5
EXTERNAL ERROR MATRIX.    NDIM=  25     NPAR=  2    ERR DEF=0.5
 1.049e-01  3.338e-04
 3.338e-04  5.739e-02
PARAMETER  CORRELATION COEFFICIENTS
      NO.  GLOBAL      1      2
        1  0.00430   1.000  0.004
        2  0.00430   0.004  1.000
```

Value of χ^2 or likelihood at minimum

(NB: χ^2 values are not **divided by $N_{d.o.f}$)**

Approximate Error matrix And covariance matrix

Theorems on $L(\theta; X)$

The mean value of the Score Function is zero:

$$\left\langle \frac{\partial}{\partial\theta} \ln p(\boldsymbol{X};\theta) \right\rangle = 0 \ .$$

The variance of the Score Function is the Fisher information:

$$\begin{aligned}\mathrm{Var}\left[\frac{\partial}{\partial\theta} \ln p(\boldsymbol{X};\theta)\right] &= \left\langle \left(\frac{\partial}{\partial\theta} \ln p(\boldsymbol{X};\theta) - \left\langle \frac{\partial}{\partial\theta} \ln p(\boldsymbol{X};\theta) \right\rangle \right)^2 \right\rangle \\ &= \left\langle \left(\frac{\partial}{\partial\theta} \ln p(\boldsymbol{X};\theta) \right)^2 \right\rangle \equiv I(\theta)\end{aligned}$$

These remarkable relations hold:

$$I(\theta) = \left\langle \left(\frac{\partial}{\partial\theta} \ln p(\boldsymbol{X};\theta) \right)^2 \right\rangle = -\left\langle \frac{\partial^2}{\partial\theta^2} \ln p(\boldsymbol{X};\theta) \right\rangle \ .$$

$$\left\langle \left(\frac{\partial}{\partial\theta} \ln L \right)^2 \right\rangle = \left\langle \left(\frac{\partial}{\partial\theta} \sum_i \ln p(\boldsymbol{X}_i;\theta) \right)^2 \right\rangle = n \left\langle \left(\frac{\partial}{\partial\theta} \ln p \right)^2 \right\rangle = nI(\theta) \ ,$$

The Cramér Rao theorem:

If T_n is an unbiased estimator

$$\mathrm{Var}[T_n] \geq \frac{1}{n \left\langle \left(\frac{\partial}{\partial\theta} \ln p(\boldsymbol{X};\theta) \right)^2 \right\rangle} = \frac{1}{nI(\theta)}$$

Minuit fun

- Purpose: find minimu

Status: Should be 'converged' but can be 'failed'

Estimated Distance to Minimum should be small $O(10^{-6})$

Error Matrix Quality should be 'accurate', but can be 'approximate' in case of trouble

```
**********
**   13 **MIGRAD       1000
**********
(some output omitted)
MIGRAD MINIMIZATION HAS CONVERGED
MIGRAD WILL VERIFY CONVERGENCE AND E      MATRIX.
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=257.304 FROM MIGRAD    STATUS=CONVERGED      31 CALLS          32 TOTAL
                    EDM=2.36773e-06     STRATEGY= 1       ERROR MATRIX ACCURATE
 EXT PARAMETER                                      STEP         FIRST
 NO.   NAME      VALUE            ERROR             SIZE      DERIVATIVE
  1  mean         8.84225e-02   3.23862e-01   3.58344e-04  -2.24755e-02
  2  sigma        3.20763e+00   2.39540e-01   2.78628e-04  -5.34724e-02
                            ERR DEF= 0.5
EXTERNAL ERROR MATRIX.    NDIM=  25    NPAR=  2    ERR DEF=0.5
 1.049e-01  3.338e-04
 3.338e-04  5.739e-02
PARAMETER  CORRELATION COEFFICIENTS
      NO.  GLOBAL      1       2
       1  0.00430   1.000  0.004
       2  0.00430   0.004  1.000
```

Minuit function HESSE

- Purpose: calculate error matrix from $\frac{d^2L}{dp^2}$

```
**********
**    18 **HESSE        1000
**********
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=257.304 FROM HESSE     STATUS=OK                                   OTAL
                    EDM=2.36534e-06    STRAT                           CURATE
 EXT PARAMETER                                     I    RNAL      INTERNAL
 NO.   NAME      VALUE            ERROR          STEP SIZE       VALUE
  1  mean         8.84225e-02   3.23861e-01   7.16689e-05   8.84237e-03
  2  sigma        3.20763e+00   2.39539e-01   5.57256e-05   3.26535e-01
                                   ERR DEF= 0.5
EXTERNAL ERROR MATRIX.    NDIM=  25    NPAR=  2    ERR DEF=0.5
 1.049e-01  2.780e-04
 2.780e-04  5.739e-02
PARAMETER  CORRELATION COEFFICIENTS
      NO.  GLOBAL      1       2
        1  0.00358   1.000   0.004
        2  0.00358   0.004   1.000
```

Symmetric errors calculated from 2nd derivative of −ln(L) or χ^2

Minuit function HESSE

$$\frac{d^2L}{dp^2}$$

```
**********
**
***
COV                              UCCESSFULLY
FCN                              TUS=OK                10 CALLS            42 TOTAL
                                 4e-06     STRATEGY= 1       ERROR MATRIX ACCURATE
 EX                                                    INTERNAL       INTERNAL
 NO                                  ERROR             STEP SIZE        VALUE
  1                                3.23861e-01         7.16689e-05    8.84237e-03
  2  si               3.20763e+00  2.39539e-01         5.57256e-05    3.26535e-01
                                 ERR DEF= 0.5
EXTERNAL ERROR MATRIX.      NDIM=  25     NPAR=  2     ERR DEF=0.5
 1.049e-01  2.780e-04
 2.780e-04  5.739e-02
PARAMETER  CORRELATION COEFFICIENTS
      NO.  GLOBAL      1       2
       1  0.00358   1.000   0.004
       2  0.00358   0.004   1.000
```

Error matrix (Covariance Matrix) calculated from

$$V_{ij} = \left(\frac{d^2(-\ln L)}{dp_i dp_j} \right)^{-1}$$

Minuit function HESSE

$$\frac{d^2L}{dp^2}$$

```
**********
**   18 **HESSE        1000
**********
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=257.304 FROM HESSE     STATUS=OK                  10 CALLS            42 TOTAL
                        EDM=2.36534e-06    STRATEGY= 1       ERROR MATRIX ACCURATE
 EXT PARAMETER                                       INTERNAL        INTERNAL
 NO.   NAME        VALUE           ERROR             STEP SIZE         VALUE
  1  mean          8.84225e-02                                       8.84237e-03
  2  sigma         3.20763e+00                                       3.26535e-01

EXTERNAL ERROR MATRIX.    NDIM                                     F=0.5
 1.049e-01  2.780e-04
 2.780e-04  5.739e-02
PARAMETER  CORRELATION COEFFICIENTS
      NO.  GLOBAL      1       2
        1  0.00358   1.000   0.004
        2  0.00358   0.004   1.000
```

Correlation matrix ρ_{ij} calculated from

$$V_{ij} = \sigma_i \sigma_j \rho_{ij}$$

Minuit function HESSE

$$\frac{d^2L}{dp^2}$$

```
 **********
 **   18 **HESSE        1000
 **********
 COVARIANCE MATRIX CALCULATED SUCCESSFULLY
 FCN=257.304 FROM HESSE     STATUS=OK              10 CALLS          42 TOTAL
                     EDM=2.36534e-06    STRATEGY= 1      ERROR MATRIX ACCURATE
  EXT PARAMETER                                   INTERNAL       INTERNAL
  NO.   NAME      VALUE            ERROR          STEP SIZE        VALUE
   1  mean                                        7.16689e-05   8.84237e-03
   2  sigma                                       5.57256e-05   3.26535e-01

 EXTERNAL ERROR                                   2    ERR DEF=0.5
  1.049e-01  2.7
  2.780e-04  5.739e-
 PARAMETER  CORRELA  N COEFFICIENTS
       NO.  GLOBAL       1       2
        1  0.00358    1.000   0.004
        2  0.00358    0.004   1.000
```

Global correlation vector: correlation of each parameter with *all other* parameters

Minuit function MINOS

```
**********
**    23 **MINOS        1000
**********
FCN=257.304 FROM MINOS     STATUS=SUCCESSFUL     52 CALLS          94 TOTAL
                    EDM=2.36534e-06    STRATEGY= 1      ERROR MATRIX ACCURATE
 EXT PARAMETER                PARABOLIC         MINOS ERRORS
 NO.   NAME      VALUE          ERROR      NEGATIVE      POSITIVE
  1   mean        8.84225e-02  3.23861e-01 -3.24688e-01   3.25391e-01
  2   sigma       3.20763e+00  2.39539e-01 -2.23321e-01   2.58893e-01
                           ERR DEF= 0.5
```

Symmetric error

(repeated result from HESSE)

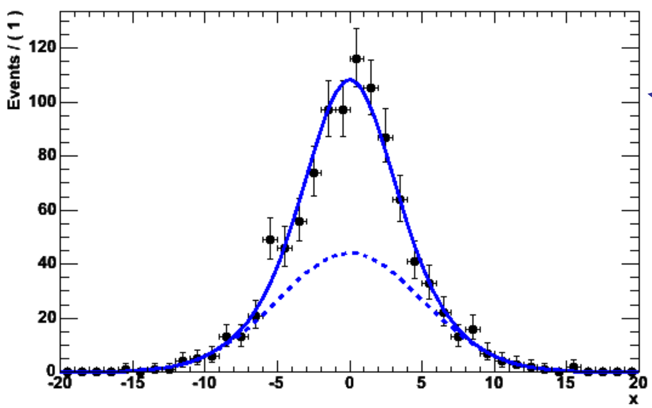
MINOS error
Can be asymmetric

(in this example the 'sigma' error is slightly asymmetric)

Mitigating fit stability problems

- Strategy I – More orthogonal choice of parameters
  - Example: fitting sum of 2 Gaussians of similar width

$$F(x; f, m, s_1, s_2) = fG_1(x; s_1, m) + (1 - f)G_2(x; s_2, m)$$



HESSE correlation matrix

PARAMETER NO.	CORRELATION COEFFICIENTS GLOBAL	[f]	[m]	[s1]	[s2]
[f]	0.96973	1.000	-0.135	0.918	0.915
[m]	0.14407	-0.135	1.000	-0.144	-0.114
[s1]	0.92762	0.918	-0.144	1.000	0.786
[s2]	0.92486	0.915	-0.114	0.786	1.000

Widths s_1, s_2 strongly correlated fraction f

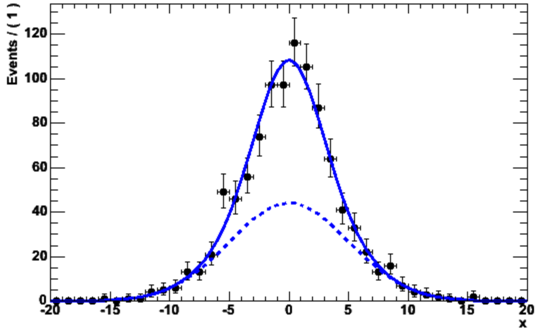
Mitigating fit stability problems

- Different parameterization:

$$fG_1(x;s_1,m_1)+(1-f)G_2(x;\underline{s_1\cdot s_2},m_2)$$

PARAMETER NO.	GLOBAL	[f]	[m]	[s1]	[s2]
[f]	0.96951	1.000	-0.134	0.917	-0.681
[m]	0.14312	-0.134	1.000	-0.143	0.127
[s1]	0.98879	0.917	-0.143	1.000	-0.895
[s2]	0.96156	-0.681	0.127	-0.895	1.000

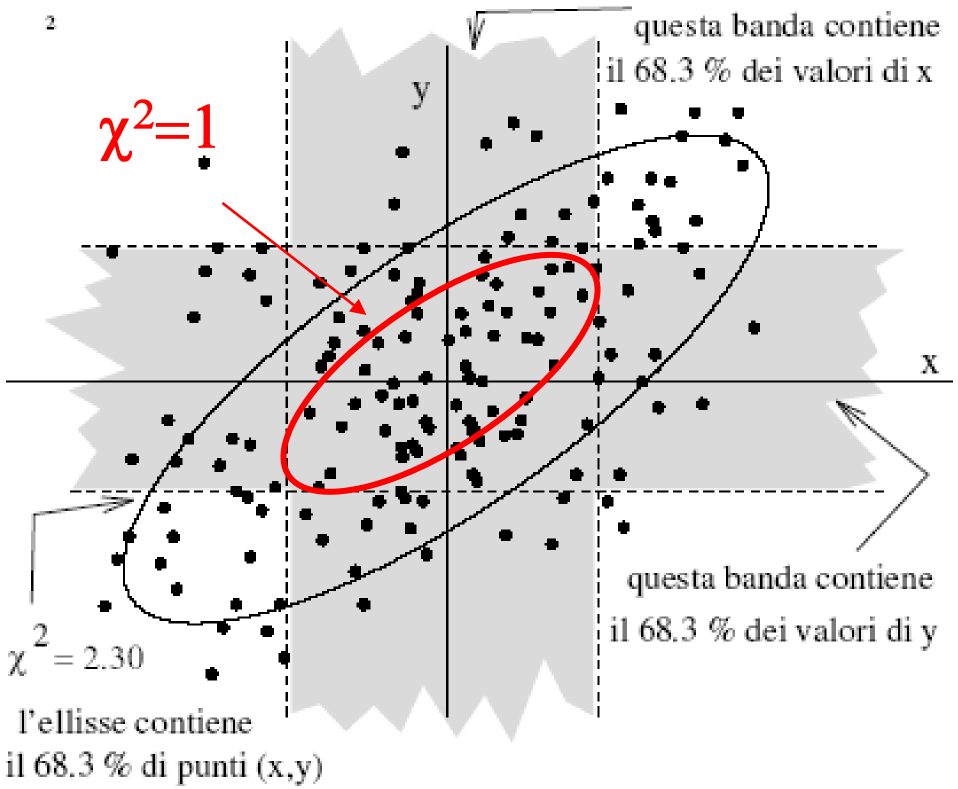
CORRELATION COEFFICIENTS



- Correlation of width s2 and fraction f reduced from 0.92 to 0.68
- Choice of parameterization matters!

- Strategy II – Fix all but one of the correlated parameters
  - If floating parameters are highly correlated, some of them may be redundant and not contribute to additional degrees of freedom in your model

Non linear fits (MINUIT)



If the model is linear, the error matrix $(FWF^{\dagger})^{-1}$ gives variances and covariances. The diagonal terms are reported as the errors on the parameters.

When the model is not linear, the errors can be found wih the MINOS option:

make a grid to each parameter in turn, by minimizing on the remaining ones. The error on the blocked parameter is given when $\Delta\chi^2 = 1$ w.r.t the minimum:

$$\Delta\chi^2 \equiv \chi^2[\hat{\theta}_k \pm s(\hat{\theta}_k)] = 1 \ .$$

END

We have been using other estimators:

$N_S/\sqrt{N_B}$,

$N_S/\sqrt{N_B + N_S}$,

$2(\sqrt{N_B + N_S} - \sqrt{N_B})$

Finaly, we calculate the signal statistical significance as:

$$S_L = \sqrt{2\,(\ln L_{B+S} - \ln L_B)}$$

where compute two likelihoods:

$$\ln L_B = \sum_{i=1}^{10} (-b_i + n_i \cdot \ln b_i)$$

$$\ln L_{B+S} = \sum_{i=1}^{10} (-b_i - s_i + n_i \cdot \ln(b_i + s_i))$$

b_i, s_i, n_i are the number of predicted background and signal events and observed data events in the i-th bin

Proof of the Effective Variance formula:

Case III

$$f(X) = f(x_0) + (X - x_0) f'(x_0) + o(X - x_0)^2 \ .$$

If $\mathrm{Var}[X_R]$ is small

$$\mathrm{Var}[f(X)] \simeq f'^2(x_0)\sigma_x^2$$

By substituting the unknown x_0 with x:,

$$\mathrm{Var}[Y - f(X)] = \mathrm{Var}[Y] + \mathrm{Var}[f(X)] \simeq \sigma_y^2 + f'^2(x)\,\sigma_x^2 \equiv \sigma_E^2 \ .$$

Hence:

$$\chi^2(\boldsymbol{\theta}) = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma_{Ei}^2} = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma_{y_i}^2 + f'^2(x_i, \boldsymbol{\theta})\sigma_{x_i}^2} \ ,$$

where only the measured values (x_i, y_i) are present.
This is a non linear minimiztion (MINUIT)

Alternatvely, one can use a two-step method
1)set $\sigma_{Ei}^2 = \sigma_{y_i}^2$
2) and in the k-th cycle put

$$\sigma_{Ei}^2 = \sigma_{y_i}^2 + f'^2(x_i, \boldsymbol{\theta}^{(k-1)})\sigma_{x_i}^2$$

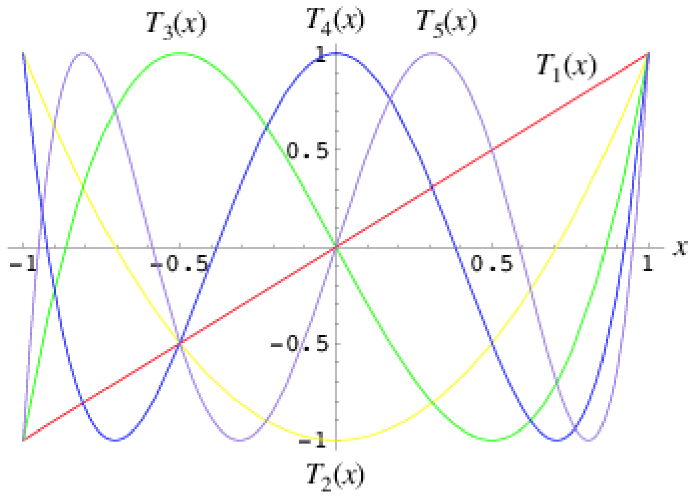
by using the estimations $\boldsymbol{\theta}^{(k-1)}$ of the previous cycle.

CAUTION: the iterative method is distorted!

Fit stability with polynomials

- **Warning:** Regular parameterization of polynomials $a_0+a_1x+a_2x^2+a_3x^3$ nearly always results in strong correlations between the coefficients a_i.
  - *Fit stability problems, inability to find right solution common at higher orders*
- **Solution:** Use existing parameterizations of polynomials that have (mostly) uncorrelated variables
  - ***Example: Chebychev polynomials***

$$
\begin{aligned}
T_0(x) &= 1 \\
T_1(x) &= x \\
T_2(x) &= 2x^2 - 1 \\
T_3(x) &= 4x^3 - 3x \\
T_4(x) &= 8x^4 - 8x^2 + 1 \\
T_5(x) &= 16x^5 - 20x^3 + 5x \\
T_6(x) &= 32x^6 - 48x^4 + 18x^2 - 1.
\end{aligned}
$$



Browsing fit results

- As fits grow in complexity (e.g. 45 floating parameters), number of output variables increases
  - Need better way to navigate output that MINUIT screen dump
- **RooFitResult** holds complete snapshot of fit results
  - Constant parameters
  - Initial and final values of floating parameters
  - Global correlations & full correlation matrix
  - Returned from **RooAbsPdf::fitTo()** when "**r**" option is supplied
- Compact & verbose printing mode

Compact Mode

Constant parameters omitted in compact mode

Alphabetical parameter listing

```
fitres->Print() ;

  RooFitResult: min. NLL value: 1.6e+04, est. distance to min: 1.2e-05

    Floating Parameter    FinalValue +/-  Error
  --------------------  --------------------------
                argpar   -4.6855e-01 +/-  7.11e-02
                g2frac    3.0652e-01 +/-  5.10e-03
                 mean1    7.0022e+00 +/-  7.11e-03
                 mean2    1.9971e+00 +/-  6.27e-03
                 sigma    2.9803e-01 +/-  4.00e-03
```

Browsing fit results

Verbose printing mode

```
fitres->Print("v") ;

  RooFitResult: min. NLL value: 1.6e+04, est. distance to min: 1.2e-05

  Constant Parameter    Value
  --------------------  ------------
                cutoff    9.0000e+00
                g1frac    3.0000e-01

    Floating Parameter  InitialValue    FinalValue +/-  Error     GblCorr.
  --------------------  ------------  --------------------------  --------
                argpar   -5.0000e-01   -4.6855e-01 +/-  7.11e-02  0.191895
                g2frac    3.0000e-01    3.0652e-01 +/-  5.10e-03  0.293455
                 mean1    7.0000e+00    7.0022e+00 +/-  7.11e-03  0.113253
                 mean2    2.0000e+00    1.9971e+00 +/-  6.27e-03  0.100026
                 sigma    3.0000e-01    2.9803e-01 +/-  4.00e-03  0.276640
```

Constant parameters listed separately

Initial,final value and global corr. listed side-by-side

Correlation matrix accessed separately

Browsing fit results

- Easy navigation of correlation matrix
  - Select single element or complete row by parameter name

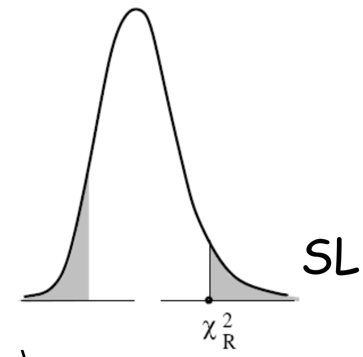
```
r->correlation("argpar","sigma")
(const Double_t)(-9.25606412005910845e-02)

r->correlation("mean1")->Print("v")
RooArgList::C[mean1,*]: (Owning contents)
  1) RooRealVar::C[mean1,argpar] :  0.11064 C
  2) RooRealVar::C[mean1,g2frac] : -0.0262487
C
  3) RooRealVar::C[mean1,mean1]  :  1.0000 C
  4) RooRealVar::C[mean1,mean2]  : -
0.00632847 C
  5) RooRealVar::C[mean1,sigma]  : -0.0339814
C
```

- **RooFitResult** persistable with ROOT I/O
  - Save your batch fit results in a ROOT file and navigate your results just as easy afterwards

Research of functional forms

We generate artificial data from the model (15% error)

$$y = \theta_0 + \theta_2 x^2 + \theta_3 x^3 = 3 + 5x^2 - 0.5x^3 \ ,$$

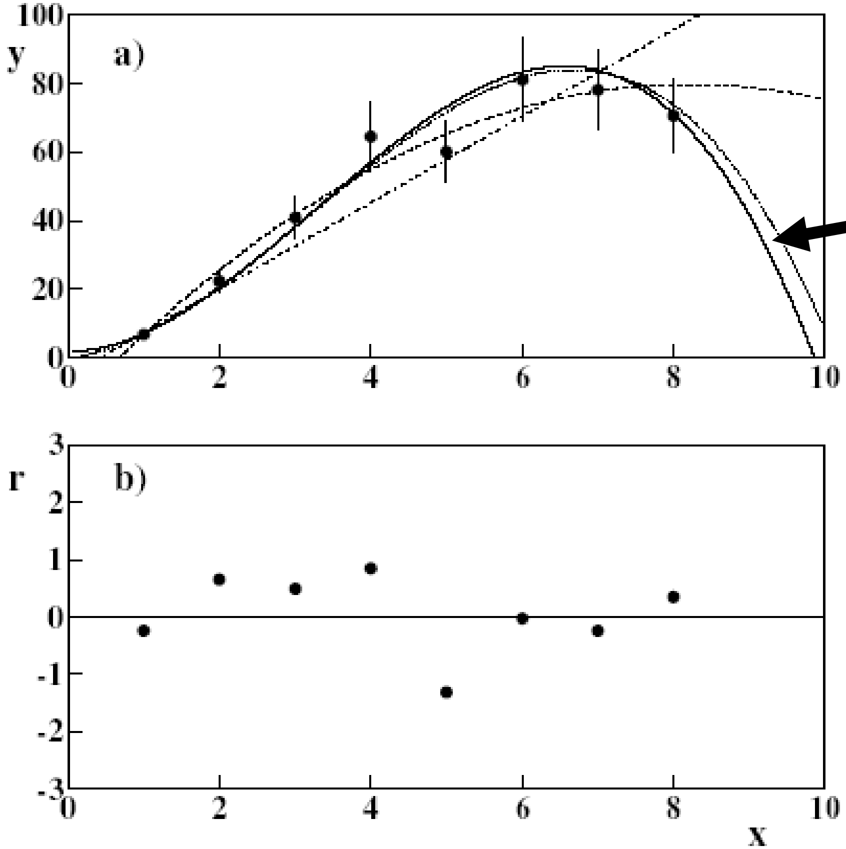


We try some fits with the functions

$$y = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \ldots$$

and the results are in the table:

	FIT1	FIT2	FIT3	FIT4	FIT5
θ_0	-5.1 ± 1.5	-14.8 ± 3.4	-7.0 ± 7.6		1.9 ± 1.2
θ_1	12.5 ± 0.9	23.0 ± 3.4	12.0 ± 10.0	2.8 ± 1.7	
θ_2		-1.4 ± 0.5	2.1 ± 3.1	4.8 ± 1.0	5.8 ± 0.6
θ_3			-0.3 ± 0.3	-0.5 ± 0.1	-0.6 ± 0.1
χ^2/ν	$\frac{13.2}{6} = 2.20$	$\frac{3.2}{5} = 0.64$	$\frac{1.9}{4} = 0.47$	$\frac{2.8}{5} = 0.55$	$\frac{3.4}{5} = 0.67$
SL	8%	50%	40%	50%	70%



$1.9 + 5.8\, x^2 - 0.6\, x^3$

Applications of Least Squares

- research of physics models $f(x_i, \boldsymbol{\theta})$

$$\chi^2(\boldsymbol{\theta}) = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma_i^2} \ ,$$

  $(n-p)$ degrees of freedom (DoF)

  - function determination
  - fit of histograms

- research of correlations when X and Y are both random

- determination of measured variables y with constraints (kinematical fit) by the use of Lagrange Multpliers

$$\chi^2(\boldsymbol{\theta}) = \sum_i \frac{[y_i - x_i]^2}{\sigma_i^2} + \lambda \Phi(x_i, z) \ ,$$

  DoF equal to the number of equations

- minimization with a regularization term (bayesian minimization)

$$\mathrm{Max}\ [p(\boldsymbol{\theta}|\boldsymbol{x}) \propto L(\boldsymbol{x}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})]$$

  which gives the constrained minimzation

$$\mathrm{Min}\ [-2 \ln L(\boldsymbol{x}|\boldsymbol{\theta}) - 2 \ln p(\boldsymbol{\theta})]$$

  variable DoF

Search for correlations

$Y = f(X) + Y_R \ , \qquad X = x_0 + X_R$

Often $\mathrm{Var}[Y|x] = \sigma_y^2 - \mathrm{Var}[f(X)] \equiv \sigma^2$ is constant. In this case the minimization of

$$\chi^2(\boldsymbol{\theta}) = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma^2}$$

must determine the correlation function f The following equalities hold:

$$(n-1)\sigma_y^2 \simeq \sum_i (y_i - \langle y \rangle)^2 = \sum_i (y_i - f(x_i) + f(x_i) - \langle y \rangle)^2$$

$$= \sum_i \left[(y_i - f(x_i))^2 + (f(x_i) - \langle y \rangle)^2 + 2\,(y_i - f(x_i))\,(f(x_i) - \langle y \rangle) \right]$$

Due to the normal equations, the cross term is zero. Hence

$$\sum_i (y_i - \langle y \rangle)^2 = \sum_i (y_i - \hat{y}_i)^2 + \sum_i (\hat{y}_i - \langle y \rangle)^2 \ ,$$

$$\text{total deviation} = \text{residual deviation} \ + \ \text{explained deviation} \ .$$

$$s_y^2 = s_{y_R}^2 + s^2\left(f(x, \hat{\boldsymbol{\theta}})\right) = s_{y_R}^2 + s^2(\hat{y}) \ ,$$

$$r^2 = \frac{s^2(\hat{y})}{s_y^2} = \frac{\sum_i (\hat{y}_i - \langle y \rangle)^2}{\sum_i (y_i - \langle y \rangle)^2} = \frac{\text{explained deviation}}{\text{total deviation}}$$

$$s_{y_R}^2 = s_y^2(1 - r^2)$$

One has to fin the function f that maximize r^2

a) for a given $f(x, \boldsymbol{\theta})$, linear in the parameters, one minimizes

Search for correlations

$$\chi^2(\boldsymbol{\theta}) = \sum_i [y_i - f(x_i, \boldsymbol{\theta})]^2 \ .$$

Then, the estimation of $\sigma[Y|x]$ is given by:

$$s^2_{y_R} = \chi^2_R(\hat{\boldsymbol{\theta}}) = \frac{1}{n-p} \sum_i [y_i - f(x_i, \hat{\boldsymbol{\theta}})]^2 \ .$$

b) the form with maximum $r^2$is selected. The solution with the smallest number of parameters is selcted among the equivalent ones.

c) also the residuals

$$r_i = y_i - f(x_i, \hat{\boldsymbol{\theta}}) \qquad (2)$$

must have a random behaviour

Test with smulated data (g_1 and g_2 are $N(0,1)$ random variables):

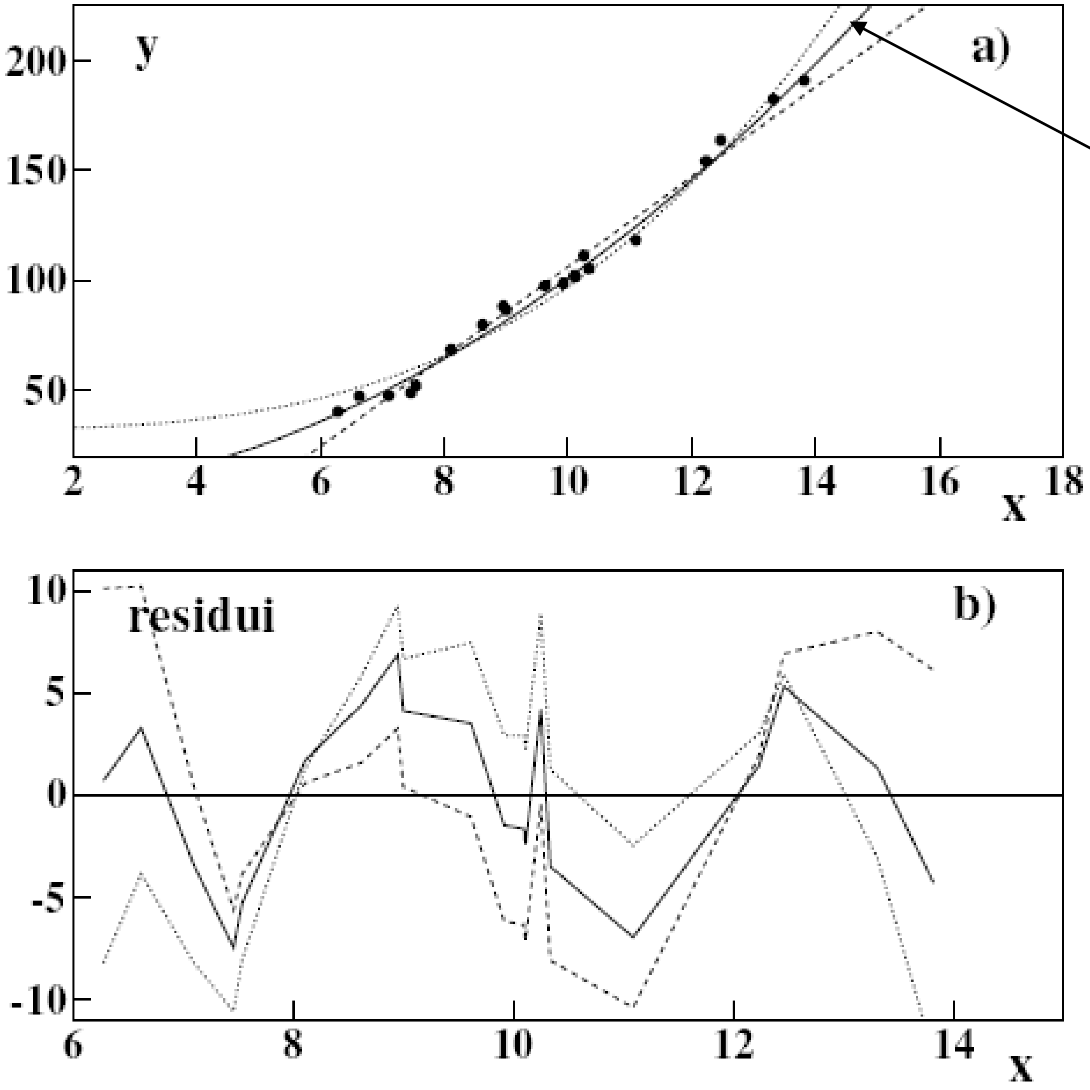
$$x = x_0 + x_R = 10 + 2\,g_1 \;, \qquad f(x) = 2 + x^2$$
$$y = f(x) + y_R = 2 + x^2 + 5\,g_2 \;,$$

Search for correlations

χ^2 minimization:

$$\chi^2_{(\sigma=1)} = \sum_i (y_i - \theta_0 - \theta_1 x_i - \theta_2 x_i^2 - \theta_3 x_i^3)^2$$

	FIT1	FIT2	FIT3	FIT4
θ_0	-98.7 ± 6.6	12.2 ± 7.7	-0.95 ± 2.45	32.4 ± 2.8
θ_1	20.5 ± 0.7	2.3 ± 1.5		
θ_2		0.91 ± 0.07	1.03 ± 0.02	
θ_3				0.065 ± 0.002
$s_{y_R} = \sqrt{\chi^2_\nu}$	6.3	4.4	4.4	6.9
r^2	98.11%	99.12%	99.10%	97.72%


y
a)
200
150
100
50
2
4
6
8
10
12
14
16
18
x
b)
residui
10
5
0
-5
-10
6
8
10
12
14
x

$-0.95 + 1.03\ x^2$

Minimization with constraints

In this case one minimizes

$$\chi^2(\boldsymbol{\mu}) = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\mu})]^2}{\sigma_i^2} \quad \text{or} \quad \chi^2(\boldsymbol{\mu}) = \sum_{\mathrm{i}} \frac{[\mathrm{y_i} - \mu_\mathrm{i}]^2}{\sigma_\mathrm{i}^2}$$

with M equations of constraint

$$\Phi_k(\boldsymbol{n}, \boldsymbol{\mu}, \boldsymbol{\varphi}) = 0 \qquad k = 1, 2, \ldots, M \ , \qquad (3)$$

where there are the p parameters and sometimes also q non-measured parameters $\boldsymbol{\varphi}$. We must have $M < p$

By minimizing χ^2 and from (3)

$$\mathrm{d}(\chi^2 + \Phi_k) = \sum_{j=1}^{p} \frac{\partial}{\partial \mu_j}[\chi^2 + \Phi_k]\, \mathrm{d}\mu_j + \sum_{n=1}^{q} \frac{\partial \Phi_k}{\partial \varphi_n}\, \mathrm{d}\varphi_n = 0\,, \quad k = 1, 2 \ldots, M \qquad (4)$$

μ = p measured

φ = q non measured

However, the differentials $\mathrm{d}\mu_j$ are not independent due to constraint equations. To restore the independence one introduces the M Lagrange Multipliers

$$\sum_{j=p-M}^{p} \frac{\partial}{\partial \mu_j} [\chi^2 + \lambda_k \Phi_k] = 0\ . \qquad (5)$$

which make the μ_j independent.

Kinematic fit

where $\boldsymbol{\lambda}$ is a vector of r unknowns, the Lagrange multipliers. Minimizing the χ^2 with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ yields two vector equations which can be solved for the parameters $\boldsymbol{\alpha}$ and their covariance matrix:

$$\begin{aligned} \mathbf{V}_{\alpha 0}^{-1}(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)+\mathbf{D}^T\boldsymbol{\lambda}=0 \\ \mathbf{D}\delta\boldsymbol{\alpha}+\mathbf{d}=0 \end{aligned} \qquad (5)$$

The latter equation demonstrates clearly that the solution satisfies the constraints. The solution can be written [3]

$$\begin{aligned} \boldsymbol{\alpha} &= \boldsymbol{\alpha}_0-\mathbf{V}_{\alpha_0}\mathbf{D}^T\boldsymbol{\lambda} \\ \boldsymbol{\lambda} &= \mathbf{V}_D(\mathbf{D}\delta\boldsymbol{\alpha}_0+\mathbf{d}) \\ \mathbf{V}_D &= \left(\mathbf{D}\mathbf{V}_{\alpha_0}\mathbf{D}^T\right)^{-1} \\ \mathbf{V}_\alpha &= \mathbf{V}_{\alpha_0}-\mathbf{V}_{\alpha_0}\mathbf{D}^T\mathbf{V}_D\mathbf{D}\mathbf{V}_{\alpha_0} \\ \chi^2 &= \boldsymbol{\lambda}^T\mathbf{V}_D^{-1}\boldsymbol{\lambda} \\ &= \boldsymbol{\lambda}^T(\mathbf{D}\delta\boldsymbol{\alpha}_0+\mathbf{d}) \end{aligned} \qquad (6)$$

where $\delta\boldsymbol{\alpha}_0=\boldsymbol{\alpha}_0-\boldsymbol{\alpha}_A$. It can be shown that the diagonal elements of the $\boldsymbol{\alpha}$ covariance matrix are reduced in size, as expected by intuition.

Kinematic fit

V.1 Invariant mass constraint

The constraint equation which forces a track to have an invariant mass m_c is

$$\mathbf{D}\delta\alpha + \mathbf{d} = 0 \qquad E^2 - p_x^2 - p_y^2 - p_z^2 - m_c^2 = 0, \tag{15}$$

Expanding about the initial parameters $(p_x, p_y, p_z, E, x, y, z)$, we get for **D** and **d**:

$$\begin{aligned} \mathbf{D} &= \begin{pmatrix} -2p_x & -2p_y & -2p_z & 2E & 0 & 0 & 0 \end{pmatrix} \\ \mathbf{d} &= E^2 - p_x^2 - p_y^2 - p_z^2 - m_c^2 \end{aligned} \tag{16}$$

V.2 Total energy constraint

The constraint that a track must have a total energy E_c can be written $E - E_c = 0$. The **D** and **d** matrices are trivially computed to be

$$\begin{aligned} \mathbf{D} &= \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \\ \mathbf{d} &= E - E_c \end{aligned} \tag{17}$$

V.3 Total momentum constraint

The constraint that a track must have a total momentum p_c can be written

$$\sqrt{p_x^2 + p_y^2 + p_z^2} - p_c = 0, \tag{18}$$

Expanding about an initial set of parameters $(p_x, p_y, p_z, E, x, y, z)$, we compute the **D** and **d** matrices to be

$$\begin{aligned} \mathbf{D} &= \begin{pmatrix} p_x / p & p_y / p & p_z / p & 0 & 0 & 0 & 0 \end{pmatrix} \\ \mathbf{d} &= \sqrt{p_x^2 + p_y^2 + p_z^2} - p_c \end{aligned} \tag{19}$$

Kinematic fit

Lagrange multipliers [3] in which the χ^2 is written as a sum of two terms, e.g.

$$\chi^2 = (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^T \mathbf{V}_{\alpha_0}^{-1} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + 2\boldsymbol{\lambda}^T (\mathbf{D}\delta\boldsymbol{\alpha} + \mathbf{d}), \quad (4)$$

where $\boldsymbol{\lambda}$ is a vector of r unknowns, the Lagrange multipliers. Minimizing the χ^2 with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ yields two vector equations which can be solved for the parameters $\boldsymbol{\alpha}$ and their covariance matrix:

$$\chi^2 = \sum_{j=1}^{3N} \sum_{i=1}^{3N} (p_i^{meas} - p_i^{fit}) W_{ij} (p_j^{meas} - p_j^{fit}) + 2\lambda\phi(\vec{p}, \vec{p}_u)$$

$$\phi(\vec{p}, \vec{p}_u) = 0, \qquad D\ \delta\alpha + d = 0$$

Look at measured and unmeasured variables!

Least Squares properties

Table 2: Properties of the χ^2 estimator. (∞) refers to asymptoic roperties

Problem		Property		
DATA GAUSSIAN?	MODEL LINEAR?	ESTIMATOR EFFICIENT?	ESTIMATION GAUSSIAN?	χ^2 test POSSIBLE?
YES	YES	YES	YES	YES (∞)
YES	NO	YES	YES (∞)	YES (∞)
NO	YES	YES	NO	NO
NO	NO	?	NO	NO