Karlsruher Institut für Technologie

Teilchenphysik 2 — W/Z/Higgs an Collidern

Sommersemester 2019

Matthias Schröder und Roger Wolf | Exercises No. 4

INSTITUT FÜR EXPERIMENTELLE TEILCHENPHYSIK (ETP)

Exercise 1

Imagine you are walking along the road but you get tired and consider hitchhiking. There are on average four cars per hour going in your direction. The chance that a car would stop and give you a lift is 5 %.

a) What is the probability that at least one car appears during the next hour?

b) Assume that four cars appear. What is the probability that you get a lift?

Exercise 1 — Solution

a) The appearance of cars can be assumed to be Poisson distributed with an average rate of $\lambda = 4$ cars per hour (counting experiment in intervals of hours). Then, the probability of at least one car is

$$P(x \geq 1; \lambda = 4) = 1 - P(x = 0; \lambda = 4) \approx 1 - 0.02 = 0.98$$

b) Here we are interested in the probability of $k \geq 1$ successes out of $n = 4$ trials if the success probability is $p = 0.05$ at each trial. This can be computed using Binomial statistics as

$$\begin{aligned} P(k \geq 1; n = 4, p = 0.05) &= 1 - P(k = 0; n = 4, p = 0.05) \\ &= 1 - \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \\ &= 1 - 1 \cdot 1 \cdot 0.95^4 \approx 0.19 \end{aligned}$$

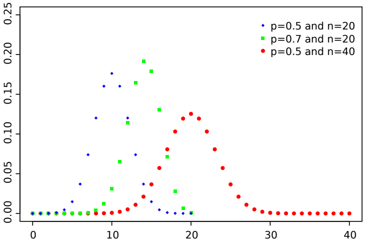
Important Probability Density Functions

- Statistical processes underlying the data are generally assumed to follow a probability density function (pdf) $\mathcal{P}$

  - **Binomial**:

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

e. g. drawing N times with fixed success probability



wikipedia

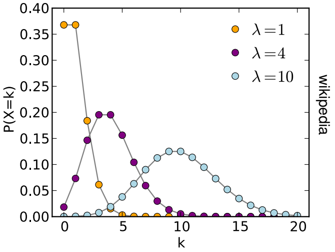
Important Probability Density Functions

- Statistical processes underlying the data are generally assumed to follow a probability density function (pdf) $\mathcal{P}$

- **Poisson**:

$$\mathcal{P}(x;\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}$$

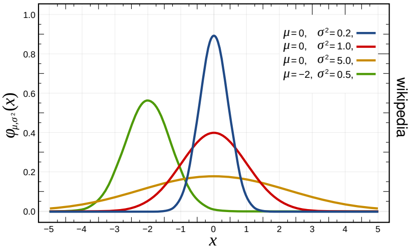
e. g. counting experiments
(like cross-section measurements)



- **Gaussian**:

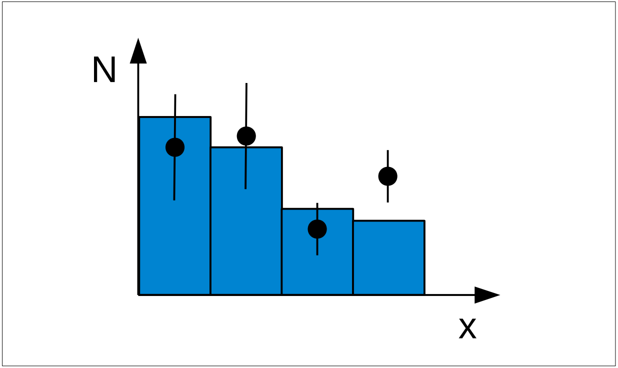
$$\mathcal{P}(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

e. g. parameter estimates
(like mass measurements)

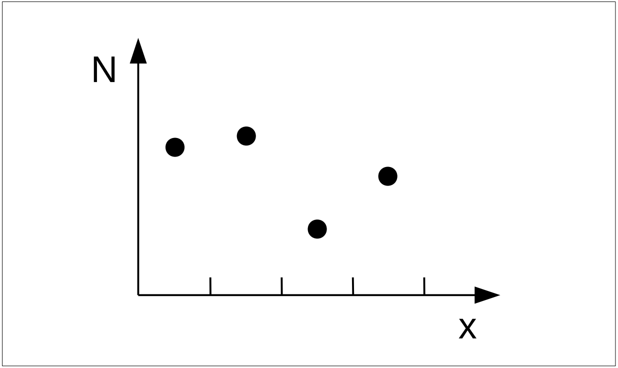


NB: both are in fact limits of the Binomial distribution for a large number of tries

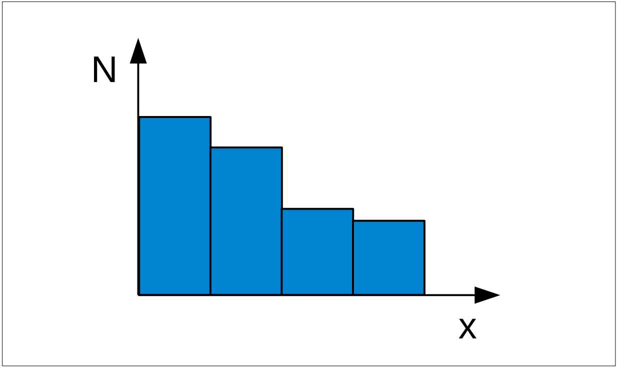
Histogram



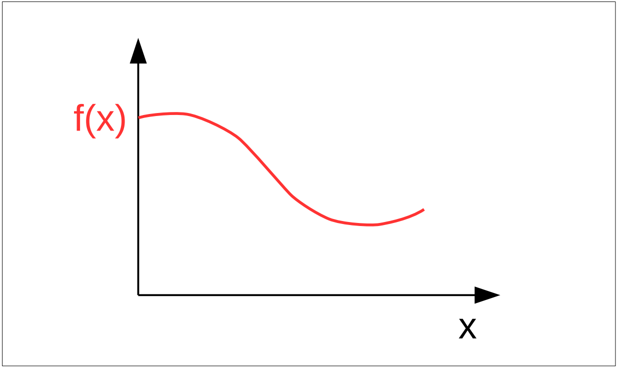
Histogram



Histogram

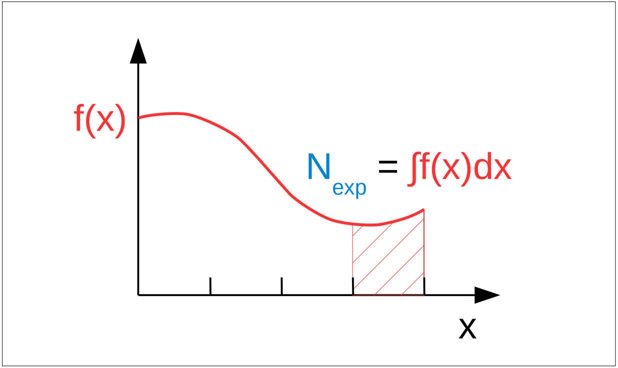


Histogram



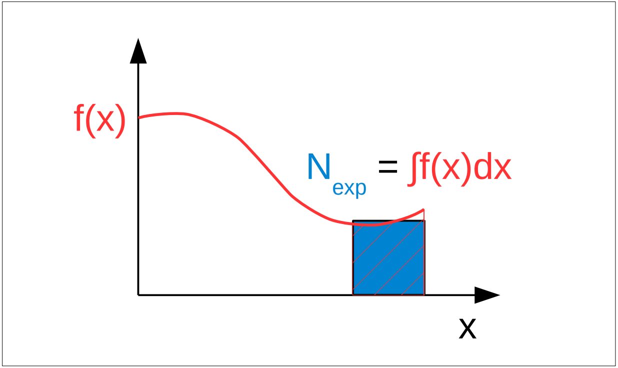
Probability density function f(x)

Histogram



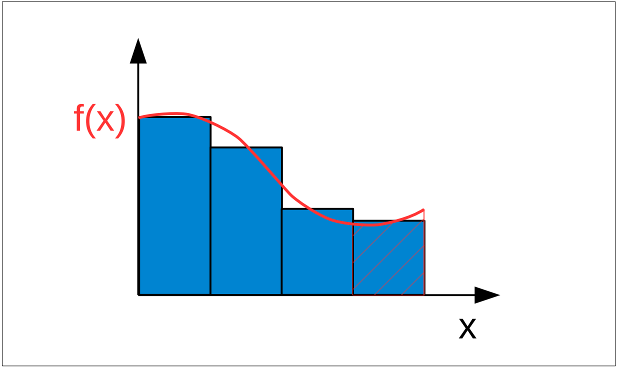
Probability density function f(x)

Histogram



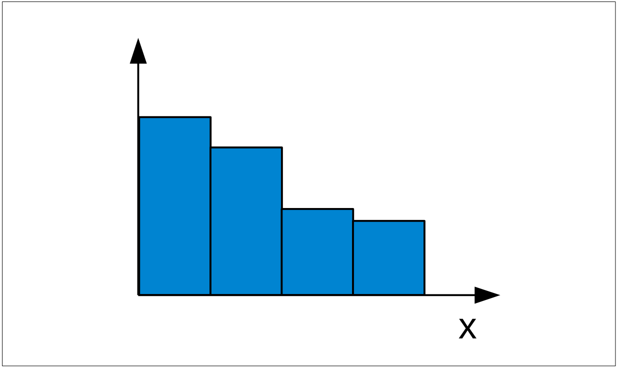
Probability density function f(x)

Histogram

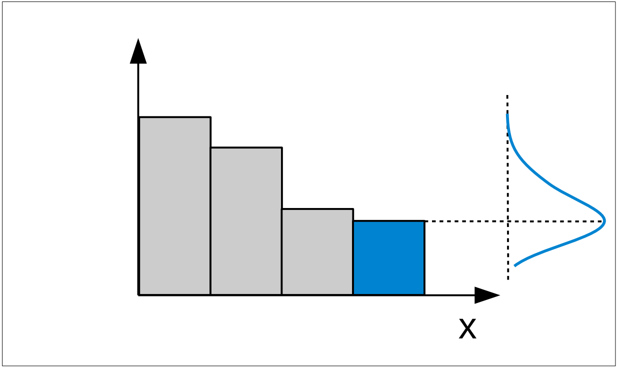


Probability density function f(x)

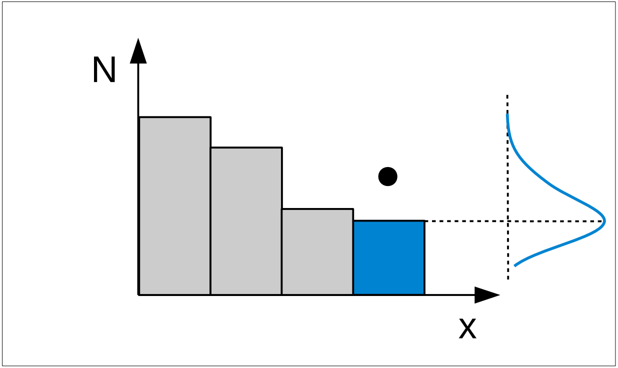
Histogram



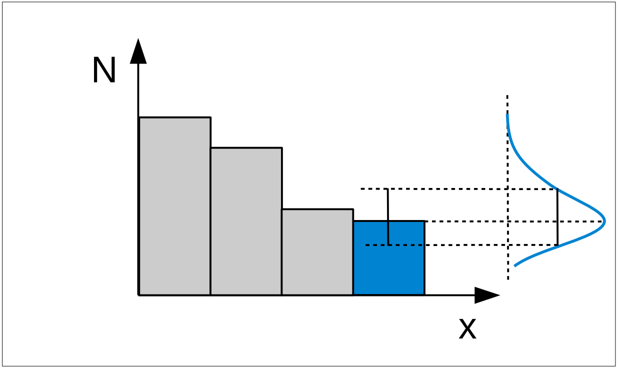
Histogram



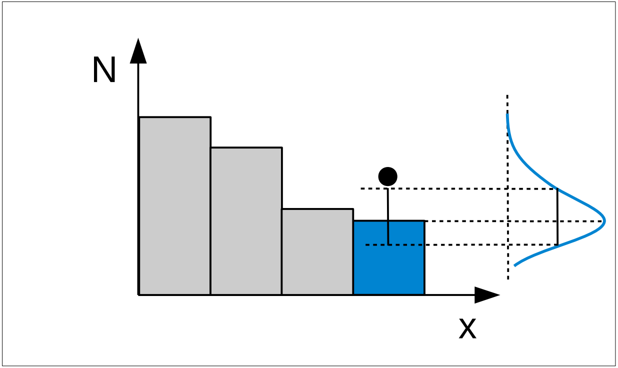
Histogram



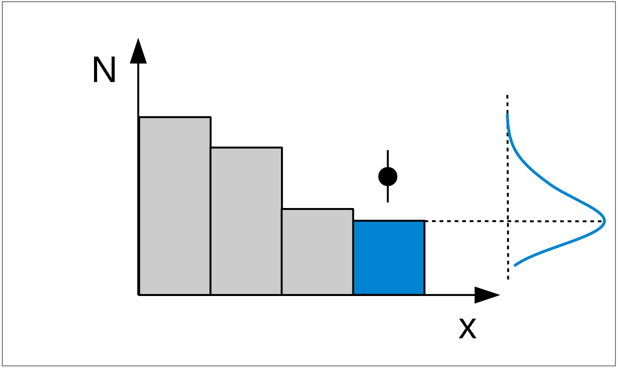
Histogram



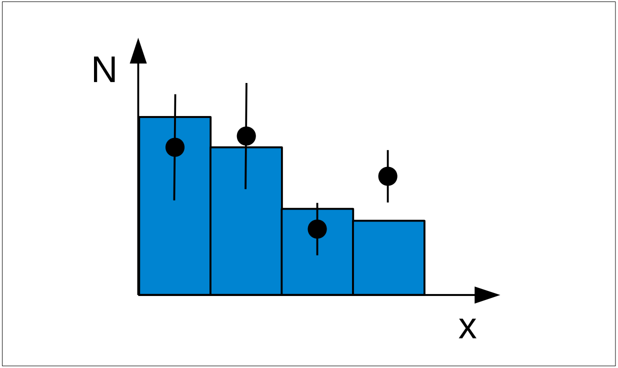
Histogram



Histogram



Histogram



Exercise 2

You are searching for events from a new-physics process in 1 fb^{-1} pp collision events recorded by CMS during 2018. The signal process is expected to have a cross section of 2 fb and feature two jets of an extremely large invariant mass above 10 TeV.

You select the events with invariant dijet mass > 10 TeV. In each of the following cases, did you discover new physics?

a) You observe 2 events.

b) You observe 0 events.

c) You observe 6 events, where you expect a background from Standard Model processes of 4 events on average.

Exercise 2 — Solutions

You expect 2 signal events. Did you discover new physics?

a) Yes (this is assuming you expect 0 background events!)

b) No, but you also did not rule out the new physics: the probability to observe 0 events at an expectation of 2 is 0.14 (Poission distribution).

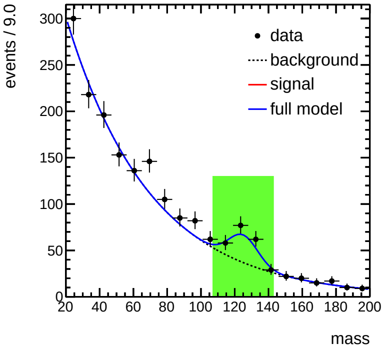
c) No, this could be signal or an upward fluctuation of the background. The probability to observe 6 background events is 0.10 (Poission distribution).

Significance

- In particle physics, usually counting experiments
- Allows often **simple approximation of significance** of an effect:

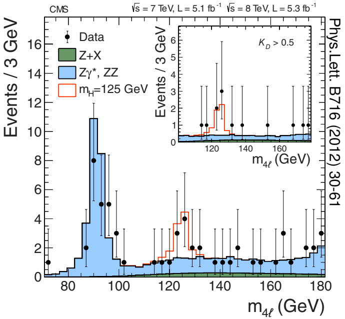
1. Total number b of expected background events in interesting region
2. Background-only hypothesis: expect data to be Poisson distributed with mean b and standard deviation $\sqrt{b}$
3. For large number of events: Poisson $\rightarrow$ Gaussian

$$Z = \frac{n_{\text{obs}} - b}{\sqrt{b}} \rightarrow \frac{S}{\sqrt{B}}$$



p Value

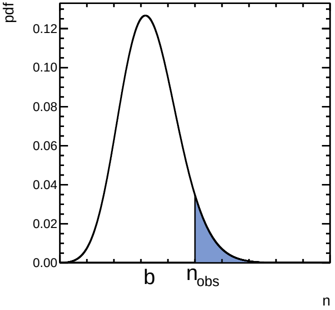
- How well is the data described by my hypothesis (model)?
- For example, **how likely that observed peak in data just upward fluctuation** of the background, i. e. no Higgs boson present?
- Assume simple counting experiment
  - b: expected number of background events (=model)
  - n_{obs}: number of observed events



p Value

- How well is the data described by my hypothesis (model)?
- For example, **how likely that observed peak in data just upward fluctuation** of the background, i. e. no Higgs boson present?
- Assume simple counting experiment
  - b: expected number of background events (=model)
  - n_{obs}: number of observed events
- "*p* value": **probability of upward fluctuation as large as or larger than observed in data**

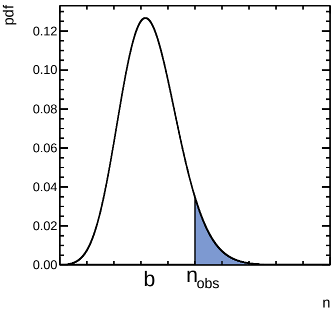
$$p \equiv \mathrm{P}(n \geq n_{\text{obs}} | b) = \int_{n_{\text{obs}}}^{\infty} \mathrm{d}n \, \mathcal{P}(n|b)$$



p Value

- How well is the data described by my hypothesis (model)?
- For example, **how likely that observed peak in data just upward fluctuation** of the background, i. e. no Higgs boson present?
- Assume simple counting experiment
  - b: expected number of background events (=model)
  - n_{obs}: number of observed events
- "p value": **probability of upward fluctuation as large as or larger than observed in data**

$$p \equiv \mathrm{P}(n \geq n_{\text{obs}} | b) = \int_{n_{\text{obs}}}^{\infty} \mathrm{d}n \, \mathcal{P}(n|b)$$

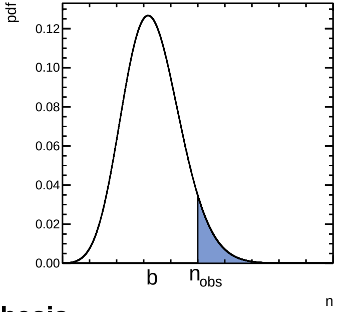


  - p depends on model and on data
  - If hypothesis is true, p is uniformly distributed between 0 and 1

p Value

- How well is the data described by my hypothesis (model)?
- For example, **how likely that observed peak in data just upward fluctuation** of the background, i. e. no Higgs boson present?
- Assume simple counting experiment
  - b: expected number of background events (=model)
  - n_{obs}: number of observed events
- "p value": **probability of upward fluctuation as large as or larger than observed in data**

$$p \equiv \mathrm{P}(n \geq n_{\text{obs}}|b) = \int_{n_{\text{obs}}}^{\infty} \mathrm{d}n\, \mathcal{P}(n|b)$$



p value is **not the probability of a hypothesis**

p quantifies level of (dis-)agreement between model and data:

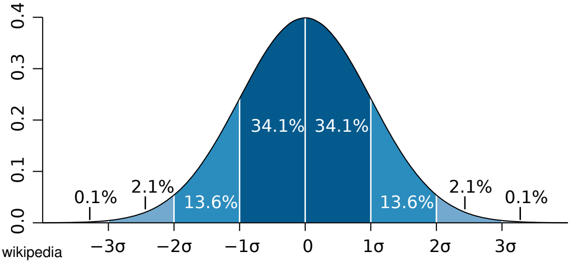
$\rightarrow$ judgement call whether to keep model or reject it

Significance

- Often, p value converted into equivalent **significance Z**: *upward fluctuation from 0 by Z of normal-distributed variable corresponding to same p value*
  - Corresponds to Z standard deviations σ of the Gaussian distribution

$$Z = \Phi^{-1}(1 - p)$$

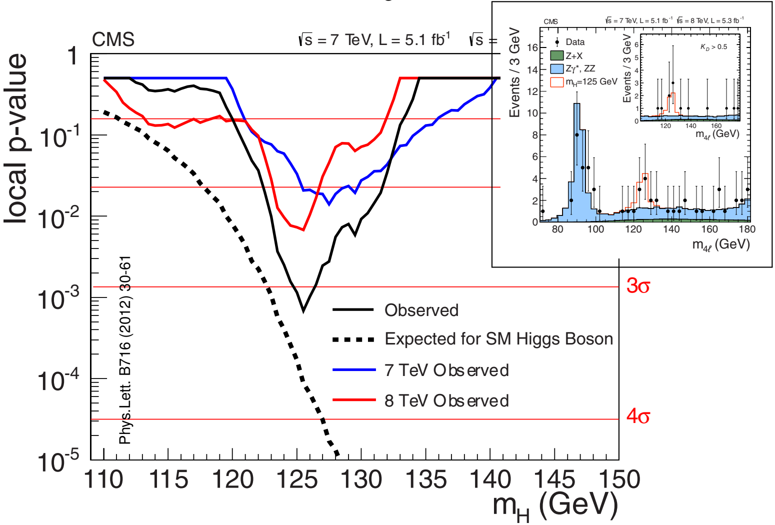
Φ: cumulative (=quantile) function of normal distribution



- **Convention** to classify effects by significance
  - 3σ: *evidence* for signal (0.3% chance of background fluctuation)
  - 5σ: *discovery* of signal (0.00006% chance of background fluctuation)

2012: The Higgs-Boson Discovery?

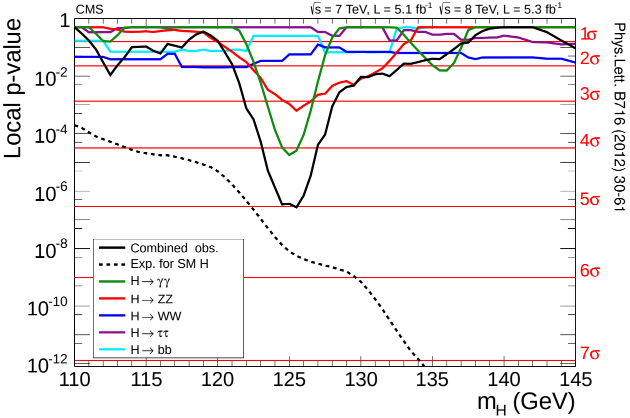
NB: by now much more data available, increased significance



$\approx 3\sigma$ significance in 4-leptons final state channel

2012: The Higgs-Boson Discovery!

NB: by now much more data available, increased significance



Different channels combined: 5σ significance

Still a $\approx 0.00006\%$ chance it is just a background fluctuation!

Exercise 3

You are searching for a new particle that is expected to decay into two photons. Thus, you select events with two photons and histogram their invariant mass using 40 equidistant bins.

a) What is the probability that you observe an excess of $2\,\sigma$ or more above the background expectation in exactly 1 bin just because of an upward fluctuation of the background?

b) What is the probability for at least one $2\,\sigma$ excess due to a background fluctuation?

c) What is the probability that you observe a $2\,\sigma$ excess in two adjacent bins?

d) Your colleague performs the same search at another experiment. What is the probability that she also measures a $2\,\sigma$ excess in the same mass bins?

Exercise 3 — Solutions

a) The probability of a $2\,\sigma$ fluctuation of the background is $p \approx 0.046$. Using Binomial statistics, the probability of exactly $k = 1$ excess by $2\,\sigma$ in any of the $N = 40$ bins is

$$P(k = 1; p = 0.046, N = 40) = N \cdot p^1 \cdot (1 - p)^{N-1} = 0.29\ .$$

b) Similarly, the probability at least one $2\,\sigma$ fluctuation is

$$P(k \geq 1; p = 0.046, N = 40) = 1 - P(k = 0) = 1 - 1 \cdot 1 \cdot (1 - p)^N = 0.85\ .$$

c) The probability of an excess in two adjacent bins with all other bins within $2\,\sigma$ can be computed as

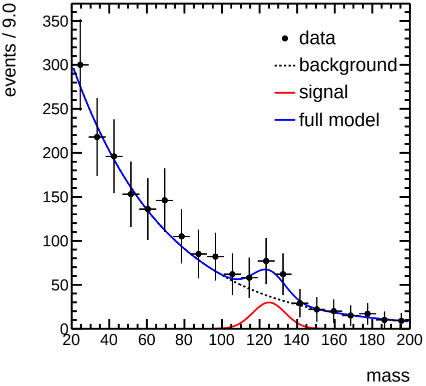
$$P(p = 0.046, N = 40) = 39 \cdot p^2 \cdot (1 - p)^{N-2} = 0.014\ .$$

The combinatorial factor is 39 because of the edges of the histogram.

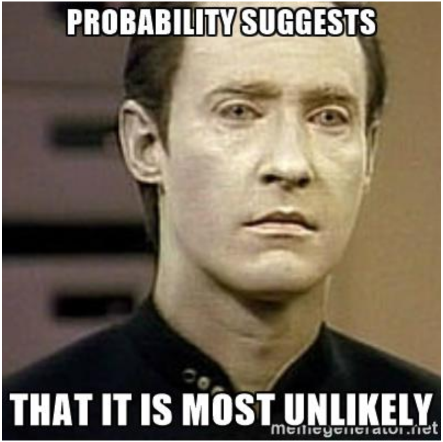
d) The probability of exactly one 2-bin excess in the same mass bins in both experiments follows as

$$P(p = 0.046, N = 40) = \underbrace{39 \cdot p^2 \cdot (1 - p)^{N-2}}_{\text{anywhere in exp. 1}} \cdot \underbrace{p^2 \cdot (1 - p)^{N-2}}_{\text{same two bins in exp. 2}} = 4.9 \cdot 10^{-6}\ .$$

What Is Wrong With This Histogram?



What Is Wrong With This Histogram?



Look Elsewhere Effect

- Example with invariant mass: we are searching for a deviation from the background-only model in any of the bins
  - i. e. we are performing several independent measurements
- Repeating measurement 3 times, expect one fluctuation by 1σ
- Repeating measurement 20 times, expect one fluctuation by 2σ
- Repeating measurement 330 times, expect one fluctuation by 3σ

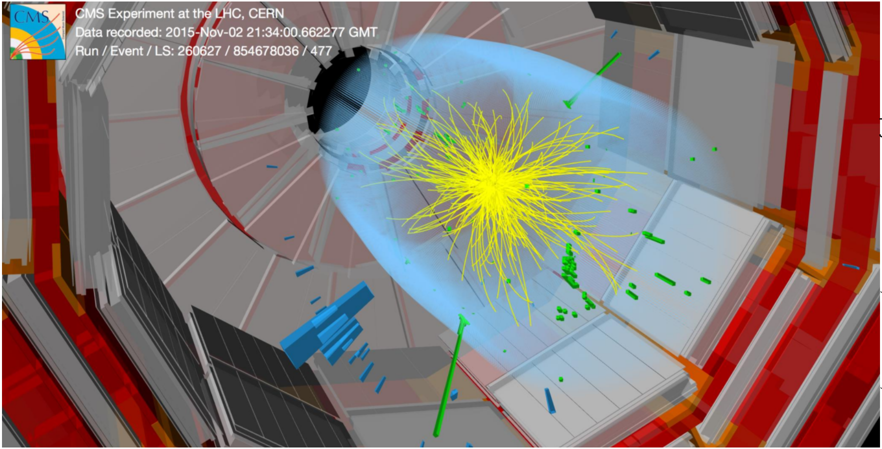
$\rightarrow$ "**look-elsewhere effect**"

  - Here we considered bins, but can be channels etc. as well
- Deviation in one bin/channel: **local significance**
- **Global significance** takes into account number of bins/channels

Look Elsewhere Effect

- Applies also more generally to e. g. LHC physics programme as a whole
  - Huge number of searches for new phenomena (and generally measurements), which are effectively a test of the SM hypothesis
- Expect significant deviations to appear by chance
- In fact, too few deviations!
  - Deviations not reported?
  - Experimentalists being overly conservative when assigning systematic uncertainties? [Eur. Phys. J. Plus 127 (2012) 157]

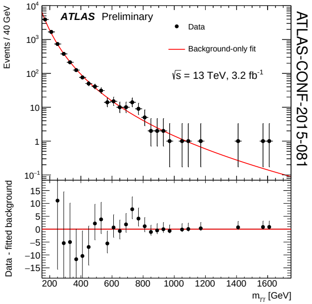
High-Mass Higgs Boson Search in $\gamma\gamma$

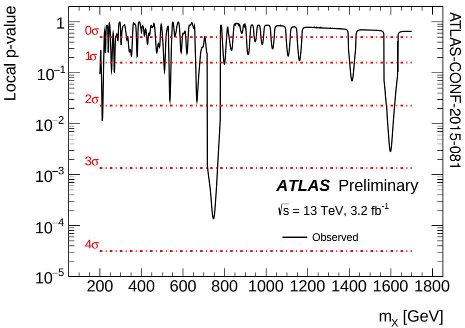


ATLAS and CMS physics results from Run 2, December 15, 2015

di-photon event with $m_{\gamma\gamma} = 745\,\text{GeV}$

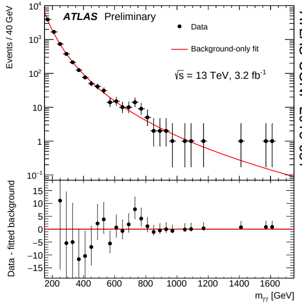
High-Mass Higgs Boson Search in $\gamma\gamma$

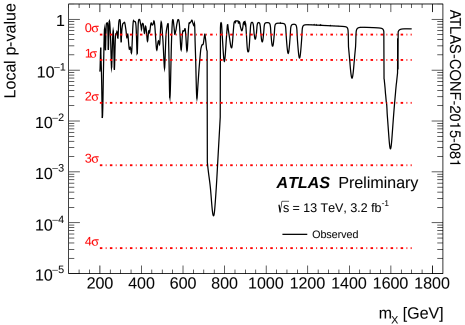




- With first LHC Run-II data at $\sqrt{s} = 13$ TeV: both ATLAS and CMS observe excess in $m_{\gamma\gamma}$ at ≈ 750 GeV
  (reported at CERN seminar on December 15, 2015)
  - ATLAS: 3.6 σ local significance ($1 : 10^4$), 2 σ global significance
    [ATLAS-CONF-2015-081]
  - CMS: 2.6 σ local significance, 1.2 σ global significance
    [CMS-PAS-EXO-15-004]

High-Mass Higgs Boson Search in $\gamma\gamma$

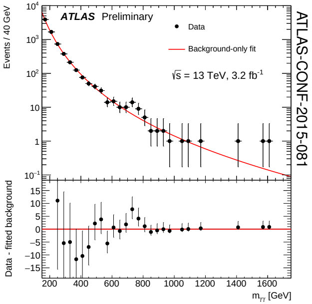


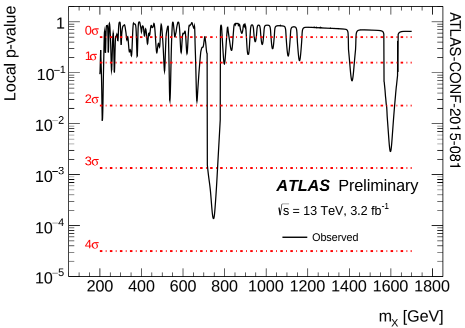


- Both experiments observed small excess around 750 GeV in LHC Run-I data at $\sqrt{s} = 8$ TeV
  - If $F(750)$ produced in gluon-gluon-fusion: $\sigma_F(8\,\text{TeV})/\sigma_F(13\,\text{TeV}) \approx \frac{1}{5}$
  - Amount of data: $\text{L}(8\,\text{TeV})/\text{L}(13\,\text{TeV}) \approx 8$

$\rightarrow$ naively expect higher significance in Run-I. But bkg. may grow less fast with $\sqrt{s}$ or signal have mild downward (upward) fluctuation at 8 TeV (13 TeV) (or heavier resonance decaying to 750 GeV particle)

High-Mass Higgs Boson Search in $\gamma\gamma$





- Both experiments observed small excess around 750 GeV in LHC Run-I data at $\sqrt{s} = 8$ TeV
- Some tension between 8 and 13 TeV data but compatible

→ **ATLAS+CMS Run I+II combined local significance of 4.4 σ**

A New Particle "F"?

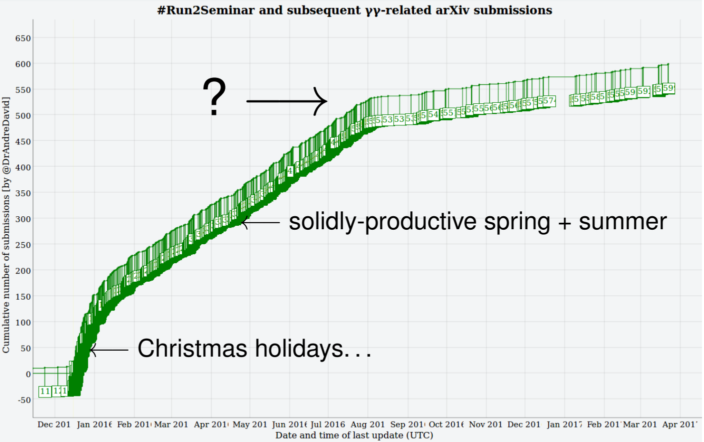
- Scalar particle, coupling to vector-like quarks in gluon-gluon fusion production and decay to $\gamma\gamma$
- Composite state bound by new strong interaction
- Kaluza-Klein graviton in Randall-Sundrum (extra dimension) models
- Additional Higgs boson of models with extended Higgs sectors (later)

…a tiny, tiny fraction of the proposed models

(review from Aug 2016: arXiv:1605.09401 [hep-ph])

Huge Attention by Theory Community

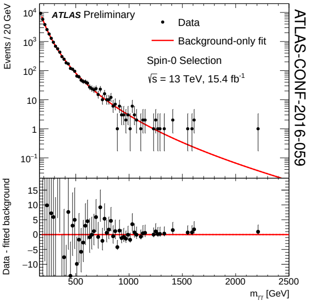
Cumulative number of related submissions to arXiv

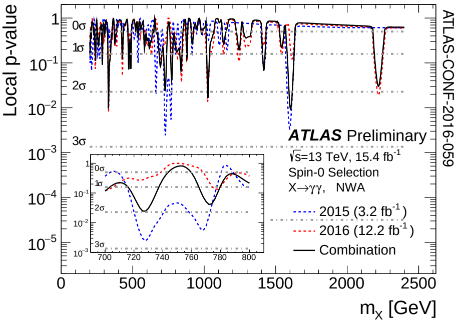


December 2015 — April 2017

http://jsfiddle.net/adavid/bk2tmc2m/show/

High-Mass Higgs Boson Search in $\gamma\gamma$





- Adding further $12\,\text{fb}^{-1}$ of data: significance decreased ($\approx 2.3\,\sigma$)
- Further decreased after analysing even more data of 2016

We were so close…
This is statistics at work