Statistical Methods in Particle Physics

(Discovery and Exclusion)

Saeid Paktinat
IPM

IPM National School on the Phenomenological And Experimental aspects of the Elementary Particle Physics
16-19 Oct 2007

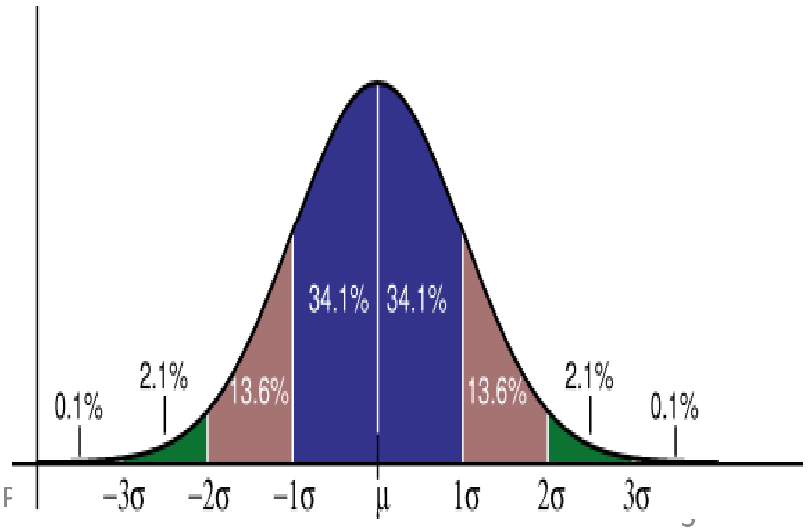
5-sigma significance

- 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)

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



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!

50% discovery probability

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

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

More formal

- n_b is predicted by SM, if n_b is large, probability to observe n_{obs} is:

$$P(n \geq n_{obs}) = \int_{n_{obs}}^{\infty} P_G(x; n_b, n_b)dx = \frac{1}{\sqrt{2\pi}} \int_{S_1}^{\infty} exp(-x^2/2)dx,$$

$$S_1 = \frac{n_{obs} - n_b}{\sqrt{n_b}} \equiv \frac{n_s}{\sqrt{n_b}}$$

Another example from hep

- $H \rightarrow \gamma\gamma$ (m_H = 110 GeV/c^2) @ CMS
  with IL = 30 (20) fb^{-1}
- S = 357 (238), B = 2893 (1929)
- S_1 = 6.6 (5.4)

- It means that CMS can achieve 5σ discovery for H (110 GeV) with the probability of 93(60)%.

  CMS CR 2002/05

Confidence Interval

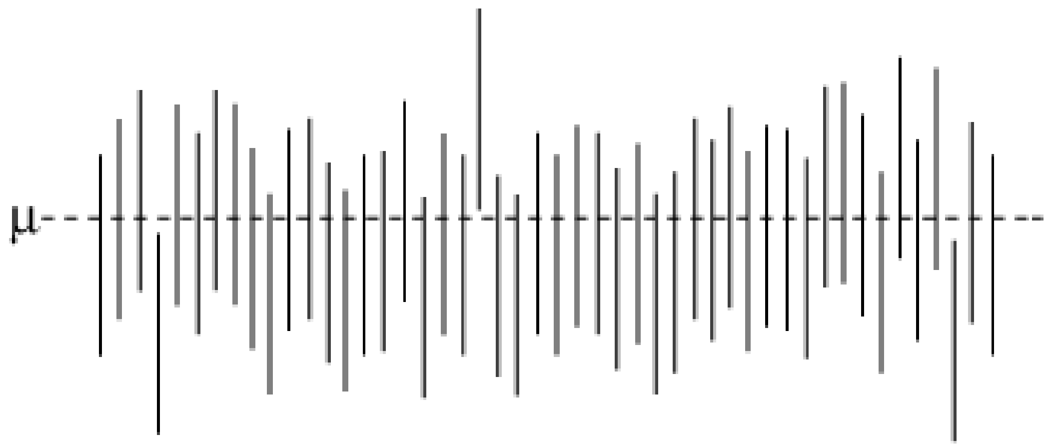
- A quantity X is measured with error σ, the result is used as the estimator of the real value.

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

- 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 $B +/- \sigma_B$ events from SM is expected. N events are observed:

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

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

Confidence Interval

- A model with a new physics predicts $S+B+/- \sigma_{S+B}$; 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.

Back up

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.

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