Statistical Analysis in HEP and Beyond

Chicago HEP Lunch Seminar

Maximilian Swiatlowski[1]
[1] *Enrico Fermi Institute, University of Chicago*

October 31, 2017

Our Goals

- Statistics and data interpretation is a huge topic
- Very hard to do it justice in just a few minutes
- So we instead will have a very limited set of goals today:
  - Explain the basics of stats as used in HEP (and other areas of physics)
  - Show you some tools we use in ATLAS for statistical analysis of counting experiments
  - Hope you get a slightly better sense of what tools are available for your stats analysis!
- NB: slides will be in large part borrowed from S. Pataraia and W. Verkerkes: many thanks to them!

The Basic Question

- We have a dataset. We want to be able to answer the questions:

Is our data compatible with our background expectation?

Is our data compatible with signal?

- This is a super general question: many ways to answer this!
- Our main approach today is frequentist, using profile likelihoods
  - Many other approaches are possible! Advantages/disadvantages to each... Only time to cover one approach, though!

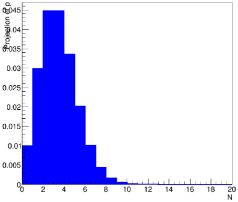
Simple statistical example

- Central concept in statistics is the 'probability model' : assigns a probability to each possible experimental outcome

- Example: a HEP counting experiment
  - Count number of events in your signal region (SR) in your data (specific lumi): Poisson distribution $P(N|\mu) = \frac{\mu^N e^{-\mu}}{N!}$
  - Given the *expected(MC)* event count, the probability model is fully specified

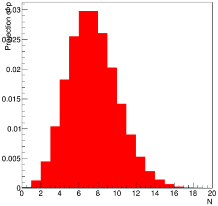
PROBABILITY

μ=3 ("bkg only")

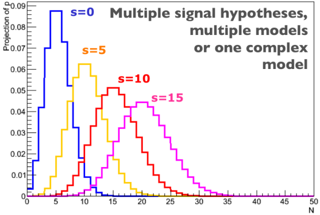


Poisson(N| b)

μ=7 ("bkg+signal")



Poisson(N| s + b)



Poisson(N| s + b)

- Suppose we measure N = 7 events (Nobs), then can calculate the probability
- P(Nobs|hypothesis) is called **LIKELIHOOD** - **L(Nobs|b), L(Nobs|s+b), L(observed data|theory)**

p(Nobs|b) = 2.2% **p(Nobs|s+b) = 14.9%**

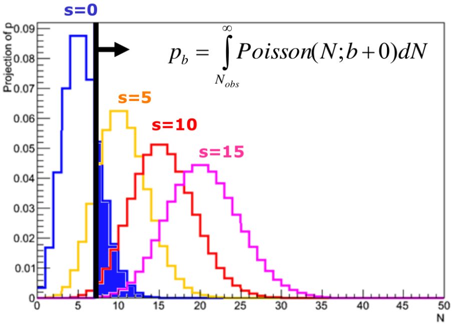
  - Data is more likely under s+b hypothesis than bkg-only

W. Verkerke

p-value

- **P-VALUE:** probability to obtain observed data, or more extreme, given the hypothesis in future repeated identical experiments

- For our example from previous page:
  - For the bkg-only hypothesis: $\mathbf{p_b}$ = Fraction of future measurements with N=Nobs (or larger) if s=0



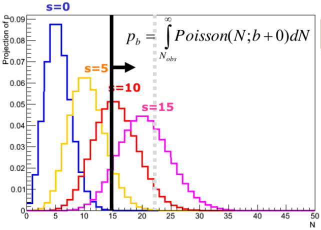
$$p_b = \int_{N_{obs}}^{\infty} Poisson(N; b+0)\,dN \quad (=0.23)$$

- Frequentist p-values (apologies to Bayesians) -- see links later

Excess over background

- p_b or p-values of background hypothesis is used to quantify 'discovery'
- 'discovery' = excess of events over background expectation

- One more example:
  - Nobs=15 for same model, what is p_b?



$(=0.00022)$

- Results customarily expressed as odds of a Gaussian fluctuation with equal p-value: **significance, Zn, z-value**

- Nobs = 15 → Zn = 3.5σ
- Nobs = 22 → **Zn = 5σ** or $\mathbf{p_b < 2.87 \times 10^{-7}}$

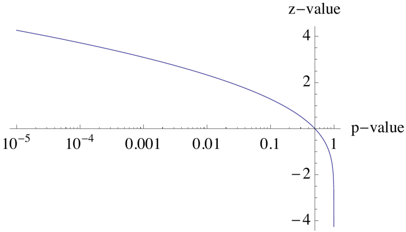


Fig. 1. Relationship between p-value and z-value.

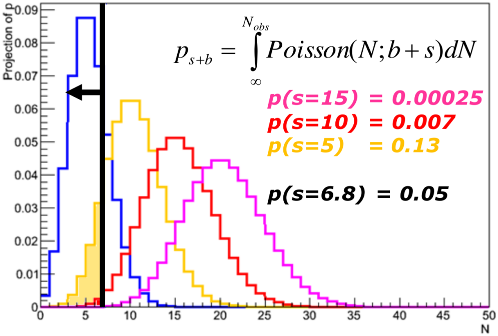
z-value =

`sqrt(2.) * TMath :: ErfInverse(1. - 2. * pvalue)`

$$p\text{-value} = \int_{z\text{-value}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx,$$

Upper limits

- Can also define p-value for s+b hypothesis $\mathbf{p_{s+b}}$
  - Note convention change: integration range in $\mathbf{p_{s+b}}$ is flipped



- Convention: express result as value (upper limit) of **s** for which $\mathbf{p_{s+b}} = 5\%$
  or excluded at 95% confidence level (95% C.L.)
- Our example:
  - s>6.8 is excluded at 95% C.L.

Modified Upper limits : CLs

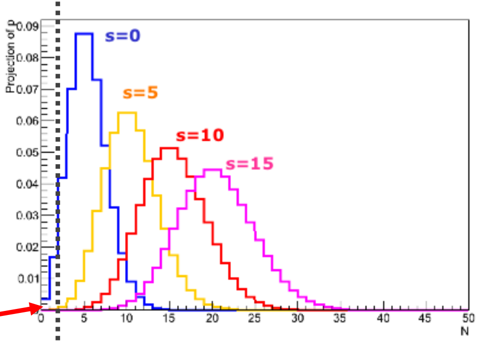
- Interpretation of $\mathbf{p_{s+b}}$ in terms of inference on signal only is problematic
  - Since $\mathbf{p_{s+b}}$ quantifies consistency with data of signal + background
  - Problem apparent when observed data has downward fluctuation wrt background expectation

- Example: Nobs = 2 → $\mathbf{p_{s+b}(s=0) = 0.04}$
  - **s≥0 excluded at 95% C.L. ???**

- Modified approach to protect against such inference on signal (LHC convention):
  - Instead of requiring $\mathbf{p_{s+b}}$ = 5%,

    require

$$\mathrm{CL}_s \equiv \frac{p_{s+b}}{\boxed{1 - p_b}} = 5\%$$



- Example: Nobs = 2 → **s>3.4 excluded at 95% CLs**
- For large Nobs effect on limit is small as $\mathbf{p_b} \to 0$
- https://twiki.cern.ch/twiki/pub/AtlasProtected/StatisticsTools/CLsInfo.pdf

Generalizing This a Bit...

- What we've been doing, in practicle, is evaluating a (Poisson) likelihood:

$$L(n|\mu, b) = \frac{(\mu s + b)^n}{n!} e^{-(\mu s + b)}$$

- We can modify the Likelihood to also describe uncertainties with a Gaussian G and auxiliary measurements θ:

$$L(n, \theta^0|\mu, b, \theta) = \frac{(\mu s + b)^n}{n!} e^{-(\mu s + b)} G(\theta^0 - \theta)$$

- Can also have products of individual "bins", and more than one G for more uncertainties

Profile Likelihood

- Now that we have a more general likelihood, we can make a new test statistics in a variety of ways
- The LHC standard is the *profile likelihood*:

$$q(\mu) = -2\frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})}$$

- This is a bit complicated, but it means:
  - $\hat{\mu}$ and $\hat{\theta}$ maximize the likelihood function
  - $\hat{\hat{\theta}}$ maximizes the likelihood function for that particular μ
  - So, we are evaluating a test statistic and comparing it the the *maximum value possible*
  - This maximization dramatically simplifies the evaluation of the likelihood in the case of many θ
- To evaluate a p-value, we generate a distribution $f(q|\mu, \theta)$ by throwing toys (vary the observed n)
- Then just integrate: $p = \int_x^\infty f(q|\mu, \theta)$

Asymptotics Vs Toys

- Turns out the test statistic q_μ has one very nice property:
- For large enough N, $f(q|\mu, \theta)$ has a known analytic distribution
- This is called the "asymptotic" or "Asimov" approximation
- Instead of using time-consuming pseudoexperiments to determine shape of f, can use the analytic form
  - Huge speed up!

How Do We Do This in Practice? HistFitter

- ATLAS uses a piece of software called *HistFitter* to do much of this
- HistFitter is a wrapper around RooStats: a sophisticated and powerful framework to define and evaluate likelihoods, but not very friendly
- Website is here, arXiv paper describing everything is here and the general purpose tutorial is here
  - Today's exercises borrowed mostly from this :)

Installation

- Need a few quick steps to install HistFitter:
  1. Open a terminal in Jupyter (File $\rightarrow$ New $\rightarrow$ Terminal)
  2. Setup an old version of ROOT: `source bin/thisroot.sh`
  3. Download HF: `wget http://histfitter.web.cern.ch/histfitter/Files/HistFitter-2.0.tar.gz`
  4. Untar: `tar -xzvf HistFitter-2.0.tar.gz`
  5. Setup and make: `cd HistFitter_branch46 && source setup.sh && cd src && make && cd ..`
- All OK?

Some Simple Examples Now...

- Much of this from the HF tutorial!
- NB: to visualize outputs, copy results to `/ML_platform_tests/tutorial/mydir`
- Some things to try:
  1. Very basic fit: `HistFitter.py -w -f -D "before,after,corrMatrix" analysis/tutorial/MyUserAnalysis.py`
  2. Play with changing uncertainties, ndata, nbkgd, etc.
  3. Change from asymptotics to toys: `calculatorType=0` (and set `NToys`)
  4. Add an exclusion fit to the calculation: option `-p`
  5. Add an upper limit calculation to the calculation: option `-l`
- Lots more to go— let me know if you want more!