Statistical Methods in Particle Physics

Saeid Paktinat
Yazd University

2nd Workshop for the Collaboration of Iranian Universities with the CMS Experiment at CERN
13 September 2021

Some material from talks by Glen Cowan

Outline

- Definitions
- Famous pdf's
- χ^2 and its applications
- Signal/Background Separation
- Discovery/Exclusion

Definitions(1)

- Kolmogorov axioms (1993):

S is a sample space, A and B are two subset of S, probability is a real value and

$$P(A) \geq 0$$

$$A \cap B = 0 \Rightarrow P(A \cup B) = P(A) + P(B)$$

$$P(S) = 1$$

- Random variable is a numerical characteristic assigned to an element of S. e.g, people's height, weight, …

Definitions(2)

- If x is a continuous variable f(x)dx is the probability to have a measurement which lies between x and x+dx
- f(x) is called the *probability density function* (pdf)

$$\langle u(x) \rangle = \int_{-\infty}^{+\infty} u(x)f(x)dx$$

- Different moments are defined as

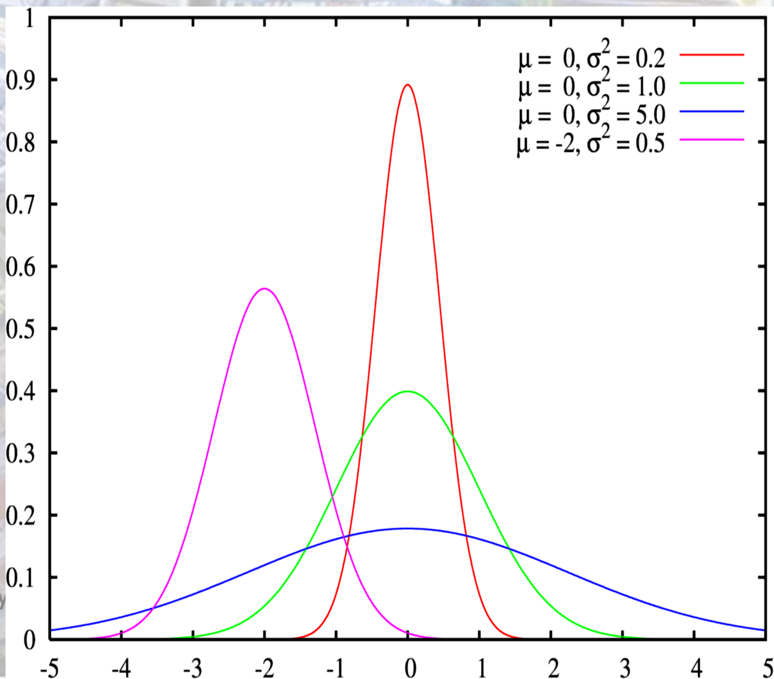
$$\alpha_n = \int_{-\infty}^{+\infty} x^n f(x)dx \qquad \langle x \rangle = \alpha_1 \text{ and } \sigma^2 = \alpha_2 - \langle x \rangle^2$$

- Variance is the square of the standard deviation (Root Mean Square)

Famuos pdf's (Gaussian pdf)

When many small, independent effects are additively contributing to each observation the result follows the Gaussian (normal) distribution, e.g, people's height.

$$f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

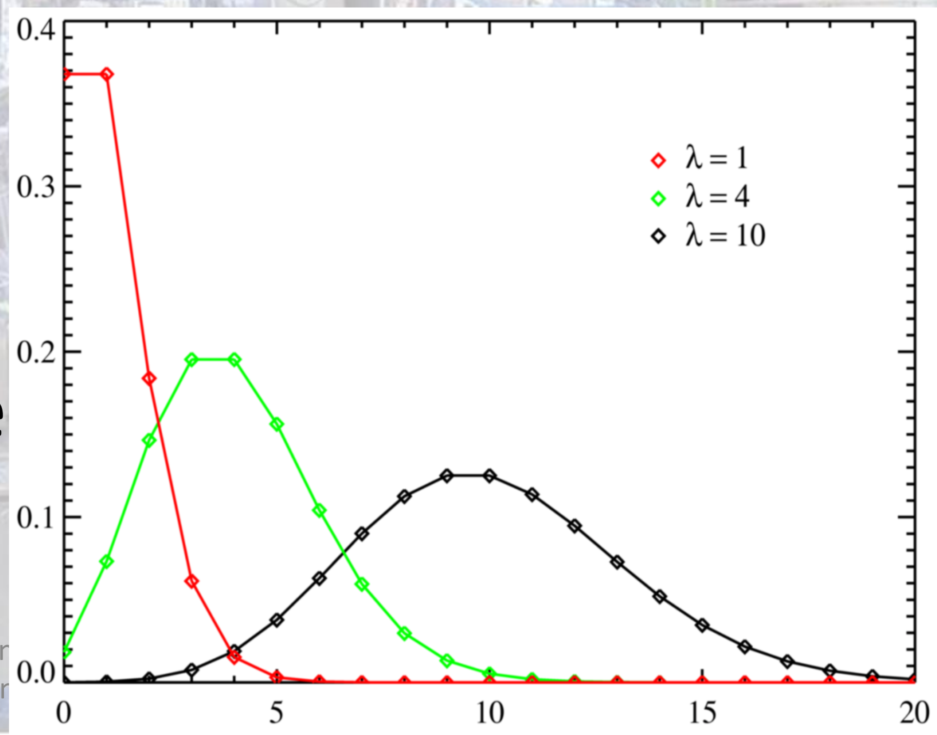


Poisson distribution

- Probability to find n events in a special range when the mean is v.
- variance is equal to v.

$$f(\mathrm{n}, \mathrm{v}) = \frac{\mathrm{v}^{\mathrm{n}} e^{-\mathrm{v}}}{\mathrm{n}!}$$

- Large v approaches the Gaussian pdf.



Chi-square (χ^2) pdf

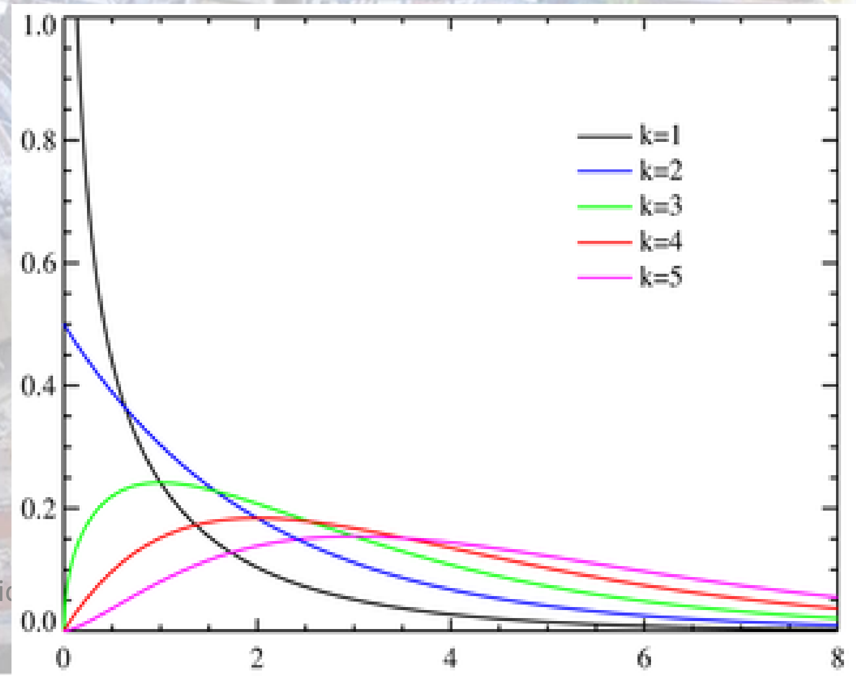
- k independent, normally distributed variables x_i:

$$z = \sum_{i=1}^{k} \frac{(x_i - \mu_i)^2}{\sigma_i^2}$$

- z follows a chi-square pdf with k degrees of freedom.

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

- For the large k, it approaches Gaussian pdf with mean k and variance 2k.



χ^2 application

Chi-square distribution is used as a test

- To estimate the unknown parameters of a model
- To evaluate the unknown mean value of a distribution
- To quantify the goodness of fit

Least squares

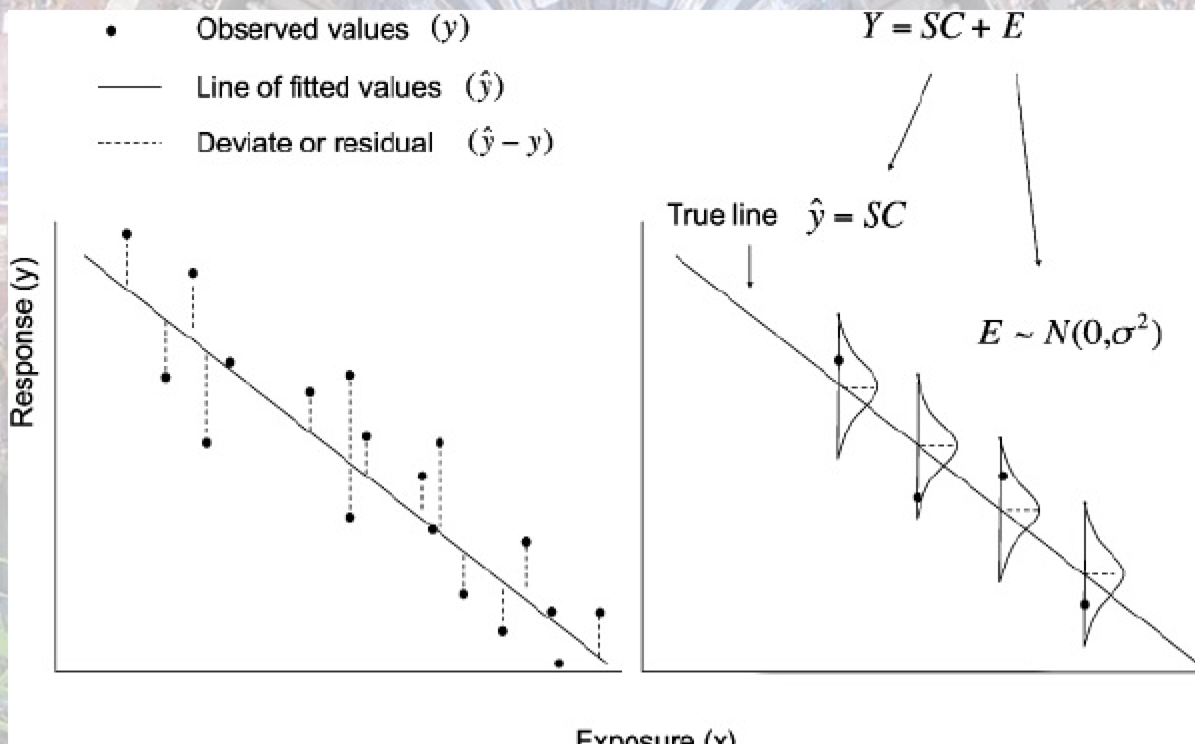
- (x_i, y_i) are the results of an experiment. If y_i has to follow a gaussian pdf with mean $F(x_i,T)$ and a known variance σ_i^2, the correct T will minimize

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - F(x_i, T))^2}{\sigma_i^2}$$

- E.g, V = RI

An example

- extract the slope and the offset of a line…



Parameter estimation

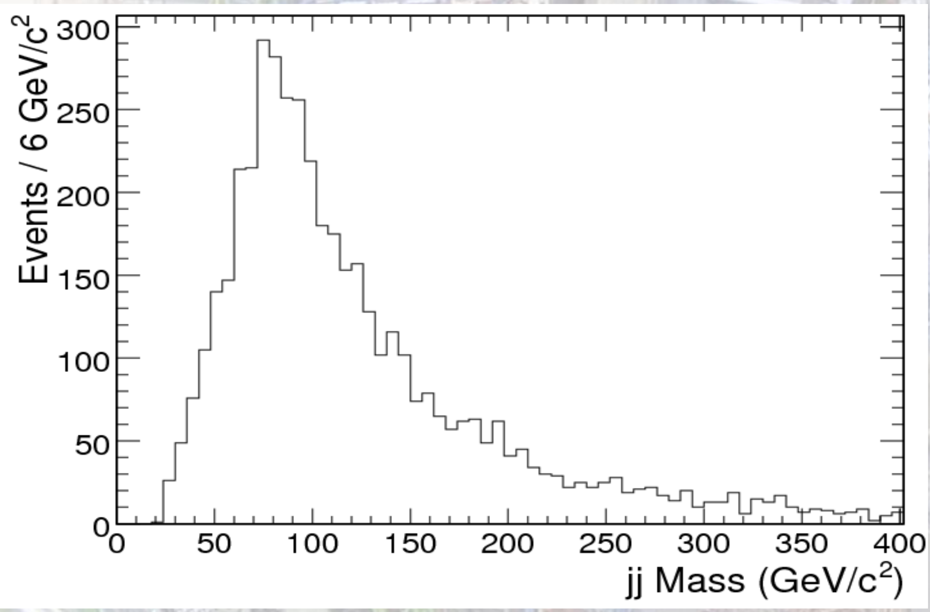
- Maximum likelihood
- If $x_1, x_2, \dots x_n$ are the independent measurements which follow pdf f(x,T) with $T(T_1, T_2, \dots T_m)$ a vector of unknown parameters:
- Likelihood → $L = f(x_1,T) * f(x_2,T) * \dots f(x_n,T)$
- The correct T will maximize the L.

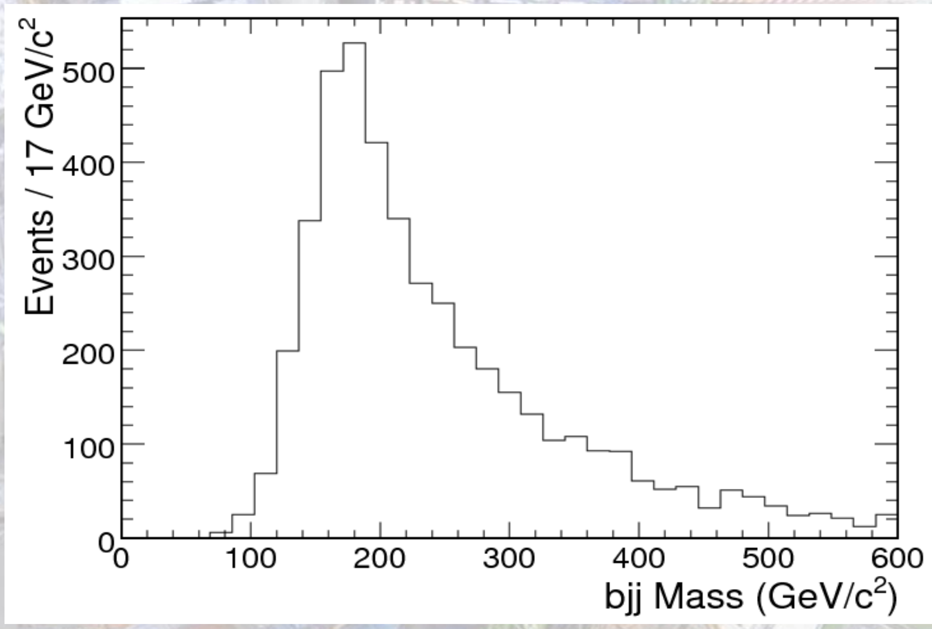
Goodness of fit

- **Goodness of fit** means how well a statistical model fits a set of observations. Measures of goodness of fit typically summarize the discrepancy between observed values and the values expected under the model in question.
- In ROOT   TMath::Prob(χ^2,ndf) gives the probability to find χ^2 with ndf.

mean value of a distribution (Kinematic Fit)

- t→ bW → bjj
- W can be made of two non b-jets
- Top can be made of the extracted W and b-jet.





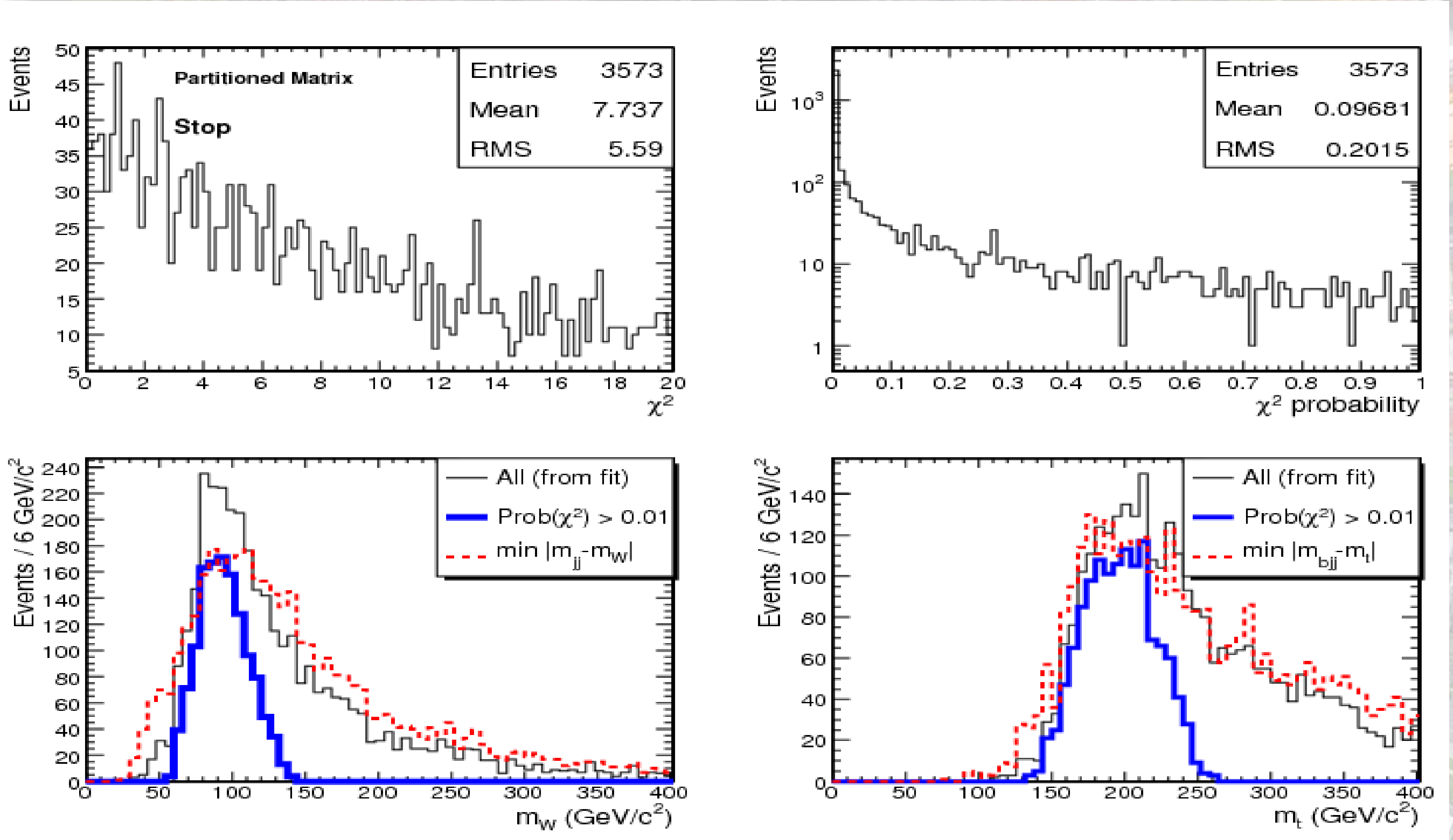
Top quark extraction

- The purpose of the analysis is not to measure the top mass→ top mass can be used with W mass as 2 constraints to find the best jet combination

$$\chi^2 = \sum_{i=1}^{3} \frac{(E_i - E_i^m)^2}{\sigma_i^2} + \frac{(m_W - M_W)^2}{(\Gamma_W/2)^2} + \frac{(m_{Top} - M_{Top})^2}{(\Gamma_{Top}/2)^2}$$

- E_i^m is a measured energy with a gaussian distribution (E_i, σ_i).

The least χ^2 in every event



Selecting events

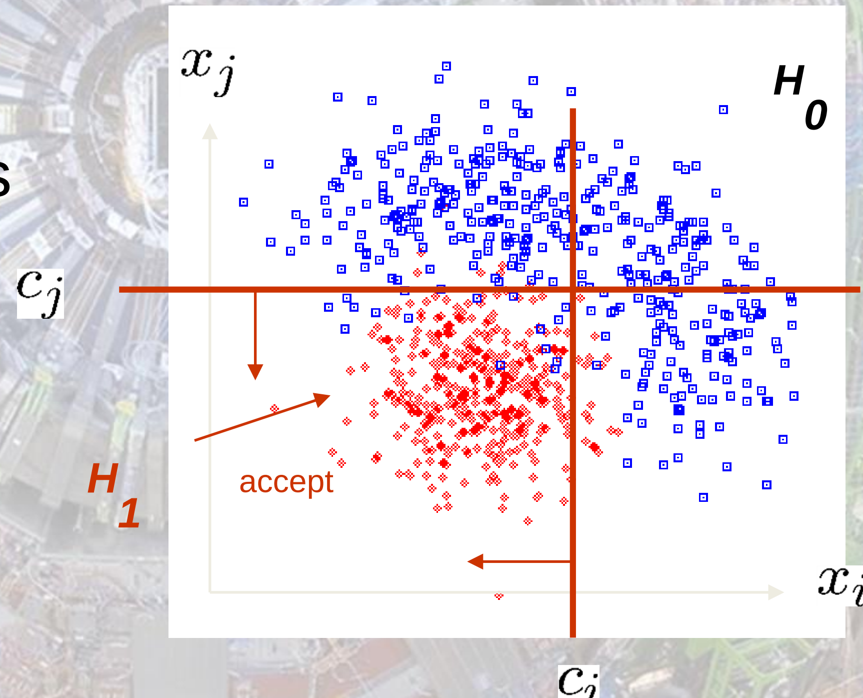
Suppose we have a data sample with two kinds of events, corresponding to hypotheses H_0 and H_1 and we want to select those of type H_1.

Each event is a point in $\vec{x}$ space. What 'decision boundary' should we use to accept/reject events as belonging to event type H_1?

Perhaps select events with 'cuts':

$$x_i < c_i$$

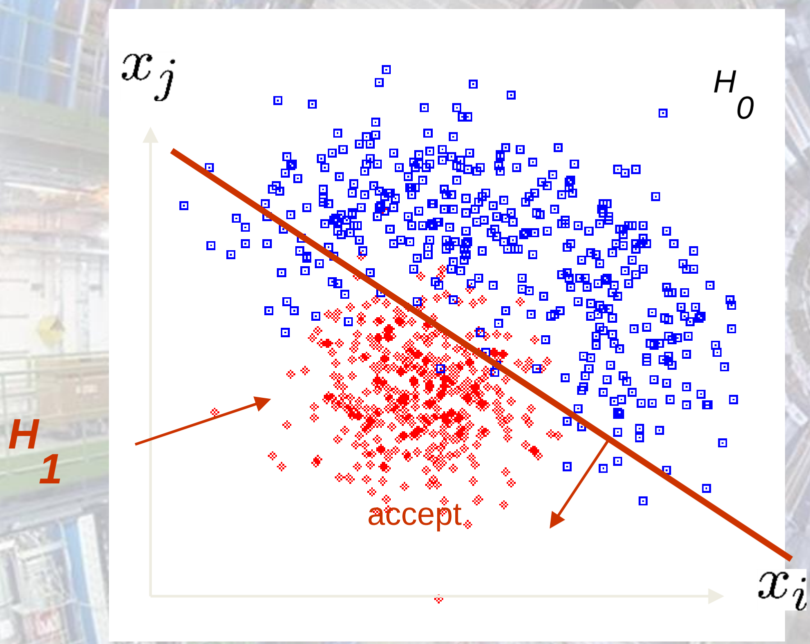
$$x_j < c_j$$



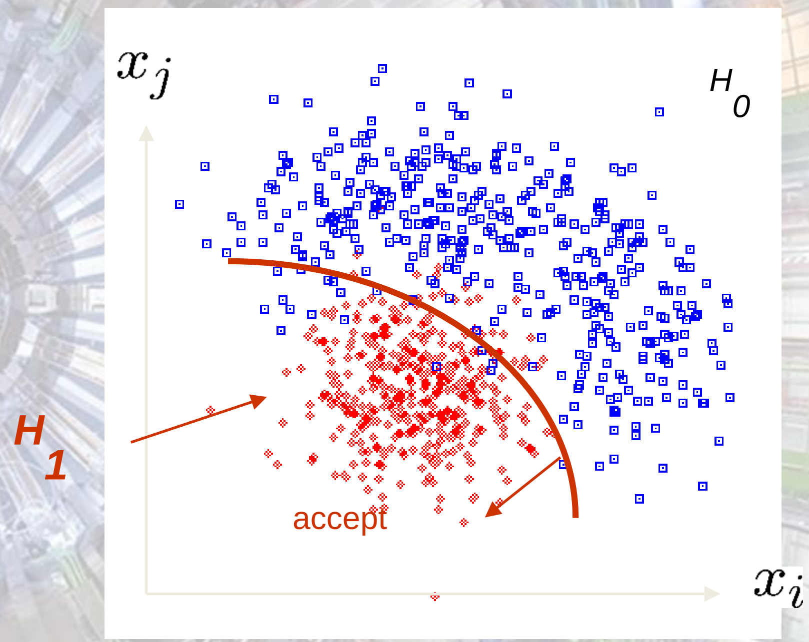
Other ways to select events

Or maybe use some other sort of decision boundary:

linear



or nonlinear



How can we do this in an 'optimal' way?

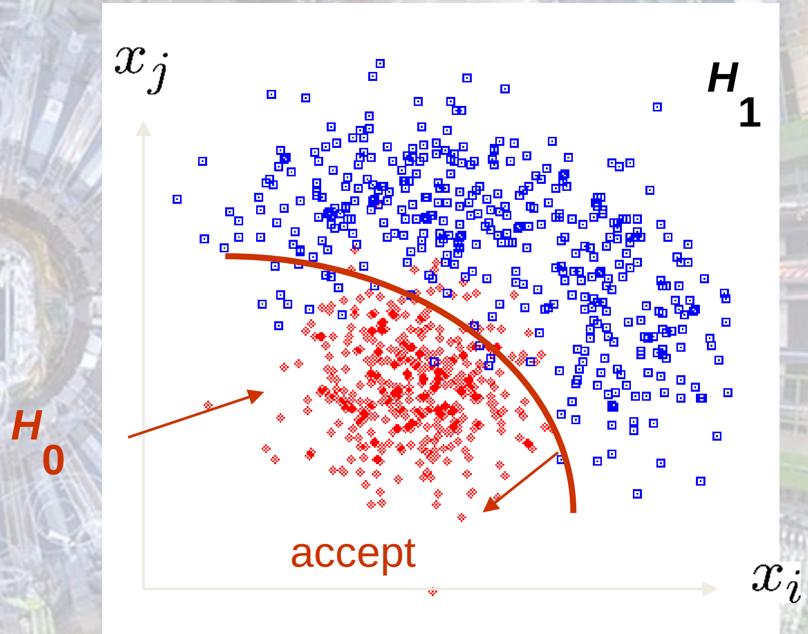
Nonlinear test statistics

The optimal decision boundary may not be a hyperplane,

$\rightarrow$ nonlinear test statistic $t(\vec{x})$

Multivariate statistical methods are a Big Industry:

Neural Networks,
Support Vector Machines,
Kernel density estimation,
Boosted decision trees, ...



Softwares for HEP, e.g.,

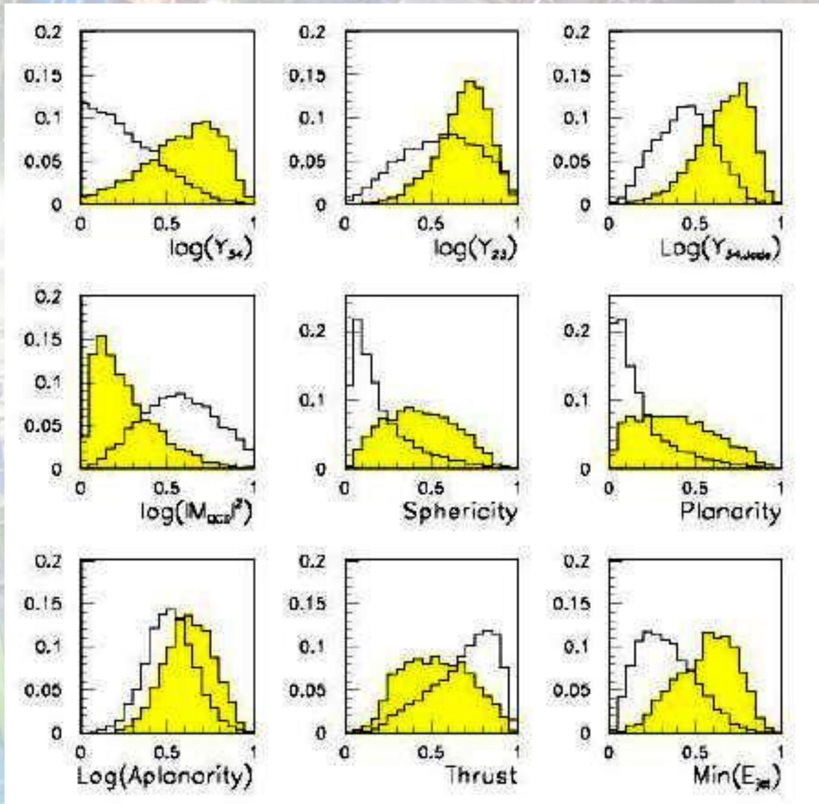
TMVA , Höcker, Stelzer, Tegenfeldt, Voss, Voss, physics/0703039

StatPatternRecognition, I. Narsky, physics/0507143

Neural network example from LEP II

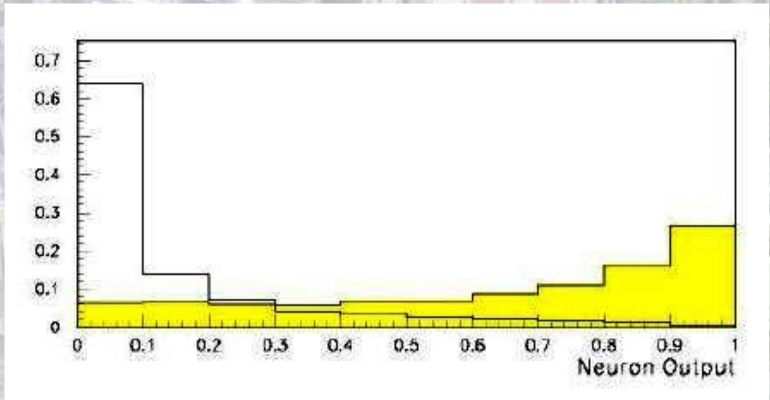
Signal: $e^+e^- \to W^+W^-$ (often 4 well separated hadron jets)

Background: $e^+e^- \to qqgg$ (4 less well separated hadron jets)



← input variables based on jet structure, event shape, ...
none by itself gives much separation.

Neural network output does better...



(Garrido, Juste and Martinez, ALEPH 96-144)

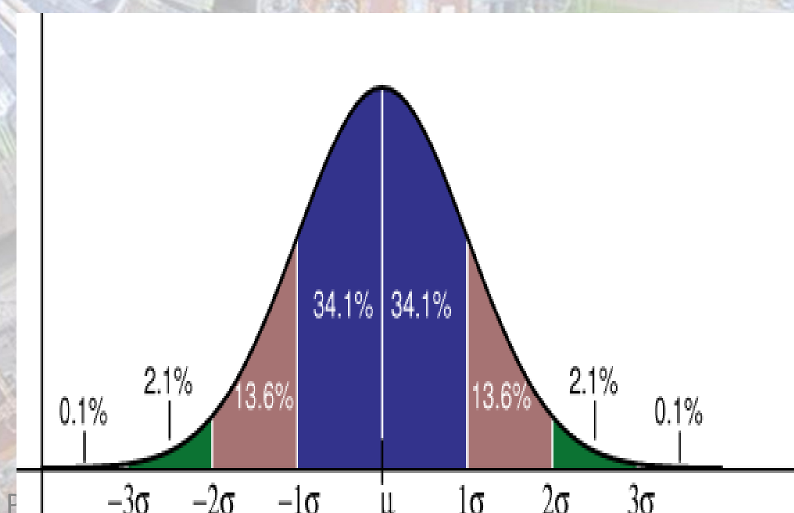
5-sigma significance

- Imagine there is a theory that pdf of the people's height is Gaussian with mean 165cm and RMS 15cm.
- There is somebody as high as 200 cm, is theory violated? What about a man with 280 cm?
- How to quantify this violation.

Some conventions

- A theory is considered to be close to violation if something happens with a probability less than 10^{-3} (3-sigma)
- A theory is excluded if something happens with a probability less than 10^{-7} (5-sigma)

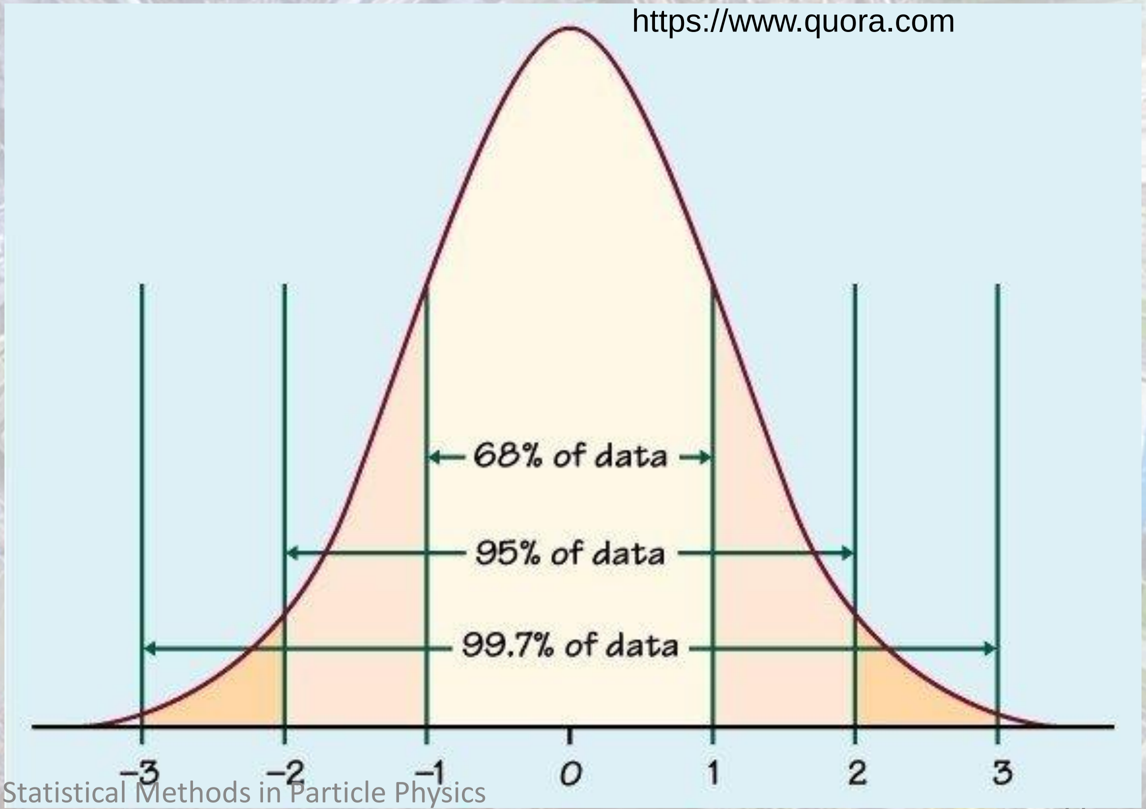
$$p = \int_{x}^{+\infty} f(x)dx$$



5 σ Discovery

- N_{bkg} = n +- σ (Pure SM)
- N_{obs} (Observed)
- Significance

= $(N_{obs} - N_{bkg})/\ \sigma$



An example from hep

- SM without higgs predicts that after a set of cuts 64 events will remain, but in reality 80 events are found, is it the higgs signal?
- No, because 80-64 = 16 = 2 * sqrt(64)
- Remember that RMS for a poisson distribution is sqrt(mean)

$$\text{significance} = \frac{N_{\text{observed}} - N_{\text{predicted}}}{\sigma_{N_{\text{predicted}}}}$$

Significance(s)?!

$$S_1 = \frac{S}{\sqrt{B}}$$

- Statistical uncertainty ONLY!

$$S_2 = \frac{S}{\sqrt{S+B}}$$

50% discovery probability

$$S_{12} = 2(\sqrt{S+B} - \sqrt{B})$$

Confidence Interval

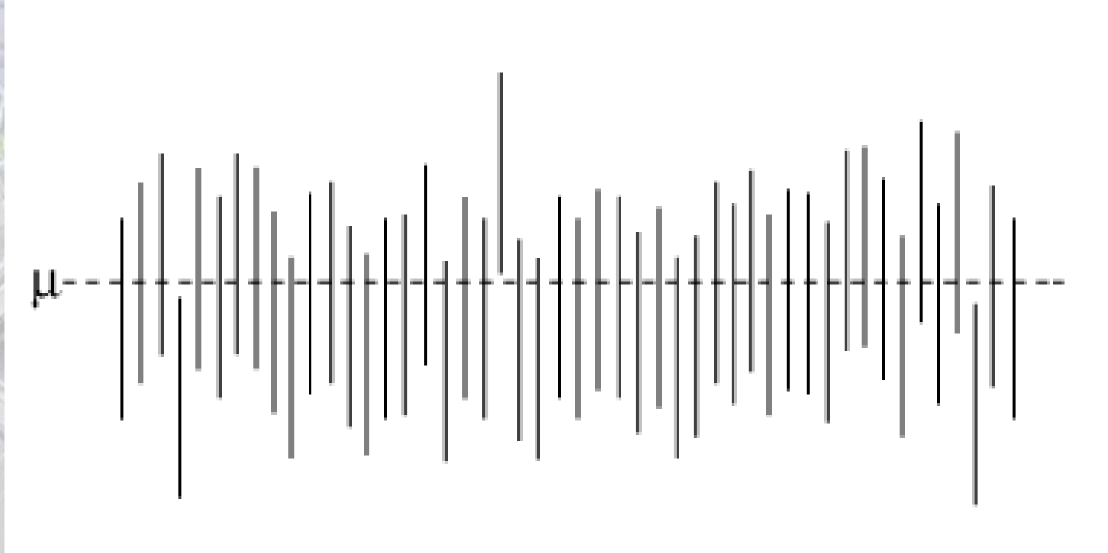
- A quantity X is measured with error σ, the result is used as the estimator of the real value.

$$\bar{x} \pm \sigma \quad \text{or even} \quad \bar{x}^{+\sigma_+}_{-\sigma_-}$$

- It does not mean that $P(\bar{x} - \sigma_- < X < \bar{x} + \sigma_+) = 0.68$
- $(\bar{x} - \sigma_-, x + \sigma_+)$ is the 68% confidence interval for X.

Confidence Interval

- Or if the same method is repeated 100 times in 68 times X will lie inside the confidence intervals.



Confidence Interval

- In an experiment B +/- σ_B events from SM is expected. N events are observed:

  |N - B| < 1.64 (1.96) * σ_B means that SM agrees at 90(95)% C.L. with the experiment.

  |N - B| > 5 * σ_B means that SM is excluded, $P(N;B, \sigma_B) < 10^{-7}$

Confidence Interval

- A model with a new physics predicts $S+B \pm \sigma_{S+B}$; N events are observed:

$|N - S - B| < 1.64\ (1.96) * \sigma_{S+B}$ means that the model with a new physics agrees

90(95)% C.L. with the experiment.

Exclusion

- If $|N - B| < 1.64\ (1.96) * \sigma_B$ (SM agrees at 90(95)% C.L. with the experiment.) Some exclusion limit on new physics (limit on S) can be obtained by solving:

  $|N - S - B| < 1.96 * \sigma_{S+B}$

An example: N = B = 100, σ_{S+B} = sqrt(S+B)

S < 22 at 95% C.L.

95% Exclusion Limit

$$N_{obs} = N_{bkg} + N_{sig} \quad \Rightarrow N_{sig} = N_{obs} - N_{bkg}$$