Models and Methods for Beyond Standard Model Physics at colliders

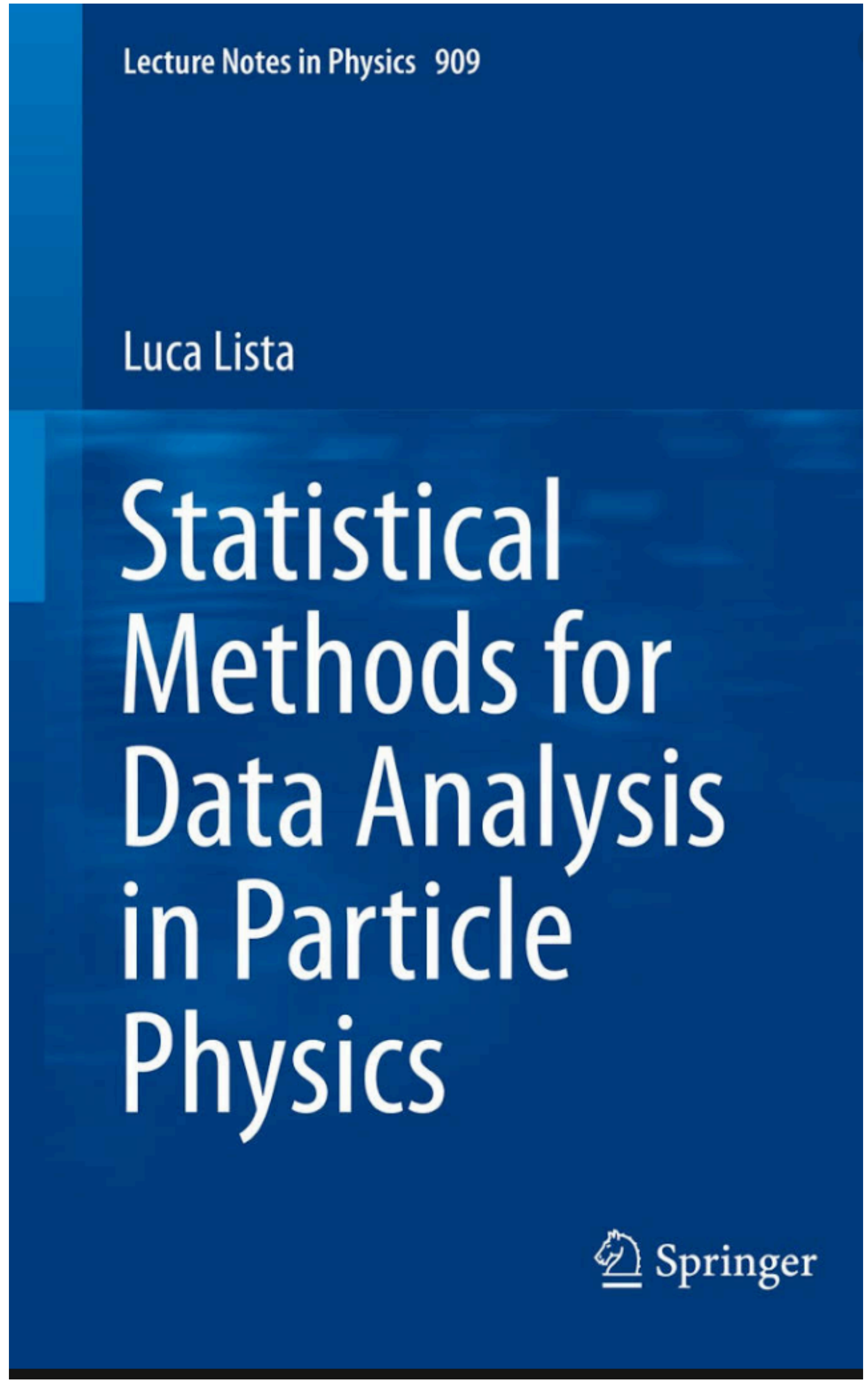
Lectures for the Ph.D. Program in Physics, XXXVI Cycle

19/04/2021

Vieri Candelise
University of Trieste

Chapter III

Statistics for HEP



Lectures taken from “Statistical Methods for Data Analysis in Particle Physics” by Luca Lista

statìstica

Vocabolario on line

Crea un ebook con questa voce | Scaricalo ora (0)   Condividi

Nel significato originario, da cui trae il nome, essa rappresenta "la scienza che si occupa della raccolta e la classificazione di certi fatti concernenti la popolazione di uno *Stato*" (Webster's). Detto con le parole di Trilussa, "*È 'na cosa / che serve pe' fa' un conto in generale / de la gente che nasce, che sta male, / che more, che va in carcere e che sposa.*". In questa accezione essa è più propriamente nota come *statistica descrittiva*.

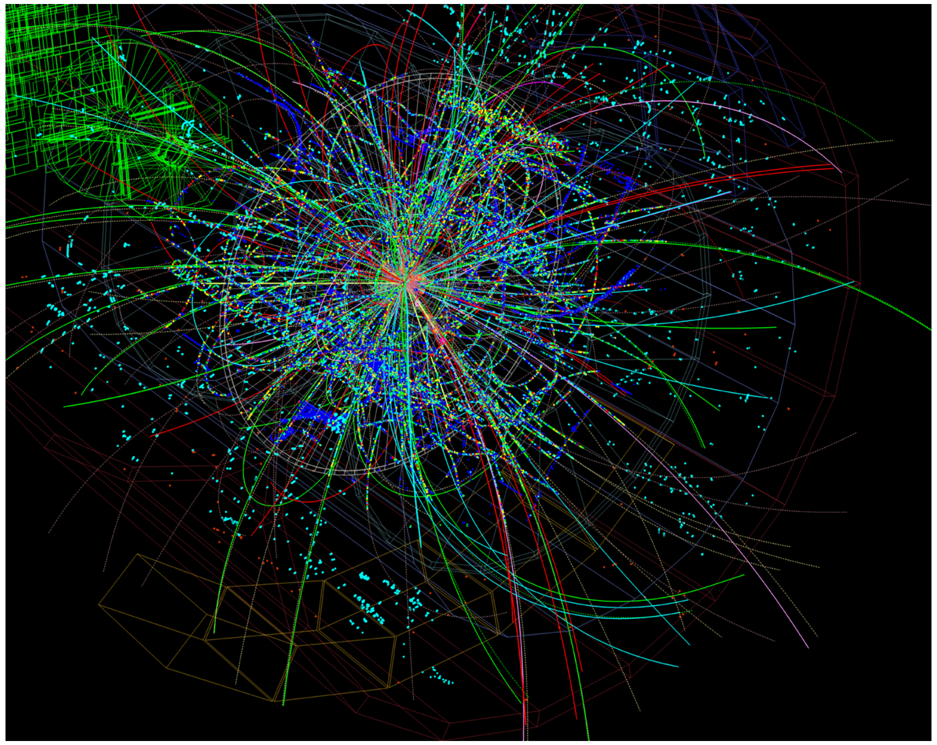
(D'Agostini)

Statistics in HEP

Particle collisions are recorded in form of data delivered by detectors
– Measurements of particle position in the detector, energy, time, ...

Usually a large number of collision events are collected by an experiment, each event usually containing large amounts of data

Intrinsic randomness of physics process



Collision event data are all different from each other (Quantum Mechanics: $P \propto |M|^2$)

- **Detector response** is somewhat random
- Fluctuations, resolution, efficiency,....

Statistics in HEP

Distributions of measured quantities in data:

– are predicted by a theory model,

– depend on some theory parameters,

– e.g.: particle mass, cross section, etc.

$$
\begin{gathered}
\mathcal{L}_{SM} = -\tfrac{1}{2}\partial_\nu g^a_\mu \partial_\nu g^a_\mu - g_s f^{abc}\partial_\mu g^a_\nu g^b_\mu g^c_\nu - \tfrac{1}{4}g_s^2 f^{abc} f^{ade} g^b_\mu g^c_\nu g^d_\mu g^e_\nu - \partial_\nu W^+_\mu \partial_\nu W^-_\mu - \\
M^2 W^+_\mu W^-_\mu - \tfrac{1}{2}\partial_\nu Z^0_\mu \partial_\nu Z^0_\mu - \tfrac{1}{2c_w^2} M^2 Z^0_\mu Z^0_\mu - \tfrac{1}{2}\partial_\mu A_\nu \partial_\mu A_\nu - igc_w(\partial_\nu Z^0_\mu (W^+_\mu W^-_\nu - \\
W^+_\nu W^-_\mu) - Z^0_\nu (W^+_\mu \partial_\nu W^-_\mu - W^-_\mu \partial_\nu W^+_\mu) + Z^0_\mu (W^+_\nu \partial_\nu W^-_\mu - W^-_\nu \partial_\nu W^+_\mu)) - \\
igs_w(\partial_\nu A_\mu (W^+_\mu W^-_\nu - W^+_\nu W^-_\mu) - A_\nu (W^+_\mu \partial_\nu W^-_\mu - W^-_\mu \partial_\nu W^+_\mu) + A_\mu (W^+_\nu \partial_\nu W^-_\mu - \\
W^-_\nu \partial_\nu W^+_\mu)) - \tfrac{1}{2}g^2 W^+_\mu W^-_\mu W^+_\nu W^-_\nu + \tfrac{1}{2}g^2 W^+_\mu W^-_\nu W^+_\mu W^-_\nu + g^2 c_w^2 (Z^0_\mu W^+_\mu Z^0_\nu W^-_\nu - \\
Z^0_\mu Z^0_\mu W^+_\nu W^-_\nu) + g^2 s_w^2 (A_\mu W^+_\mu A_\nu W^-_\nu - A_\mu A_\mu W^+_\nu W^-_\nu) + g^2 s_w c_w (A_\mu Z^0_\nu (W^+_\mu W^-_\nu - \\
W^+_\nu W^-_\mu) - 2A_\mu Z^0_\mu W^+_\nu W^-_\nu) - \tfrac{1}{2}\partial_\mu H \partial_\mu H - 2M^2 \alpha_h H^2 - \partial_\mu \phi^+ \partial_\mu \phi^- - \tfrac{1}{2}\partial_\mu \phi^0 \partial_\mu \phi^0 - \\
\beta_h \left(\tfrac{2M^2}{g^2} + \tfrac{2M}{g} H + \tfrac{1}{2}(H^2 + \phi^0\phi^0 + 2\phi^+\phi^-) \right) + \tfrac{2M^4}{g^2}\alpha_h - \\
g\alpha_h M \left(H^3 + H\phi^0\phi^0 + 2H\phi^+\phi^-\right) - \\
\tfrac{1}{8}g^2 \alpha_h \left(H^4 + (\phi^0)^4 + 4(\phi^+\phi^-)^2 + 4(\phi^0)^2\phi^+\phi^- + 4H^2\phi^+\phi^- + 2(\phi^0)^2 H^2\right) - \\
gM W^+_\mu W^-_\mu H - \tfrac{1}{2}g\tfrac{M}{c_w^2} Z^0_\mu Z^0_\mu H - \\
\tfrac{1}{2}ig \left(W^+_\mu (\phi^0 \partial_\mu \phi^- - \phi^- \partial_\mu \phi^0) - W^-_\mu (\phi^0 \partial_\mu \phi^+ - \phi^+ \partial_\mu \phi^0) \right) + \\
g \left(W^+_\mu (H \partial_\mu \phi^- - \phi^- \partial_\mu H) + W^-_\mu (H \partial_\mu \phi^+ - \phi^+ \partial_\mu H) \right) + \tfrac{1}{2}g\tfrac{1}{c_w}(Z^0_\mu (H \partial_\mu \phi^0 - \phi^0 \partial_\mu H) + \\
(\tfrac{1}{c_w} Z^0_\mu \partial_\mu \phi^0 + W^+_\mu \partial_\mu \phi^- + W^-_\mu \partial_\mu \phi^+) - ig\tfrac{s_w^2}{c_w} M Z^0_\mu (W^+_\mu \phi^- - W^-_\mu \phi^+) + igs_w M A_\mu (W^+_\mu \phi^- - \\
W^-_\mu \phi^+) - ig\tfrac{1-2c_w^2}{2c_w} Z^0_\mu (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) + igs_w A_\mu (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) - \\
\tfrac{1}{4}g^2 W^+_\mu W^-_\mu \left(H^2 + (\phi^0)^2 + 2\phi^+\phi^-\right) - \tfrac{1}{8}g^2 \tfrac{1}{c_w^2} Z^0_\mu Z^0_\mu \left(H^2 + (\phi^0)^2 + 2(2s_w^2 - 1)^2 \phi^+\phi^-\right) - \\
g^2 \tfrac{s_w^2}{c_w} Z^0_\mu \phi^0 (W^+_\mu \phi^- + W^-_\mu \phi^+) - \tfrac{1}{2}ig^2 \tfrac{s_w^2}{c_w} Z^0_\mu H (W^+_\mu \phi^- - W^-_\mu \phi^+) + \tfrac{1}{2}g^2 s_w A_\mu \phi^0 (W^+_\mu \phi^- + \\
W^-_\mu \phi^+) + \tfrac{1}{2}ig^2 s_w A_\mu H (W^+_\mu \phi^- - W^-_\mu \phi^+) - g^2 \tfrac{s_w}{c_w}(2c_w^2 - 1) Z^0_\mu A_\mu \phi^+ \phi^- - \\
g^2 s_w^2 A_\mu A_\mu \phi^+ \phi^- + \tfrac{1}{2}ig_s \lambda^a_{ij} (\bar{q}^\sigma_i \gamma^\mu q^\sigma_j) g^a_\mu - \bar{e}^\lambda (\gamma\partial + m^\lambda_e) e^\lambda - \bar{\nu}^\lambda (\gamma\partial + m^\lambda_\nu) \nu^\lambda - \bar{u}^\lambda_j (\gamma\partial + \\
m^\lambda_u) u^\lambda_j - \bar{d}^\lambda_j (\gamma\partial + m^\lambda_d) d^\lambda_j + igs_w A_\mu \left(-(\bar{e}^\lambda \gamma^\mu e^\lambda) + \tfrac{2}{3}(\bar{u}^\lambda_j \gamma^\mu u^\lambda_j) - \tfrac{1}{3}(\bar{d}^\lambda_j \gamma^\mu d^\lambda_j) \right) + \\
\tfrac{ig}{4c_w} Z^0_\mu \{ (\bar{\nu}^\lambda \gamma^\mu (1+\gamma^5) \nu^\lambda) + (\bar{e}^\lambda \gamma^\mu (4s_w^2 - 1 - \gamma^5) e^\lambda) + (\bar{d}^\lambda_j \gamma^\mu (\tfrac{4}{3}s_w^2 - 1 - \gamma^5) d^\lambda_j) + \\
\bar{u}^\lambda_j \gamma^\mu (1 - \tfrac{8}{3}s_w^2 + \gamma^5) u^\lambda_j) \} + \tfrac{ig}{2\sqrt{2}} W^+_\mu \left((\bar{\nu}^\lambda \gamma^\mu (1+\gamma^5) U^{lep}{}_{\lambda\kappa} e^\kappa) + (\bar{u}^\lambda_j \gamma^\mu (1+\gamma^5) C_{\lambda\kappa} d^\kappa_j) \right) + \\
\tfrac{ig}{2\sqrt{2}} W^-_\mu \left((\bar{e}^\kappa U^{lep\dagger}{}_{\kappa\lambda} \gamma^\mu (1+\gamma^5) \nu^\lambda) + (\bar{d}^\kappa_j C^\dagger_{\kappa\lambda} \gamma^\mu (1+\gamma^5) u^\lambda_j) \right) + \\
\tfrac{ig}{2M\sqrt{2}} \phi^+ \left(-m^\kappa_e (\bar{\nu}^\lambda U^{lep}{}_{\lambda\kappa} (1-\gamma^5) e^\kappa) + m^\lambda_\nu (\bar{\nu}^\lambda U^{lep}{}_{\lambda\kappa} (1+\gamma^5) e^\kappa \right) + \\
\tfrac{ig}{2M\sqrt{2}} \phi^- \left(m^\lambda_e (\bar{e}^\lambda U^{lep\dagger}{}_{\lambda\kappa} (1+\gamma^5) \nu^\kappa) - m^\kappa_\nu (\bar{e}^\lambda U^{lep\dagger}{}_{\lambda\kappa} (1-\gamma^5) \nu^\kappa \right) - \tfrac{g}{2}\tfrac{m^\lambda_\nu}{M} H (\bar{\nu}^\lambda \nu^\lambda) - \\
\tfrac{g}{2}\tfrac{m^\lambda_e}{M} H (\bar{e}^\lambda e^\lambda) + \tfrac{ig}{2}\tfrac{m^\lambda_\nu}{M} \phi^0 (\bar{\nu}^\lambda \gamma^5 \nu^\lambda) - \tfrac{ig}{2}\tfrac{m^\lambda_e}{M} \phi^0 (\bar{e}^\lambda \gamma^5 e^\lambda) - \tfrac{1}{4}\bar{\nu}_\lambda M^R_{\lambda\kappa} (1-\gamma_5) \hat{\nu}_\kappa - \\
\tfrac{1}{4}\bar{\nu}_\lambda M^R_{\lambda\kappa} (1-\gamma_5) \hat{\nu}_\kappa + \tfrac{ig}{2M\sqrt{2}} \phi^+ \left(-m^\kappa_d (\bar{u}^\lambda_j C_{\lambda\kappa} (1-\gamma^5) d^\kappa_j) + m^\lambda_u (\bar{u}^\lambda_j C_{\lambda\kappa} (1+\gamma^5) d^\kappa_j \right) + \\
\tfrac{ig}{2M\sqrt{2}} \phi^- \left(m^\lambda_d (\bar{d}^\lambda_j C^\dagger_{\lambda\kappa} (1+\gamma^5) u^\kappa_j) - m^\kappa_u (\bar{d}^\lambda_j C^\dagger_{\lambda\kappa} (1-\gamma^5) u^\kappa_j \right) - \tfrac{g}{2}\tfrac{m^\lambda_u}{M} H (\bar{u}^\lambda_j u^\lambda_j) - \\
\tfrac{g}{2}\tfrac{m^\lambda_d}{M} H (\bar{d}^\lambda_j d^\lambda_j) + \tfrac{ig}{2}\tfrac{m^\lambda_u}{M} \phi^0 (\bar{u}^\lambda_j \gamma^5 u^\lambda_j) - \tfrac{ig}{2}\tfrac{m^\lambda_d}{M} \phi^0 (\bar{d}^\lambda_j \gamma^5 d^\lambda_j) + \bar{G}^a \partial^2 G^a + g_s f^{abc} \partial_\mu \bar{G}^a G^b g^c_\mu + \\
{}^+(\partial^2 - M^2) X^+ + \bar{X}^- (\partial^2 - M^2) X^- + \bar{X}^0 (\partial^2 - \tfrac{M^2}{c_w^2}) X^0 + \bar{Y} \partial^2 Y + igc_w W^+_\mu (\partial_\mu \bar{X}^0 X^- - \\
\partial_\mu \bar{X}^+ X^0) + igs_w W^+_\mu (\partial_\mu \bar{Y} X^- - \partial_\mu \bar{X}^+ Y) + igc_w W^-_\mu (\partial_\mu \bar{X}^- X^0 - \\
\partial_\mu \bar{X}^0 X^+) + igs_w W^-_\mu (\partial_\mu \bar{X}^- Y - \partial_\mu \bar{Y} X^+) + igc_w Z^0_\mu (\partial_\mu \bar{X}^+ X^+ - \\
\partial_\mu \bar{X}^- X^-) + igs_w A_\mu (\partial_\mu \bar{X}^+ X^+ - \\
\bar{X}^- X^-) - \tfrac{1}{2}gM \left(\bar{X}^+ X^+ H + \bar{X}^- X^- H + \tfrac{1}{c_w^2} \bar{X}^0 X^0 H \right) + \tfrac{1-2c_w^2}{2c_w} igM \left(\bar{X}^+ X^0 \phi^+ - \bar{X}^- X^0 \phi^- \right) + \\
\tfrac{1}{2c_w} igM \left(\bar{X}^0 X^- \phi^+ - \bar{X}^0 X^+ \phi^- \right) + igM s_w \left(\bar{X}^0 X^- \phi^+ - \bar{X}^0 X^+ \phi^- \right) + \\
\tfrac{1}{2} igM \left(\bar{X}^+ X^+ \phi^0 - \bar{X}^- X^- \phi^0 \right) .
\end{gathered}
$$

Given our data sample, we want to:

– measure theory parameters and answer questions about the nature of data

- Is the Higgs boson real? (strong evidence? Quantify!)

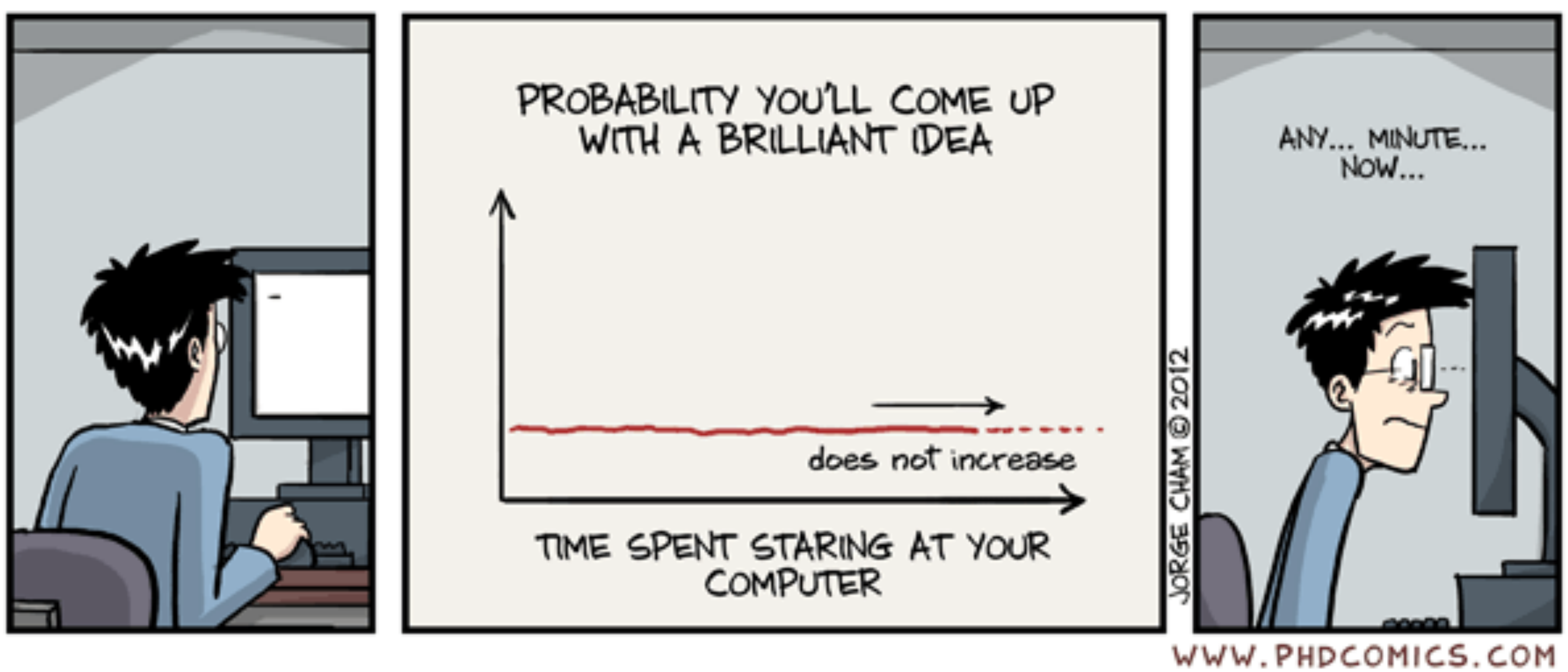
- Is Dark Matter real? (No evidence, so far… Quantify!)

- What is the range of theory parameters compatible with the observed data? What parameter range can we exclude?

What is probability?

Probability doesn't have a unique, Universal definition!

- The applicability of each definition depends on the kind of *claim* we are considering to applying the concept of probability
- One *subjective* approach expresses the degree of belief of the claim, which may vary from subject to subject
- For repeatable experiments, probability *may be* a measure of how frequently the claim is true



The importance of being repeatable

Repeatable experiments

- What's the probability to extract one ace in a deck of cards?
- What is the probability to win a lottery?
- What is the probability that a pion is incorrectly identified as a muon in a particle detector?

more complicated:

What is the probability that *a fluctuation in the background can produce a peak* in the $\gamma\gamma$ spectrum with a magnitude at least equal to what has been observed by a given experiment?

Note: different question w.r.t.: what is the probability that the peak is due to a background fluctuation? (non repeatable!)

Unreplicable claims

Could be about *future events*:

– what's the probability that tomorrow it will rain in Trieste?
– what's the probability of your favourite team will win next championship?

But also past events:

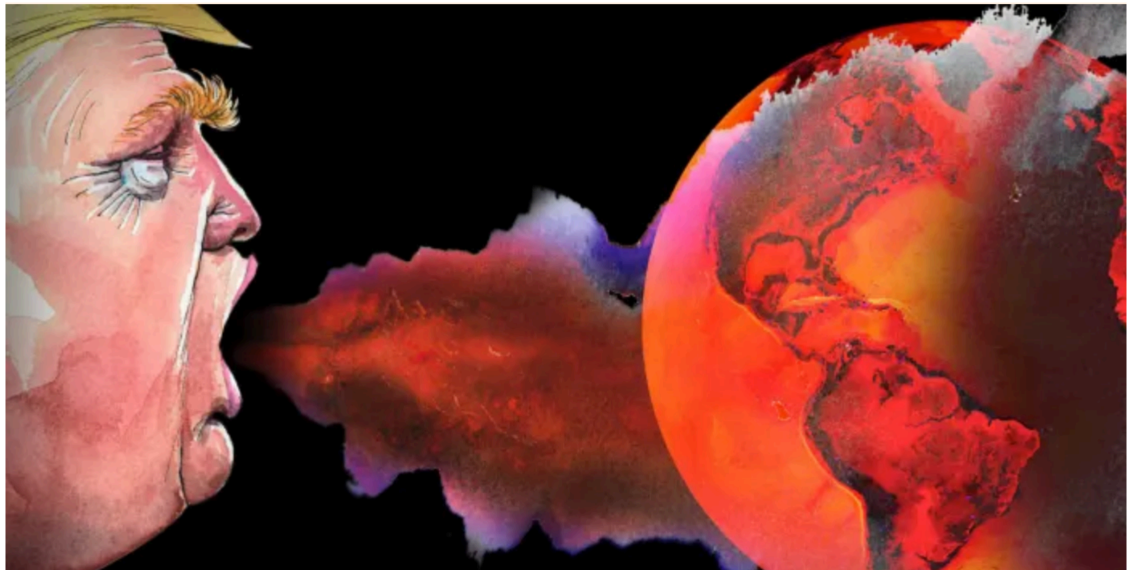
– what's the probability that dinosaurs went extinct because of an asteroid?

More in general,
it's about *unknown* events:

– what is the probability that matter is made of particles heavier than 1 eV?

– what is the probability that

climate changes are mainly due to human intervention?



Maths basics of probability

- Probability determined by symmetry properties of a random device
- “Equally undecided” about event outcome, according to Laplace definition

P = 1/6 (each dice)

P = 1/2

P = 1/4 P = 1/10

Composite cases

- Reduce the (composite) event of interest into elementary equiprobable events (sample space)

- Composite cases are managed via combinatorial analysis

- Statements about an event can be defined via set algebra – and/or/not ⇒ intersection/union/complement

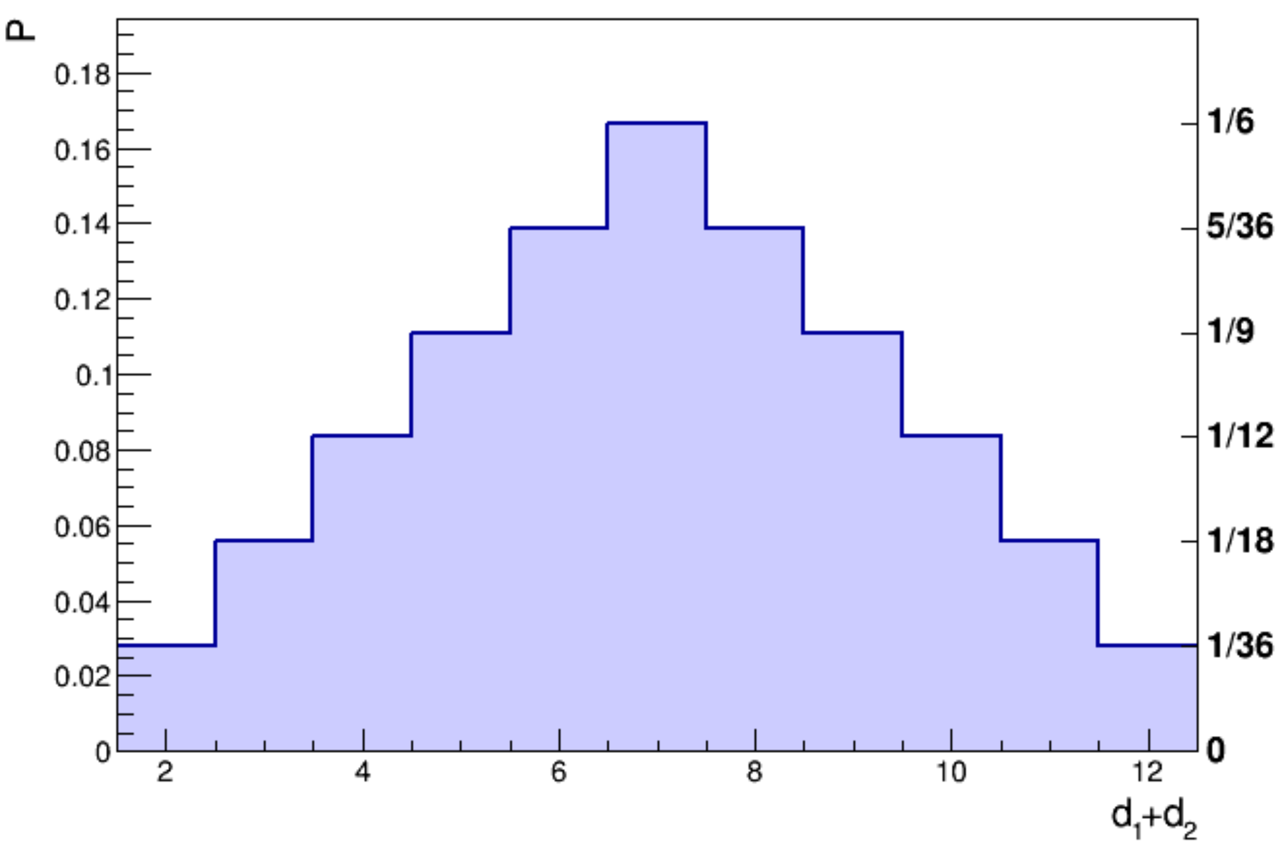
E.g:

2 = {(1,1)}
3 = {(1,2), (2,1)}
4 = {(1,3), (2,2), (3,1)}
5 = {(1,4), (2,3), (3,2), (4,1)} etc. ...

– E.g.:

"sum of two dices is even and greater than four"

$\{(d_1,d_2):\mathrm{mod}(d_1+d_2,2)=0\}\cap\{(d_1,d_2):d_1+d_2>4\}$



Events

- Note that in physics and statistics usually the word *event* have different meanings
- Statistics: a subset in the sample space
    - E.g.: "*the sum of two dices is ≥ 5*"
- Physics: the result of a collision, as recorded by our experiment
  – E.g.: a Higgs to two-photon candidate event
- In several concrete cases, an event in statistics may correspond to many possible collision events
    - – E.g.: "$p_{T}(\gamma) > 40$ GeV", "The measured m_{H} is > 125 GeV"

Frequentist Probability

Probability P = frequency of occurrence of an event in the limit of very large number ($N \to \infty$) of repeated trials

Probability: P = $\lim_{N \to \infty}$

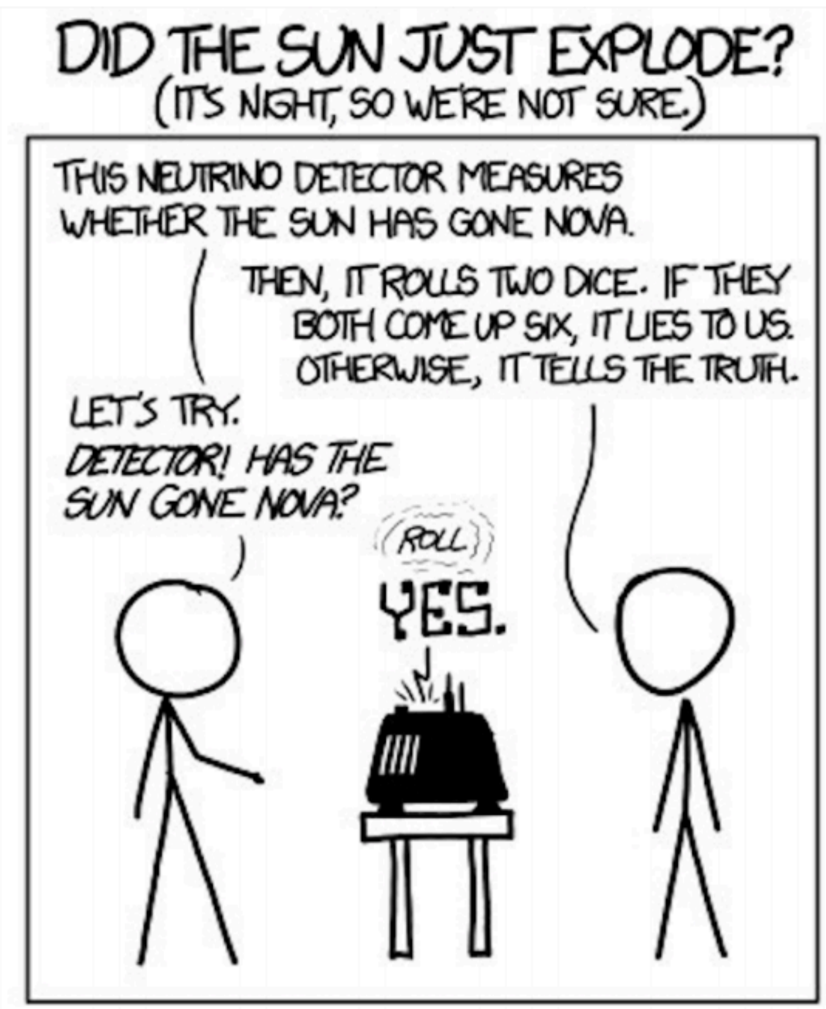
Number of favourable cases N = Number of trials

- Exactly realizable only with an infinite number of trials
  - – Conceptually is unpleasant
  - – Pragmatically acceptable by physicists
  - - Easy to compute integrals
- Only applicable to repeatable experiments

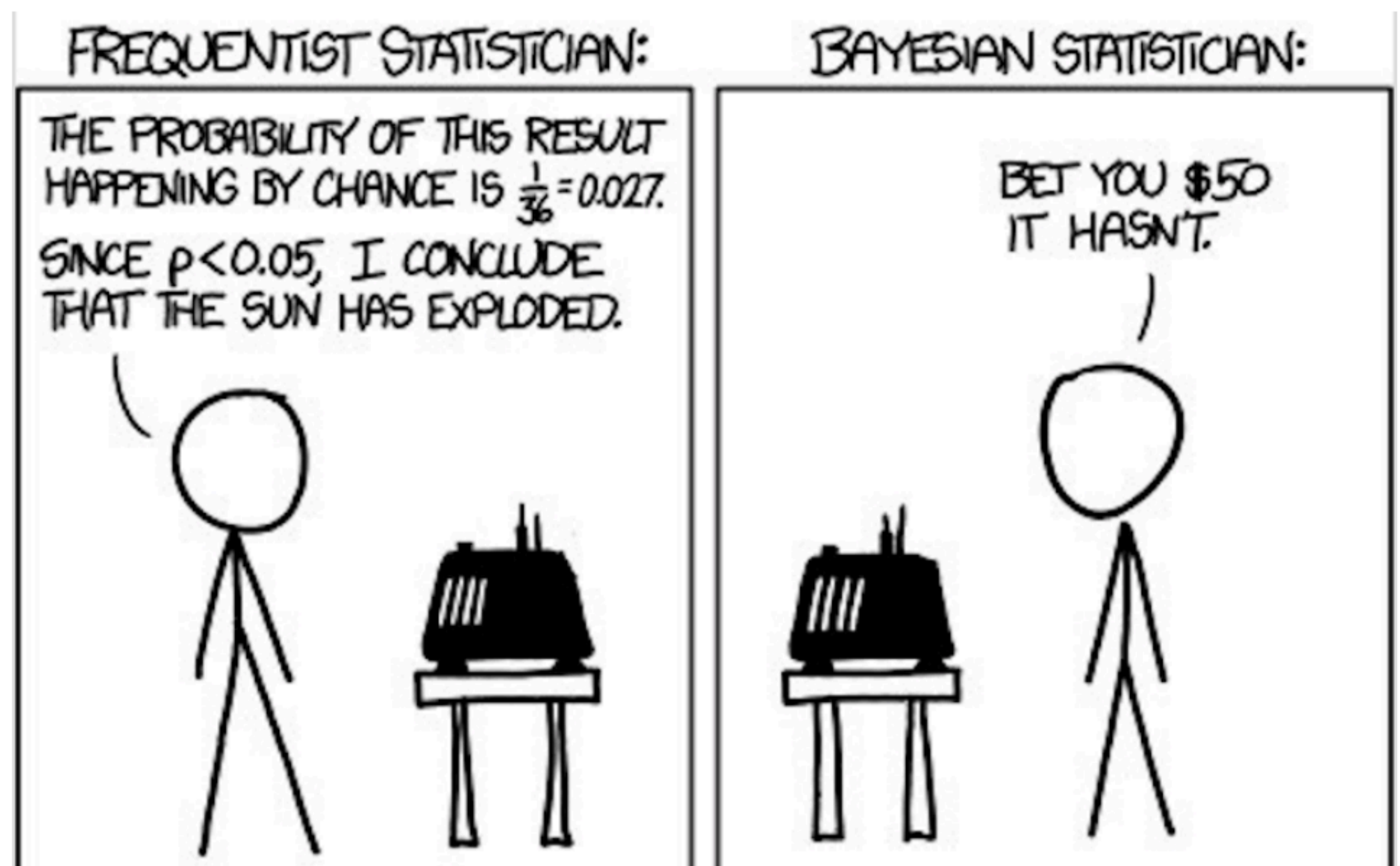
Bayesian Probability

- Expresses **one's degree of belief** that a claim is true
  - – How strong would you bet?
  - – Applicable to all unknown events/claims, not only repeatable experiments
  - – Each individual may have a different opinion/prejudice
- Quantitative rules exist about how subjective probability should be modified after learning about some observation/evidence
  - – Consistent with Bayes Theorem
  - – Prior probability and Posterior probability (following observation)
  - – The more information we receive, the more Bayesian probability is *insensitive* on prior subjective prejudice (unless pathological cases...)

Bayesian vs. Frequentist

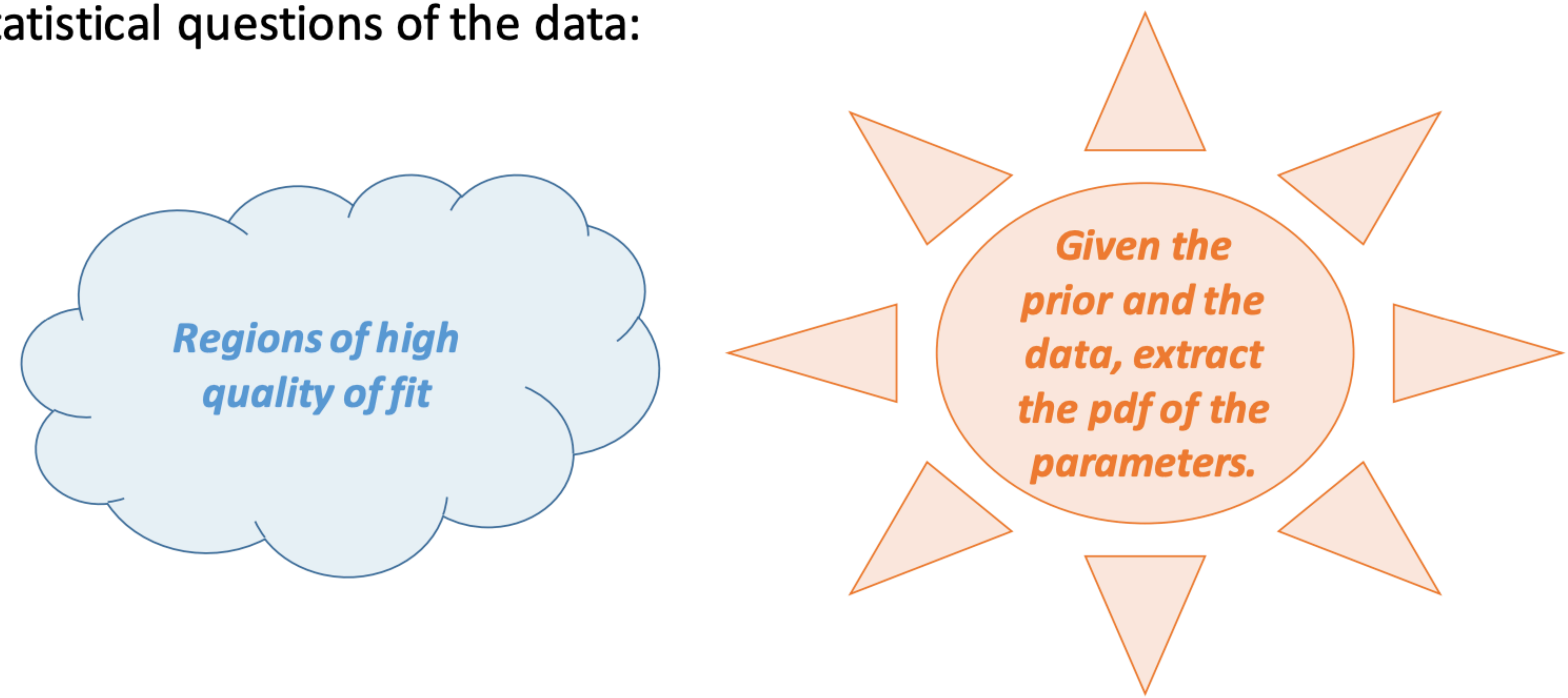


Taken from xkcd



Bayesian vs. Frequentist

The ***Frequentist likelihood*** and the ***Bayesian posterior*** ask two different statistical questions of the data:

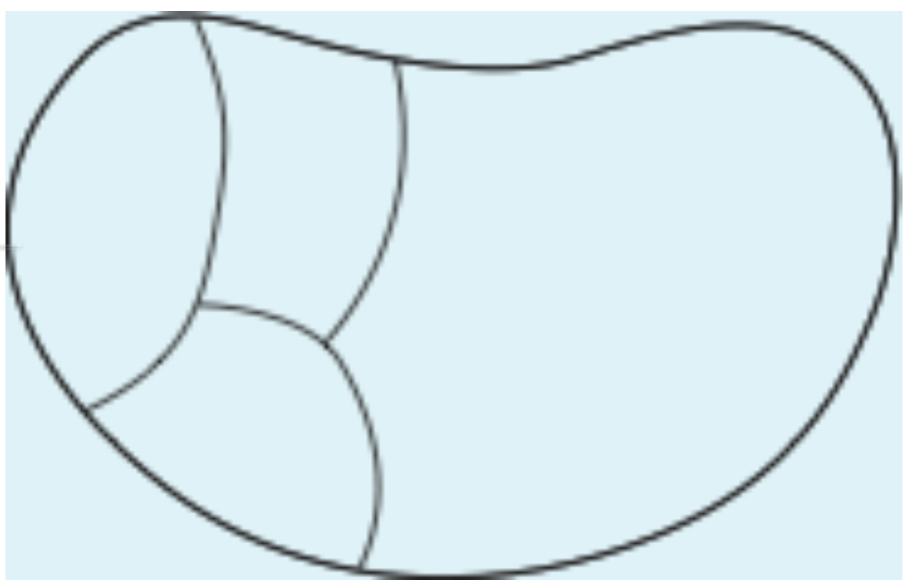


Axioms

<u>Axiomatic probability definitions</u>

– Terminology: Ω = sample space, F = event space, P = probability measure

– Let (Ω, F⊆2^Ω, P) be a measure space that satisfies:



★ $\forall(E_1, \cdots, E_n) \in F^n : E_i \cap E_j = 0$

★ $P(\Omega) = 1$

★ $P(\bigcup_{i=1,\cdots,n} E_i) = \sum_{i=1,\cdots,n} P(E_i)$

★ $P(E) \geq 0 \ \ \forall E \in F$

The same formalism applies to
either frequentist and Bayesian probability

Probability Distributions

Given a discrete random variable, we can assign a probability to each individual value:

In case of a continuous variable, the probability assigned to an individual value may be 0

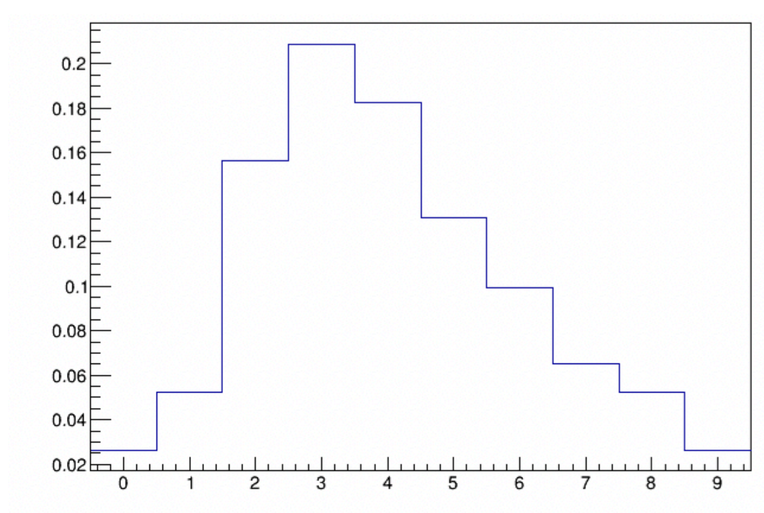
- A **probability density** better quantifies the probability content (unlike P({x}) = 0 !):

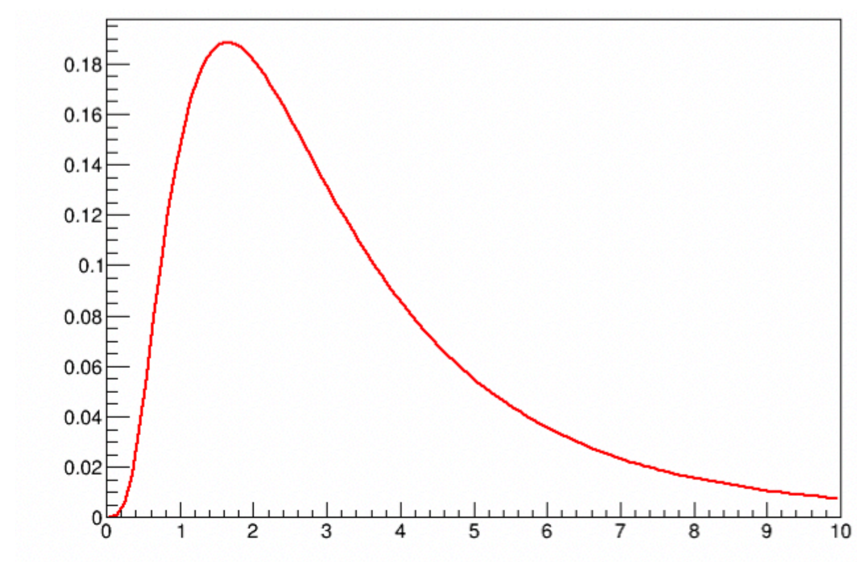
Discrete and continuous distributions can be combined using Dirac's delta functions.

$$P(x) = P(\{x\})$$

$$\frac{\mathrm{d}P(x)}{\mathrm{d}x} = f(x)$$

$$\frac{\mathrm{d}P}{\mathrm{d}x} = \frac{1}{2}\delta(x) + \frac{1}{2}f(x)$$





50% prob. to have zero (P({0}) = 0.5), 50% distributed according to f(x)

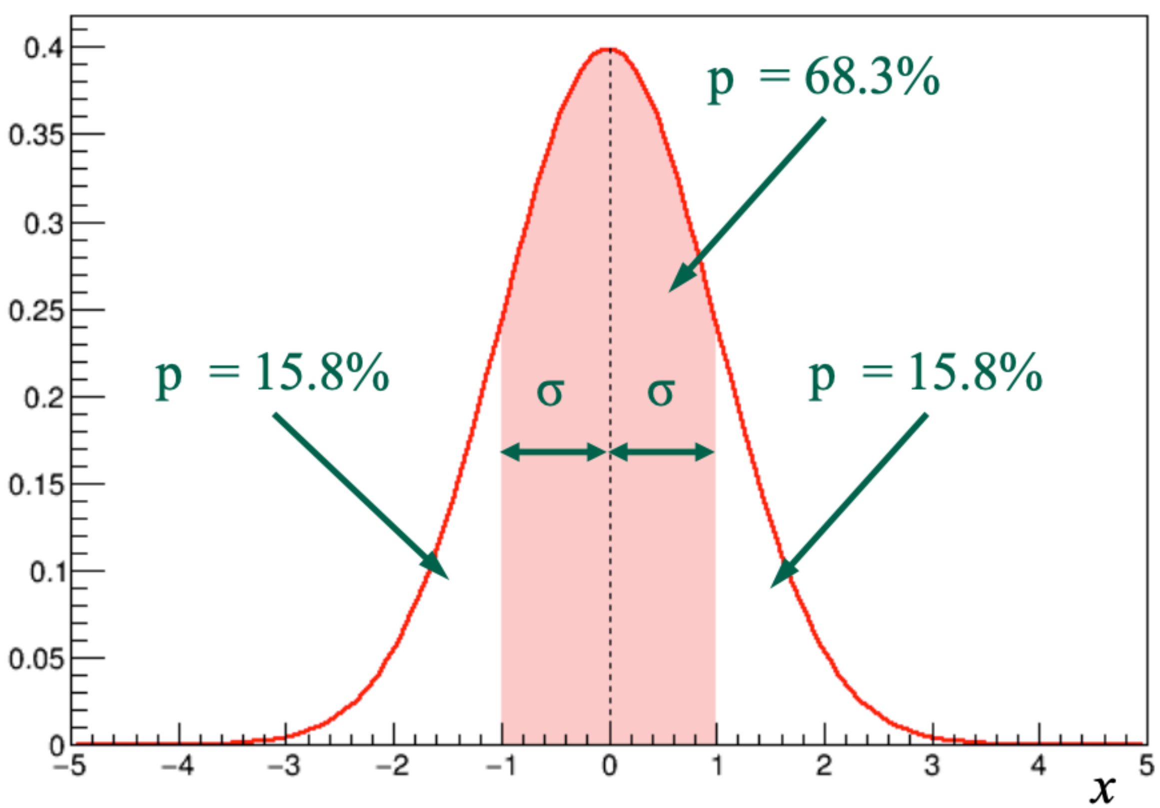
Gaussian Case

- Many random variables in real experiments follow a Gaussian distribution

Central Limit Theorem:

approximate sum of multiple random contributions, regardless of the individual distributions

- Frequently used to model detector resolution



$$g(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

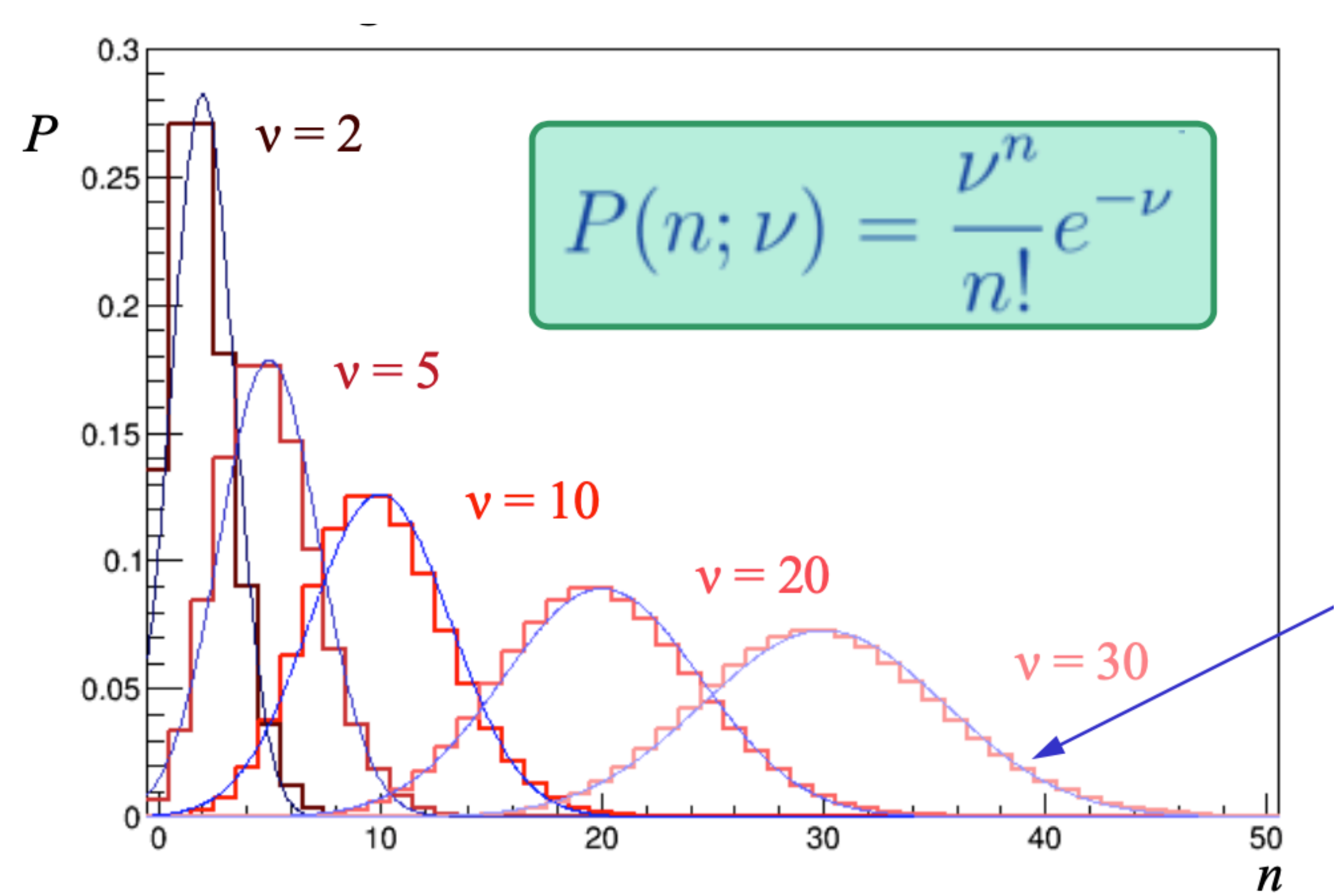
$n\sigma$	Prob.
1	0.683
2	0.954
3	0.997
4	$1 - 6.3\times10^{-5}$
5	$1 - 5.7\times10^{-7}$

Poisson Case

Distribution of the number of occurrences of random event uniformly distributed in a measurement range whose rate is known

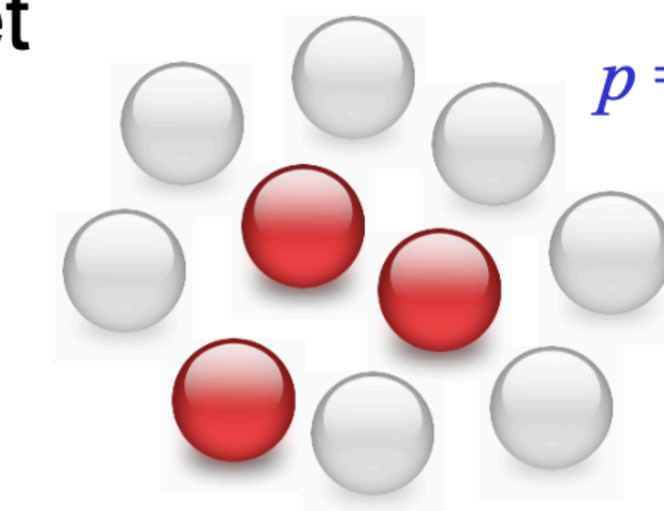
– E.g.: number of rain drops in a given area and in a given time interval, number of cosmic rays crossing a detector in a given time interval

Can be approximated with a Gaussian distribution for large values of ν.



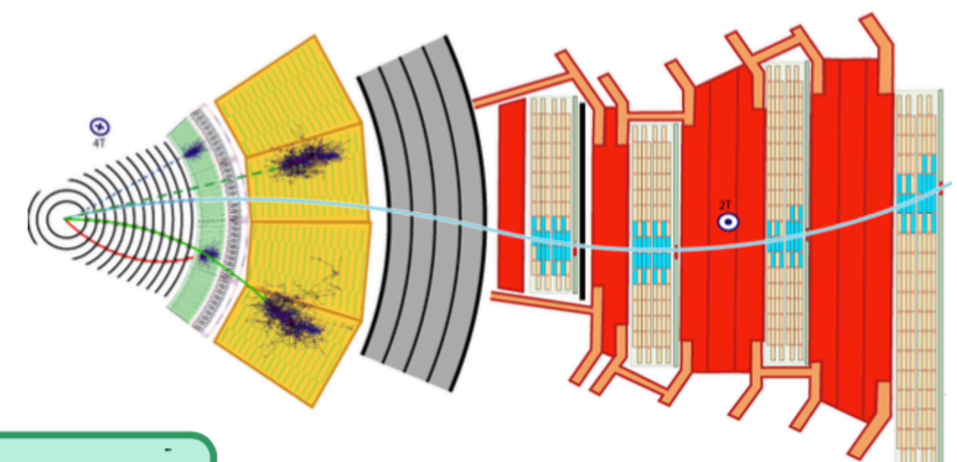
Binomial Case

- Probability to extract n red balls over N trials, given the fraction p of red balls in a basket
- Red:  $p = 3/10$
- White:  $1 - p = 7/10$



$p = 3/10$

- Typical application in physics: detector efficiency ($\varepsilon = p$)



$$P(n; N, p) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

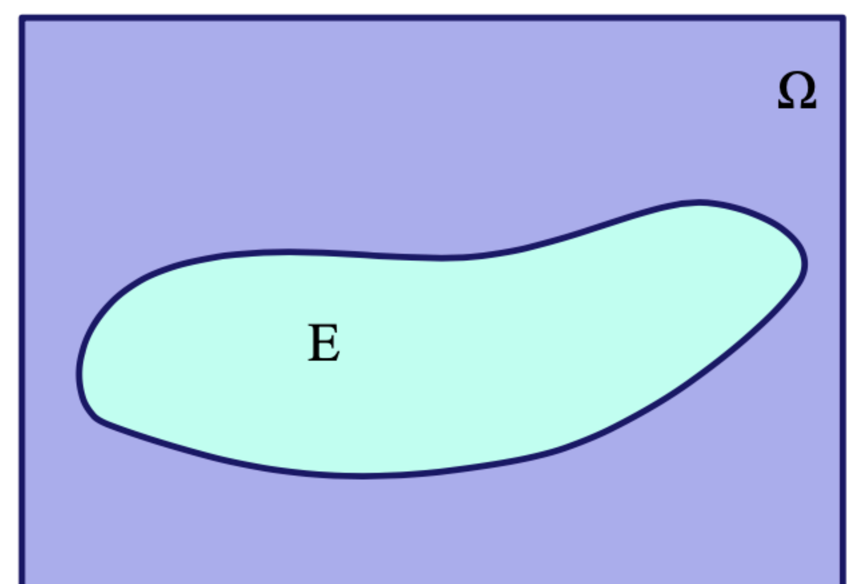
PDFs in higher dimensions

- In more dimensions (n random variables), PDF can be defined as:

$$\frac{d^n P}{dx_1 \ldots dx_n} = f(x_1 \ldots x_n)$$

- The probability associated to an event E is obtained by integrating the PDF over the corresponding set in the sample space

$$P(E) = \int_E f(x_1 \ldots x_n) dx^n$$



Mean & Variance

- Given a random variable *x* with distribution *f(x)* we can define:

  - Mean or

  expected value:

$$\mathrm{E}[g(x)] = \langle g(x) \rangle = \int g(x) f(x) \mathrm{d}x$$

$$\mathrm{E}[x] = \langle x \rangle = \int x f(x) \mathrm{d}x$$

  - Variance:

$$\mathrm{Var}[x] = \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2$$

  - Standard deviation:

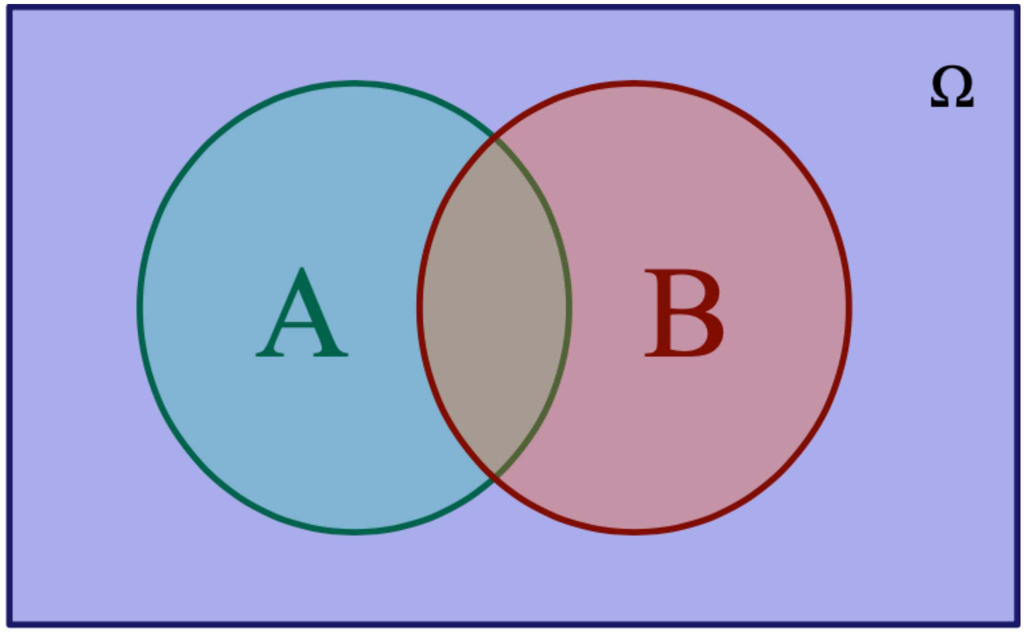
$$\sigma_x = \sqrt{\mathrm{Var}[x]} = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$$

  - Covariance and correlation coefficient of two variables x and y:

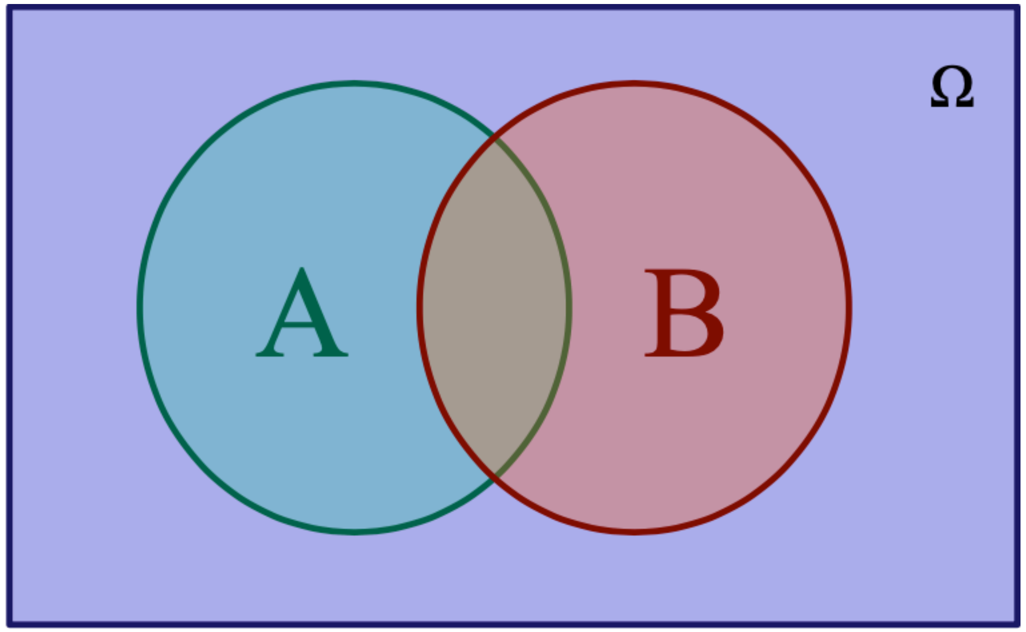
$$\mathrm{cov}(x, y) = \langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle \qquad \rho_{xy} = \frac{\mathrm{cov}(x, y)}{\sigma_x \sigma_y}$$

Conditional Probability

- Probability of A, given B: P(A | B), i.e.: probability that an event known to belong to set B also belongs to set A:
  - $P(A|B) = P(A \cap B)/P(B)$ – Notice that: $P(A|\Omega) = P(A \cap \Omega)/P(\Omega)$



- Event A is said to be independent of B if the probability of A given B is equal to the probability of A:
  - $P(A|B)=P(A)$
- If A is independent of B then $P(A \cap B) = P(A)\,P(B)$



- If A is independent on B, B is independent on A

Independent Variables

$$\frac{d^2 P}{dx\,dy} = f(x,y)$$

- 1D projections:
  (marginal distributions)

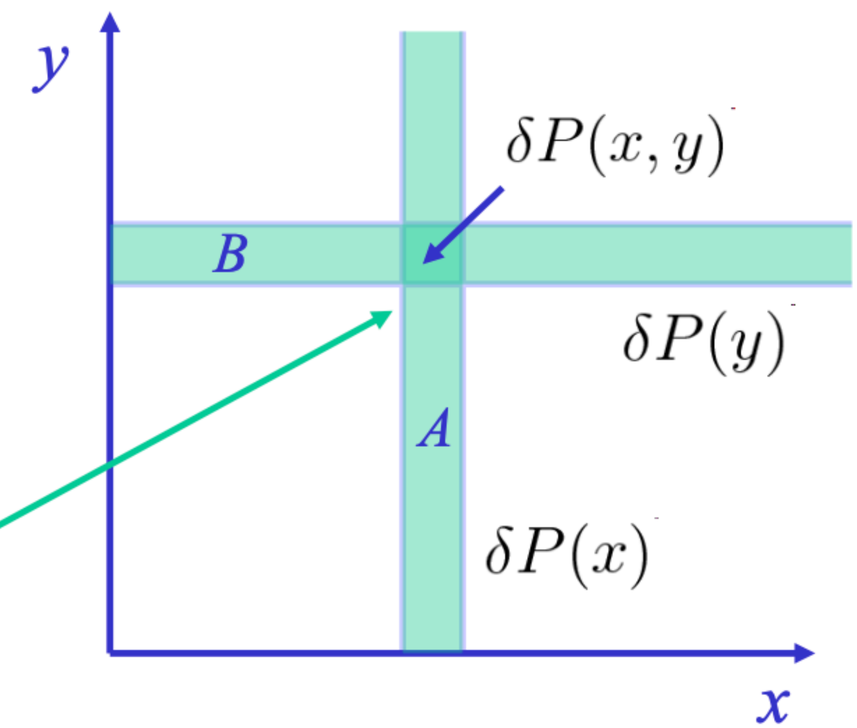
$$\begin{cases} f_x(x) = \int f(x,y)\,dy \\ f_y(y) = \int f(x,y)\,dx \end{cases}$$

- x and y are independent if:

$$f(x,y) = f_x(x) f_y(y)$$

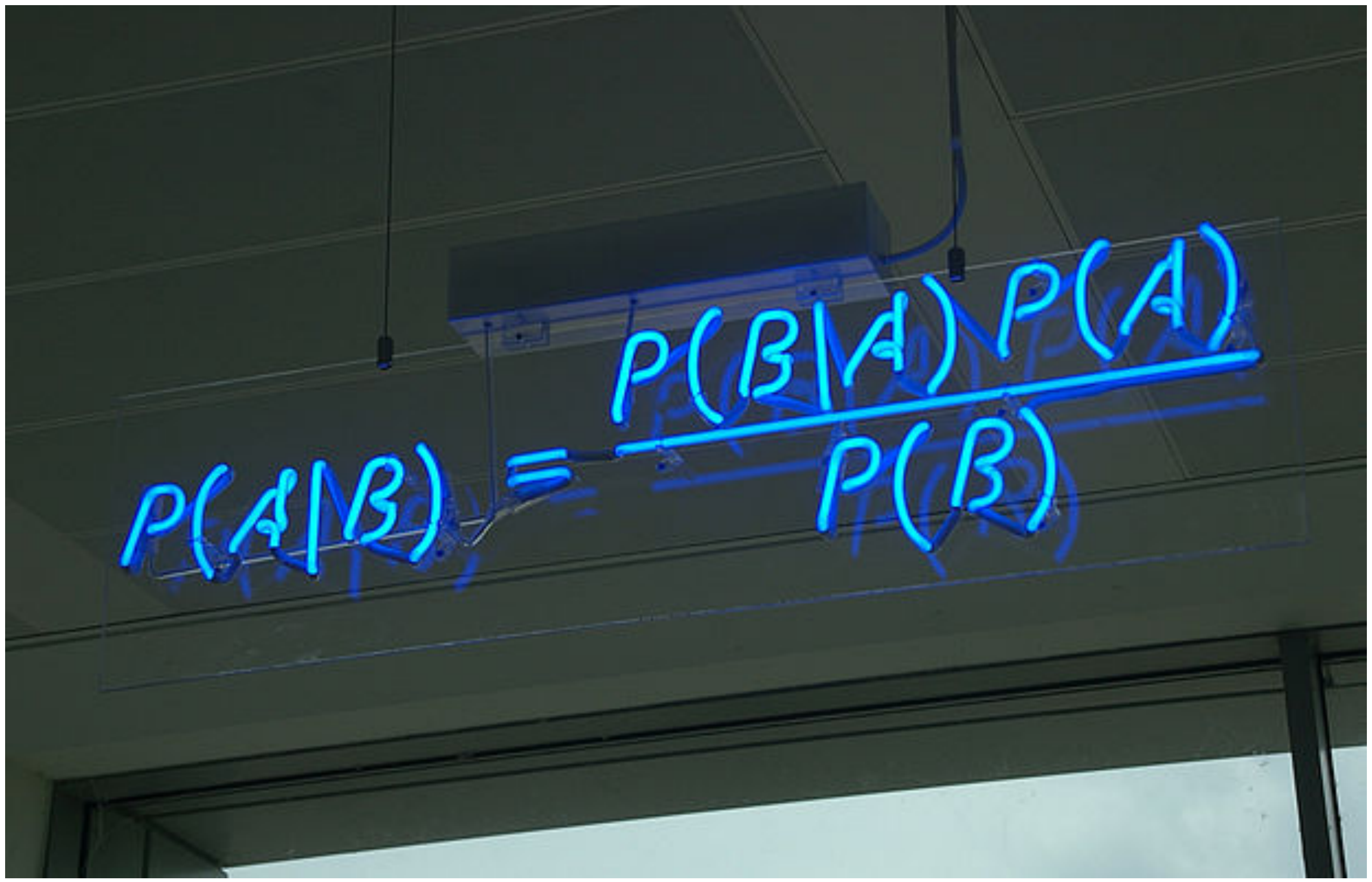
- We saw that A and B are independent events if:

$$P(A \cap B) = P(A)P(B)$$



- Where $A = \{x' : x < x' < x + \delta x\}$, $B = \{y' : y < y' < y + \delta y\}$

The Bayes Theorem



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

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

$$P(A|B)P(B) = P(B|A)P(A)$$

$P(A)$ = prior probability $\quad$ $P(A|B)$ = posterior probability

The Bayes Theorem: the role of the posterior

- Bayes theorem allows to determine probability about hypotheses or claims *H* that not related random variables, given an observation or evidence *E*:

$$\boxed{P(H|E) = \frac{P(E|H)P(H)}{P(E)}}$$

- $P(H)$ = prior probability
- $P(H \mid E)$ = posterior probability, given E

The Bayes rule allows to define a rational way to modify one's prior belief once some observation is known

Frequentist approach in practice

Let's take an example: muon fake rate estimation

- A detector identifies muons with high efficiently, $\varepsilon = 95\%$
- A small fraction $\delta = 5\%$ of pions are incorrectly identified as muons ("fakes")

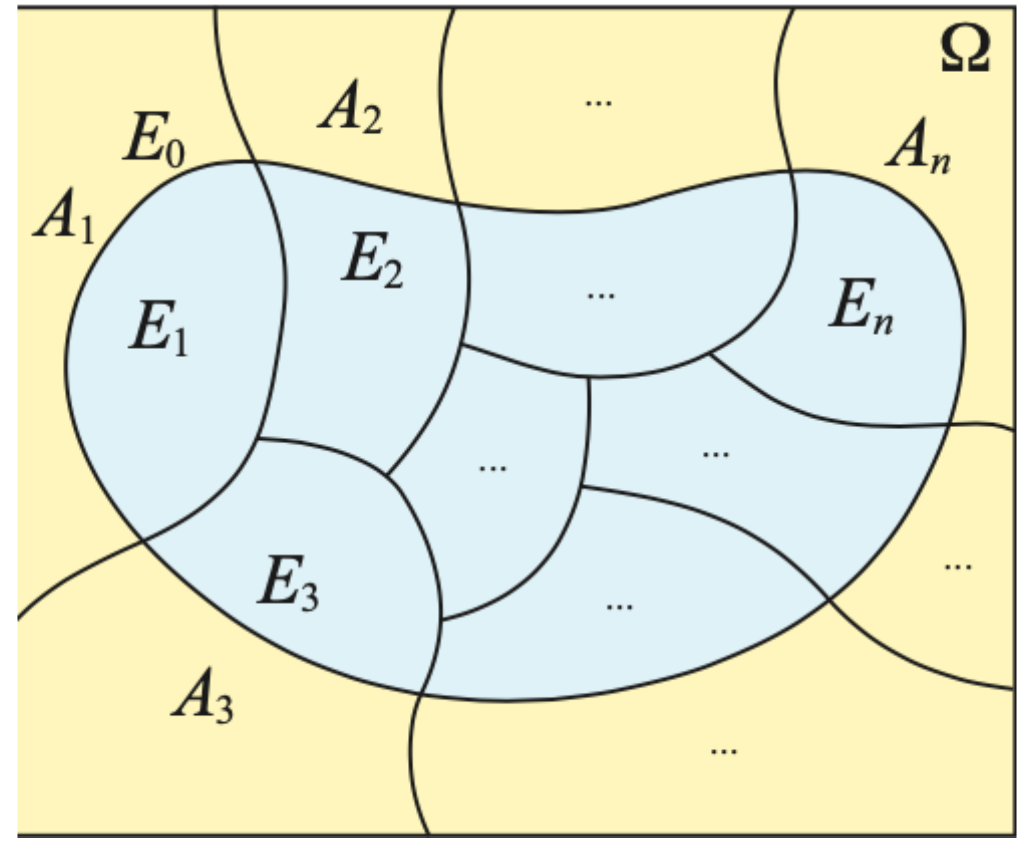
- If a particle is identified as a muon, what is the probability it is really a muon?
  - The answer also depends on the composition of the sample!
  - i.e.:the fraction of muons and pions in the overall sample

This example is usually presented as an epidemiology case.
Naïve answers about fake positive probability are often wrong!

Bayesian resolution

$$P(E_0) = \sum_{i=1}^{n} P(E_0|A_i)P(A_i)$$

$E_0 = \text{'+'}$, $A_i = \mu, \pi$



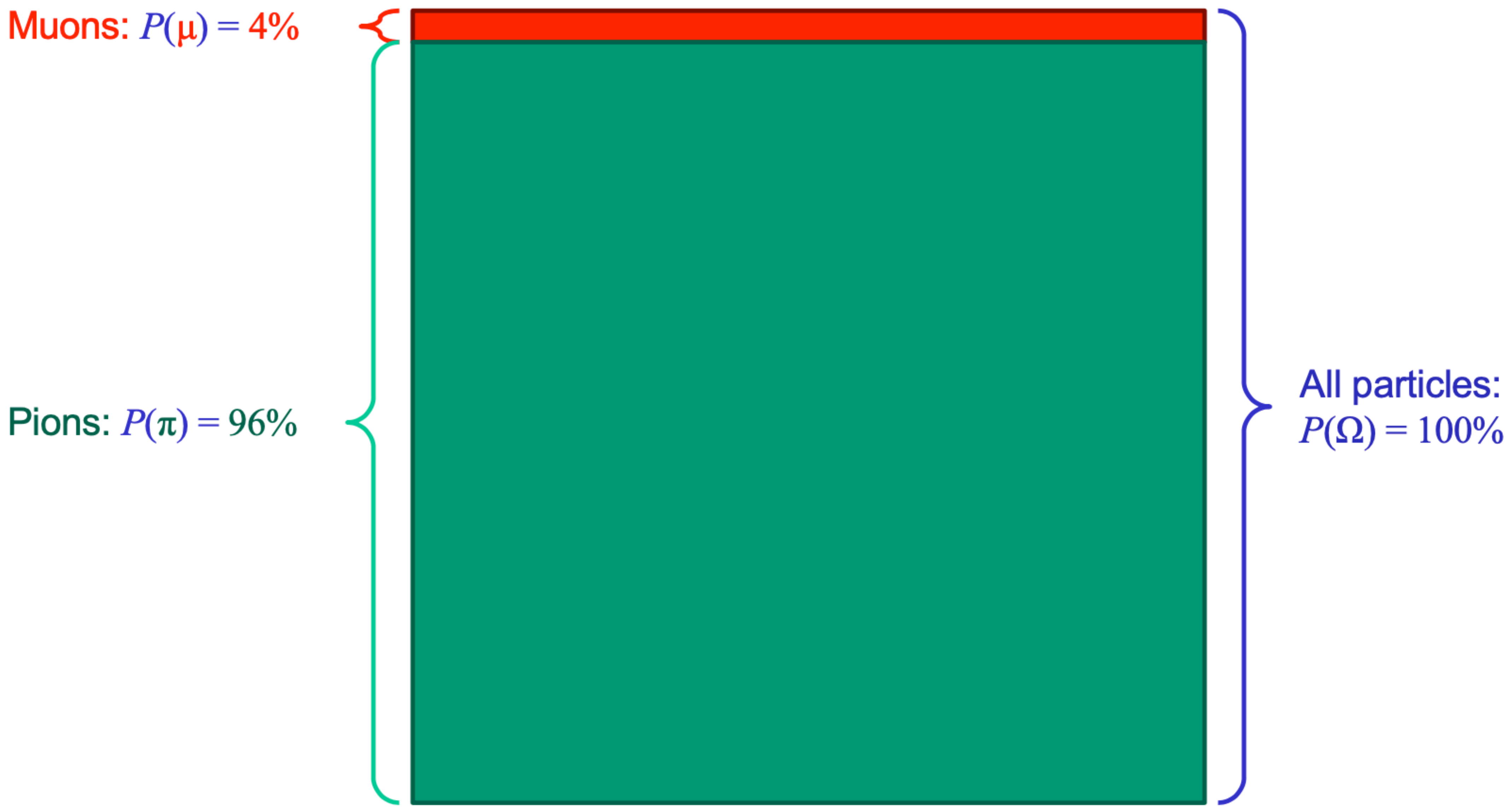
- Using Bayes theorem:
  – $P(\mu|+)=P(+|\mu)P(\mu)/P(+)$
  Where our inputs are:
  – $P(+|\mu)=\varepsilon=0.95, P(+|\pi)=\delta=0.05$
- We can decompose $P(+)$ as:
  – $P(+)=P(+|\mu)P(\mu)+P(+|\pi)P(\pi)$
- Putting all together:
  – $P(\mu|+)=\varepsilon P(\mu)/(\varepsilon P(\mu)+\delta P(\pi))$

• Assume we have a sample made of $P(\mu)$=4% muons and $P(\pi)$=96% pions, we have:

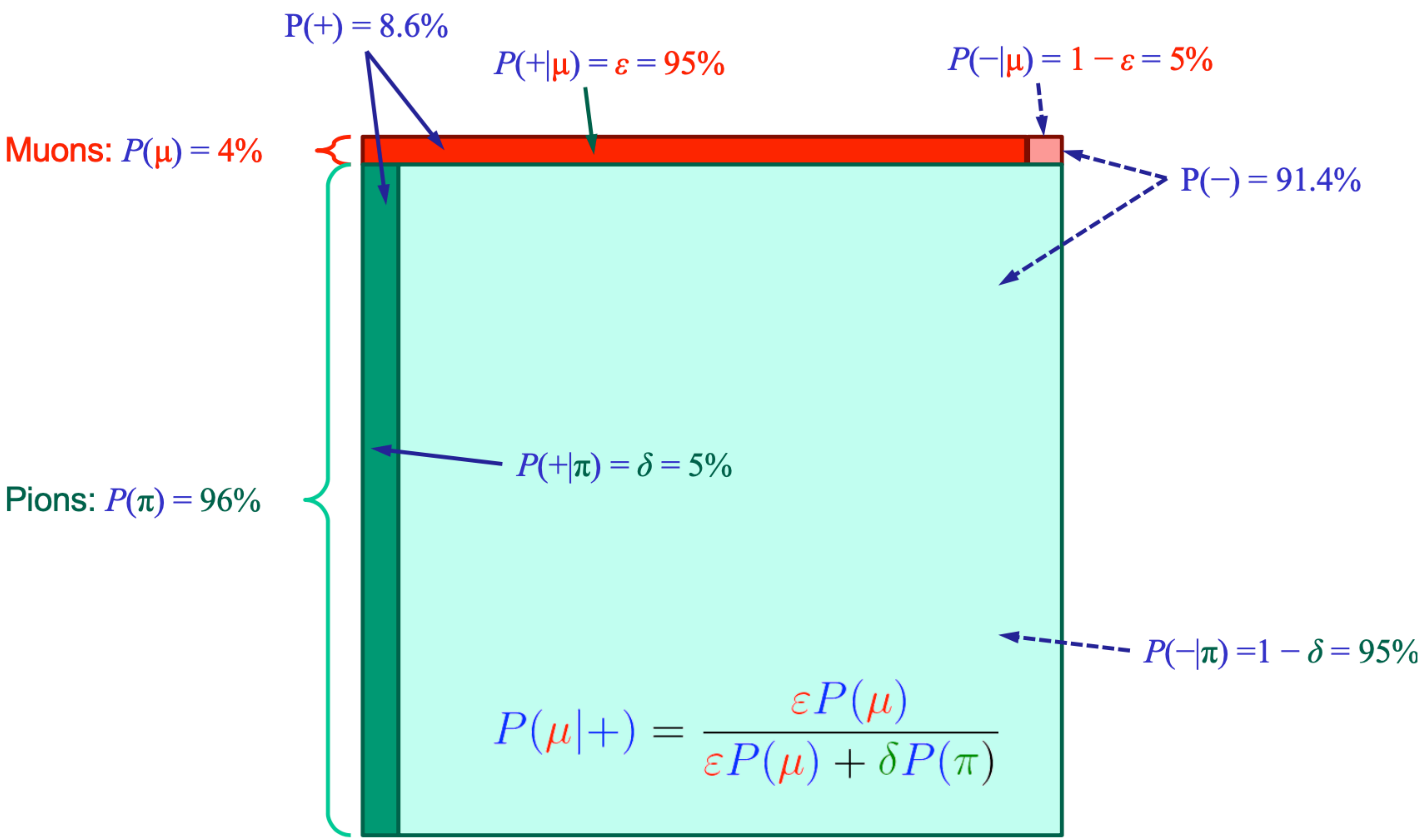
– $P(\mu|+) = 0.95 \times 0.04 / (0.95 \times 0.04 + 0.05 \times 0.96) \cong 0.44$

• Even if the selection efficiency is very high, the low sample purity makes $P(\mu|+)$ lower than 50%.

Bayesian resolution



Bayesian resolution



$$P(\mu|+) = \frac{\varepsilon P(\mu)}{\varepsilon P(\mu) + \delta P(\pi)}$$

The Likelihood Function

- In many cases, the outcome of our experiment can be modelled as a set of random variables $x_1, \ldots, x_n$ whose distribution takes into account:
  - intrinsic sample randomness (quantum physics is intrinsically random),
  - detector effects (resolution, efficiency, ...).
  - Theory and detector effects can be described according to some parameters $\theta_1, \ldots, \theta_m$, whose values are, in most of the cases, unknown
- The overall PDF, evaluated at our observation $x_1, \ldots, x_n$, is called likelihood function:

$$L = f(x_1 \ldots x_n; \theta_1 \ldots \theta_m)$$

- In case our sample consists of N independent measurements (collision events) the likelihood function can be written as:

$$L = \Pi_{i=1}^{N} f(x_1 \ldots x_n; \theta_1 \ldots \theta_m)$$

Bayes and the Likelihood Function

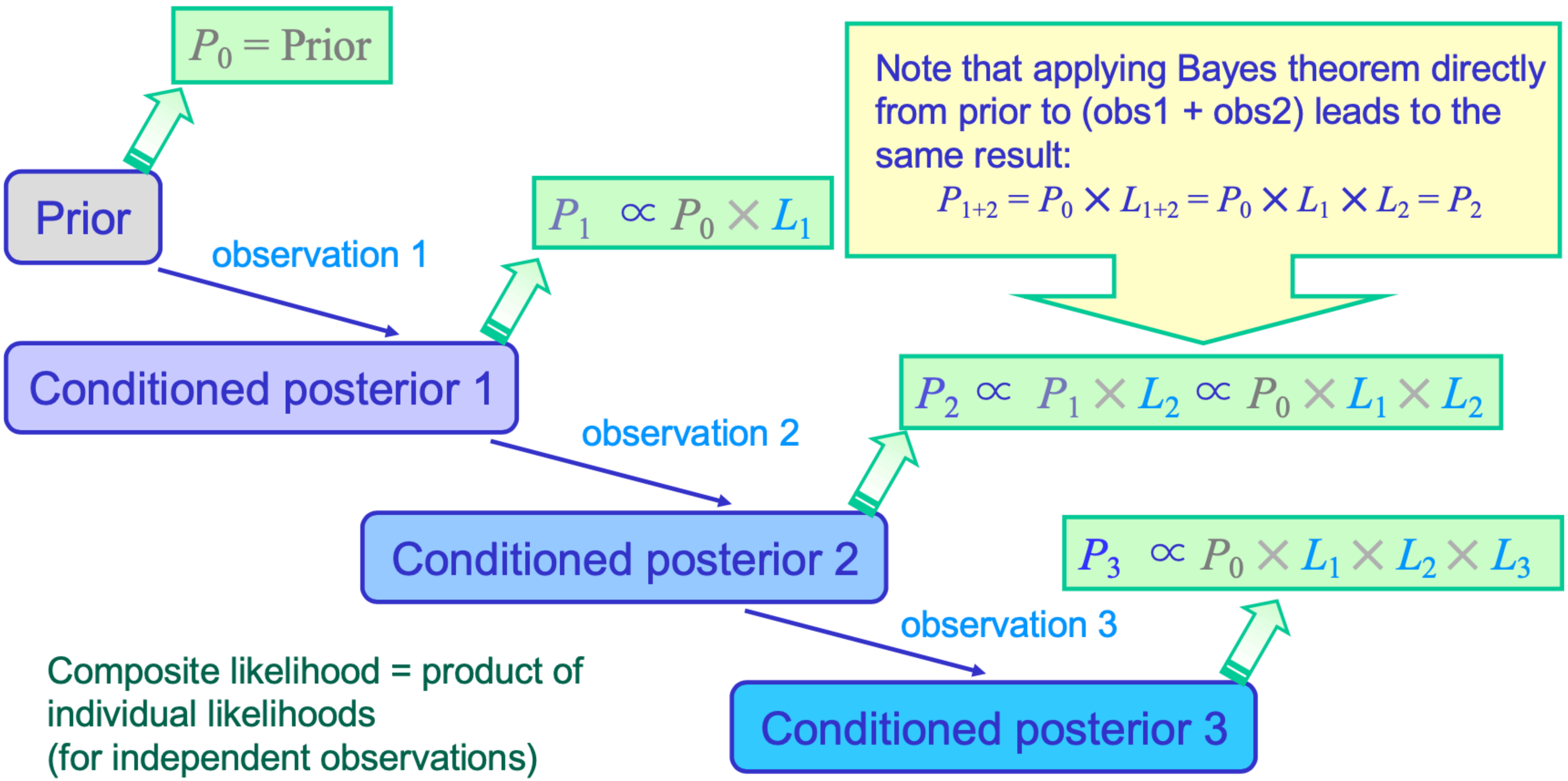
Given a set of measurements $x_1, \ldots, x_n$, Bayesian posterior PDF of the unknown parameters $\theta_1, \ldots, \theta_m$ can be determined as:

$$P(\theta_1, \cdots, \theta_m | x_1, \cdots, x_n) = \frac{L(x_1, \cdots, x_n; \theta_1, \cdots, \theta_m)\pi(\theta_1, \cdots, \theta_m)}{\int L(x_1, \cdots, x_n; \theta_1, \cdots, \theta_m)\pi(\theta_1, \cdots, \theta_m)\mathrm{d}^m\theta}$$

- Where $\pi(\theta_1, \ldots, \theta_m)$ is the subjective prior probability
- The denominator $\int L(x, \theta)\, \pi(\theta)\, d^m\theta$ is a normalization factor
- The observation of $x_1, \ldots, x_n$ modifies the prior knowledge of the unknown parameters $\theta_1, \ldots, \theta_m$
- If $\pi(\theta_1, \ldots, \theta_m)$ is sufficiently smooth and L is sharply peaked around the true values $\theta_1, \ldots, \theta_m$, the resulting posterior will not be strongly dependent on the prior's choice

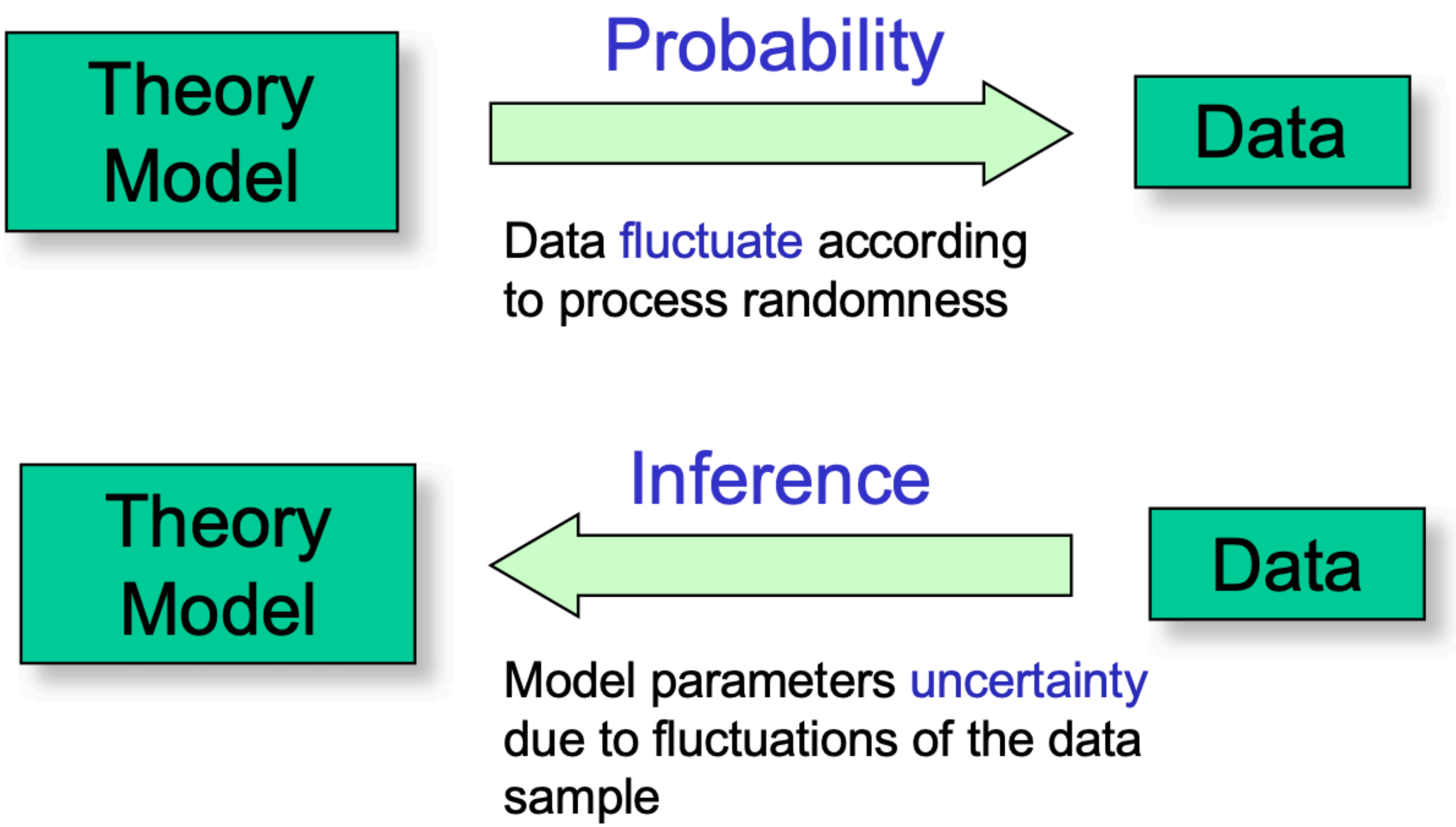
Iterating Bayes

Bayes theorem can be applied sequentially for repeated independent observations (posterior PDF = learning from experiments)



Inference

Determining information about unknown parameters using probability theory

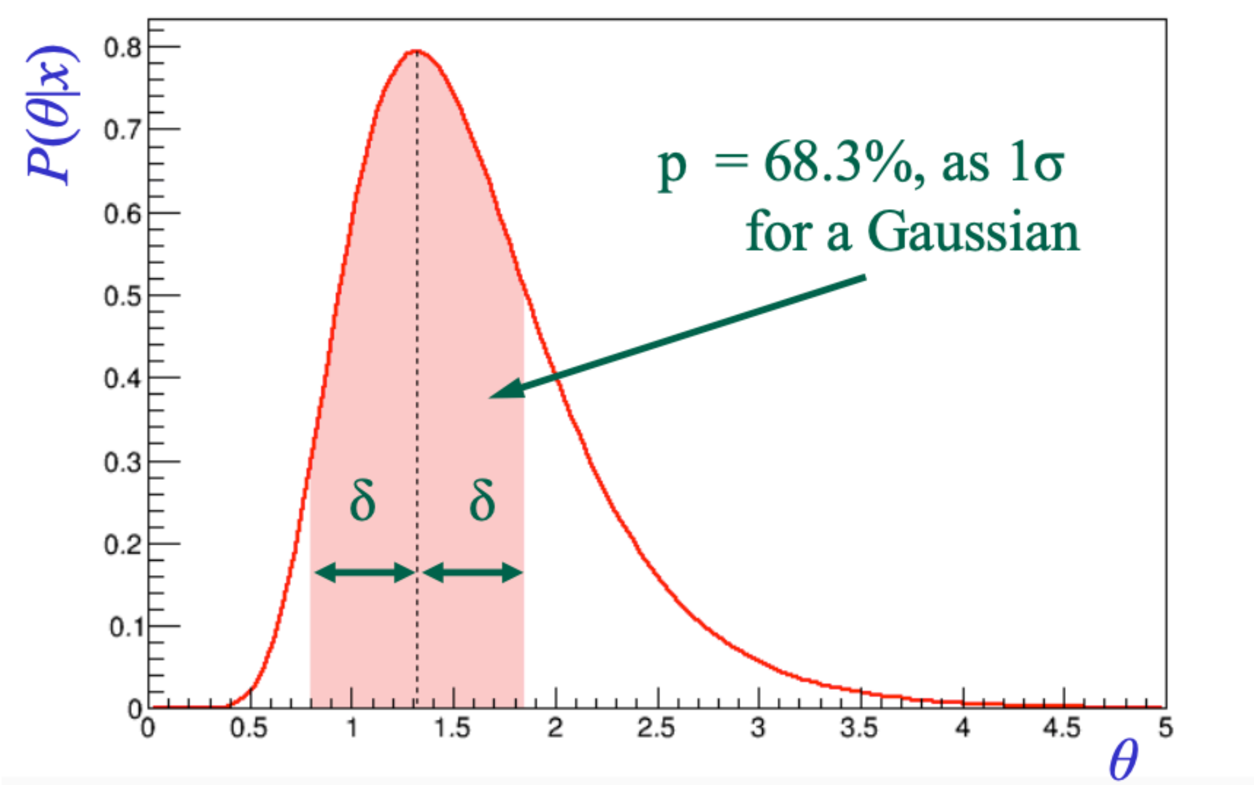


Bayesian Inference

The posterior PDF provides all the information about the unknown parameters (let's assume here it's just a single parameter θ for simplicity)

$$P(\theta|x) = \frac{L(x;\theta)\pi(\theta)}{\int L(x;\theta)\pi(\theta)\mathrm{d}\theta}$$

- Given P(θ|x), we can determine:
  – The most probable value (best estimate)
  – Intervals corresponding to a specified probability



- Notice that if π(θ) is a constant, the most probable value of θ correspond to the maximum of the likelihood function

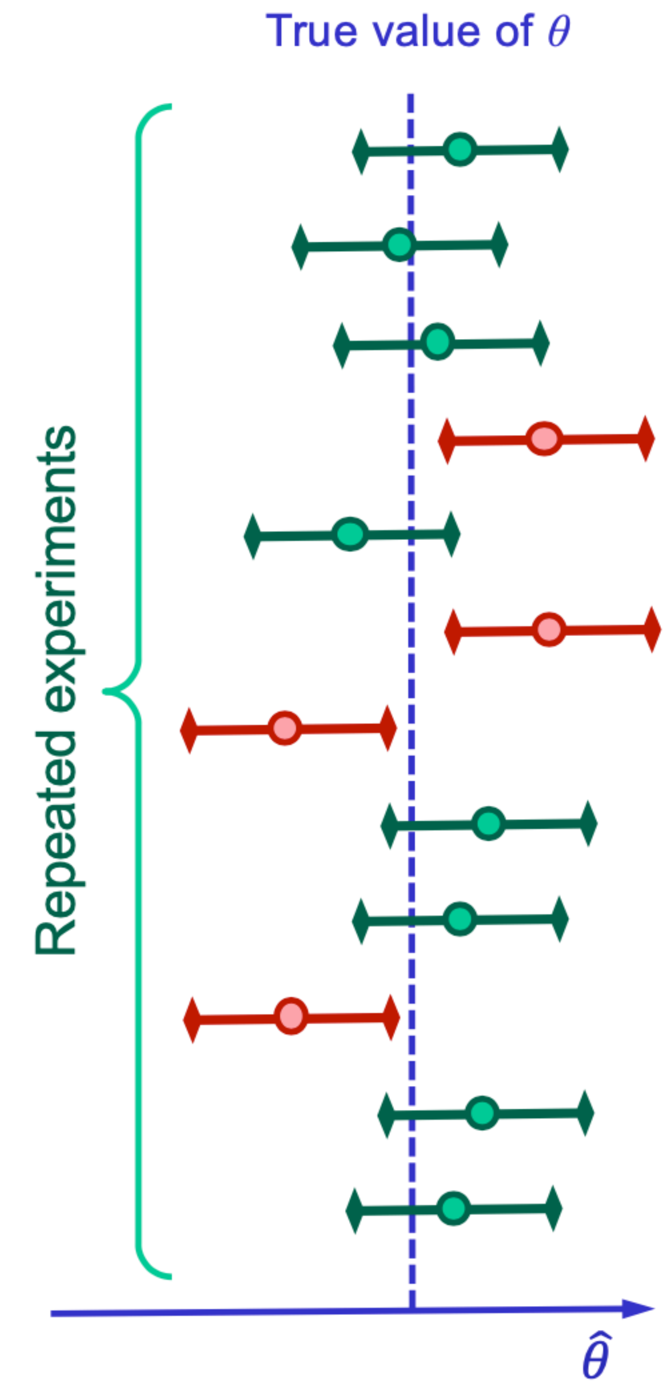
Frequentist Inference

Assigning a probability level of an unknown parameter
makes no sense in the frequentist approach
– Parameters are not random variables!

- A frequentist inference procedure determines a central value and an uncertainty interval that depend on the observed measurements
  - The central value and interval extremes are random variables
  - No subjective element is introduced in the determination
  - The function that returns the central value given an observed measurement is called estimator
  - Different estimator choices are possible, the most frequently adopted is the maximum likelihood estimator because of its statistical properties discussed in the following

Frequentist Coverage

- Repeating the experiment will result each time in a different data sample
- For each data sample, the estimator returns a different central value θ"
- An uncertainty interval [θ" – δ, θ" + δ] can be associated to the estimator's value θ"
- Some of the confidence intervals contain the fixed and unknown true value of θ, corresponding to a fraction equal to 68% of the times, in the limit of very large number of experiments (coverage)



Choice of 68% Intervals

Different interval choices are possible, corresponding to the same probability level (usually 68%, as 1σ for a Gaussian)

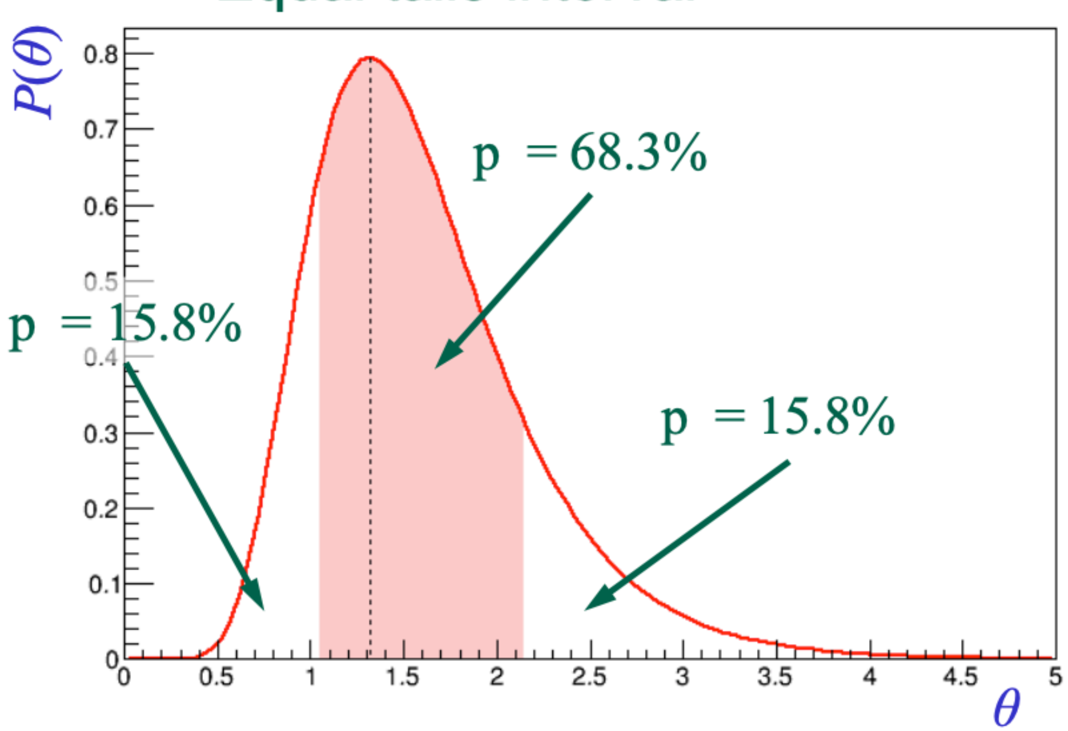
– Equal are as in the right and left tails

– Symmetric interval
– Shortest interval
–...

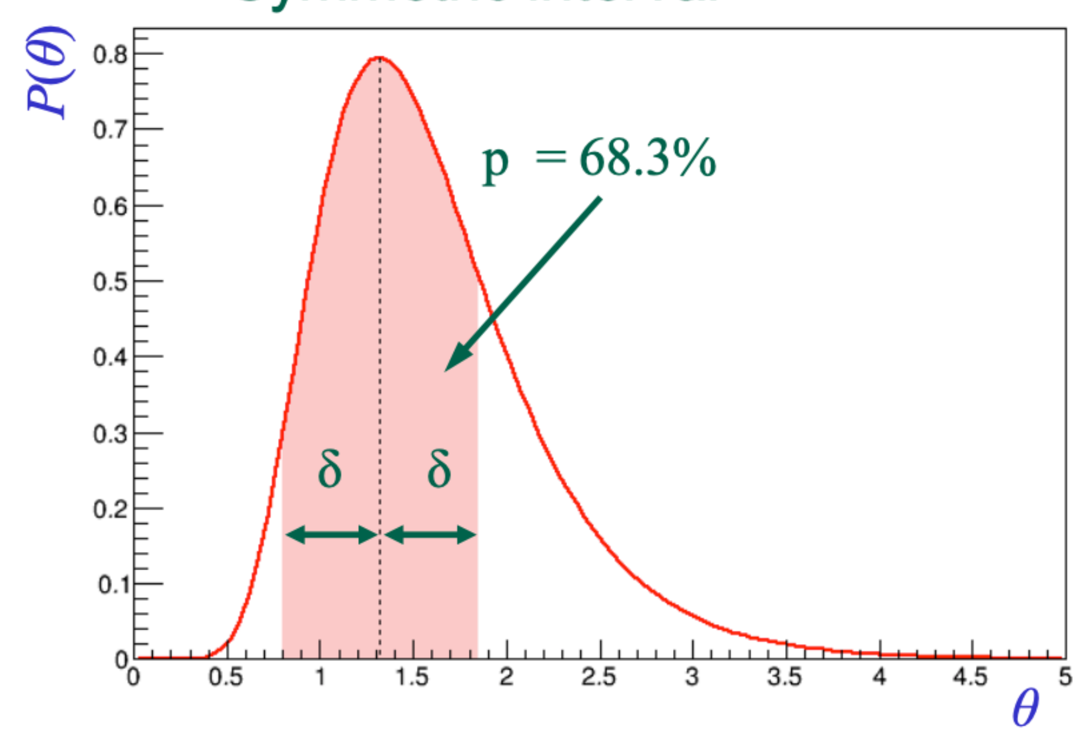
All equivalent for a symmetric distribution (e.g.Gaussian)

Reported as $\theta = \theta^{\text{up}} \pm \delta$ (sym.) or $\theta = \theta^{\text{up}}$ (asym.)

Equal tails interval

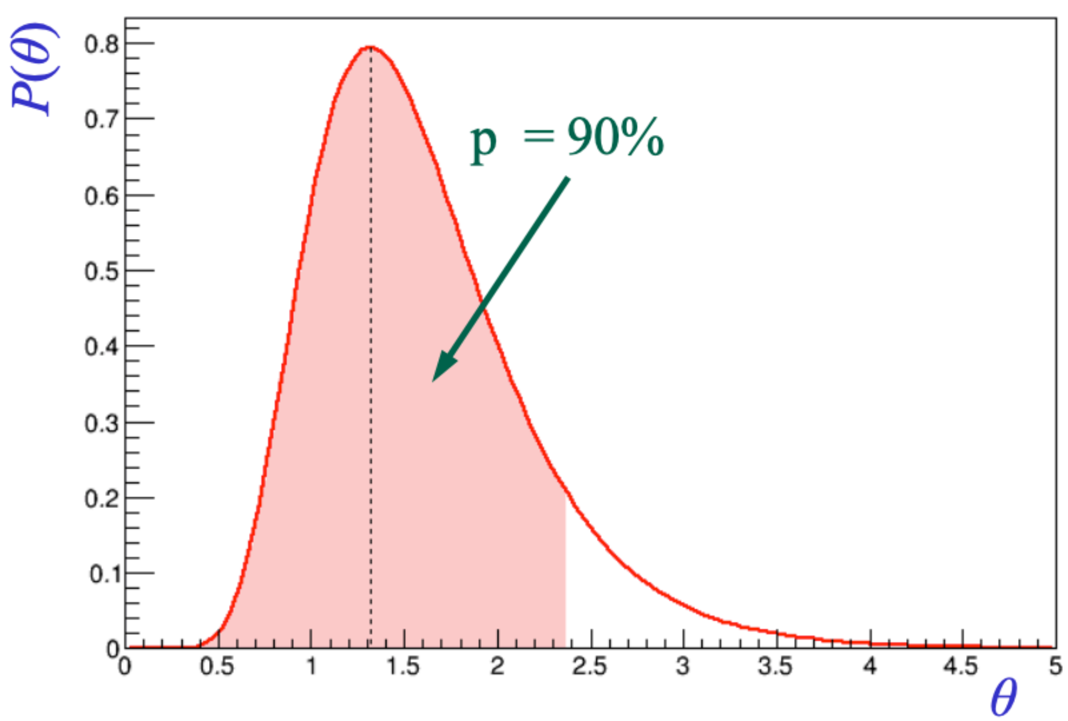


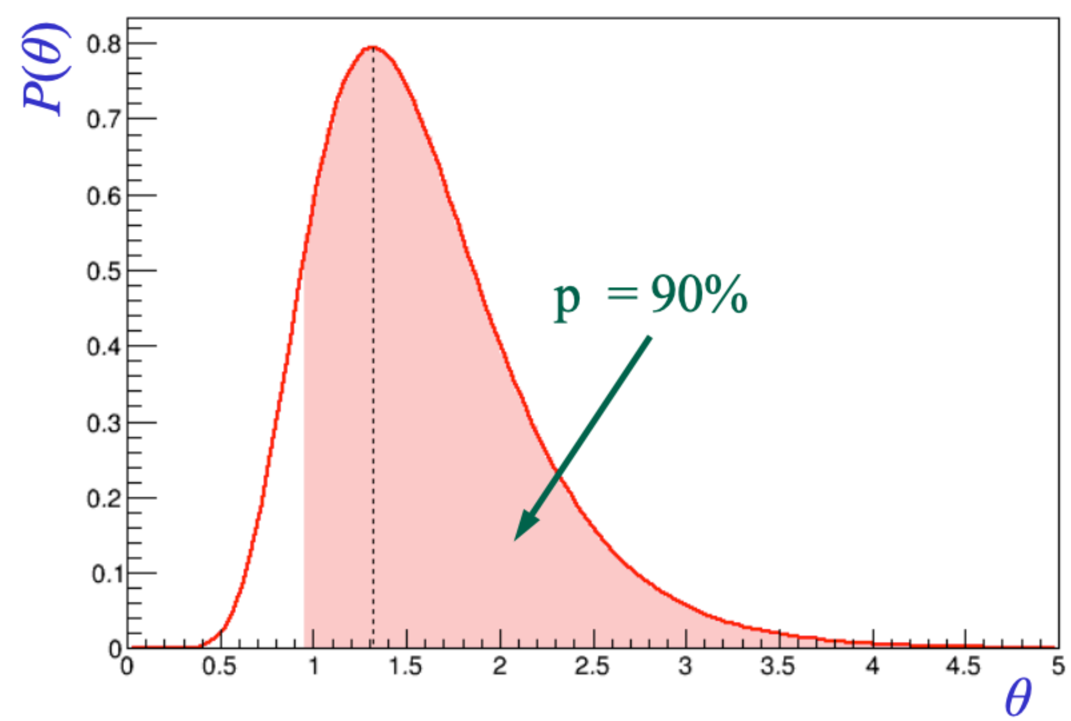
Symmetric interval



Upper and Lower Limits

- A fully asymmetric interval choice is obtained setting one extreme of the interval to the lowest or highest allowed range
- The other extreme indicates an upper or lower limits to the "allowed" range
- For upper or lower limits, usually a probability of 90% or 95% is preferred to the usual 68% adopted for central intervals
- Reported as: $\theta < \theta^{up}$ (90%CL) or $\theta > \theta^{lo}$ (90%CL)





Frequentist Inference - 2

An estimator is a function of a given set of measurements that provides an approximate value of a parameter of interest which appears in our PDF model (*best fit*)

- Simplest example:
  - Assume a Gaussian PDF with a known σ and an unknown μ
  - – A single experiment provides a measurement x
  - – We estimate μ as $\underline{\mu}$=x
  - – The distribution of $\underline{\mu}$ (repeating the experiment many times) is the original Gaussian
  - – 68.3% of the experiments (in the limit of large number of repetitions) will provide an estimate within: $\mu - \sigma < \underline{\mu} < \mu + \sigma$

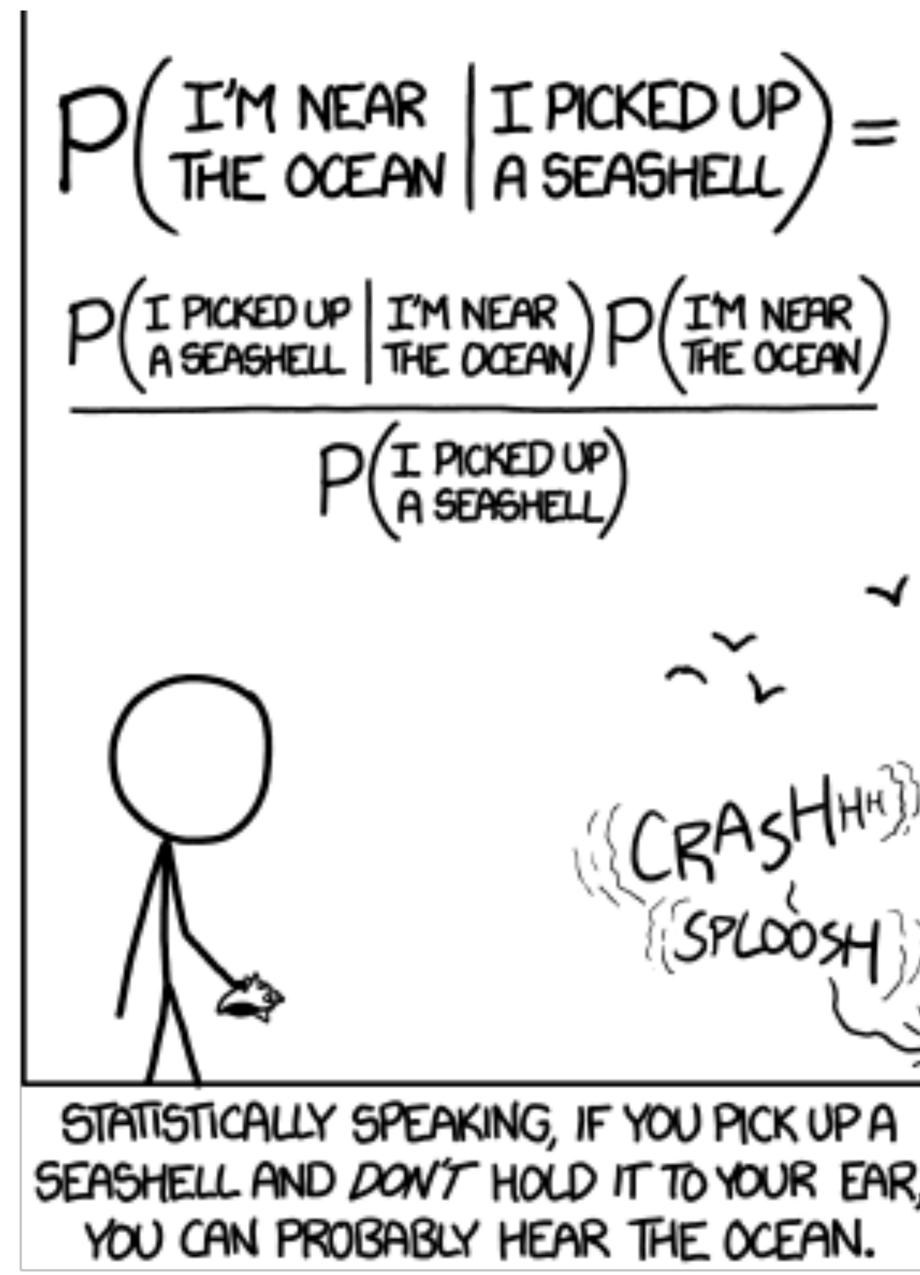
$$\mu = x \pm \sigma$$

The Maximum Likelihood Method

- The *maximum-likelihood estimator* is the most adopted parameter estimator

- The *best fit parameters* correspond to the set of values that maximizes the likelihood function

- The maximization can be performed analytically only in the simplest cases, and numerically for most of realistic cases



Meaning of parameter estimate

- We are interested in some physical unknown parameters
- Experiments provide samplings of some PDF which has among its parameters the physical unknowns we are interested in
- Experiment’s results are statistically “related” to the unknown PDF
  - PDF parameters can be determined from the sample within some approximation or uncertainty
- Knowing a parameter within some error may mean different things:

Meaning of parameter estimate

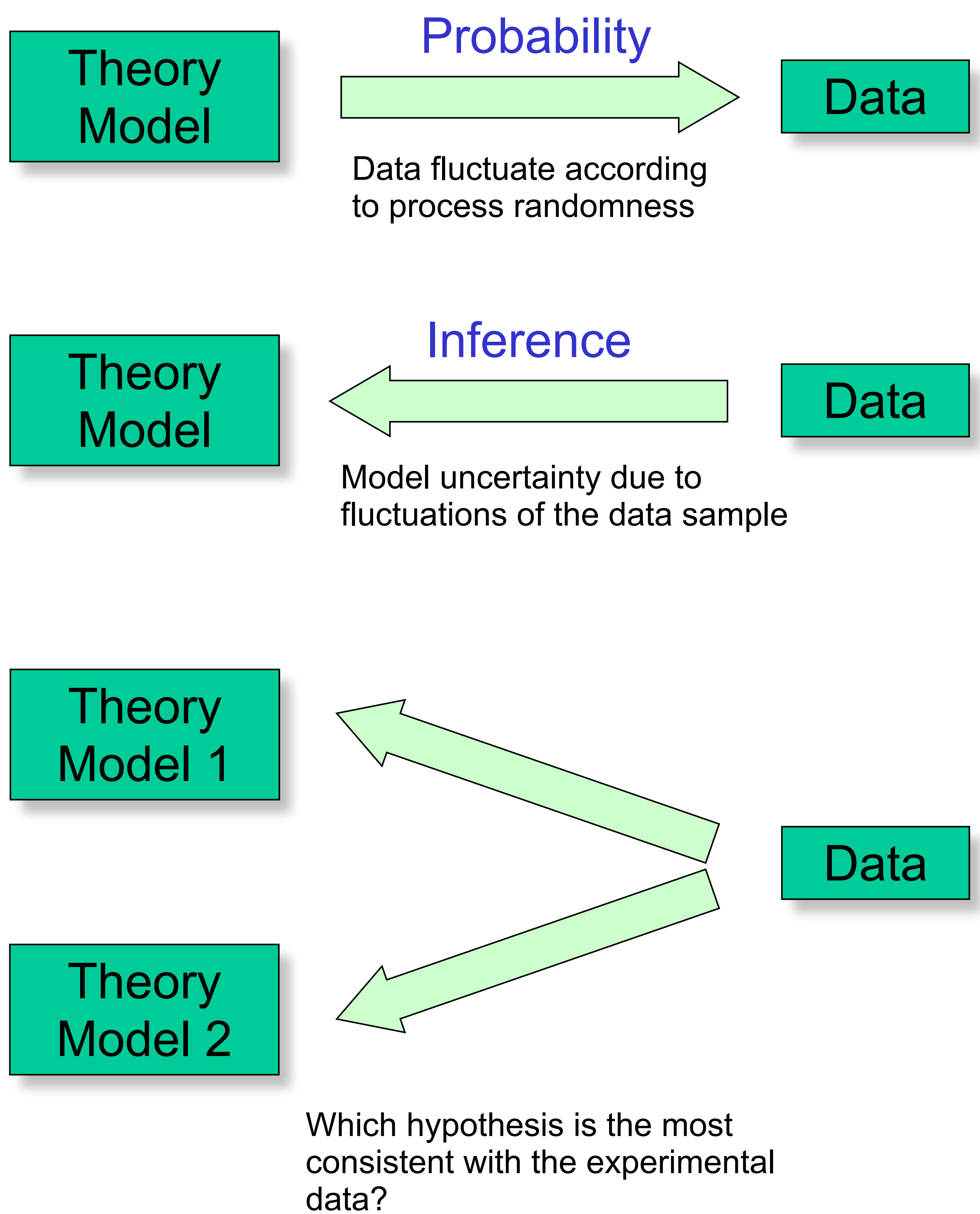
- **Frequentist**: a large fraction (68% or 95%, usually) of the experiments will contain, in the limit of large number of experiments, the (fixed) unknown true value within the quoted confidence interval, usually $[\mu - \sigma, \mu + \sigma]$ (*coverage*)

- **Bayesian**: we determine a degree of belief that the unknown parameter is contained in a specified interval can be quantified as 68% or 95%

- We will see that there is still some more degree of arbitrariness in the definition of confidence intervals…

Statistical inference vs Hypothesis testing

Statistical inference



Hypothesis testing

Parameter estimators

- An estimator is a function of a given sample whose statistical properties are known and related to some PDF parameters
  - "Best fit"

- Simplest example:
  - Assume we have a Gaussian PDF with a *known* σ and an *unknown* μ

    - A single experiment will provide a measurement x
    - We estimate μ as $\mu^{est} = x$
    - The distribution of μ^{est}

(repeating the experiment many times) is the original Gaussian

68.27%, *on average*, of the experiments will provide an estimate within: $\mu - \sigma < \mu^{est} < \mu + \sigma$

- We can determine: $\mu = \mu^{est} \pm \sigma$

Likelihood function

- Given a sample of N events each with variables $(x_1, \ldots, x_n)$, the likelihood function expresses the probability density of the sample, as a function of the unknown parameters:

$$L = \prod_{i=1}^{N} f(x_1^i, \cdots, x_n^i; \theta_1, \cdots, \theta_m)$$

- Sometimes the used notation for parameters is the same as for conditional probability:

$$f(x_1, \cdots, x_n | \theta_1, \cdots, \theta_m)$$

- If the size N of the sample is also a random variable, the extended likelihood function is also used:

$$L = p(N; \theta_1, \cdots, \theta_m) \prod_{i=1}^{N} f(x_1^i, \cdots, x_n^i; \theta_1, \cdots, \theta_m)$$

  – Where p is most of the times a Poisson distribution whose average is a function of the unknown parameters

$$\prod_i \rightarrow \sum_i$$

- In many cases it is convenient to use $-\ln L$ or $-2\ln L$:

Maximum likelihood estimates

- ML is the widest used parameter estimator
- The "best fit" parameters are the set that maximizes the likelihood function

  – "Very good" statistical properties, as will be seen in the following

- The maximization can be performed analytically, for the simplest cases, and numerically for most of the cases

- Minuit is historically the most used minimization engine in High Energy Physics
  – F. James, 1970's; rewritten in C++ recently

Extended likelihood function

- For Poissonian signal and background processes:

$$L(x_i; s, b, \theta) = \frac{(s+b)^n e^{-(s+b)}}{n!} \prod_{i=1}^{n} (f_s P_s(x_i;\theta) + f_b P_b(x_i;\theta))$$

$$\left.\begin{aligned} f_s &= \frac{s}{s+b} \\ f_b &= \frac{b}{s+b} \end{aligned}\right\} \rightarrow \quad = \frac{e^{-(s+b)}}{n!} \prod_{i=1}^{n} (s P_s(x_i;\theta) + b P_b(x_i;\theta))$$

- We can fit simultaneously s, b and θ minimizing:

$$-\ln L = s + b - \sum_{i=1}^{n} \ln(s P_s(x_i;\theta) + b P_b(x_i;\theta)) + \ln n!$$

- Sometimes s is replaced by $\mu\, s_0$, where s_0 is the theory estimate and μ is called signal strength

Gaussian Case

- If we have n independent measurements all modeled with (or approximated to) the same Gaussian PDF, we have:

$$-2\ln L = \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2} + n(\ln 2\pi + 2\ln\sigma)$$

- An analytical minimization of $-2\ln L$ w.r.t μ (assuming σ^2 is known) gives the arithmetic mean as ML estimate of μ:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- If σ^2 is also unknown, the ML estimate of σ^2 is:

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

- The above estimate can be demonstrated to have an unpleasant feature, called *bias* (→ next slide)

Estimators: Efficiency

- The variance of any consistent estimator is subject to a lower bound (Cramér-Rao bound):

$$\mathrm{Var}[\hat{\theta}] \geq \frac{\left(1 + \overbrace{\frac{\partial b(\theta)}{\partial \theta}}^{\text{bias of } \theta}\right)^2}{\underbrace{\left\langle \left(\frac{\partial \ln L(x_1, \cdots, x_n; \theta)}{\partial \theta} \right)^2 \right\rangle}_{\text{Fisher information}}} = V_{\mathrm{CR}}$$

- **Efficiency** can be defined as the ratio of Cramér-Rao bound and the estimator's variance:

$$\varepsilon(\hat{\theta}) = \frac{V_{\mathrm{CR}}}{\mathrm{Var}[\hat{\theta}]}$$

  - Efficiency for ML estimators tends to 1 for large number of measurements

  - I.e.: ML estimates have, asymptotically, the smallest possible variance

Estimators: Bias

- The bias of a parameter is the average value of its deviation from the true value

$$\mathrm{b}(\theta) = \left\langle \hat{\theta} - \theta \right\rangle = \left\langle \hat{\theta} \right\rangle - \theta$$

  – ML estimators may have a bias, but the bias decreases with large number of measurements (if the fit model is correct…!)

  – E.g.: in the case of the estimate of a Gaussian's σ^2, the unbiased estimate is the well known:

$$\hat{\sigma^2}_{\text{unbias.}} = \frac{n}{n-1}\hat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

ML method underestimates the variance σ^2

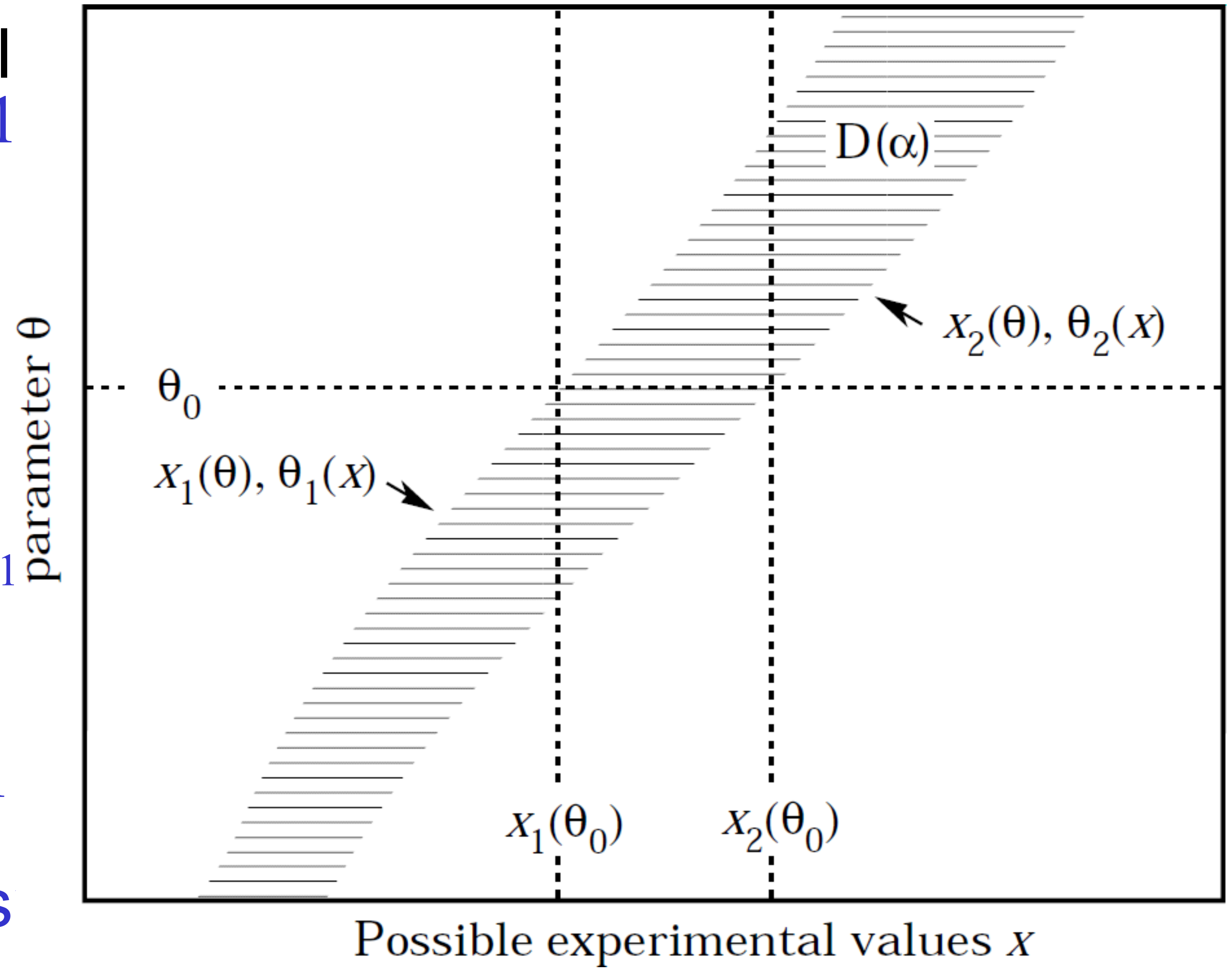
Estimators: Robustness

- If the sample distribution has (slight?) deviations from the theoretical PDF model, some estimators may deviate more or less than others from the true value
  - E.g.: unexpected tails ("outliers")
- The median is a robust estimate of a distribution average, while the mean is not
- Trimmed estimators: removing n extreme values
- Evaluation of estimator robustness:
  - Breakdown point: max. fraction of *incorrect* measurements above which the estimate may be arbitrary large
    - Trimmed observations at $x\%$ have a break point of x
    - The median has a break point of 0.5
  - Influence function:
    - Deviation of estimator if one measurement is replaced by an arbitrary (incorrect measurement)

Neyman's Confidence Intervals

Procedure to determine frequentist confidence intervals

- Scan the allowed range of an unknown parameter θ
- Given a value of θ compute the interval $[x_1, x_2]$ that contain x with a probability $1 - \alpha$ equal to 68% (or 90%, 95%)
- **Choice of interval needed!**
- Invert the confidence belt: for an observed value of x, find the interval $[\theta_1, \theta_2]$
- A fraction of the experiments equal to $1 - \alpha$ will measure x such that the corresponding $[\theta_1, \theta_2]$ contains ("covers") the true value of θ ("coverage")
- **Note**: the random variables are $[\theta_1, \theta_2]$, not θ !



α = significance level

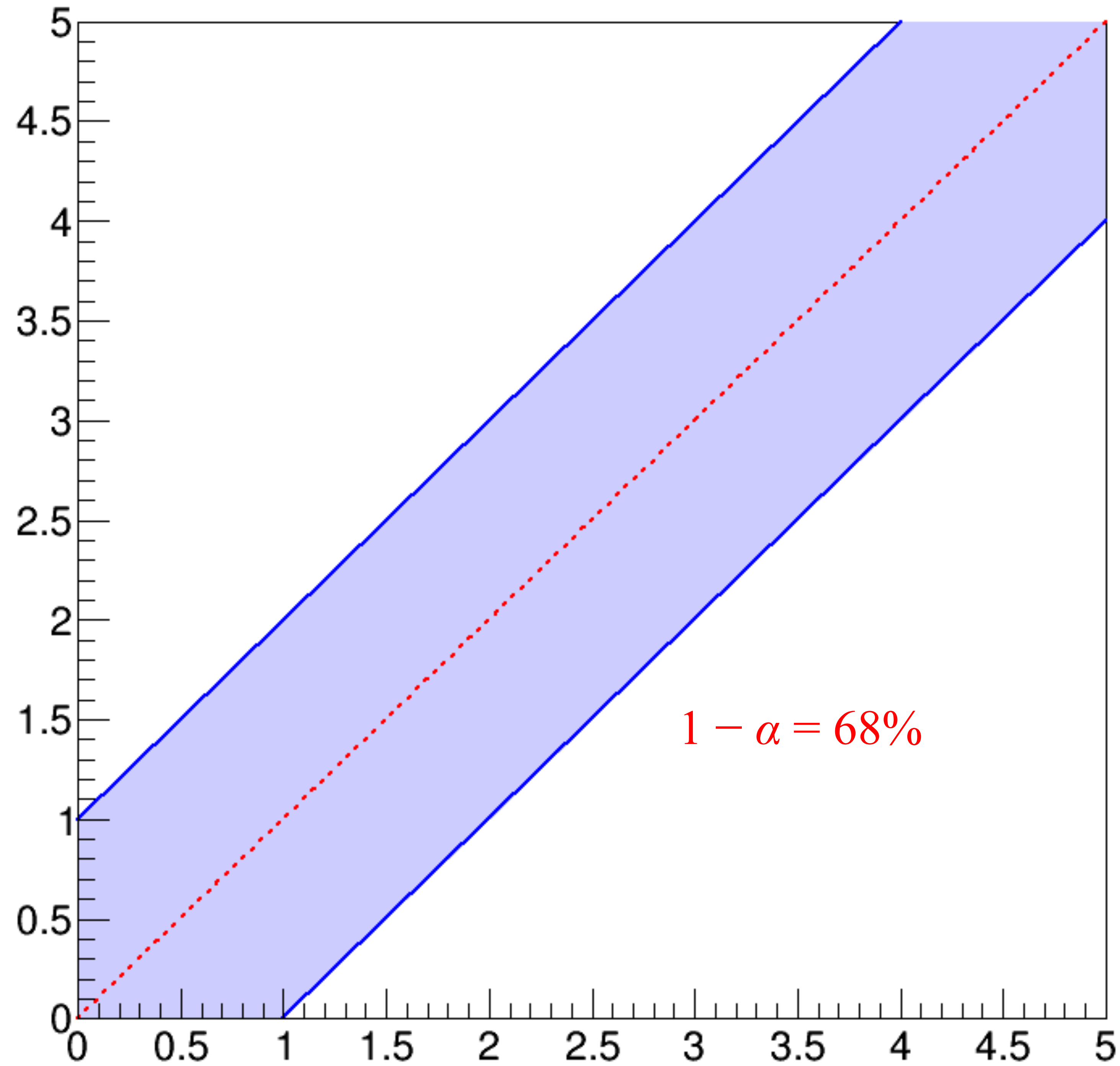
Neyman's Confidence Intervals: Gaussian case

- Assume a Gaussian distribution with unknown average μ and known $\sigma = 1$

- The belt inversion is trivial and gives the expected result:
  Central value $\hat{\mu} = x$,
  $[\mu_1, \mu_2] = [x - \sigma, x + \sigma]$

- So we can quote:

$$\mu = x \pm \sigma$$



ML Errors

- A parabolic approximation of $-2\ln L$ around the minimum is equivalent to a Gaussian approximation
  - Sufficiently accurate in many but not all cases

$$-2\ln L = \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2} + \text{const.}$$

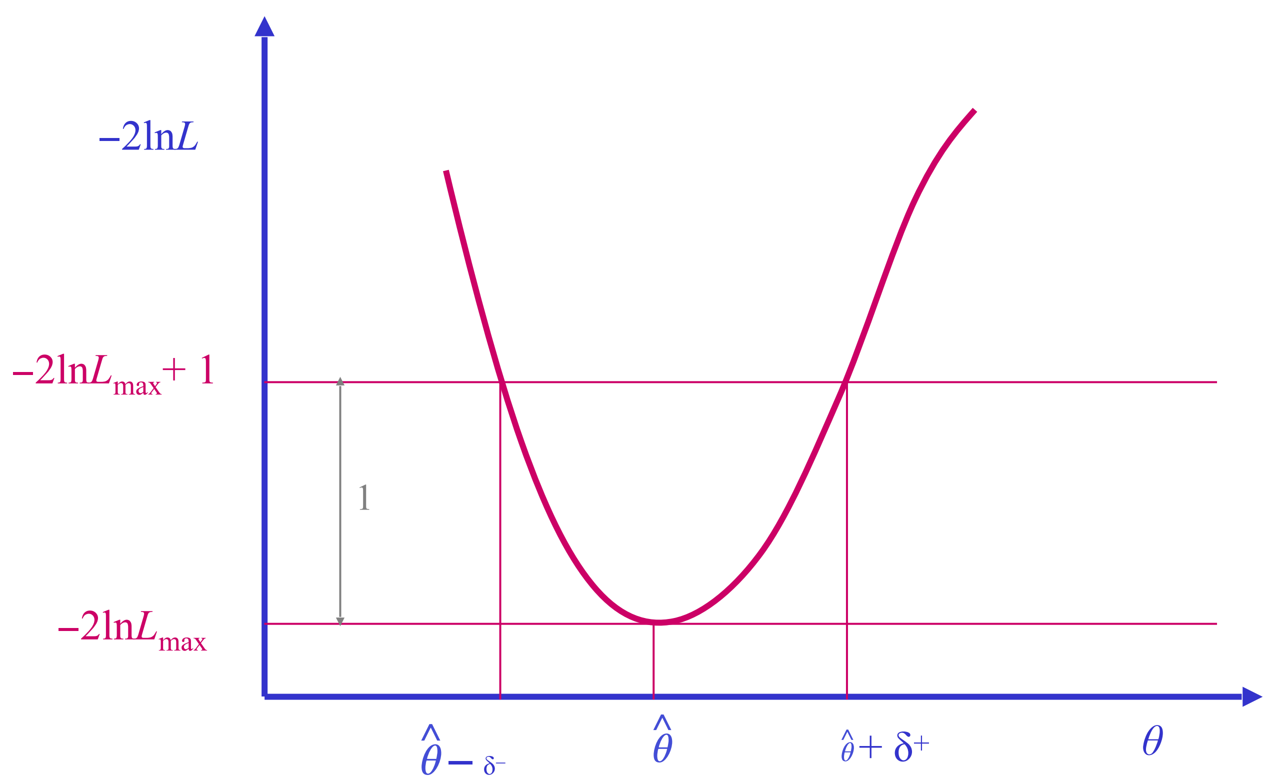
- Estimate of the covariance matrix from 2nd order partial derivatives w.r.t. fit parameters at the minimum:

$$V_{ij}^{-1} = -\left.\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}\right|_{\theta_k = \hat{\theta}_k}$$

- Implemented in Minuit as MIGRAD/HESSE function

Asymmetric Errors

- Another approximation alternative to the parabolic one may be to evaluate the excursion range of $-2\ln L$.
- Error ($n\sigma$) determined by the range around the maximum for which $-2\ln L$ increases by $+1$ ($+n^2$ for $n\sigma$ intervals)



- Errors can be asymmetric
- For a Gaussian PDF the result is identical to the 2nd order derivative matrix
- Implemented in Minuit as MINOS function
-

Asymmetric Errors (Gaussian case)

- We have the previous log-likelihood function:

$$-2\ln L = \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2} + n\left(\ln 2\pi + 2\ln\sigma\right)$$

- The error on μ is given by:

$$\frac{1}{\sigma_\mu^2} = \frac{\partial^2(-\ln L)}{\partial\mu^2} = \frac{n}{\sigma^2}$$

- I.e.: the error on the average is:

$$\sigma_\mu = \frac{\sigma}{\sqrt{n}}$$

Error Propagation

- Assume we estimate from a fit the parameter set:
  $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ and we know their covariance matrix Θ_{ij}
- We want to determine a new set of parameters that are functions of $\boldsymbol{\theta}$:
  $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_m)$.
- For small uncertainties, a linear approximation maybe sufficient
- A Taylor expansion around the central values of $\boldsymbol{\theta}$ gives, using the error matrix Θ_{ij}:

$$H_{ij} = \sum_{k,l} \frac{\partial \eta_i}{\partial \theta_k} \frac{\partial \eta_j}{\partial \theta_l} \Theta_{kl}$$

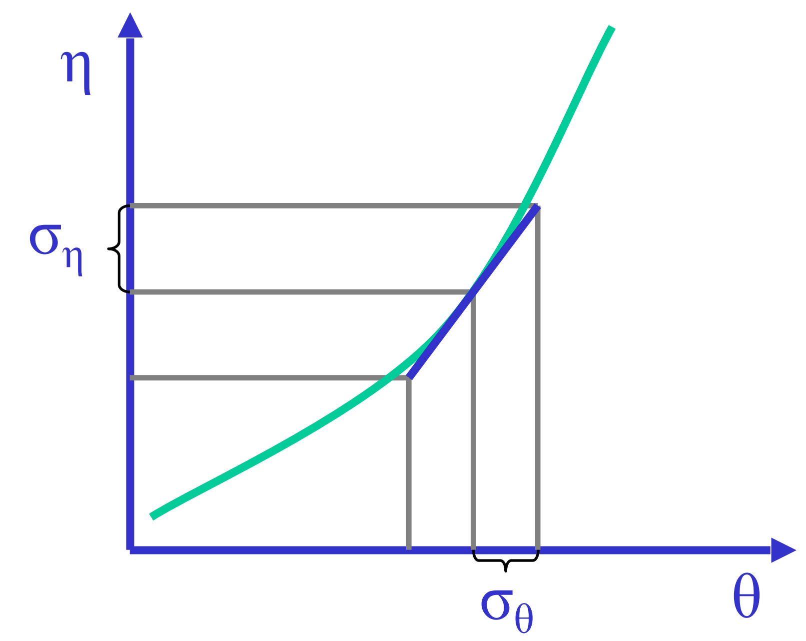
- Few examples in case of no correlation:

$$\sigma_{x+y} = \sigma_{x-y} = \sqrt{\sigma_x^2 + \sigma_y^2}$$

$$\frac{\sigma_{xy}}{xy} = \frac{\sigma_{x/y}}{x/y} = \sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

$$\sigma_{x^2} = 2x\sigma_x$$

$$\sigma_{\ln x} = \frac{\sigma_x}{\sqrt{x}}$$



Asymmetric Errors: warnings

- Much better to know the original PDF and propagate/combine the information properly!
  - Be careful about interpreting the meaning of the result

- Average value and Variance propagate linearly, while most probable value (mode) does not add linearly

- Whenever possible, use a single fit rather than multiple cascade fits, and quote the final asymmetric errors only

Asymmetric Errors: warnings

- Be careful about:
  - Asymmetric error propagation
  - Combining measurements with asymmetric errors
  - Difference of “most likely value” w.r.t. “average value”
- Naïve quadrature sum of σ_+ and σ_- lead to wrong answer
  - Violates the central limit theorem: the combined result should be more symmetric than the original sources!
  - A model of the non-linear dependence may be needed for quantitative calculations
  - Biases are very easy to achieve (depending on $\sigma_+ - \sigma_-$, and on the non-linear model)

Binned Likelihood

- Sometimes data are available as binned histogram
  - Most often each bin obeys Poissonian statistics (event counting)

- The likelihood function is the product of Poisson PDFs corresponding to each bin having entries n_i

- The expected number of entries n_i depends on some unknown parameters: $\mu_i = \mu_i(\theta_1, \ldots, \theta_m)$

- The function to minimize is the following $-2 \ln L$:

$$
\begin{aligned}
-2\ln L &= -2\ln \prod_{i=1}^{n_{\mathrm{bins}}} \mathrm{Poiss}(n_i; \mu_i(\theta_1, \cdots, \theta_m)) \\
&= -2\ln \prod_{i=1}^{n_{\mathrm{bins}}} \frac{e^{-\mu_i(\theta_1, \cdots, \theta_m)}\mu_i(\theta_1, \cdots, \theta_m)^{n_i}}{n_i!}
\end{aligned}
$$

- The expected number of entries μ_i is often approximated by a continuous function $\mu(x)$ evaluated at the center x_i of the bin
- Alternatively, μ_i can be a combination of other histograms ("templates")
  - E.g.: sum of different simulated processes with floating yields as fit parameters

Binned Likelihood

- Bin entries can be approximated by Gaussian variables for sufficiently large number of entries with standard deviation equal to n_i (Neyman's χ^2)

- Maximizing L is equivalent to minimize:

$$\chi^2 = \sum_{i=1}^{n_{\text{bins}}} \frac{(n_i - \mu(x_i; \theta_1, \cdots, \theta_m))^2}{n_i}$$

- Sometimes, the denominator n_i is replaced (Pearson's χ^2) by:

$$\mu_i = \mu\,(x_i;\, \theta_1,\, \ldots,\, \theta_m)$$

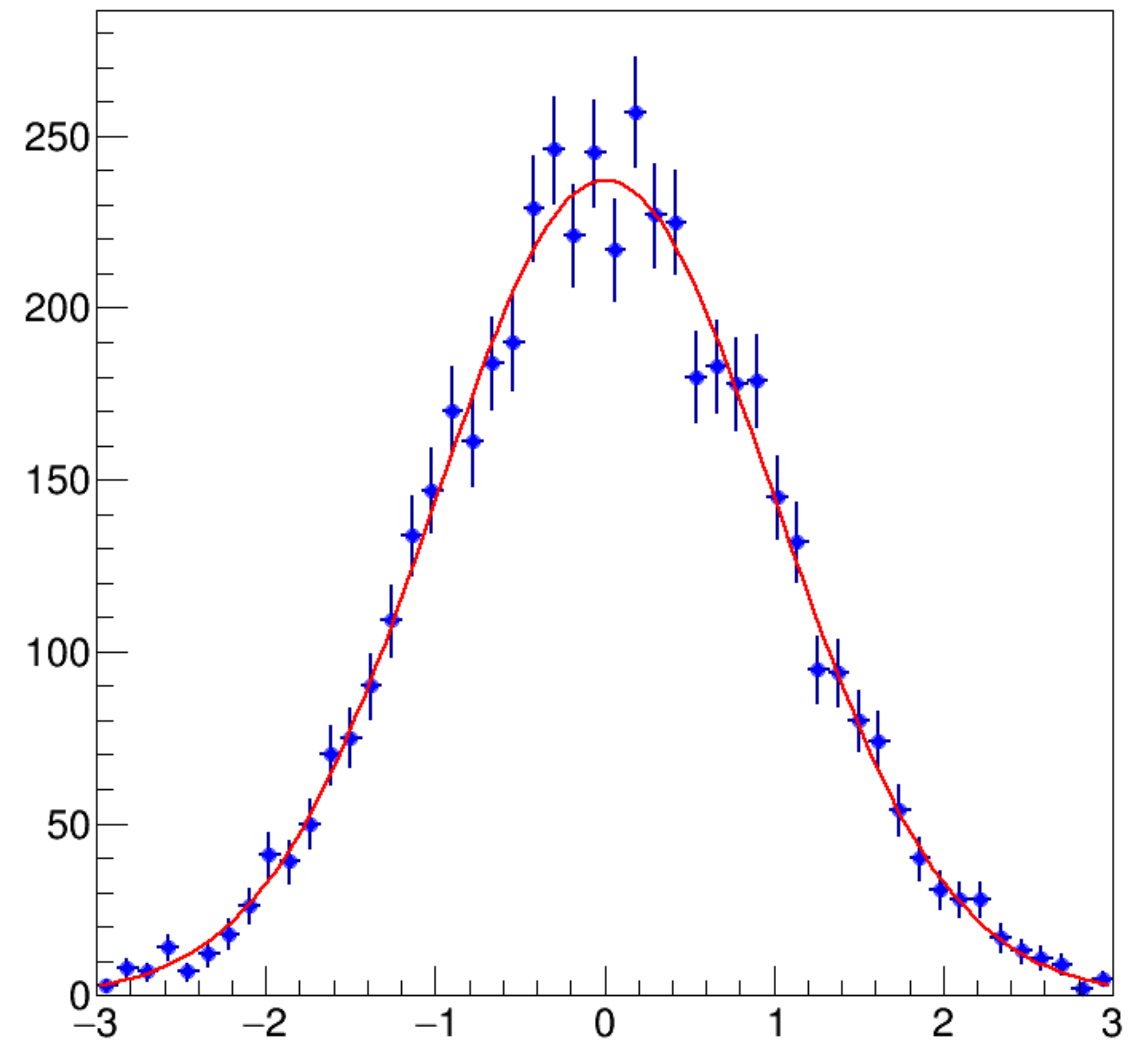
  in order to avoid cases with zero or small n_i

- Analytic solution exists for linear and other simple problems
  – E.g.: linear fit model
- Most of the cases are treated numerically, as for unbinned ML fits

Binned fit: example

- Binned fits are convenient w.r.t. unbinned fits because the number of input variables decreases from the number of entries to the number of bins
  - –Usually simpler and faster numerically
  - –Unbinned fits become unpractical for very large number of entries
- A fraction of the information is lost, hence a possible loss of precision may occur for small number of entries
- Treat correctly bins with small number of entries!

Gaussian fit (determine yield, μ and σ)



Binned fit quality: the p-value

- The maximum value of the likelihood function obtained from the fit doesn't usually give information about the goodness of the fit

- The χ^2 of a fit with a Gaussian underlying model is distributed according to a known PDF

$$P(\chi^2; n) = \frac{2^{-\frac{n}{2}}}{\Gamma(\frac{n}{2})} \chi^{n-2} e^{-\frac{\chi^2}{2}}$$

n is the number of degrees of freedom (n. of bins − n. of params.)

  - The cumulative distribution of $P(\chi^2; n)$ follows a uniform distribution between 0 and 1 (p-value)
  - If the model deviates from the assumed distribution, the distribution of the p-value will be more peaked around zero

- Note! p-values are not the "*probability of the fit hypothesis*"
  - This would be a Bayesian probability, with a different meaning, and should be computed in a different way

Likelihood Ratio

- A better alternative to the (Gaussian-inspired, Neyman and Pearson's) χ^2 has been proposed by Baker and Cousins using the following *likelihood ratio*:

$$\begin{aligned}
\chi^2_\lambda &= -2\ln\prod_i \frac{L(n_i;\mu_i)}{L(n_i;n_i)} = -2\ln\prod_i \frac{e^{-\mu_i}\mu_i^{n_i}}{\cancel{n_i!}}\frac{\cancel{n_i!}}{e^{-n_i}n_i^{n_i}} \\
&= 2\sum_i \left[\mu_i(\theta_i,\cdots,\theta_m) - n_i + n_i \ln\left(\frac{n_i}{\mu_i(\theta_1,\cdots,\theta_m)}\right)\right]
\end{aligned}$$

- Same minimum value as from Poisson likelihood function, since a constant term has been added to the log-likelihood function

- In addition, it provides goodness-of-fit information, and asymptotically obeys chi-squared distribution with $n - m$ degrees of freedom (Wilks' theorem, see following slides)

Combinations

- Assume two measurements with different uncorrelated (Gaussian) errors: $m_1 \pm \sigma_1, \quad m_2 \pm \sigma_2$

- Build the χ^2: $$\chi^2 = \frac{(m - m_1)^2}{\sigma_1^2} + \frac{(m - m_2)^2}{\sigma_2^2}$$

- Minimize the χ^2: $$0 = \frac{\partial \chi^2}{\partial m} = 2\frac{(m - m_1)}{\sigma_1^2} + 2\frac{(m - m_2)}{\sigma_2^2}$$

Weighted average, $w_i = \sigma_i^{-2}$

- Estimate m as: $$m = \frac{\frac{m_1}{\sigma_1^2} + \frac{m_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} = \frac{w_1 m_1 + w_2 m_2}{w_1 + w_2}$$

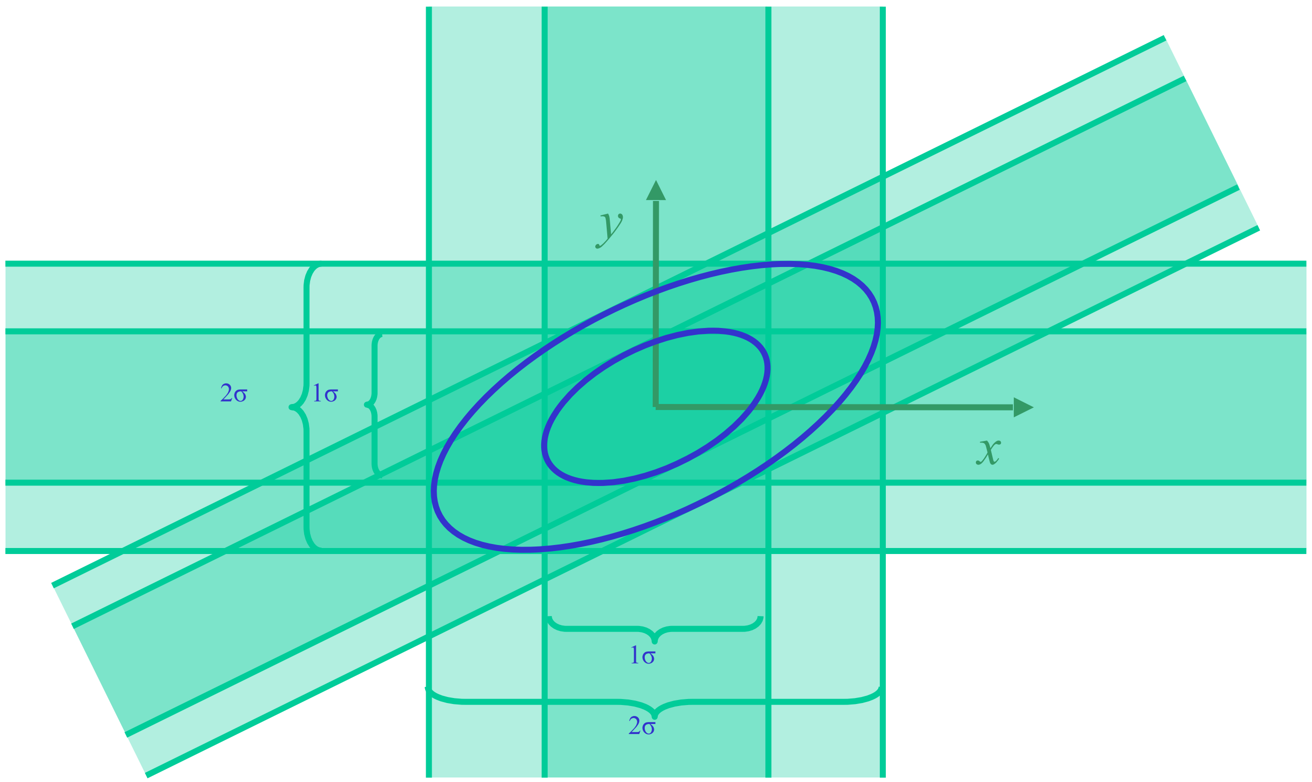
- Error estimate: $$\frac{1}{\sigma_m^2} = -\frac{\partial^2 \ln L}{\partial m^2} = \frac{1}{2}\frac{\partial^2 \chi^2}{\partial m^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

Higher dimensions: 2D-intervals

In more dimensions one can determine 1σ and 2σ contours
Note: different probability content in 2D compared to one dimension
68% and 95% contours are usually preferable

$$P_{1D}(n\sigma) = \sqrt{\frac{2}{\pi}} \int_0^n e^{-\frac{x^2}{2}} \mathrm{d}x = \mathrm{erf}\left(\frac{n}{\sqrt{2}}\right) \quad P_{2D}(n\sigma) = \int_0^n e^{-\frac{r^2}{2}} r \mathrm{d}r = 1 - e^{-\frac{n^2}{2}}$$

Width	P_{1D}	P_{2D}
1σ	0.6827	0.3934
2σ	0.9545	0.8647
3σ	0.9973	0.9889
1.515σ		0.6827
2.486σ		0.9545
3.439σ		0.9973



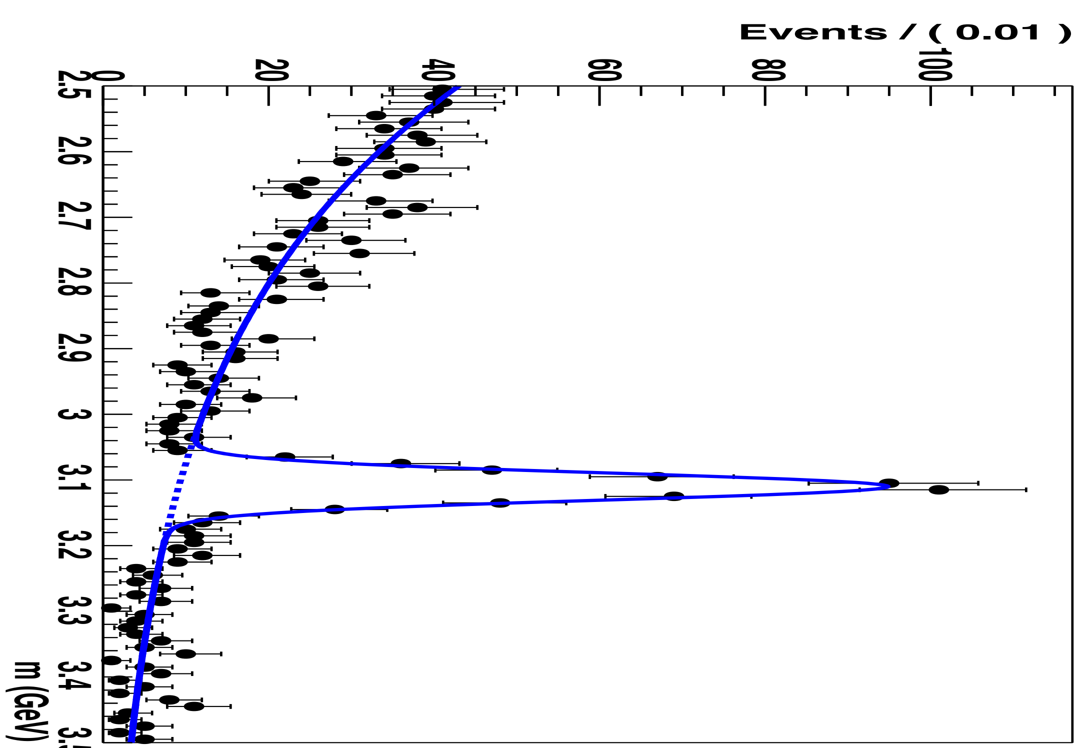
Higher dimensions: 2D-intervals

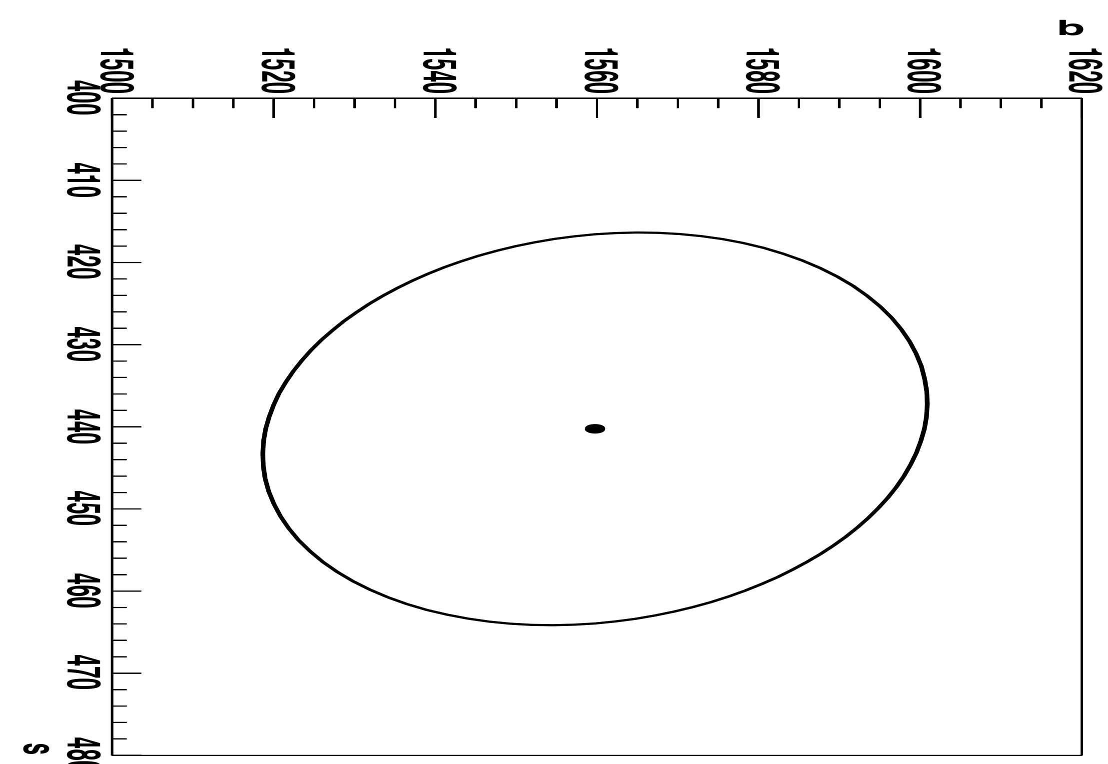
In more dimensions one can determine 1σ and 2σ contours
Note: different probability content in 2D compared to one dimension
68% and 95% contours are usually preferable

- From previous fit example:
  - $P_s(m)$: Gaussian peak
  - $P_b(m)$: exponential shape

Exponential decay parameter, Gaussian mean and standard deviation are fit together with s and b yields.

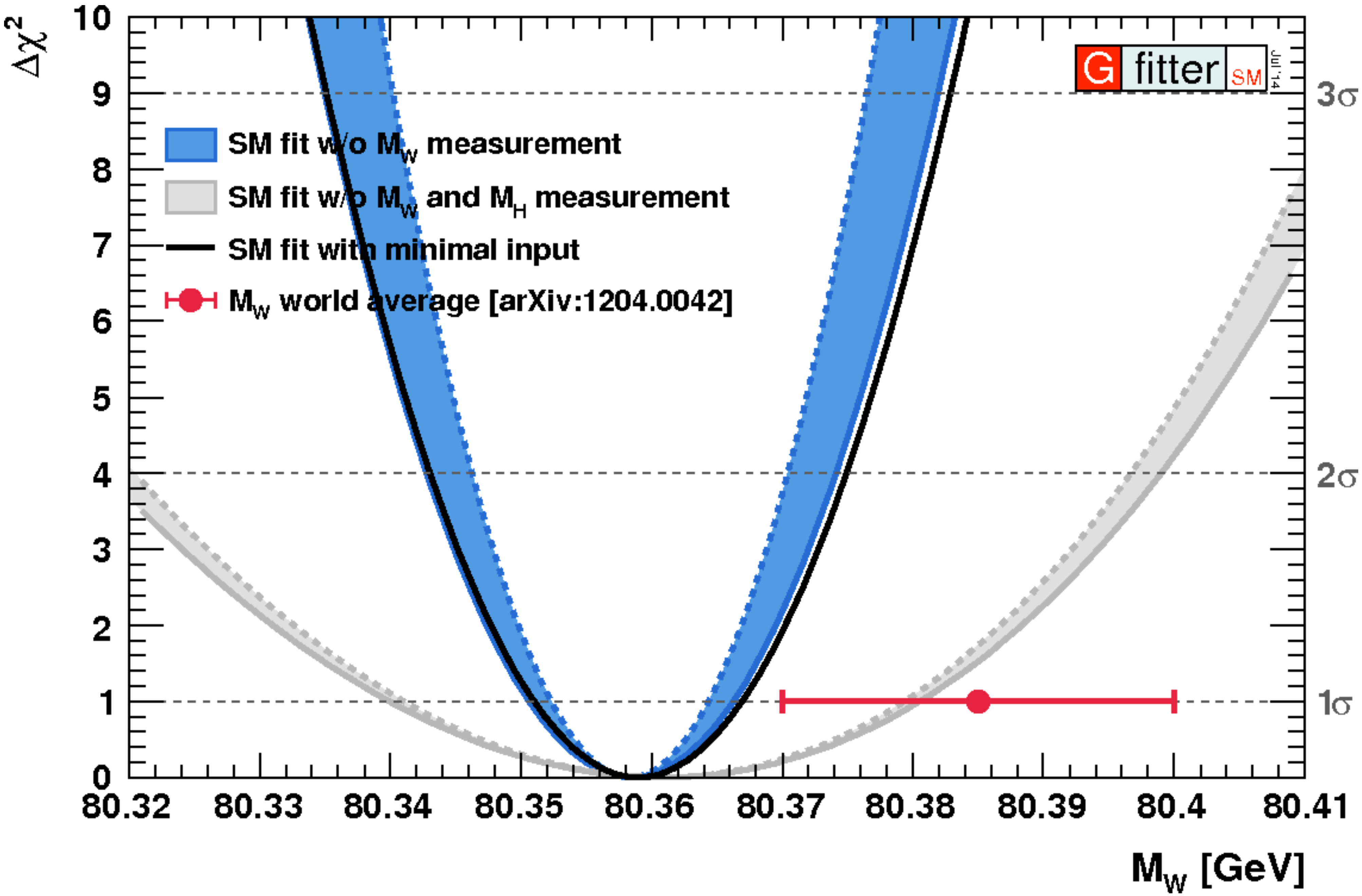
The contour shows for this case a mild correlation between s and b





Global Electroweak Fit

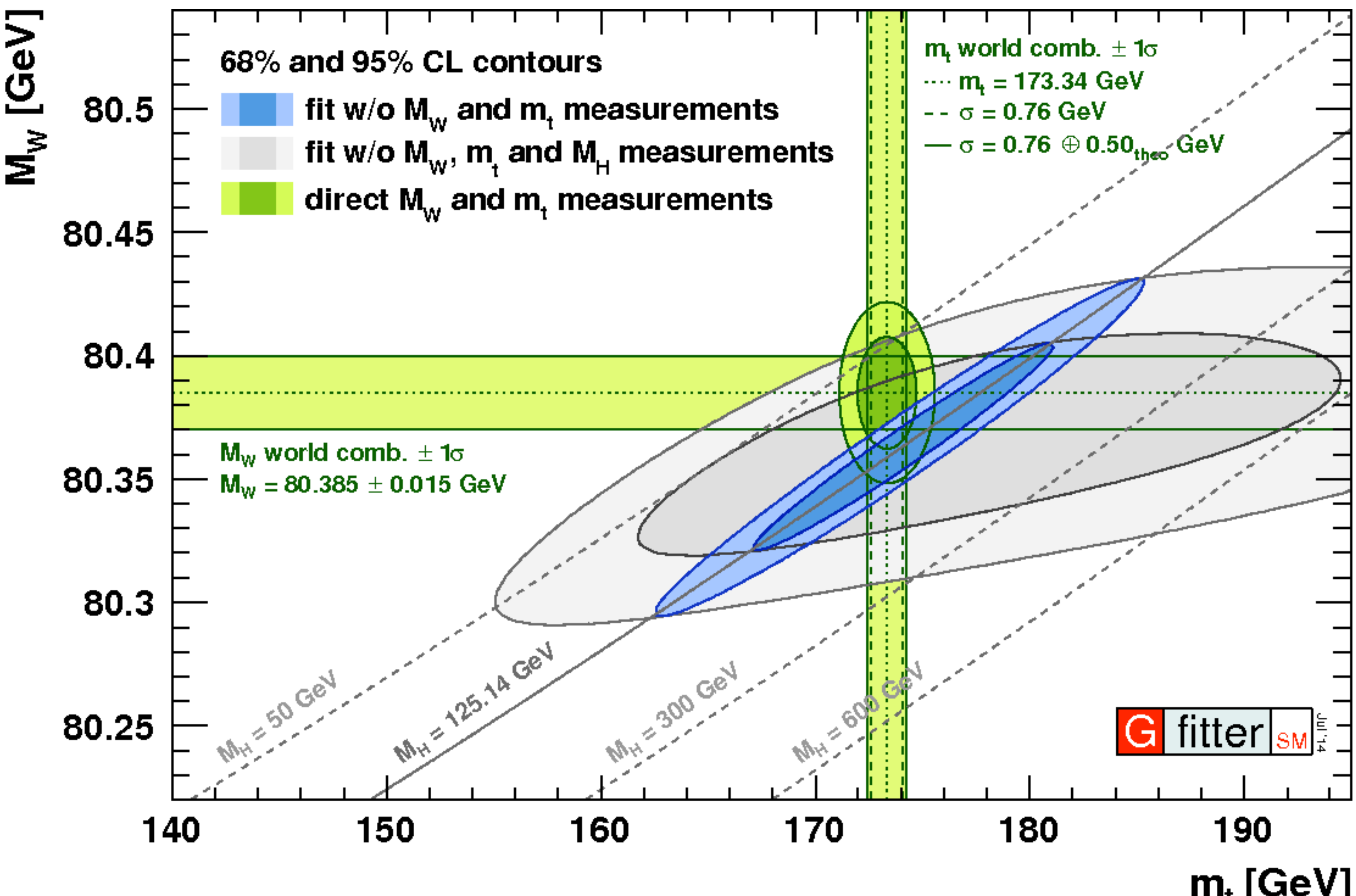
- A Global χ^2 fit to electroweak measurements predicts the W mass allowing a comparison with direct measurements



Details on:
http://gfitter.desy.de/Standard_Model/

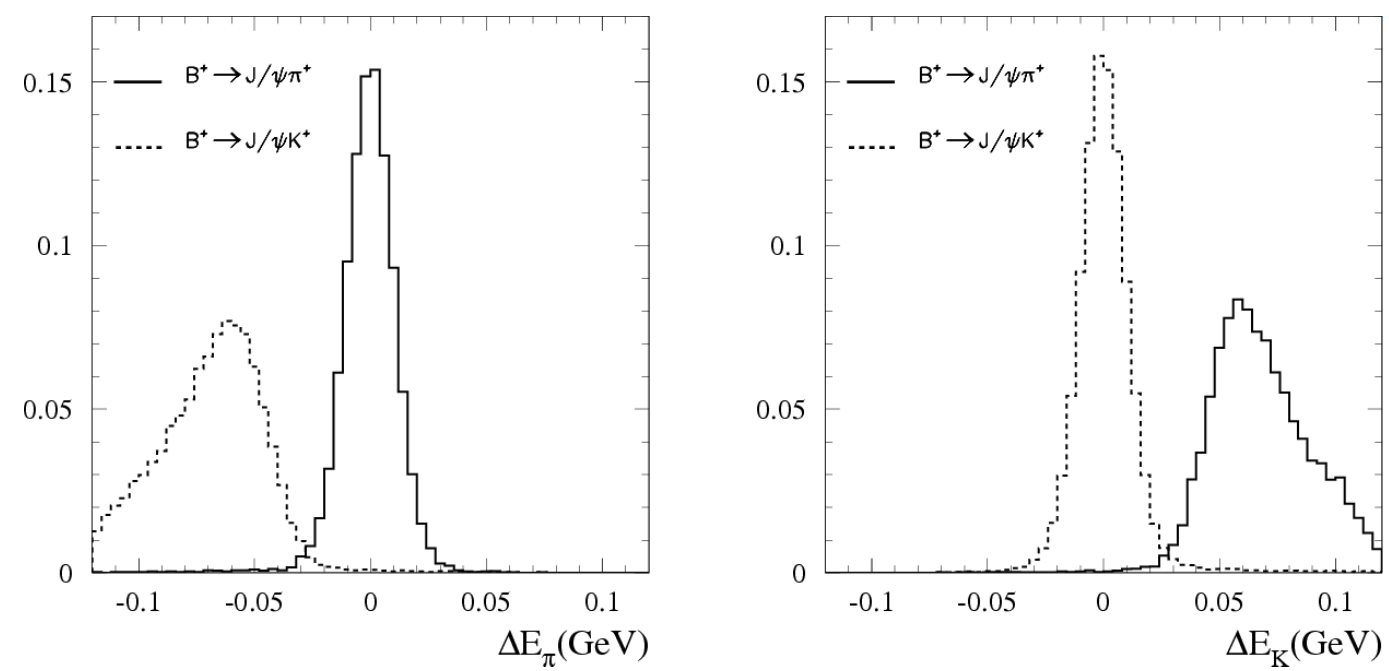
Higher dimensions: 2D-intervals

W mass vs top-quark mass from global electroweak fit



Example: Fitting $\mathcal{B}(B^+ \to J/\psi\pi^+)$ / $\mathcal{B}(B^+ \to J/\psi K^+)$

- Four variables:
  - m = B reconstructed mass as J/ψ + charged hadron invariant mass
  - ΔE_π = Beam – B energy in the π^+ mass hypothesis
  - ΔE_K = Beam – B energy in the K^+ mass hypothesis
  - q = B meson charge
- Two samples:
  - $J/\psi \to \mu^+\mu^-$, $J/\psi \to e^+e^-$



- Simultaneous fit of:
  - Total yield of $B^+ \to J/\psi\pi^+$, $B^+ \to J/\psi K^+$ and background
  - Resolutions separately for $J/\psi \to \mu^+\mu^-$, $J/\psi \to e^+e^-$
  - Charge asymmetry (direct CP violation)

$$m_{ES} = \sqrt{E_{\text{beam}}^2 - p_B^2}$$

Example: Fitting $\mathcal{B}(B^+ \to J/\psi\pi^+)$ / $\mathcal{B}(B^+ \to J/\psi K^+)$

- To extract the ratio of BR:

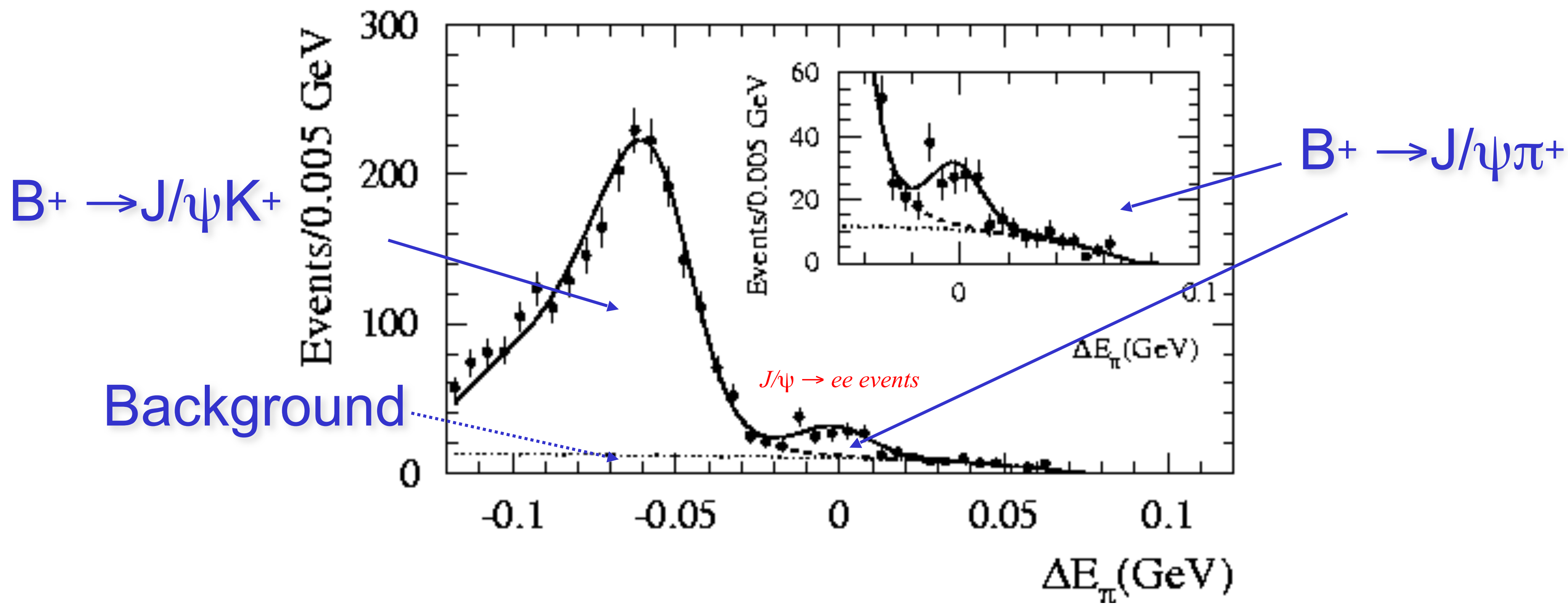
$$\begin{aligned} -\ln L &= && n_\pi + n_K + n_{bkg} \\ &= -\textstyle\sum_i \ln\,[&& n_\pi P_\pi(\Delta E_{\pi i}, \Delta E_{K i}, m_i) \\ & & + & \; n_K P_K(\Delta E_{\pi i}, \Delta E_{K i}, m_i) \\ & & + & \; n_{bkg} P_{bkg}(\Delta E_{\pi i}, \Delta E_{K i}, m_i) \quad] \end{aligned}$$

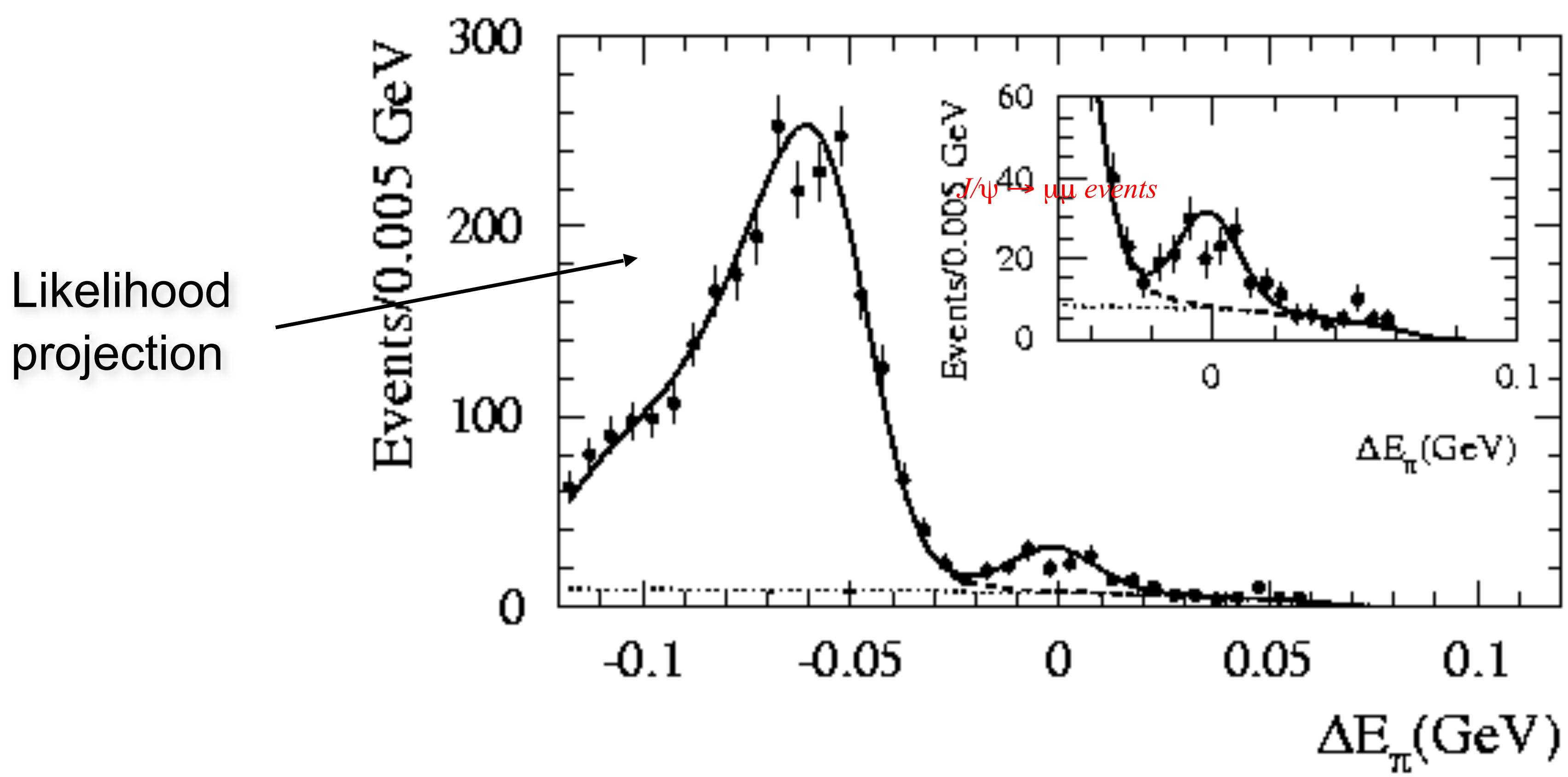
- Likelihood can be written separately, or combined for ee and $\mu\mu$ events

- Fit contains parameters of interest (mainly n_π, n_K) plus uninteresting *nuisance* parameters

- Separating $q = +1$ / -1 can be done adding A_{CP} as extra parameter

Example: Fitting $\mathcal{B}(B^+ \to J/\psi\pi^+)$ / $\mathcal{B}(B^+ \to J/\psi K^+)$

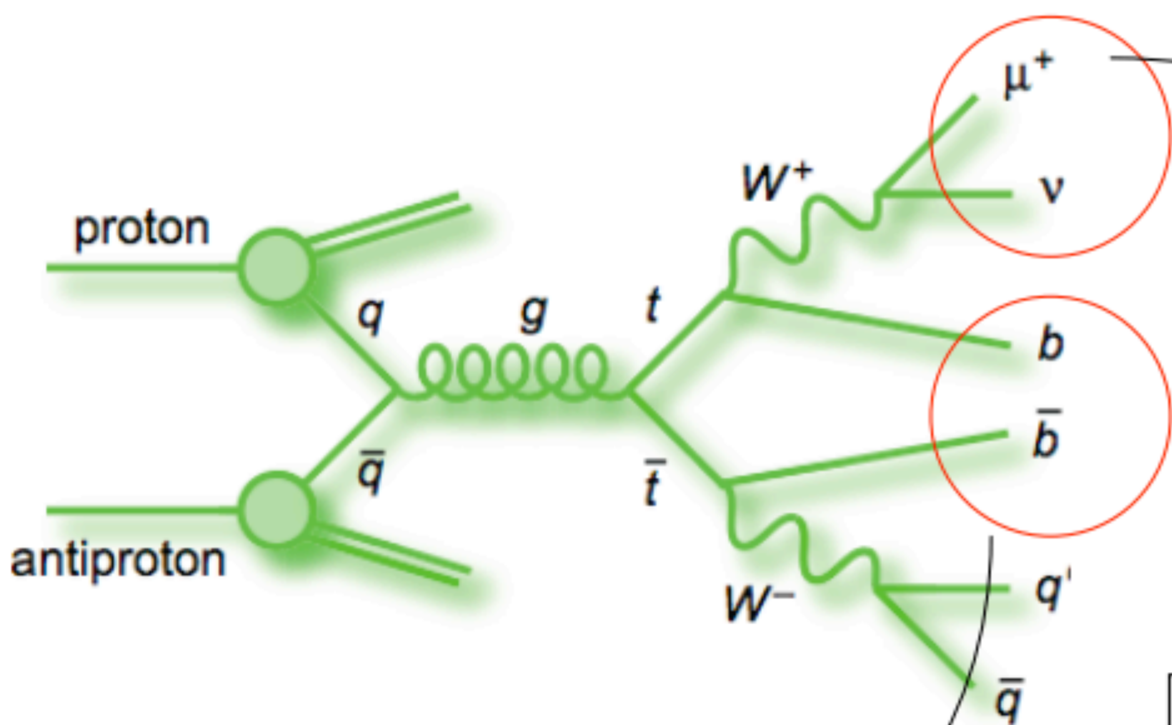




Example 2: top mass @ CDF

Il quark top al Tevatron

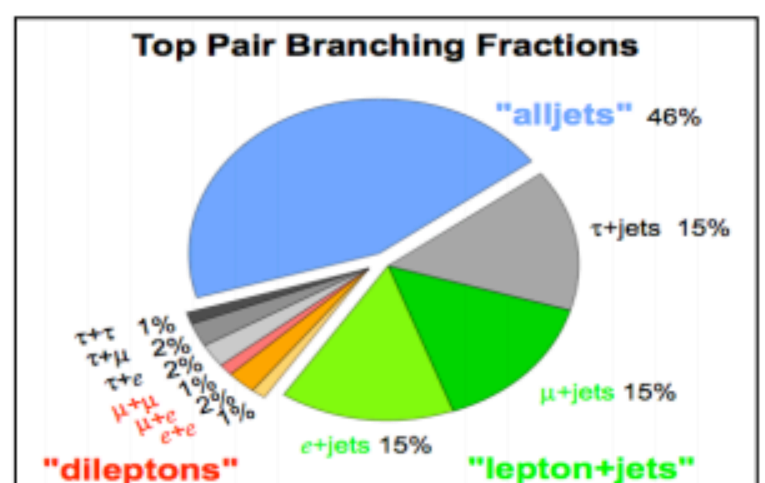
- Non riesce ad adronizzare: $\tau = 10^{-25}$s
- Decade nel canale $t \rightarrow W+b$ (BR≈100%)
- Produzione di top al Tevatron dalle collisioni pp a $\sqrt{s}$=1,96 TeV:



Tre tipologie di analisi dei prodotti di ttbar in base al decadimento del W:

- ***lνqqbb*** ***"lepton+jets"***
- *qqbb* *"all hadronic"*
- *llνν* *"pure leptonic"*

b-tagging fondamentale per ridurre il fondo

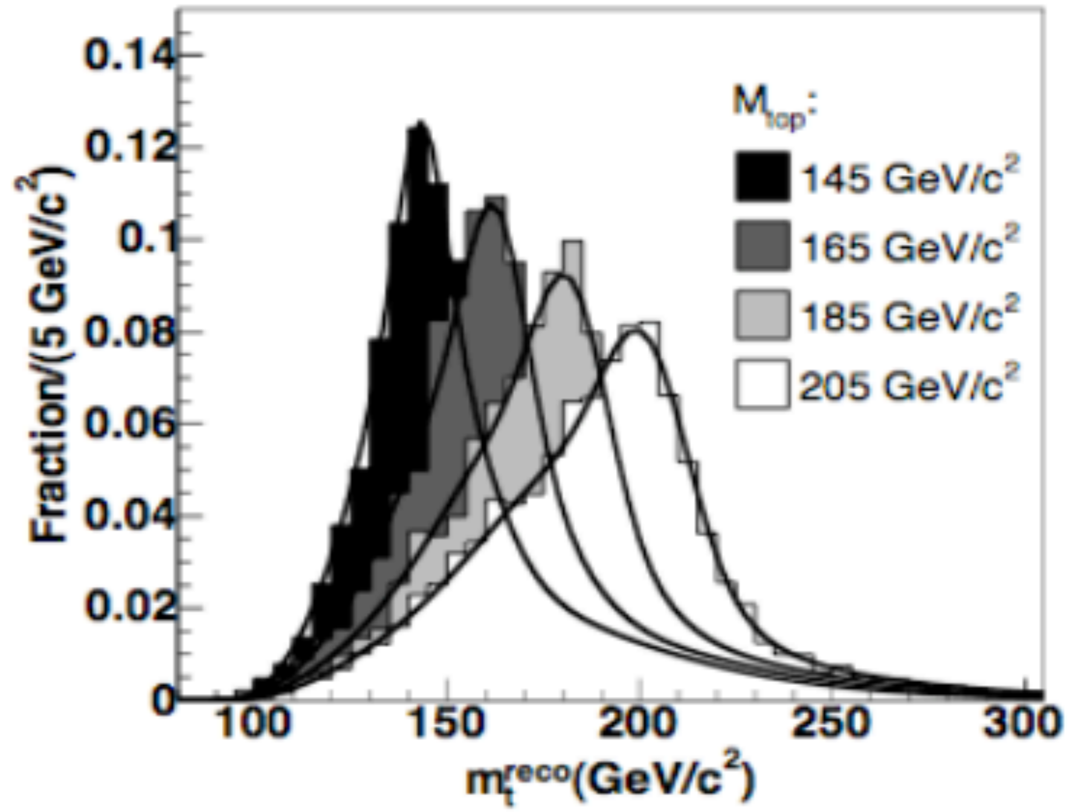


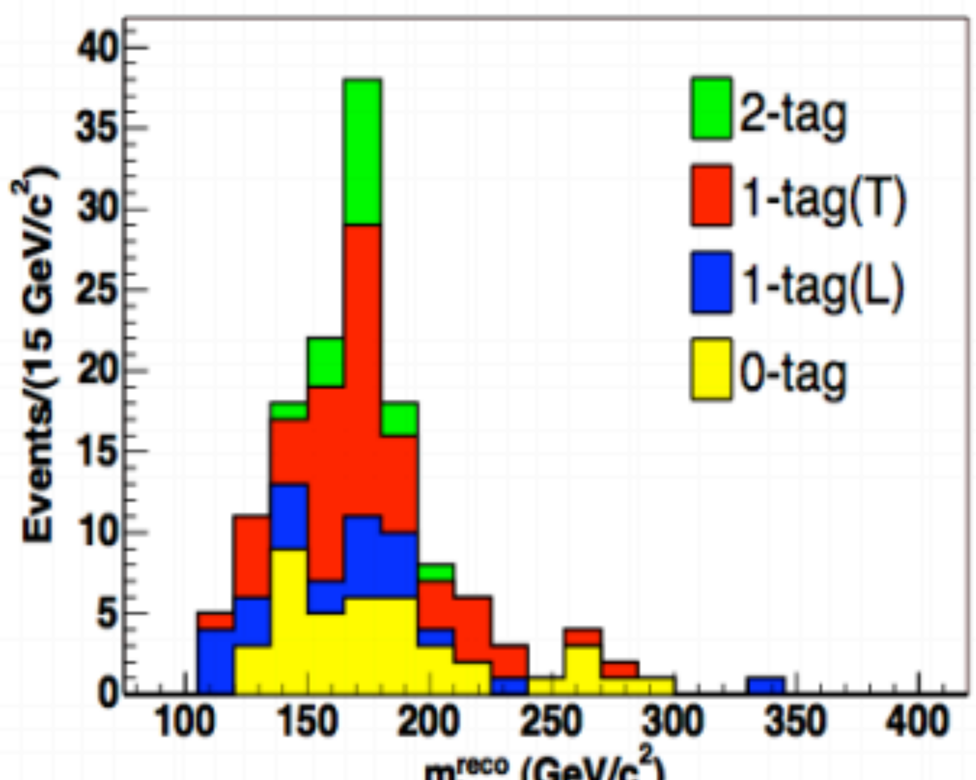
Template method nel canale lepton+jets

- Modeling degli eventi ttbar e del fondo tramite simulazioni MC
- Si genera un set di simulazioni MC a valori definiti della massa del top e della JES
- Si ottiene una buona stima della massa ***ricostruita*** del top e dei prodotti del W
- Per ogni campione un fit del $\chi 2$ estrae la massa ricostruita del top
- Questa distribuzione di m_{reco} (***template***) viene confrontata poi con la distribuzione dei dati tramite un likelihood fit

Parametrizzazione del segnale

- MC solo a valori discreti di Mtop: si ottengono delle forme funzionali dalle distribuzioni m_{reco} in funzione di Mtop (pdf's), costituite da due gaussiane e una gamma-dis.





Likelihood Fit

$$\boxed{L = L_{2tag} \times L_{1tagT} \times L_{1tagL} \times L_{0tag} \times L_{JES}}$$

- La massa ricostruita dai dati viene confrontata con le simulazioni e col fondo tramite un likelihood fit, in cui, per ogni sample:

$$L_{sample} = L^{m_t{}^{reco}}{}_{shape} \times L^{mjj^{reco}}{}_{shape} \times L_{n.ev.} \times L_{bg}$$

•Mtop/JES parametri liberi del fit

Sensibile a M_{top}

Sensibile a JES

Correl.#eventi Mtop/JES

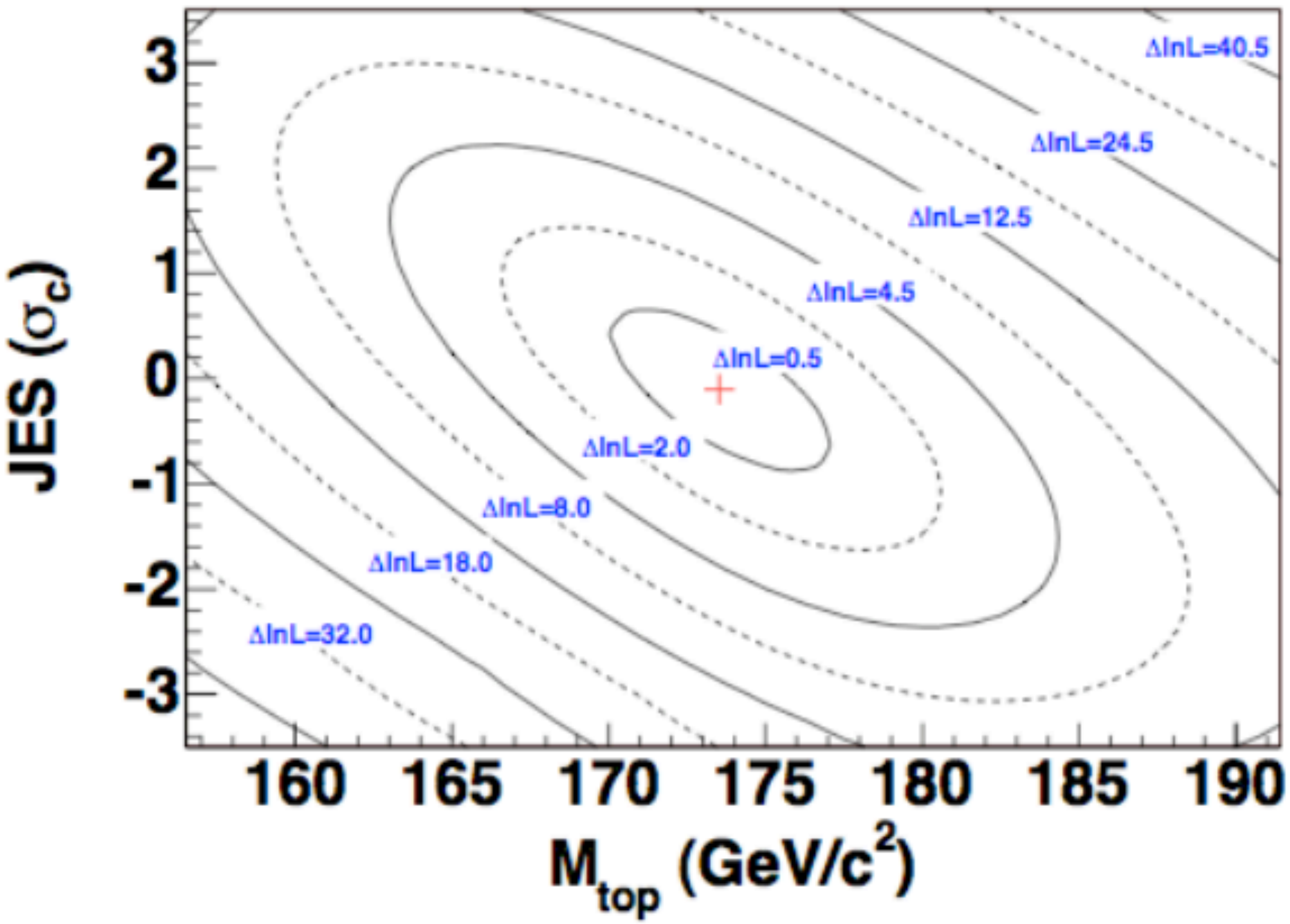
Normalizzazione del fondo

•L'errore statistico del fit è dato dai punti $M^{+/-}$ per cui **ΔlogL=-1/2**

•Per una serie di M_{top} fissati, la curva di L è massimizzata rispetto a tutti i suoi parametri

Method	M_{top} fit result [GeV/c^2]	JES fit result [σ_c]
Default	$173.5\ ^{+3.7}_{-3.6}$ (stat. + JES)	$-0.10\ ^{+0.78}_{-0.80}$
No JES constr.	174.0 ± 4.5 (stat. + JES)	-0.25 ± 1.22
M_{top}-only	$173.2\ ^{+2.9}_{-2.8}$ (stat.) $\oplus$ 3.1 (JES)	N/A
+ JPB	$173.0\ ^{+2.9}_{-2.8}$ (stat.) $\oplus$ 3.0 (JES)	N/A

La Massa del Quark Top



04/03/2006

CDF II detector @ Fermilab:

$\sqrt{s}$= 1,96 TeV

$\int L$ = 318 pb^{-1}

W boson in situ

$$M_{top} = 173.5^{+3.7}_{-3.6}(stat + JES) \pm 1.3(other\ syst) \quad GeV/c^2$$

$$= 173.5^{+3.9}_{-3.8}\, GeV/c^2 \quad .$$