Modern Particle Physics Experiments

https://xkcd.com/892/

Hypothesis testing

Many aspects of ~~particle~~ physics are of stochastic nature, e.g.:

- **outputs of measurements**

The experiment results are test for various theoretical explanations.

How can we decide between theories?

→ use statistical hypothesis testing paradigm

“theory” – can be a different explanation of some phenomena:

- “SUSY particles extist” vs “SUSY particles do not exist”

or a statement about a value of some parameter:

- “electron neutrino mass is lower than 0.9 eV/c^2” vs

“electron neutrino mass is lower than 0.1 eV/c^2”

Frequentist vs Bayesian

There is an everlasting confusion about answer of question **“Probability of what?”** Two points of view exist:

- frequentist (“classic”): **probability = a fraction of desired outcomes out of many (imagined) trials**

  - one **can not** talk about “probability that a theory is correct”, or “mass of particle X is within limits” as many trials on those would require many Universes

  - one **can talk** about “probability that given (random) range contains a true value”

Frequentist example

Let us use standard ruler, with 0.1cm ticks, to measure length of a pensil.

Measurement tool reading: 15 cm

Measurement uncertainty: $0.1/\sqrt{12} = 0.03$ cm

What is the length of a pensil? 15±0.03 cm

The full sentence: “There is gaussian 1σ probability (68.2%) that a range [15-0.03, 15+0.03] cm covers the true length of a pensil”.

Never, say: “The probability that the length of a pensil …” as there is only one pencil.

Likelihood function

Assume some experiment result: x_0

Assume statistical model, or hypothesis H(θ), for probability density function (pdf) of x:

$$P(x|H) = P(x|\theta)$$

Likelihood: the value of the probability for given experiment result (this is not a probability!):

$$\mathcal{L}(x_0, \theta) = P(x_0|H) = P(x_0|\theta)$$

Hypothesis testing

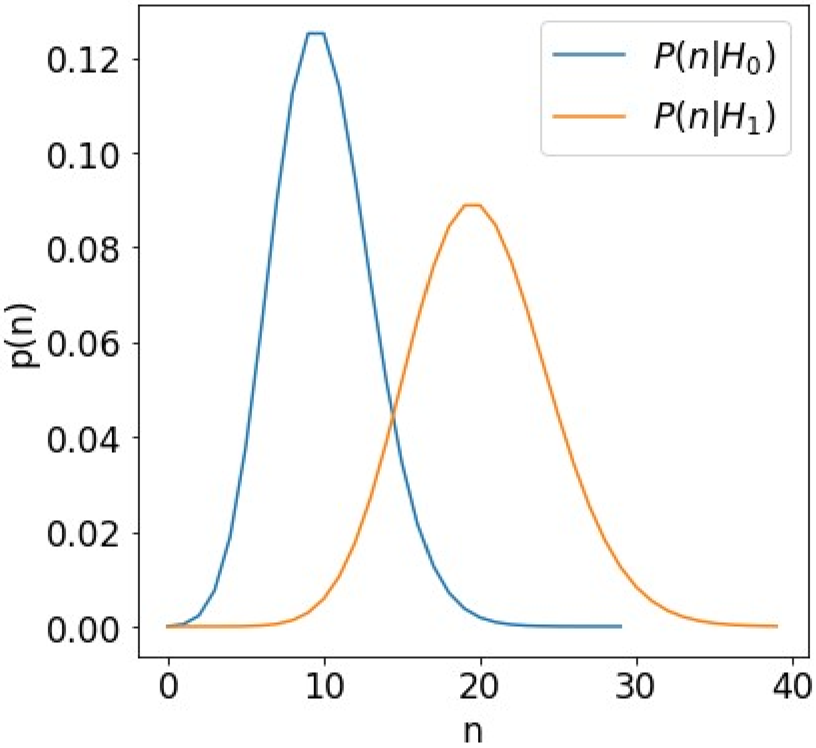
Assume some there are two models, or hypotheses:

$$H_0 - \text{null hypothesis}$$
$$H_1 - \text{alternative hypothesis}$$

Hypotheses provide statistical models for experiment outcome:

$$P(x|H_0)$$
$$P(x|H_1)$$

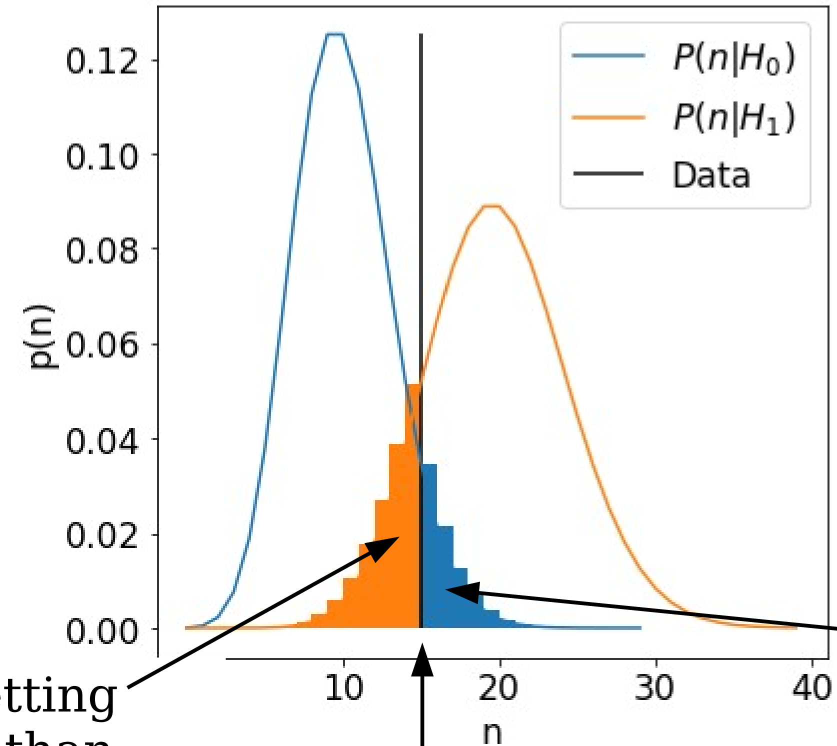
Question: can we reject H_0 given the observed experiment outcome?



Hypothesis testing

Definitions:

- α (Type I error) – probability of rejecting H_0 when is is true. α is called a **significance of the test**.
- β (Type II error) – probability of accepting H_0 when it is false



β - probability of getting this, or less events than some threshold, when H_1 when is true

the decision threshold

α - probability of getting this, or more events than some threshold, when H_0 when is true

Hypothesis testing

Optimization task:

- minimize β (Type II error) for given value of α (Type I error)

Solution:

- do not look at observable (e.g. event count) p.d.f, but at the p.d.f for the observable likelihood ratio:

$$\lambda = \frac{\mathcal{L}(x_0|H_0)}{\mathcal{L}(x_0|H_1)}$$

- or more computationally convenient:

$$\boxed{t = -2\ln\frac{\mathcal{L}(x_0|H_0)}{\mathcal{L}(x_0|H_1)}}$$

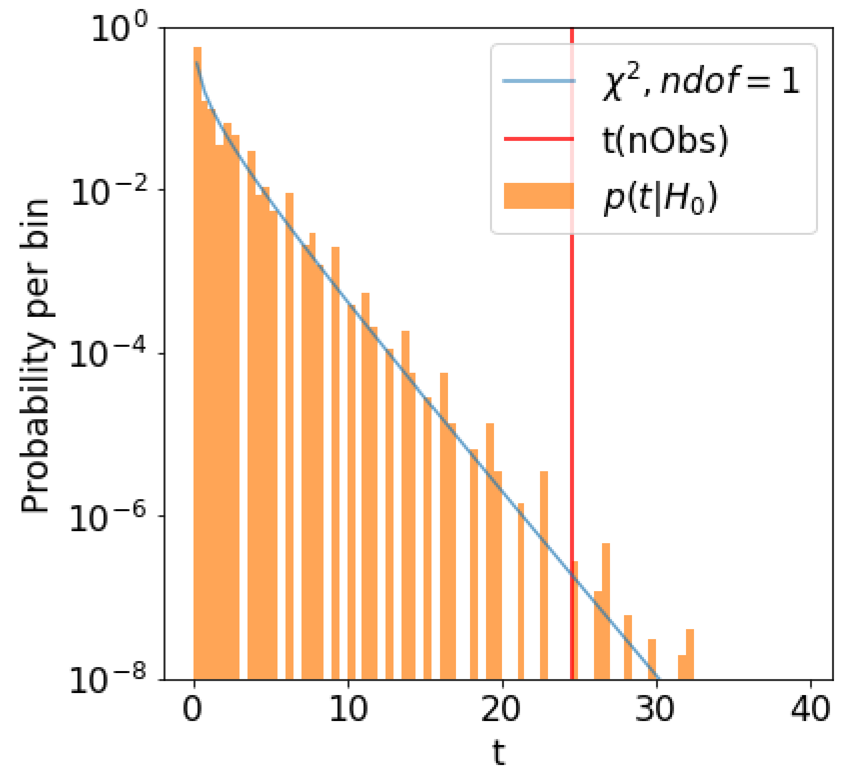
Hypothesis testing

Algorithm:

- choose the α value (Type I error) **before** the experiment.

- calculate the likelihood ratio distribution, and find t_0:

$$t_0 : \int_{t_0}^{\infty} P(t|H_0)\, dt = \alpha$$



(luckily, often the distribution of t can be approximated by Gaussian distribution, and one can use analytical approximations)

- look at the data, and calculate t_{DATA}
- reject H_0 if $t_{\mathrm{DATA}} > t_0$

Asymptotic formulas

Wilks theorem: in the limit of large amount of data the likelihood ratio for data drawn from H_0 model approaches the χ^2 distribution with number of degrees of freedom ("dof", or "ndof") equal to difference of number of free parameters in H_0 and H_1:

$$P(t|H_0) \rightarrow \chi^2_{dof\,H_1 - dof\,H_0}$$

If there is only one free parameter (e.g. the number of signal events, S) we have:

$$P(t|H_0) \rightarrow \chi^2_1$$

$$p_0 = \int_{t_{obs}}^{\infty} P(t|H_0)\,dt = 2\left(1 - \Phi\left(\sqrt{t_{obs}}\right)\right)$$

often called p-value

Gaussian cumulative probability function (cdf)

Asymptotic formulas

The t variable is sensitive for both up- and downward fluctuations. In case of looking for a new signal, one is interested only in additional events on top of known background. In that case a modified variable is better:

$$\boxed{q_0 = \begin{cases} -2\ln\dfrac{\mathcal{L}(x_0|H_0=B)}{\mathcal{L}(x_0|H_1=\hat{S}+B)} & \hat{S}>0 \\ 0 & \hat{S}<0 \end{cases}}$$

data best fit S estimate

$$p_0 = \int_{q_{0,obs}}^{\infty} P(q|H_0)\,dt = 1-\Phi\left(\sqrt{q_{0,obs}}\right)$$

Gaussian cumulative probability function (cdf)

$$z = \Phi^{-1}(1-p_0) = \sqrt{q_0}$$

Number of single sided Gaussian sigma corresponding to p_0

Asymptotic formulas

Poisson case:

H_0 – expected number of events is B

H_1 – expected number of events is μ_{H1} – the data best fit value: μ_{H1} = nObs

$$q_0(nObs) = -2\ln\frac{\mathcal{L}(x_0|H_0)}{\mathcal{L}(x_0|H_1)} = -2\ln\left(\frac{B^{nObs}}{nObs!}e^{-B} \Big/ \frac{(\mu_{H_1})^{nObs}}{nObs!}e^{-\mu_{H_1}}\right) =$$

$$-2\ln\left(\frac{B^{nObs}}{(nObs)^{nObs}}e^{-B+nObs}\right) = -2[nObs(\ln(B)-\ln(nObs))+nObs-B] =$$

$$2[nObs(\ln(nObs)-\ln(B))+B-nObs]$$

Asymptotic formulas

- if B is known, then S = nObs – B
- if we plan an experiment for some known S and B, then expected nObs = S + B

$$z \simeq \sqrt{q_0} = \sqrt{2[nObs[\ln(nObs) - \ln(B)] + B - nObs]} = |nObs = S + B| =$$

$$= \sqrt{2[(S+B)\ln(\frac{S+B}{B}) + B - B - S]} = \boxed{\sqrt{2[(S+B)\ln(1+\frac{S}{B}) - S]}} =$$

$$\sqrt{2[B(1+\frac{S}{B})\ln(1+\frac{S}{B}) - B\frac{S}{B}]} \simeq |x = \frac{S}{B}, S \ll B \rightarrow x \ll 1| \simeq$$

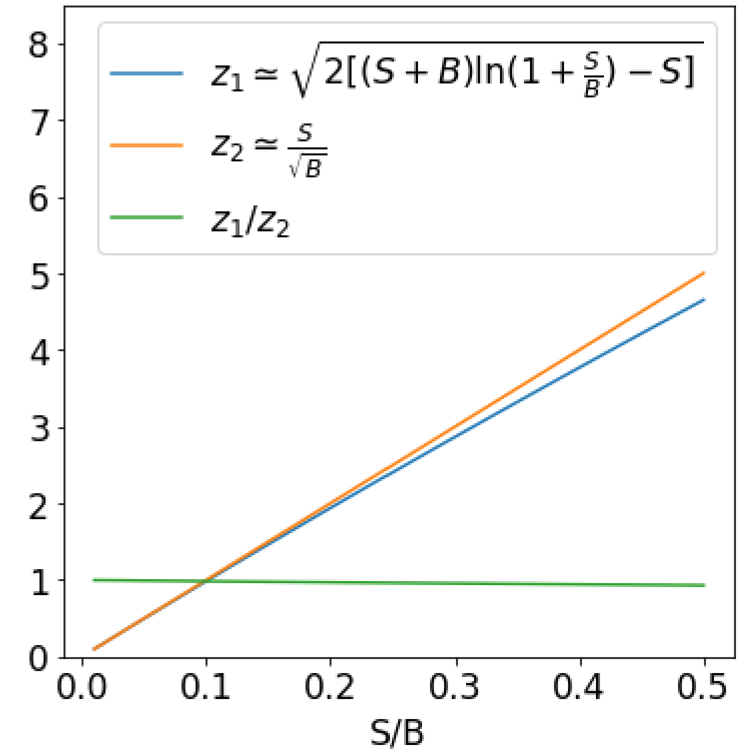
$$\sqrt{2B[(1+x)\log(1+x) - x]} \simeq \sqrt{2B[(1+x)(x - \frac{x^2}{2}) - x]} \simeq |\text{terms up to } x^2|$$

$$\sqrt{2B[x - \frac{x^2}{2} + x^2 - x]} = \sqrt{2B[\frac{x^2}{2}]} = \sqrt{\frac{S^2}{B}} = \boxed{\frac{S}{\sqrt{B}}}$$

Asymptotic formulas

- large amount of data: $$z \simeq \sqrt{q_0} = \sqrt{2\left[nObs\left[\ln\left(nObs\right) - \ln\left(B\right)\right] + B - nObs\right]}$$

- in addition S≪B: $$z \simeq \frac{S}{\sqrt{B}}$$



A dataset for which the best fit values reproduce the assumed values, e.g $S_{\text{best fit}}$ = S is called

"Asimov dataset"

Similarly putting nObs = S + B is called using Asimov values for observation.

Discovery/Exclusion

Discovery:

H_0 – background only hypothesis (B)

H_1 – best fit signal + background hypothesis ($\hat{S}$+B)

test statistics - q_0

Usual significance of the test: **"5σ" - α = 2.87·10^{-7}**

Exclusion:

H_0 – nominal signal+background hypothesis (S+B)

H_1 – best fit signal +background hypothesis ($\hat{S}$+B)

test statistics - q_μ

Usual significance of the test: **"95 CL" - α = 0.05**

$q_{0/\mu}$ test statistics

Discovery:

$$\boxed{q_0 = \begin{cases} -2\ln\dfrac{\mathcal{L}(x_0|H_0=B)}{\mathcal{L}(x_0|H_1=\hat{S}+B)} & \hat{S}>0 \\ 0 & \hat{S}<0 \end{cases}}$$

Exclusion:

$$\boxed{q_\mu = \begin{cases} -2\ln\dfrac{\mathcal{L}(x_0|H_0=S+B)}{\mathcal{L}(x_0|H_1=\hat{S}+B)} & \hat{S}<S \\ 0 & \hat{S}>S \end{cases}}$$

Important:

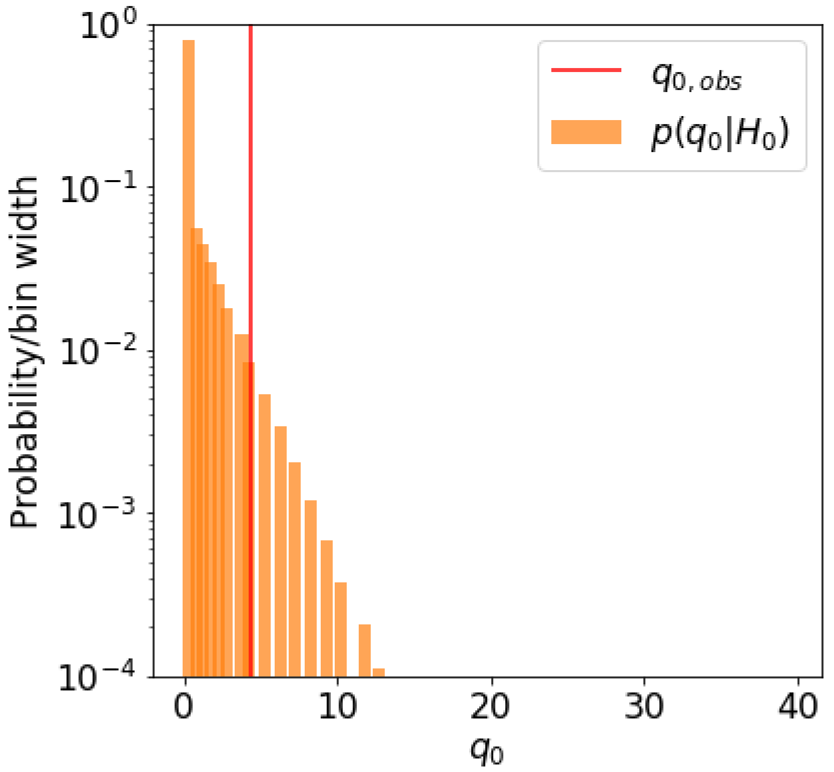
$\hat{S}$ is the best fit S value to the observed data!

Discovery example

Assumptions and observations:

$\alpha = 2.87 \cdot 10^{-7}$, $B = 20$, $N_{data} = 30$

$S_{best\ fit} = 10$



Can we reject the null hypothesis?

```
q0_Obs: 4.33
MC estimate for p(q0>q0_Obs): 1.35E-02, single sided Gaussian sigma: 2.2
Gauss esimate for p(q0>q0_Obs): 1.87E-02, single sided Gaussian sigma: 2.1

alpha: 2.87E-07
Null hypothesis (B only) is NOT REJECTED
```

Discovery: expected sensitivity

Assumptions:

$\alpha = 2.87\cdot10^{-7}$, B = 20, S = 10

What would be significance for the nominal S?

$$z \simeq \sqrt{q_0}, \quad q_{0,Asimov} = q_0(nObs = S + B)$$

Discovery: expected sensitivity

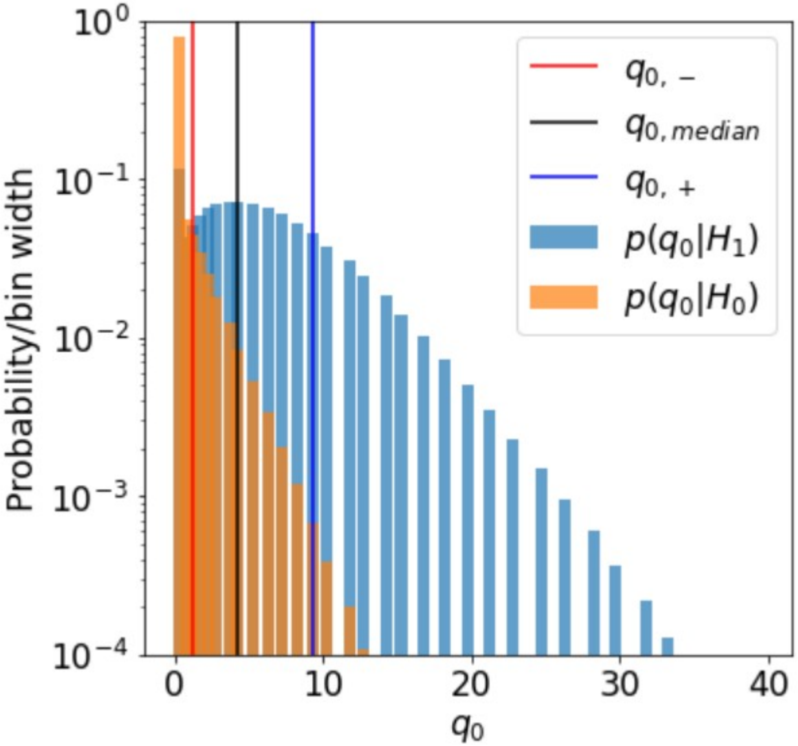
What would we observe in experiment?

q_0 values fluctuate according to number of observed events (assuming S+B), so we define:

$q_{0,median}$ – 50% of experiments will observe this or greater value

$q_{0,-}$ – 84% of experiments will observe this or greater value

$q_{0,+}$ – 16% of experiments will observe this or greater value



Discovery: expected sensitivity

What significance of planned experiment we can expect?

Take the $q_{0,\ -/median/+}$ values calculated for S+B data, and check $p(q_{0,-/median/+}>t)$ assuming H_0 (B only)

```
q0 median from Asimov dataset = 4.33
MC estimate for p(q0>q0_Asimov) under H1:        4.52E-01 (expected: 0.50)

q0 minus = 1.25, q0 median = 4.25, q0 plus = 9.25
MC estimate for p(q0>q0_minus) under H1:         7.93E-01 (expected: 0.84)
MC estimate for p(q0>q0_median) under H1:        4.52E-01 (expected: 0.50)
MC estimate for p(q0>q0_plus) under H1:          1.57E-01 (expected: 0.16)

Gauss esimate for p(q0>q0_minus):        1.32E-01,      single sided Gaussian sigma: 1.1
Gauss esimate for p(q0>q0_median):       1.96E-02,      single sided Gaussian sigma: 2.1
Gauss esimate for p(q0>q0_Asimov):       1.87E-02,      single sided Gaussian sigma: 2.1
Gauss esimate for p(q0>q0_plus):         1.18E-03,      single sided Gaussian sigma: 3.0
```

Homework: exclusion limits

Assumptions:

$\alpha = 0.05$, B = 200

What minimal number of signal events can be rejected for 50% of hypothetical experiments?

Expected answer:

```
Expectded median exclusion on S = 25
```

Additional task (for 200% points)

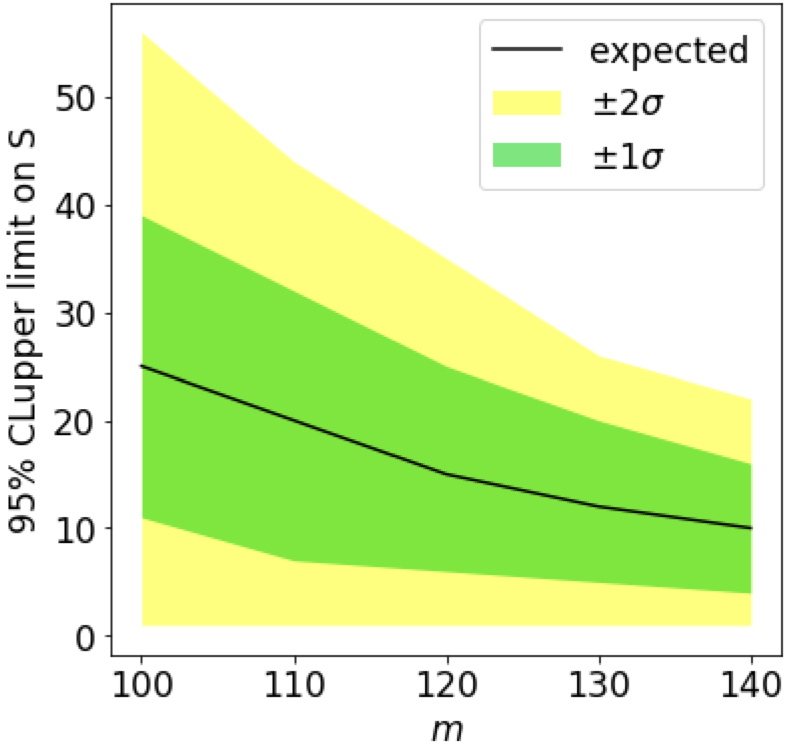
Assumptions:

- experiment is looking for a particle of unknown mass: $100<m_X<150$
- expected background is: $B(m_X) = 200 \cdot e^{-(X-100)/20}$

Task:

Plot excluded number of signal events for 97.6%, 84%, 50%, 16% 2.4% of experiments, as a function of m_X

Expected result:



Additional task (for 200% points)

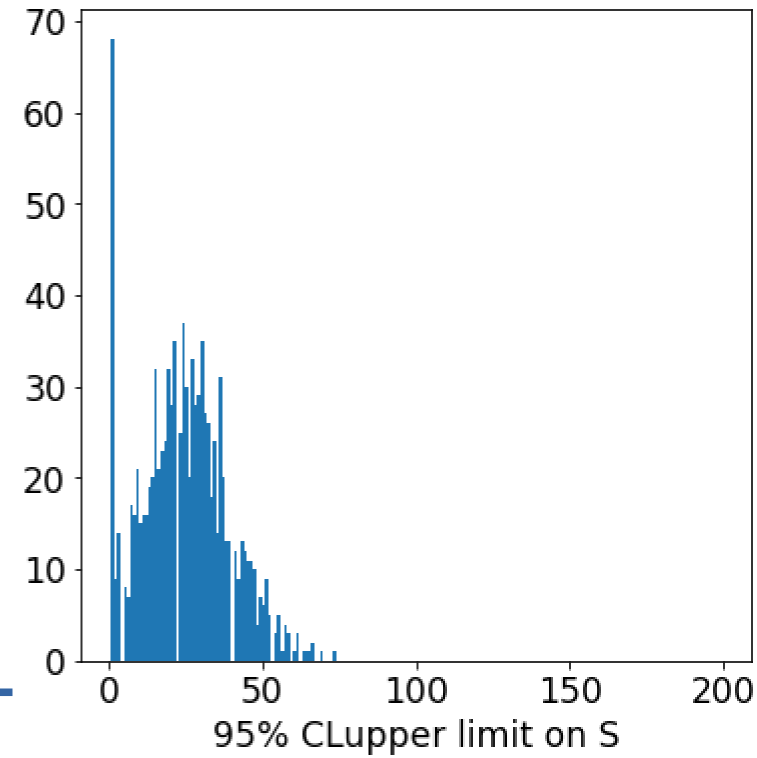
Hints:

- generate 1000 pseudo experiments according to H_1 – the number of observed events in each pseudo experiment follows Poisson(B)

- for each pseudo experiment test many S hypotheses, until S is excluded. The upper limit should be calculated using $H_0 = S + B$, where **B is the nominal B**

- create a histogram of excluded S values

- from histogram find the ±2σ, ±1σ and median values of S



Materials

Review articles of probability and statistics on PDG portal:

- *The Review of Particle Physics*

Lectures on Statistics:

- B. Cousins, *Lectures on Statistics:in Theory: Prelude to Statistics in Practice*

More details on asymptotic formulas:

- G.Cowan, et al, *Asymptotic formulae for likelihood-based tests of new physics*

Frequentist vs Bayesian

- Bayesian: **probability = a subjective degree of belief**
  - one **can talk** about probability of whatever, since this is your degree of belief that given statement if true
  - can be coherent between different people, give rigid prescription of assigning the probability

Bayesian reasoning usually used the Bayes theorem:

$$P(B|A) = \frac{P(A|B)\boxed{P(B)}}{P(A)}$$

posterior probability

prior probability