Basics in statistical analysis and hypothesis testing for physicists: Exercises

Michael Winn

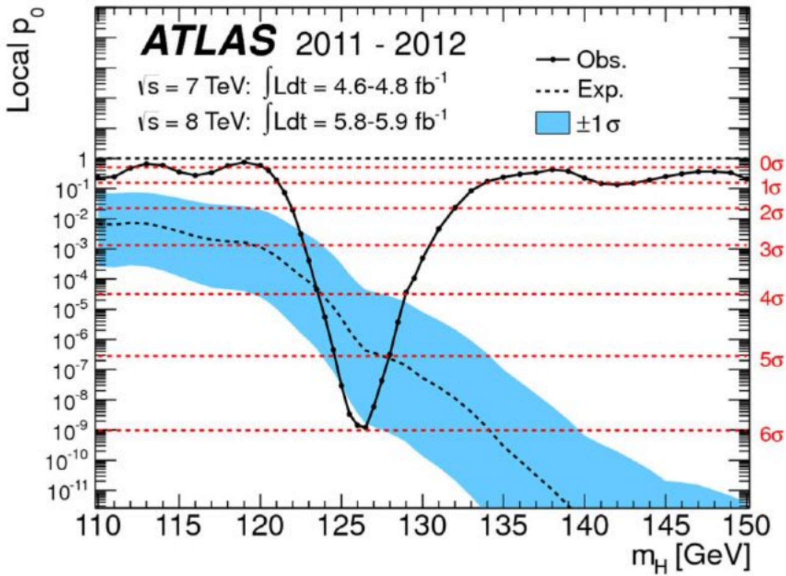
Department of Nuclear Physics IRFU/CEA, university Paris-Saclay

GRK 2149 retreat, 06.09.2023

Outline

- Warm-up: reading a discovery plot
- Ex 1: A published χ^2-fit
- Ex 2: Two cross sections and its ratio
- Ex 3: Simple exercise: How good is my model?
- Ex 4: Simple hypothesis test

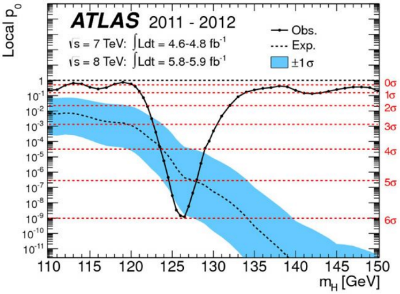
Warm-up



ATLAS 2011 - 2012
√s = 7 TeV: ∫Ldt = 4.6-4.8 fb⁻¹
√s = 8 TeV: ∫Ldt = 5.8-5.9 fb⁻¹
Obs.
Exp.
±1σ
Local p0
0σ
1σ
2σ
3σ
4σ
5σ
6σ
110
115
120
125
130
135
140
145
150
mH [GeV]

Warm-up

- Each Higgs mass hypothesis is scanned independently
- For each mass:
  - **Observed**: p-value observed in data
  - **Expected**: median of the p-value expected in the presence of the SM Higgs boson
  - Blue band: interval containing 68% of the p-values under SM Higgs hypothesis
- "Local" p_0
  - Many mass points scanned
  - **Look-elsewhere effect**: global p_0 to correct for number of trials



Ex 1: A published χ^2-fit

- a least χ^2-fit
- the reported χ^2/ndf is 1.72
- the distribution is a mass distribution containing known resonance peaks, no weights are used
- the content of each bin follows a Poisson distribution
- the data satisfy well the Gaussian approximation for a Poisson since in each bin $N_{yield,bin} >> 10$
- Is this a good fit based on the chi2-test?

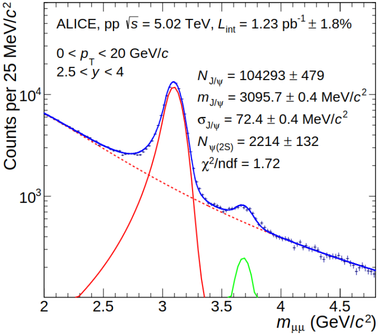
Ex 1

- we can´t tell without knowing the degrees of freedom (NDF)!
- NDF about 110
- Question: is this a good fit?

Ex 1

- for $N > 100$, χ^2-distribution approximately gaussian with mean NDF and variance $2 \cdot NDF$
- to reject the H_0, we choose a 5-sigma deviation
- sigma of Gaussian is 14.8
- we can exclude fit model at more than 5 σ
  $\rightarrow$ a new resonance in the fit, a discovery?

Ex 1



arXiv:2109.15240

- Reasons for the model doing so bad?
- Why accepted by collaboration/peer-review?

Ex 1

- Reasons: peak-parameters vary as function of integrated kinematics, would need more complicated model
- signal extraction one contributor to uncertainties, differential results available
  $\rightarrow$ certainly not ideal

Advice:

- think about it through before what makes sense
- visualise the data
- Always quote NDF if you do chi2-fits
- if the p-value of a test is very bad, but makes no sense due to systematic effects not taking into account of the model $\rightarrow$ identify the systematic effects, figure out their importance, correct them if possible

Ex2: Two cross sections and its ratio

- What is your expectation for the statistical uncertainty on a ratio vs. the uncertainties between numerator and denominator?
  $\rightarrow$ how would you guess the statistical uncertainty of the ratio between the two cross sections?

Ex2: Two cross sections and its ratio

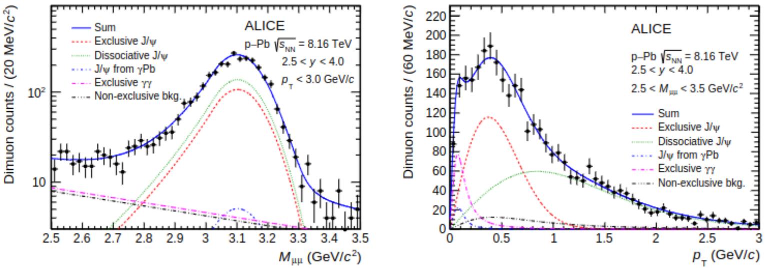


Figure 2: Projections of the two-dimensional fit on the dimuon invariant mass (left) and p_T (right).

- 2D unbinned maximum-loglikelihood fit, both cross sections from peak in the left, different shape on the right
- what does this tell you about the statistical uncertainty of the cross section ratio?

Ex2: Two cross sections and its ratio

Table 4: Ratio of dissociative-to-exclusive J/ψ photoproduction cross sections in p–Pb UPCs at $\sqrt{s_{NN}} = 8.16$ TeV. The first uncertainty is the statistical one. Its size is strongly impacted by the anti-correlation between exclusive and dissociative J/ψ components in the 2-dimensional fit. The second uncertainty is the systematic one. It is computed as the quadratic sum of the signal extraction ratio uncertainty, and the uncertainty on the V0C veto.

Rapidity range	$W_{\gamma p}$ (GeV)	$\langle W_{\gamma p}\rangle$ (GeV)	$\frac{\sigma(\gamma+p\to J/\psi+p^{(*)})}{\sigma(\gamma+p\to J/\psi+p)}$
(2.5,4)	(27,57)	39.9	$1.27\pm0.15\pm0.18$
(3.25,4)	(27,39)	32.8	$1.29\pm0.23\pm0.19$
(2.5,3.25)	(39,57)	47.7	$1.21\pm0.18\pm0.18$

- anticorrelation, e.g. for integrated cross section result:
  linear error propagation assuming independence would give 8.8%
  in reality considering anticorrelation 11.8%

Ex2: Two cross sections and its ratio

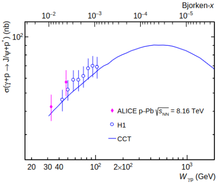
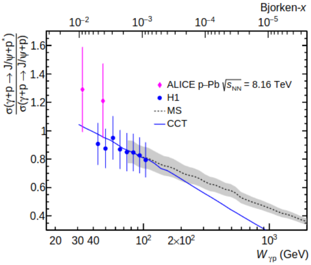


Figure 6: Dissociative J/ψ photoproduction cross section off protons measured by ALICE in p–Pb UPCs at $\sqrt{s_{NN}}$ = 8.16 TeV and compared with H1 data [40]. A comparison with the CCT model [37] is shown. The uncertainties of the data points are the quadratic sum of the statistical and systematic uncertainties.

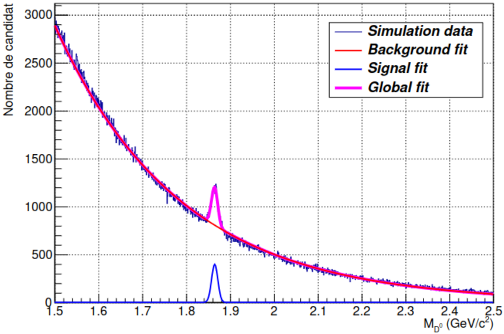


- can be very different in different experiments

Advice

- Provide correlation matrices for uncertainties if possible
- be careful with assumption transfer from one situation to another

Ex3: Simple exercise

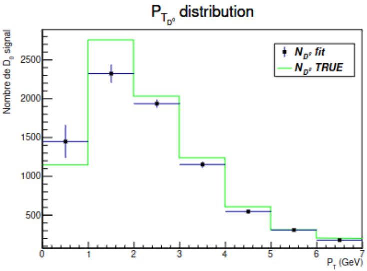


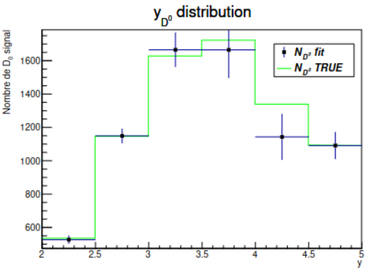
Result of a fast simulation

simple chi2-fit with exponential background, Gaussian signal; good fit quality
Side question: reasonable range and binning choice?

simulation: we know the truth!
let´s check

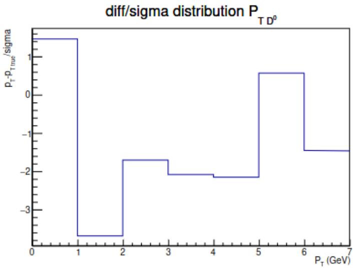
Ex3: Simple exercise

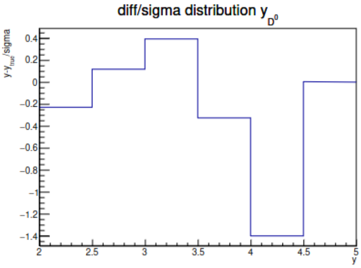




- is this worrisome?

Ex3 Simple exercise





- is this worrisome?
- what can we do to check:
  1) bias
  2) convergence
  3) correct coverage of fit?

Ex4: Simple example

Counting experiment for so far unobserved signal:

- D observed events: measurement outcome
- s expected signal events: 7.3
- b expected background events (known): 2.6

What is the correct model PDF?

Ex4: model PDF

A Poissonian distribution:

$$p(D|C)) = \frac{(C)^D e^{-(C)}}{D!}$$

Consider H_0: background only, $C = B_0$; H_1: background and signal $C = S + B_0$

Ex4: Parameter estimation

Use Maximum Likelihood to estimate signal counts provide b known and D observed events
$L(S) = p(D|S, B_0)$
$dL/dS = 0 => S = D - B_0$

Ex4: Simple example discovery

do we expect a 5σ discovery or a 3σ *evidence*?
when do we have a 5 σ discovery?
H_0: background only, $C = B_0$, H_1: signal according to simulation, $C = S + B_0$
Make simple hypothesis testing, without unspecified parameters,
see arXiv:1007.1727v3 for composite
Use the maximal likelihood ratio as test statistics due to Neyman-Pearson lemma

Ex4: The test statistics, likelihood ratio

Our data is one number: D
The parameter of interest, model expectation $\Theta = C$:
$C = B_0(H_0)$, $C = S + B_0(H_1)$

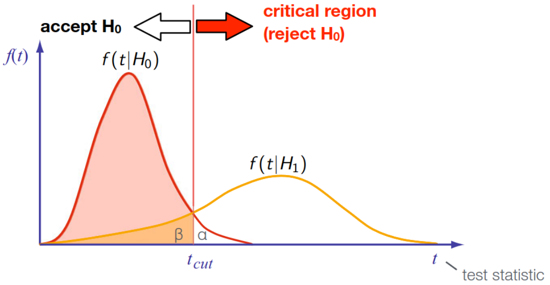
$$t(D) = \frac{L(D|H_1)}{L(D|H_0)} \tag{1}$$

$$= \frac{(S+B)^D e^{-(S+B)} D!}{D! B^D e^{-B}} \tag{2}$$

$$\propto \left(\frac{S+B_0}{B_0}\right)^D \tag{3}$$

significance of test: $2.9 \cdot 10^{-7}$ (5σ), what do we need to know determine?

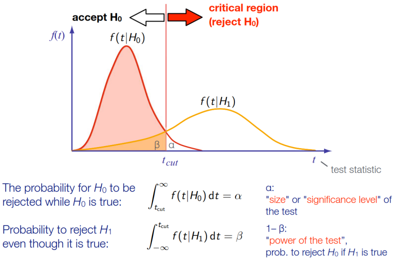
Ex4: critical region

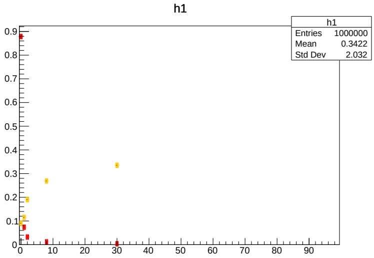


The probability for H_0 to be rejected while H_0 is true:	$\int_{t_{cut}}^{\infty} f(t\|H_0)\,dt = \alpha$	α: "size" or "significance level" of the test
Probability to reject H_1 even though it is true:	$\int_{-\infty}^{t_{cut}} f(t\|H_1)\,dt = \beta$	1− β: "power of the test", prob. to reject H_0 if H_1 is true

we need $f(t|H_1)$ and $f(t|H_0)$

Ex4: critical region



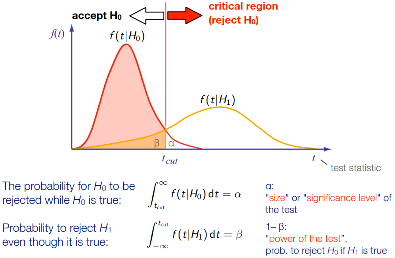


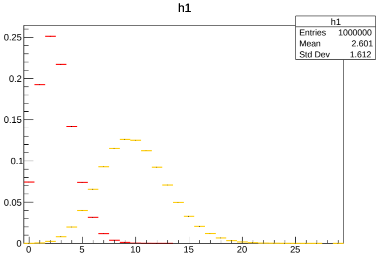
Example: light-by-light scattering, DOI:10.1038/NPHYS4208

$B_0 = 2.6$, $S = 7.3$ (expectation, neglect uncertainties)

$ln\frac{L(D|H_1)}{L(D|H_0)} = D \cdot const$: Take rather D as a t

Ex4: critical region





Example: light-by-light scattering, DOI:10.1038/NPHYS4208

$B_0 = 2.6$, $S = 7.3$ (expectation, neglect uncertainties)

$\rightarrow t = D$

t_{cut} for 5 σ: $D = 16$

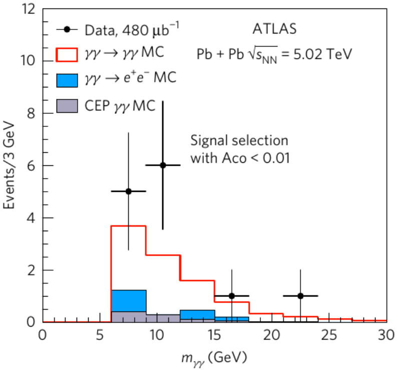
t_{cut} for 3 σ: $D = 9$

$\rightarrow$ we expect evidence, but no discovery

Ex4: Simple hypothesis testing

Example: light-by-light scattering evidence,
DOI:10.1038/NPHYS4208
$B_0 = 2.6$, $S = 7.3$ (expectation)
Observation: $D = 13$:
evidence, but no discovery with our t

Ex4: Simple hypothesis testing



Example: light-by-light scattering evidence, DOI:10.1038/NPHYS4208

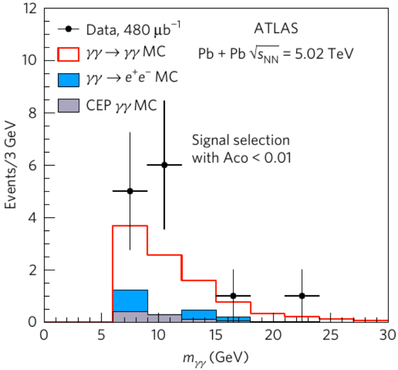
$B_0 = 2.6$, $S = 7.3$ (expectation)

Observation: $D = 13$

p-value with $t = D$ neglecting uncertainties:

$\approx 2 \cdot 10^5$

Ex4: example



To quantify an excess of events over the background expectation, a test statistic based on the profile likelihood ratio[48] is used. The p value for the background-only hypothesis, defined as the probability for the background to fluctuate and give an excess of events as large or larger than that observed in the data, is found to be 5×10^{-6}. The p value can be expressed in terms of Gaussian tail probabilities, which, given in units of standard deviation (σ), corresponds to a significance of 4.4σ. The expected p value and significance (obtained before the fit of the signal-plus-background hypothesis to the data and using standard model predictions from ref. 28) are 8×10^{-5} and 3.8σ, respectively.