Example of Bayes' Theorem.

Suppose we can identify muons with high efficiency, $\varepsilon = 95\%$.

$$P(+|\mu) = 0.95$$

Suppose that pions are incorrectly identified as muons with probability $\delta = 0.05$.

$$P(+|\pi) = 0.05$$

If we have a sample of tracks in which a fraction f are muons then the probability of a positive identification is

$$P(+) = P(+|\mu)P(\mu) + P(+|\pi)P(\pi)$$
$$= \varepsilon f + \delta(1-f)$$

Question: What is the probability that a positive tag is actually a muon?

$$P(\mu|+) = \frac{P(+|\mu)\,P(\mu)}{P(+)} \quad \text{(Bayes' theorem)}$$
$$= \frac{\varepsilon f}{\varepsilon f + \delta(1-f)}$$

If $f = 0.04$ then

$$P(\mu|+) = 0.44.$$

Statistics :

Sample mean $\bar{x} = \frac{\sum_i x_i}{N}$

$\bar{x}$ is a random variable.

$$\langle \bar{x} \rangle = \langle x \rangle$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$$

(Corrected) sample standard deviation

$$s^2 = \frac{1}{N-1} \sum_i (x_i - \bar{x})^2$$

s^2 is an unbiased estimate of $Var(x)$.

$$\chi^2 = (\vec{f} - \vec{x})^T V^{-1} (\vec{f} - \vec{x})$$

$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}$ V is diagonal when x_i are uncorrelated.

Then $V^{-1} = \begin{pmatrix} 1/\sigma_1^2 & & 0 \\ & 1/\sigma_2^2 & \\ 0 & & \ddots \end{pmatrix}$

χ^2 is a random variable. When $\langle x_i \rangle = f_i$,

$$\langle \chi^2_N \rangle = N$$

$$Var(\chi^2_N) = 2N$$

PDF : $f(x;k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}$

Example: Combining independent, uncorrelated measurements.

Two measurements $y_1 \pm \sigma_1$
$y_2 \pm \sigma_2$

Assuming y_1 and y_2 are distributed as Gaussian random variables about the true value with widths σ_1 and σ_2, the χ^2 function is

$$\chi^2 = \frac{(m - y_1)^2}{\sigma_1^2} + \frac{(m - y_2)^2}{\sigma_2^2}$$

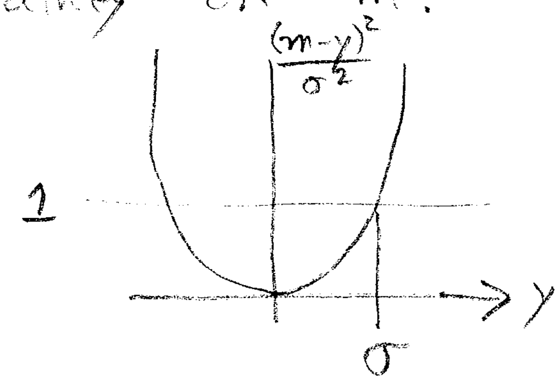
$$\frac{\partial \chi^2}{\partial m} = \frac{2(\hat{m} - y_1)}{\sigma_1^2} + \frac{2(\hat{m} - y_2)}{\sigma_2^2} = 0$$

$$\hat{m} = \frac{\frac{m_1}{\sigma_1^2} + \frac{m_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

(Weighted mean)

What is the uncertainty on $\hat{m}$?

For one variable,



The 1-sigma error is the deviation from the minimum that increases χ^2 by 1.

$$\chi^2 = \frac{(\hat{m} + \sigma_{\hat{m}} - y_1)^2}{\sigma_1^2} + \frac{(\hat{m} + \sigma_{\hat{m}} - y_2)^2}{\sigma_2^2}$$

$$= \frac{(\hat{m} - y_1)^2}{\sigma_1^2} + \frac{(\hat{m} - y_2)^2}{\sigma_2^2} + 1 = \chi^2_{min} + 1 .$$

Solve for $\sigma_{\hat{m}}$:

$$\sigma_{\hat{m}}^2 = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

What if the two measurements are correlated?

$$\chi^2 = \begin{pmatrix} m - y_1 , & m - y_2 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}^{-1} \begin{pmatrix} m - y_1 \\ m - y_2 \end{pmatrix}$$

(Algebra)

$$\hat{m} = \frac{y_1(\sigma_2 - \rho\sigma_1\sigma_2) + y_2(\sigma_1^2 - \rho\sigma_1\sigma_2)}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}$$

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

Alternatively, recognize that we can write

$$\chi^2 = (\vec{f} - \vec{x})^T V^{-1} (\vec{f} - \vec{x})$$

$$V^{-1} = \frac{\partial \chi^2}{\partial x_i \partial x_j}$$

So $\mathrm{cov}(x_i, x_j) = \left(\frac{\partial \chi^2}{\partial x_i \partial x_j}\right)^{-1}$

$$\sigma_{\hat{m}}^2 = \left(\frac{\partial^2 \chi^2}{\partial \hat{m}^2}\right)^{-1}$$

Example of a linear fit with <u>uncorrelated</u> measurements:

Suppose there are N measurements y_i made at x_i with <u>Gaussian</u> uncertainty σ_i.

Then $f_i = \alpha + \beta x_i$

$$\chi^2 = \sum_{i=1}^{N} \frac{(f_i - y_i)^2}{\sigma_i^2}$$

Note that we are assuming we can make a measurement for each x_i. This is different from the case where we fit a histogram, to be discussed later.

What are the best estimates for α and β that describe the data?

These are the values that minimize χ^2.

$$\frac{\partial \chi^2}{\partial \alpha} = 2 \sum_{i=1}^{N} \frac{(f_i - y_i)}{\sigma_i^2} = 0$$

$$\frac{\partial \chi^2}{\partial \beta} = 2 \sum_{i=1}^{N} \frac{(f_i - y_i)\, x_i}{\sigma_i^2} = 0$$

This is a set of linear equations in α, β:

$$\alpha \sum_{i=1}^{N} \frac{1}{\sigma_i^2} + \beta \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} = \sum_{i=1}^{N} \frac{y_i}{\sigma_i^2}$$

$$\alpha \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} + \beta \sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2} = \sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2}$$

Solve using, eg. Kramer's rule.

Generalizes to any polynomial function.

We can always calculate α and β even when the data is <u>not</u> linear.

Question: How well does the linear model describe the data?

Better question: Assuming the model correctly describes the data, what is the probability of obtaining a value of $\chi^2 > \chi^2_{min}$ just by chance?

χ^2 is distributed like a χ^2-distribution

$$f(x;k) = \frac{1}{2^{k/2}\,\Gamma(k/2)}\, x^{k/2-1} e^{-x/2}$$

where $x = \chi^2_{min}$.

Cumulative distribution

$$p = F(x;k) = \int_x^{\infty} f(x;k)\,dx$$

We call $F(\chi^2_{min};k)$ the p-value where $k = N-2$ (two parameters, N measurements) is the # of degrees of freedom.

If χ^2_{min} is small then $p \approx 1$.
If $\chi^2_{min} \gg k$ then $p \approx 0$.

If the data was indeed drawn from a sample described by the model, then p will be uniformly distributed between 0 and 1.

(*) Small p can occur by random chance even when the data is correctly described by the model.