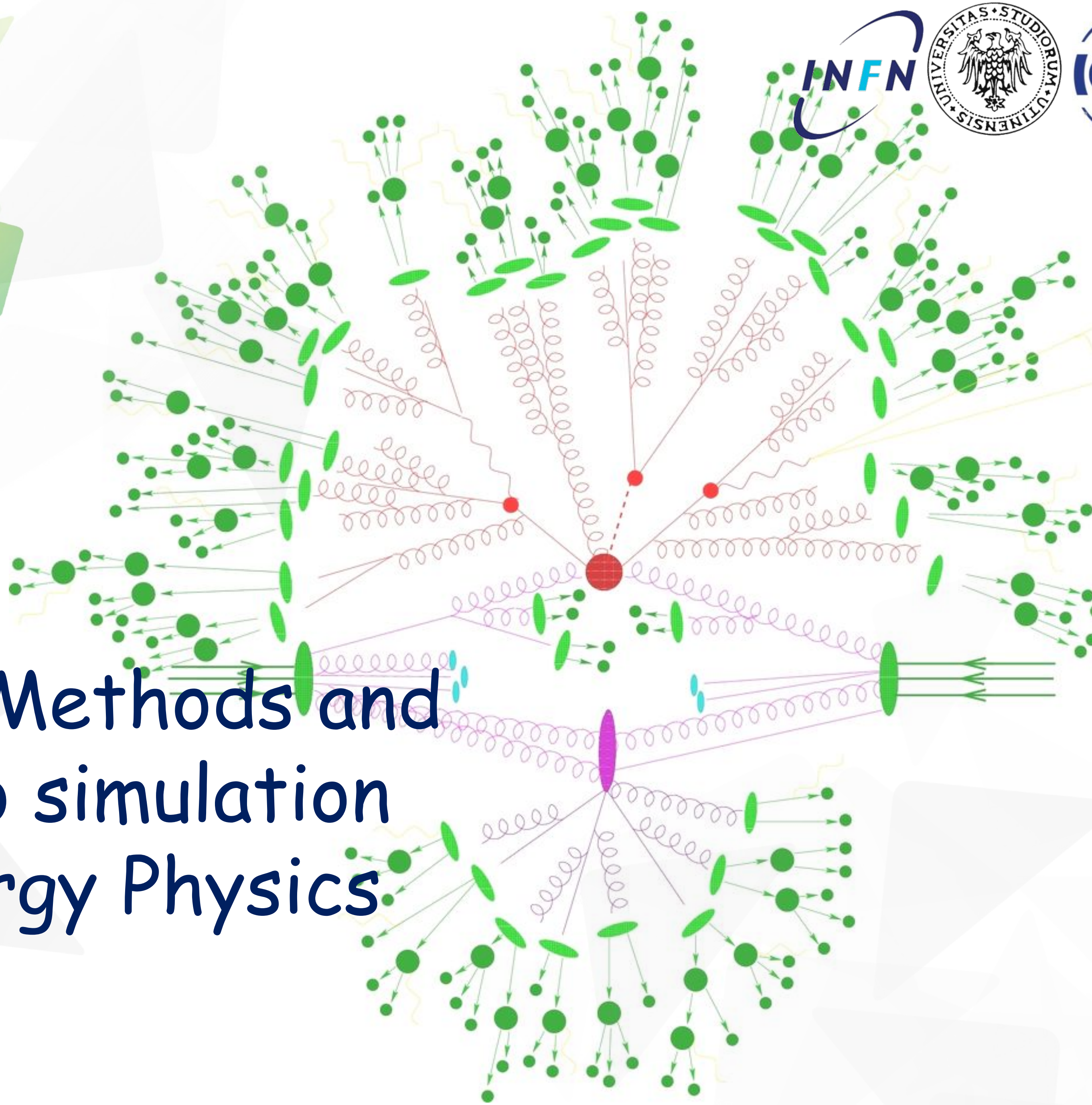


Statistical Methods and Monte Carlo simulation in High Energy Physics

Dr. Leonid Serkin (ICTP/Udine/CERN)

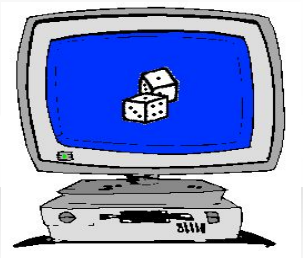
The Concept of Probability

- Many processes in nature have uncertain outcomes.

- A *random process is a process* that can be reproduced, to some extent, within some given boundary and initial conditions, but whose outcome is uncertain.

- For example, quantum mechanics phenomena have intrinsic randomness.

- *Probability* is a measurement of how *favored one of the possible outcomes of such a random* process is compared with any of the other possible outcomes.



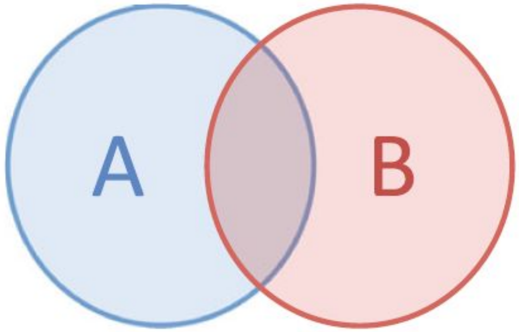
The Meaning of Probability: 2 approaches

- *Frequentist probability is defined as the fraction of the number of occurrences* of an event of interest over the total number of possible events in a repeatable experiment, in the *limit of very large number of experiments.*

- *Bayesian probability measures someone's degree of belief that a statement, and it makes u*se of an extension of the Bayes theorem: the probability of an event *A given the condition that the event B has occurred is given by:*

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

- The conditional probability is equal to the area of the intersection divided by the area if B



A Word on Simulation

- What a (computer) simulation does:
  - Applies mathematical methods to the analysis of complex, real-world problems
  - Predicts what might happen depending on various actions/scenarios

- Use simulations when
  - Doing the actual experiments is not possible
  - The cost in money, time, or danger of the actual experiment is prohibitive (e.g. nuclear reactors)
  - The system does not exist yet (e.g. an airplane)
  - Various alternatives are examined (e.g. hurricane predictions)

Why we need and have so much data at LHC?

An example for illustration

Correct dice

every number has probability 1/6

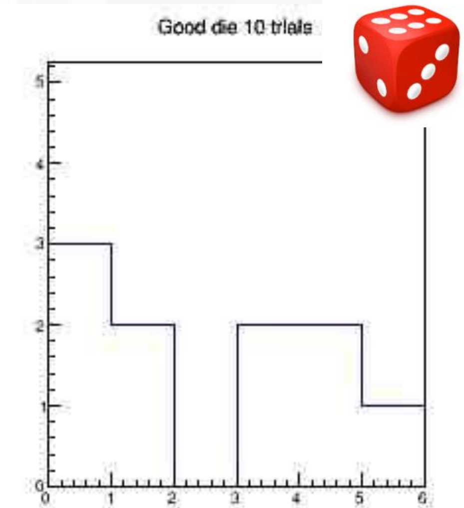
Manipulated dice

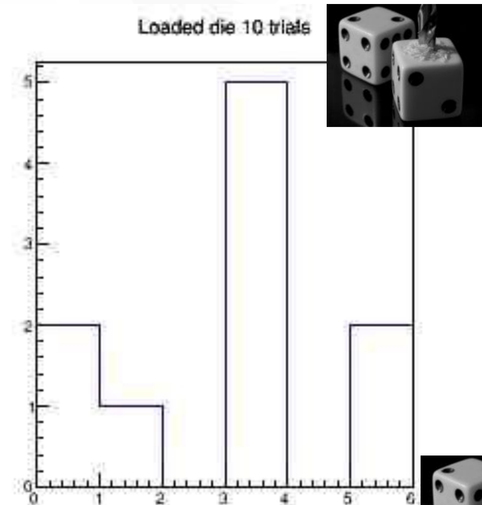
numbers 1..5 probability <1/6

number 6 probability > 1/6

Why we need and have so much data at LHC?

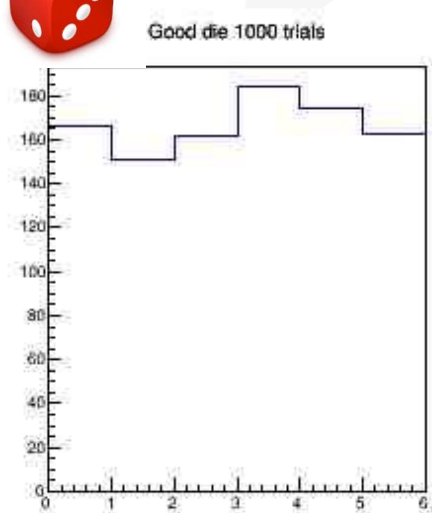
Role the dice and record the number in a bar chart

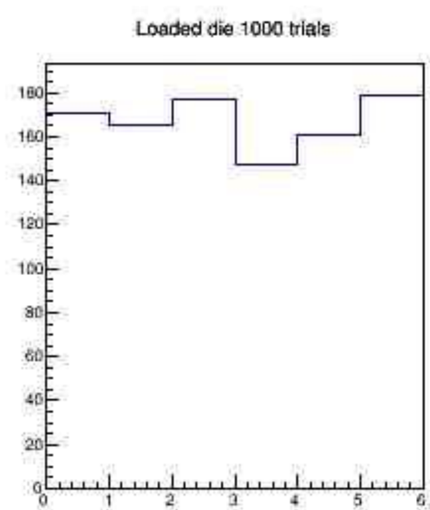




Just random fluctuations

… 10 times

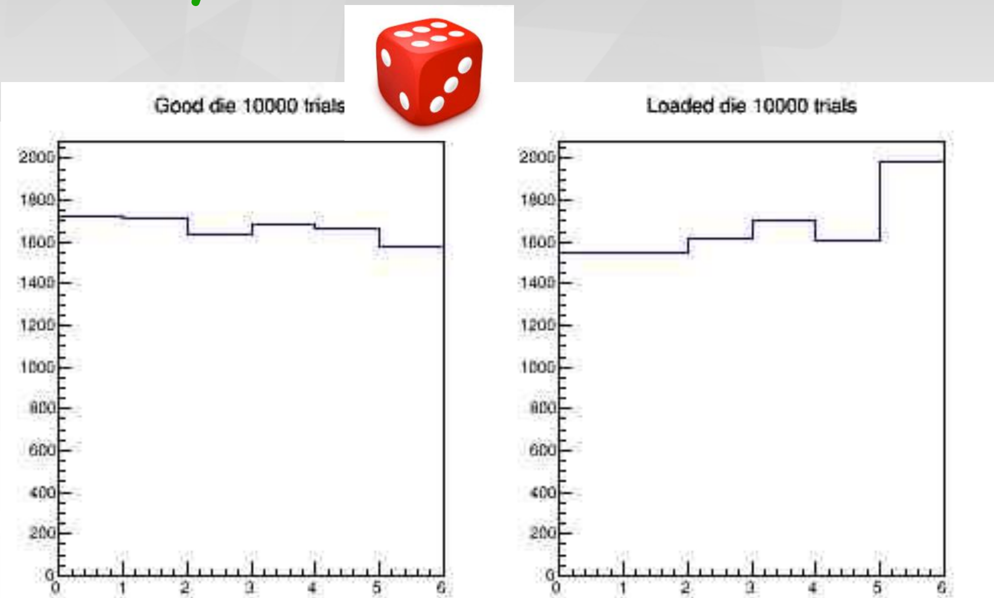




Still nothing can be concluded

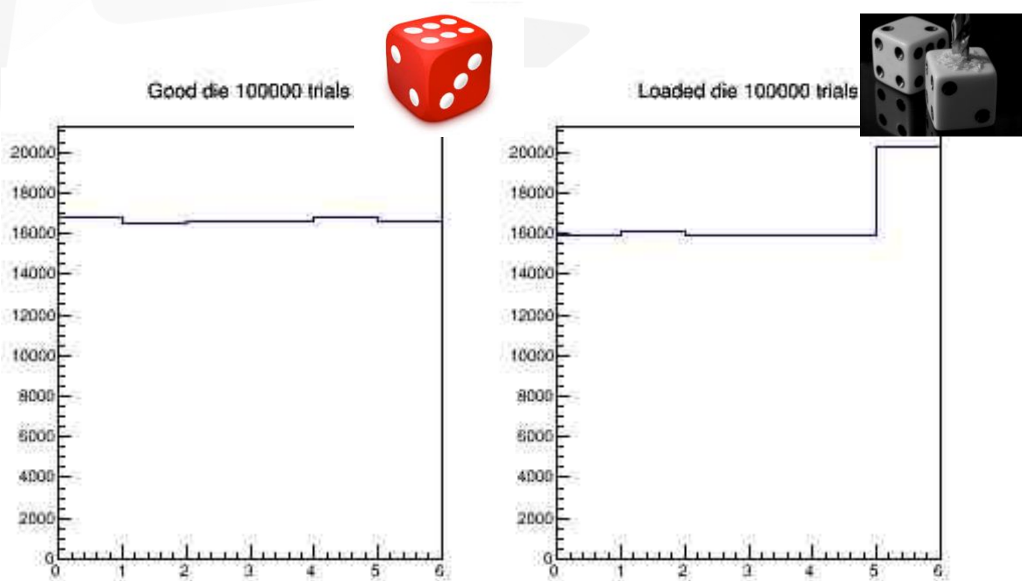
… 1000 times

Why we need and have so much data at LHC?



Evidence is rising …

… 10000 times



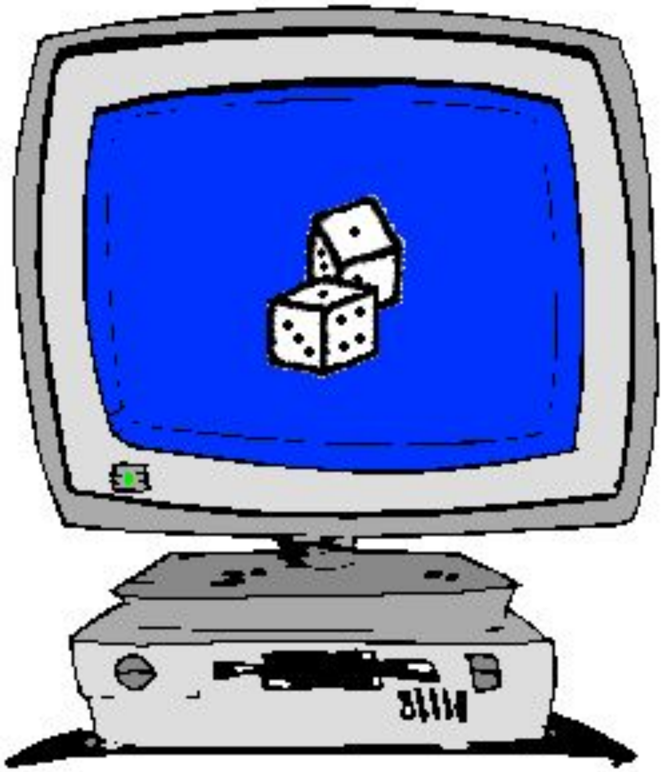
For sure there is something wrong with the dice

The more data you take the smaller your error gets (Gauss)

… 100000 times

Monte Carlo Method

- A numerical simulation method which uses sequences of random numbers to solve complex problems



What Monte Carlo does?

- MC assumes the system is described by probability density functions (PDF) which can be modeled with no need to write down equations
- These PDF are sampled randomly, many simulations are performed and the result is the average over the number of observations

Monte Carlo in High Energy Physics

- In HEP (in particular in hadron collider physics) MC are very useful:
  - To generate simulated collision events:
    - Quantum Field Theory obey probability laws
    - Proton PDF's have to be taken into account
    - Final state kinematical distributions with many alternatives (correlation of observables might be a problem...)
    - Complex soft and non-perturbative QCD (parton shower and hadronization)
  - To simulate the response of the detector:
    - Particle interaction with matter can be complicated
    - Huge number of different detector components

What to do with Monte Carlo events?

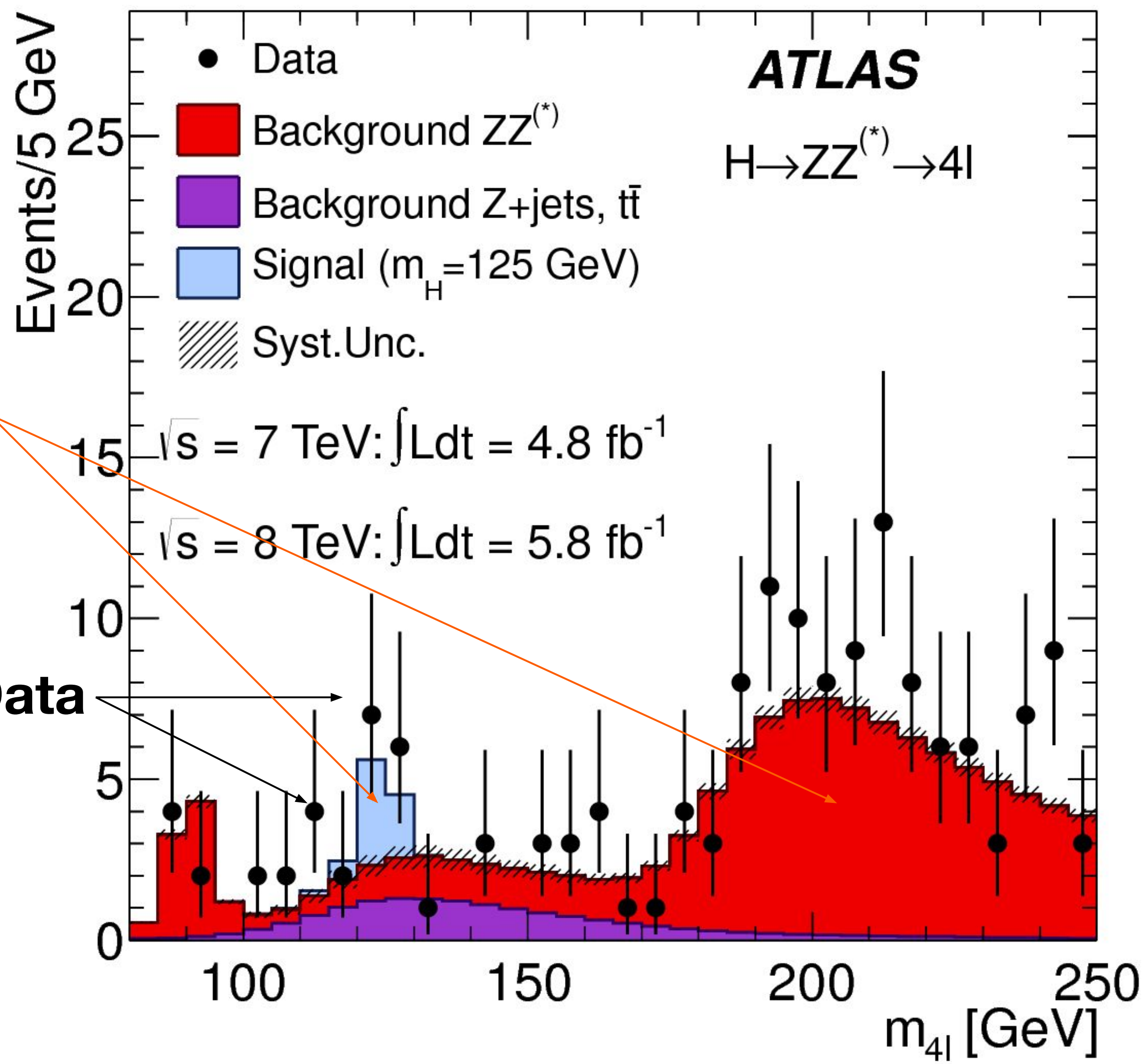
- To test performances:
  - Perform feasibility studies before looking at Data
  - Predict the performances of the detector

- To compare with real collision Data to extract physics results:
  - Background modeling
  - Signal selection efficiency (acceptance) determination

Example: Higgs discovery at ATLAS

Monte Carlo Simulation

Real Data



Collision Event Simulation

- Different steps are required:

Start by determining the hard process:

1) Choice of the interesting process to generate (start from a generic pp collision would be inefficient...)
2) Randomly generate kinematics of initial and final states (using PDF's for initial state)

Evolve the final state:

3) Decays of heavy particles according to BR's
4) Parton shower evolution
5) Hadronization of partons to form particles

Practical example: estimating the value of π using the Monte Carlo method

Practical example: estimating the value of π using the Monte Carlo method

- Q: How to estimate a value of π using the Monte Carlo method?

Practical example: estimating the value of π using the Monte Carlo method

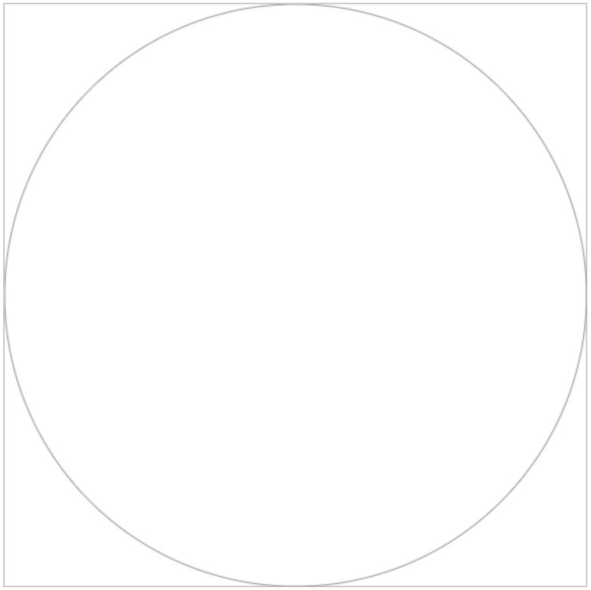
- Q: How to estimate a value of π using the Monte Carlo method?

- A: Generate a large number of random points and see how many fall in the circle enclosed by the unit square.

Practical example: estimating the value of π using the Monte Carlo method

- A: Generate a large number of random points and see how many fall in the circle enclosed by the unit square

- Build a circle of radius 0.5, enclosed by a 1 × 1 square. The area of the circle is: $\pi R^2 = \pi/4$

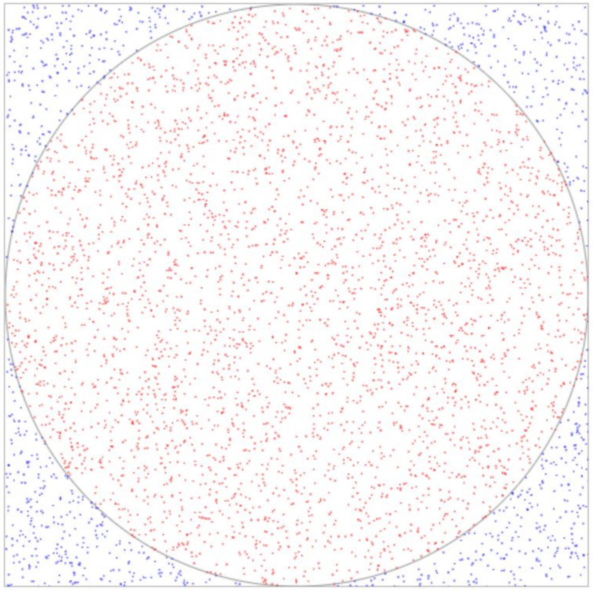


- The area of the square is 1.

- If we divide the area of the circle,
by the area of the square we get: $\pi/4$

Practical example: estimating the value of π using the Monte Carlo method

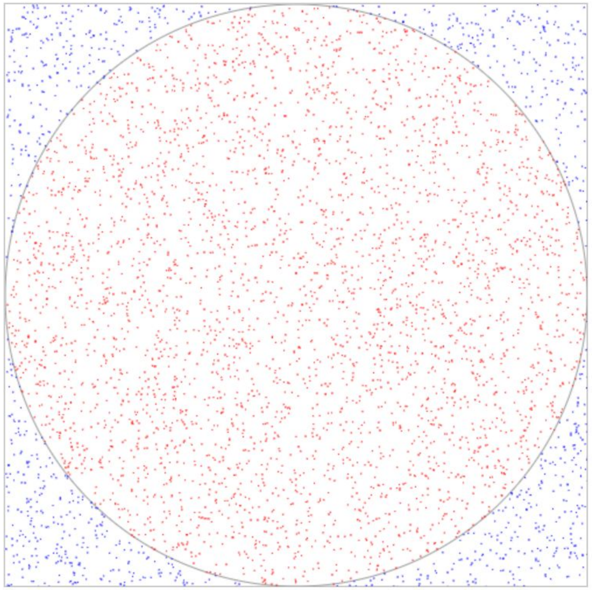
- Generate a large number of uniformly distributed random points and plot them on the graph. These points can be in any position within the square i.e. between (0,0) and (1,1).
- If they fall within the circle, they are coloured red, otherwise they are coloured blue.



Practical example: estimating the value of π using the Monte Carlo method

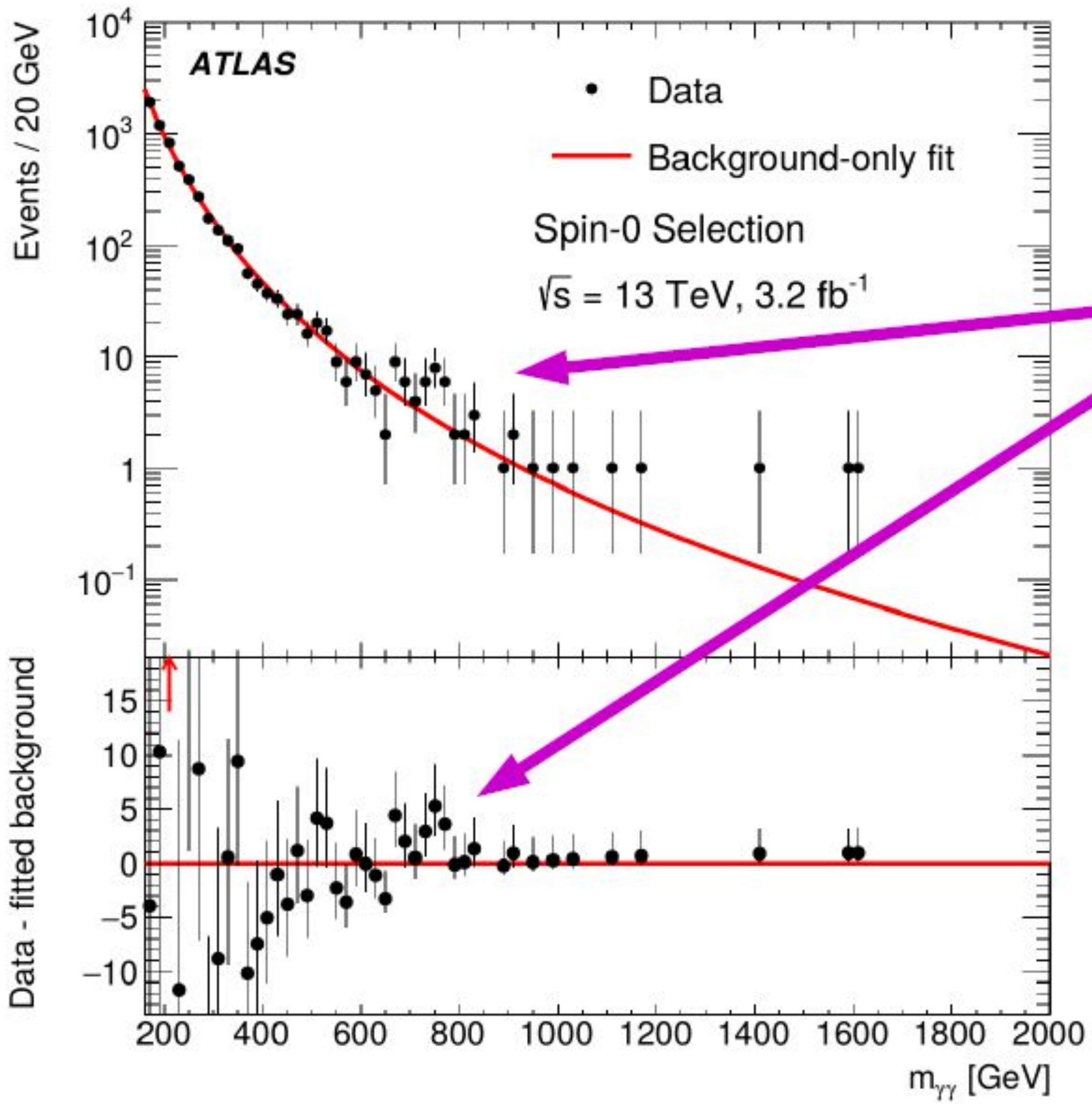
- We keep track of the total number of points, and the number of points that are inside the circle.
- If we divide the number of points within the circle, Ninner, by the total number of points, Ntotal, we should get a value that is an approximation of the ratio of the areas we calculated above, π/4

$$\pi \approx 4\frac{N_{inner}}{N_{total}}$$



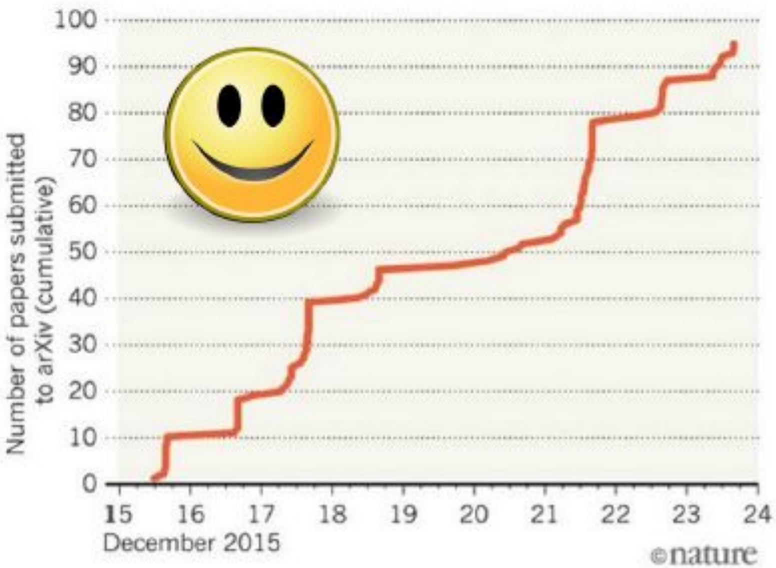
- With a small number of points, the estimation is not very accurate, but with thousands of points, we get closer to the actual value

A word on statistics



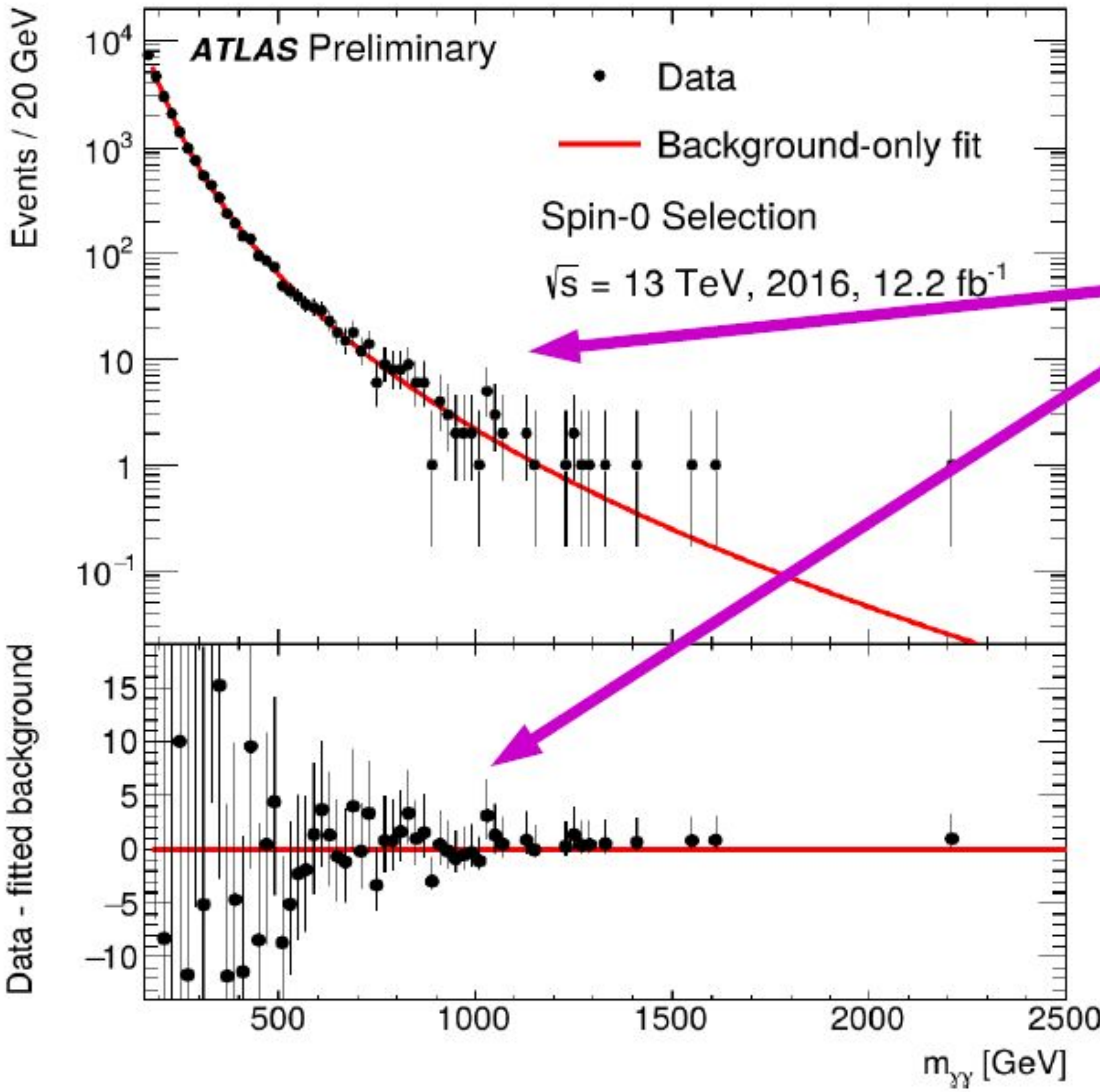
New Physics ?

3.9σ ? 2.1σ ?



JHEP 09 (2016) 1

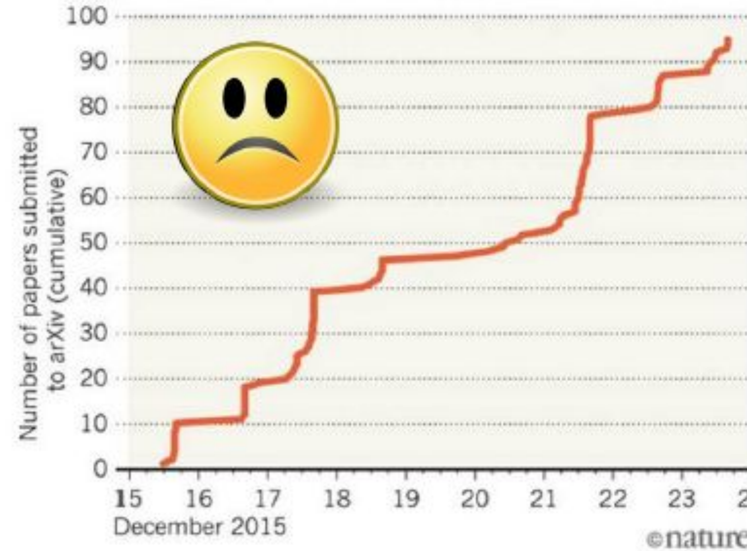
A word on statistics



A few months later...

~~**New Physics ?**~~

~~3.9σ ? 2.1σ ?~~



JHEP 09 (2016) 1

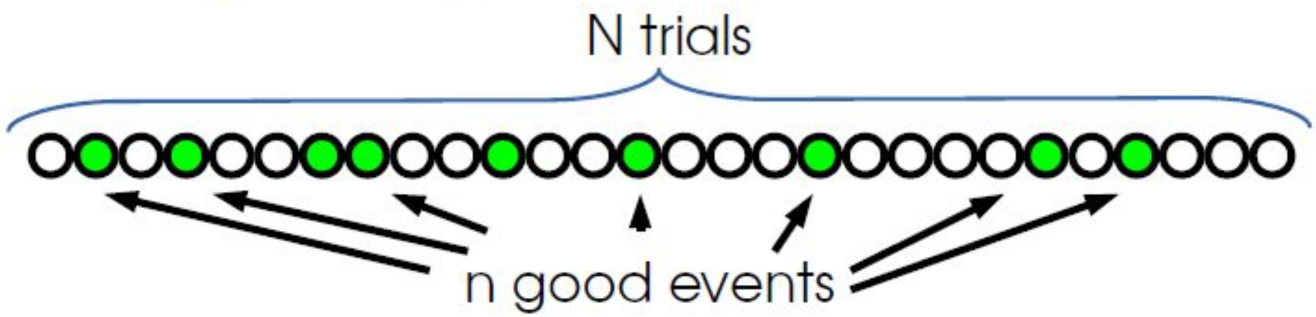
Counting events

Consider N total events, select ***good*** events with probability p.
Probability to get **n good events** ?

Binomial distribution : $$P(n;N,p) = C_N^n\, p^n (1-p)^{N-n}$$

Mean = **N·p**

Variance = **N·p(1 - p)**



However suppose **p ≪ 1, N ≫ 1**, and let **λ = N·p** :

→ *i.e.* **very rare** process, but **very many trials** so still expect to see good events

Poisson distribution: $$P(n;\lambda) = e^{-\lambda}\frac{\lambda^n}{n!}$$

Mean = **λ**

Variance = **λ** ⇒ **RMS = √λ**

$$(1-p)^{N-n} \overset{n \ll N}{\sim} \left(1-\frac{\lambda}{N}\right)^N \overset{N \gg 1}{\sim} e^{-\lambda}$$

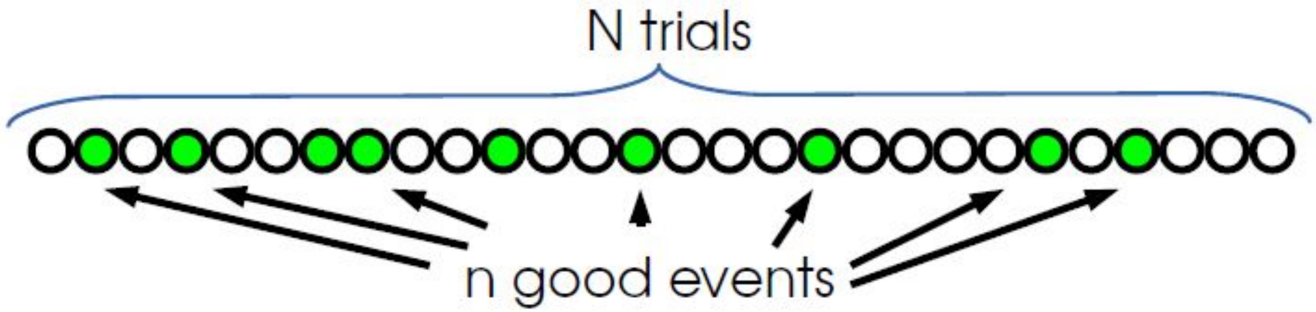
Counting events

Consider N total events, select ***good*** events with probability p.
Probability to get **n good events** ?

Binomial distribution : $$P(n;N,p)=C_N^n\, p^n (1-p)^{N-n}$$

Mean = **N·p**

Variance = **N·p(1 - p)**



However suppose **p ≪ 1, N ≫ 1**, and let **λ = N·p** :
→ *i.e.* **very rare** process, but **very many trials** so still expect to see good events

Poisson distribution: $$P(n;\lambda)=e^{-\lambda}\frac{\lambda^n}{n!}$$

Mean = **λ**

Variance = **λ** ⇒ **RMS = √λ**

$$(1-p)^{N-n} \overset{n\ll N}{\sim} \left(1-\frac{\lambda}{N}\right)^N \overset{N\gg 1}{\sim} e^{-\lambda}$$

For n expected events, the uncertainty is √n

Rare processes at the LHC

HEP : almost always use Poisson distributions. Why ?

ATLAS :

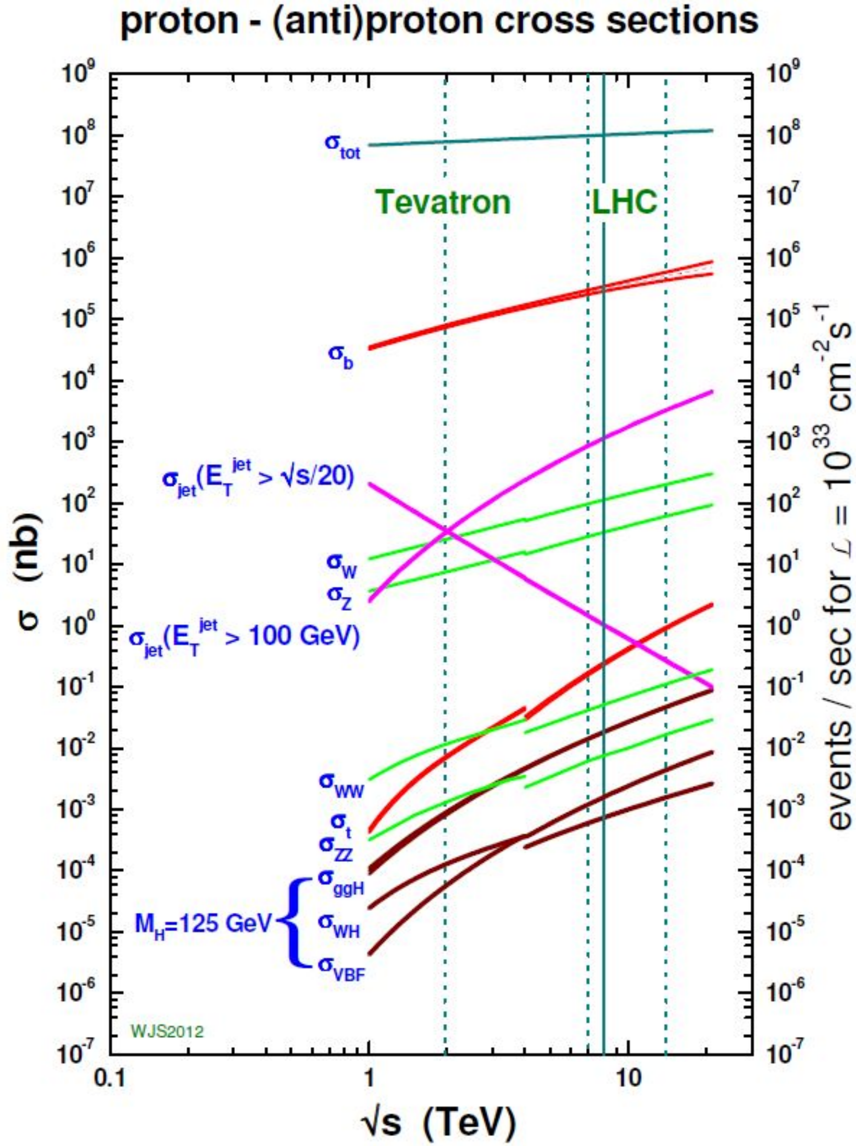
- **Event rate** ~ 1 GHz
  ($L \sim 10^{34}\ cm^{-2}s^{-1} \sim 10\ nb^{-1}/s$, $\sigma_{tot} \sim 10^8$ nb,)
- **Trigger rate** ~ 1 kHz
  (Higgs rate ~ **0.1 Hz**)

⇒ $p \sim 10^{-6} \ll 1$ ($p_{H\to\gamma\gamma} \sim 10^{-13}$)

A day of data: $N \sim 10^{14} \gg 1$

⇒ **Poisson regime!**

(Large N = design requirement, to get not-too-small λ=Np...)



Probability to find something new

In one year, the LHC provides ~10^{14} *pp* collisions

An observation of ~ 10 events could be a discovery of new physics.

Searching for a needle in a haystack?

- **typical needle: 5 mm^3**
- **typical haystack: 50 m^3**

needle : haystack = 1 : 10^{10}

Looking for new physics at the LHC is like looking for a needle in thousands of haystacks …

QUESTIONS