Signal Significance and Discoveries

Physics 252C - Lecture 14

Prof. John Conway

the logic of scientific discovery

- "Extraordinary claims require extraordinary evidence"
  - Carl Sagan
- "...when you have eliminated the impossible, whatever remains, however improbable, must be the truth."
  - Sherlock Holmes (Sir A. C. Doyle)
- "Now it is far from obvious, from a logical point of view, that we are justified in inferring universal statements from singular ones, no matter how numerous; for any conclusion drawn in this way may always turn out to be false..." - Karl Popper

when have we made a discovery?

- in high energy physics we use the probability of the null hypothesis as a measure of significance
- this is a p-value; recall that if we have H_0 and H_1
  - reject H_0 when it is true: <u>error of the first kind</u>
  - probability for this is the <u>significance</u> of the test
  - accept H_0 when it is false: <u>error of the second kind</u>
  - probability for this is the <u>power</u> of the test

gaussian tail probabilities

x	p-value
1.00	0.1587
1.28	0.10
1.64	0.05
1.96	0.025
2.00	0.0227
3.00	0.00135
5.00	2.87×10^{-7}

- 3σ “evidence” is about 1/1000
- 5σ “observation” is about 1/3.5 million
- for a <u>discovery</u> we usually demand 5σ

can one event be a discovery?

- if you know that you expect less than 2.8×10^{-7} events from background, but you see one, you have a 5σ discovery
- actually establishing that this is your background with high confidence is difficult
- you may have forgotten a background, or mismeasured it
- any examples you can think of?

“bump-hunting” or trials factor

- suppose you are looking not in one channel but in several
- equivalently, suppose we are looking for a bump somewhere in a spectrum
- an excess could show up with random probability in any of the channels, or at any of the masses considered
- how do we take this into account in determining the p-value of an observation?
- roughly speaking, need to multiply by the number of channels to obtain the p-value
- can simulate the whole process (mass scan) to get the trials factor

can several weak observations make a discovery?

- suppose you have several disparate channels with small p-values
- can we simply multiply p-values to get an overall one?
- not exactly - for example the probability of the product of two random numbers in the range 0-1 being less than 10^{-4} is much greater than the probability of them both being less than 0.01
- in fact we have a formula which tells us the probability of the product of n random numbers in the range (0,1) being less than some value :

$$\mathcal{P}(r_1 r_2 \ldots r_n < \epsilon) = \sum_{i=0}^{n-1} \frac{-1^i (\ln \epsilon)^i}{i!}$$

the importance of blind searches

- cuts should be chosen and and optimized based on the simulated, not the observed data
- it is easy (consciously or unconsciously) to alter cuts to preserve/exclude individual events after seeing the data
- this obviously results in a biased result
- for a discovery, you need an a priori hypothesis for the background
- "blind but not dumb": if you've forgotten a background, is it okay to close the box, fix the problem, and re-open it later?
- no substitute for good physics judgment

discovery bias

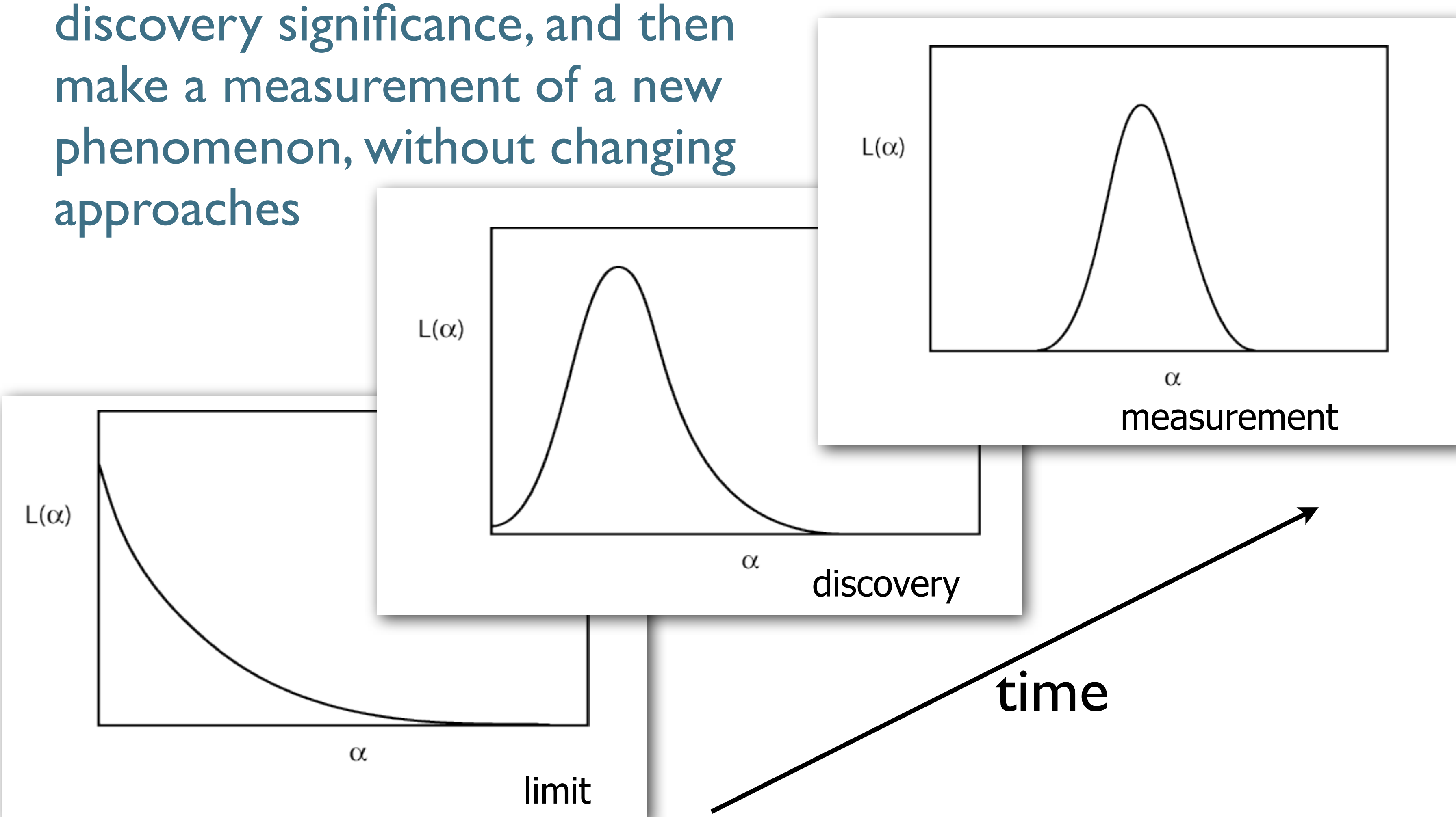
- it is interesting to note that, quite naturally, the initial estimates of cross sections, and branching ratios for rare decays, tend to be biased high
- with competing experiments, the natural tendency is to make a claim once a certain level of statistical significance is attained
- good example: D0's recent "evidence" for single top production

likelihood for discovery

- wish to base significance p-value on likelihood in some way
- remember that likelihood is not goodness of fit!
- joint likelihood provides easy way to combine measurements/channels
- with to use likelihood-based method to incorporate systematic uncertainties
- examine a few possibilities:
  - significance in CLs method
  - Bayesian prior-predictive p-values
  - likelihood ratio (or delta-log-likelihood)

a truly unified approach?

To my own mind the ultimate goal, here, is a method which would allow us in a smooth way to quote a limit, claim a discovery significance, and then make a measurement of a new phenomenon, without changing approaches



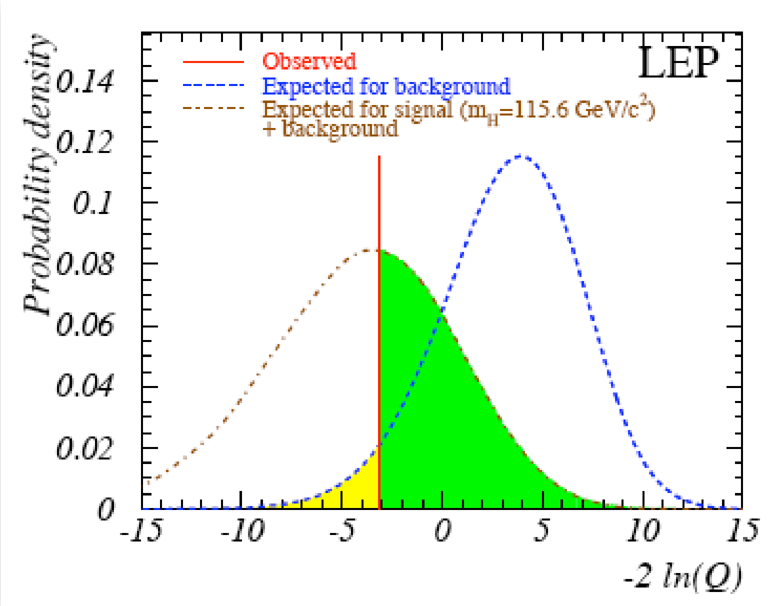
Poisson process with background

- simplest example of discovery problem
- expect $b \pm \sigma_b$ events from background
- let us assume that our background uncertainty is a truncated gaussian with mean b and width σ_b
- clearly b is an estimator for the true background rate b_{true}
- does signal efficiency or uncertainty matter?
- we will examine the behavior of several methods

Goal: a p-value telling us the probability that we observe n or more events given background only

signal significance in CL_s

- in CL_s we already have built in a p-value for the background
- $1\text{-}CL_b$ represents the p-value for the background-only case
- but remember the actual definition of $1\text{-}CL_b$:



$$Q \equiv \frac{\mathcal{L}(\bar{x}; s+b)}{\mathcal{L}(\bar{x}; b)} \qquad 1 - CL_b \equiv \int_{-\infty}^{Q_{obs}} \mathcal{P}(Q'; b)\, dQ'$$

CLs requires a signal model with a certain s in order to calculate signal significance

signal-model-free discovery?

- we seek a sort of CL_x method using only the information we possess: b and width σ_b
- how to incorporate background?
  - frequentist: p-value must be based on b_{true} and hence we arrive at a p-value depending on b_{true}!
  - Bayesian: must calculate likelihood including background uncertainty...but we are after a probability for seeing n or more events

    solution: Bayesian prior-predictive p-values

Bayesian prior-predictive p-values

$$P(\geq n; b, \sigma_b) = \sum_{k=n}^{\infty} P(k; b, \sigma_b)$$

$$P(k; b, \sigma_b) = \int_0^{\infty} \frac{b'^k e^{-b'}}{k!} G(b', b, \sigma_b) db'$$

- here we have extended the notion of Bayesian probability into probabilities calculated essentially as Bayesian posteriors
- must have that the sum normalizes to unity...good sanity check
- look at some code: psig.cpp

Bayesian prior-predictive p-values

- we are essentially averaging the Poisson probability over all possible background, weighting the background possibilities by a gaussian
- we can also do this integral by Monte Carlo:
  - generate a background (from Gaussian)
  - generate a Poisson value for the outcome
  - record how many are greater or equal to n
- are the results the same?

→ look at examples using code

likelihood ratio/delta-log-likelihood

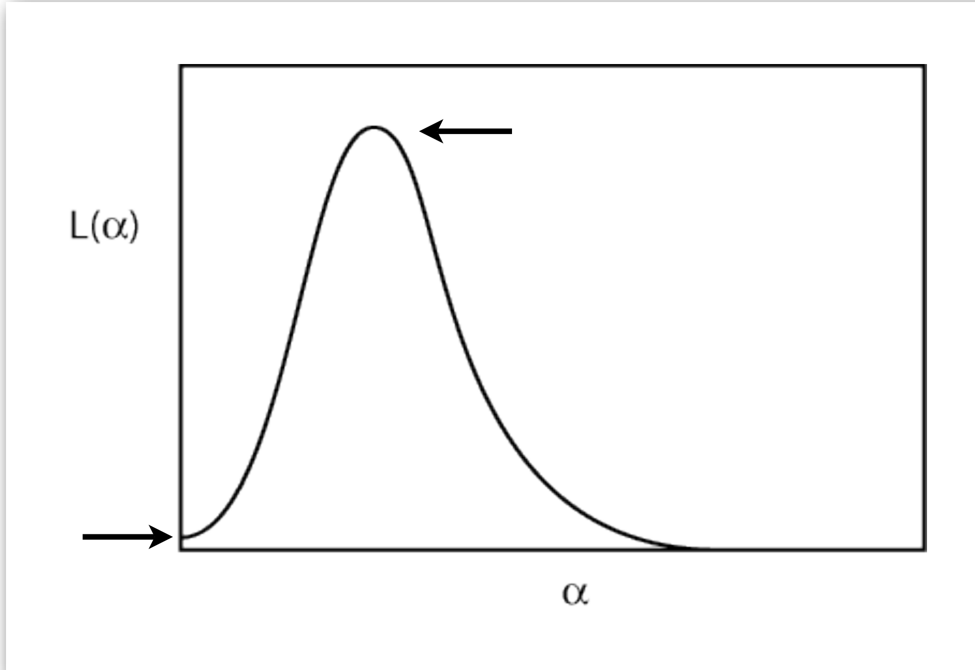
- most cases are more complex than this of course
  - spectra (many bins, or multiple channels)
  - multiple nuisance parameters
- can we use a general likelihood to claim a discovery p-value?
- given a specific signal model we can use likelihood to compare hypotheses: likelihood ratio

$$Q \equiv \frac{\mathcal{L}(\bar{x}; s+b)}{\mathcal{L}(\bar{x}; b)}$$

- we can for example equate observed Q to equivalent Gaussian LR, if $\mathcal{P}$(Q) is gaussian...

equivalent gaussian LR

- take L at maximum
- take L at α=0 (null hypothesis)
- equate ratio to gaussian case to get number of sigma



$$e^{x^2/2} = Q \quad \Rightarrow \quad x = \sqrt{2\ln Q} = \sqrt{2(\ln \mathcal{L}(\bar{x}; s+b) - \ln \mathcal{L}(\bar{x}; b))}$$

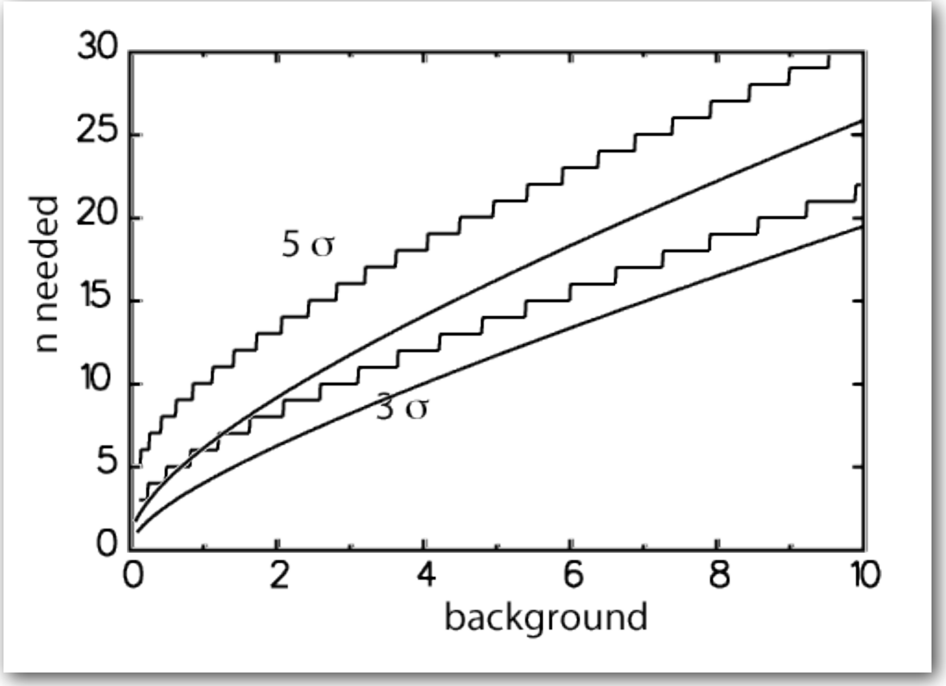
"ΔlnL"

ad hoc methods

- often we find ourselves in a situation where we have some statistic ξ which is a measure of signal-ness but is not a true likelihood
- is a p-value still obtainable?
- as long as we know $P(\xi)$ we can integrate above our observed value ξ_{obs} to get a p-value
- must take into account systematic errors in getting the distribution for ξ...but all of them?
- not always clear what is the proper ensemble; what restrictions do we put on generating pseudoexperiments?

optimizing for discovery

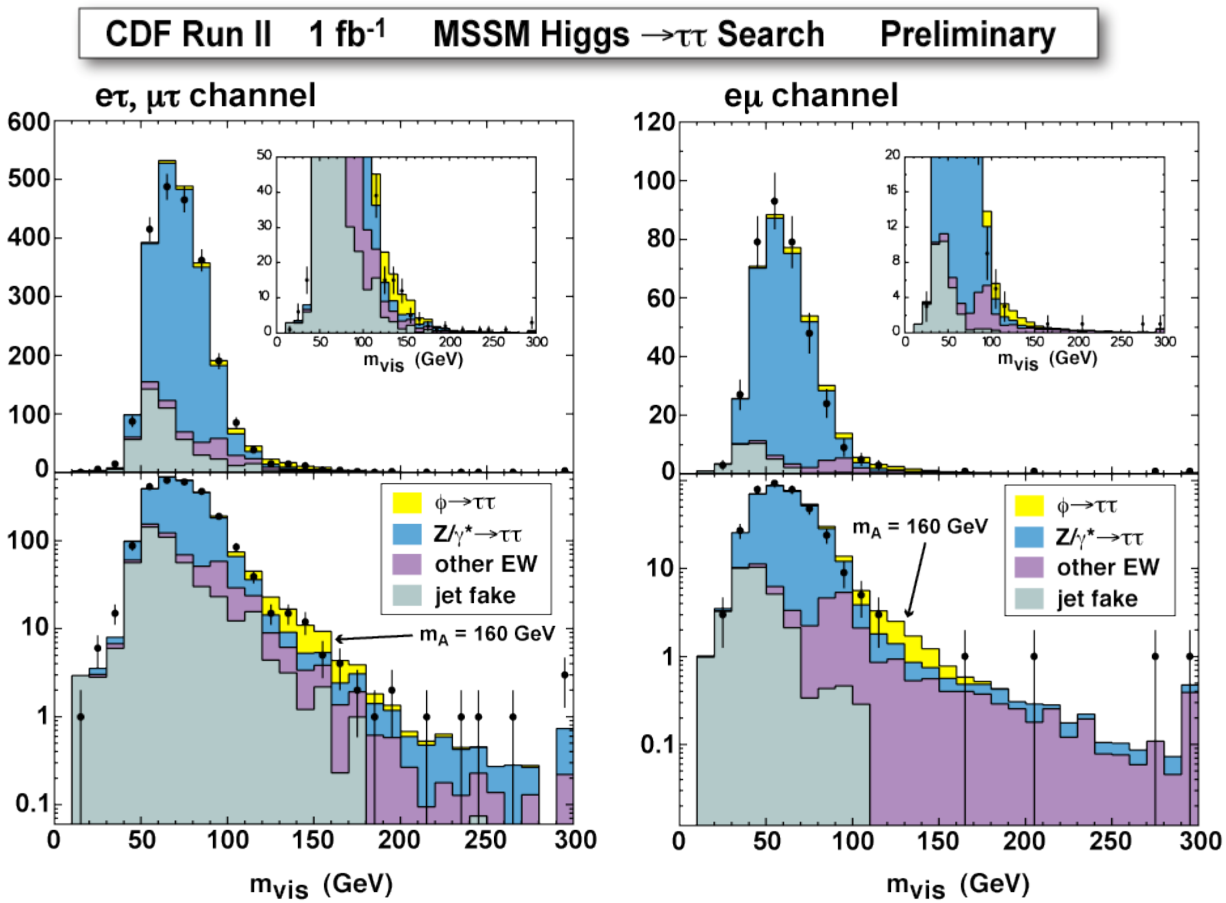
- in the Gaussian regime we've seen that the significance of a signal is (on average) $s/\sqrt{s+b}$
- does this relation persist down into the Poisson regime?
- we often face a tradeoff between background and signal efficiency
- Poisson tails really hurt in the low regime
- you need signal, not just small background!



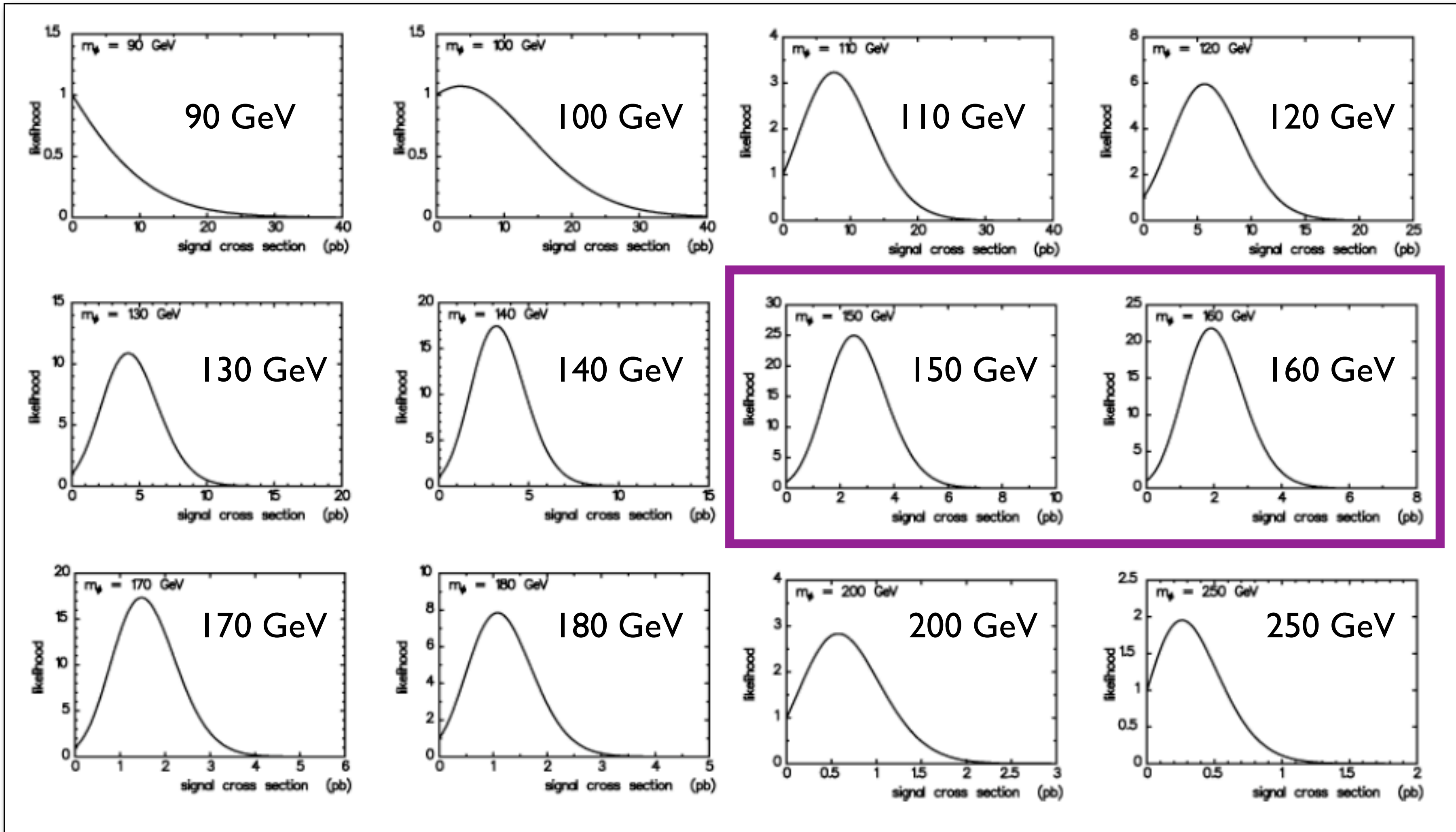
- if you can double your signal and your background, do it!

significance in the CDF H→ττ search

- real life example - and the story is not over!
- how to calculate the significance of the infamous H→ττ "bump" ?



significance in the CDF H→ττ search



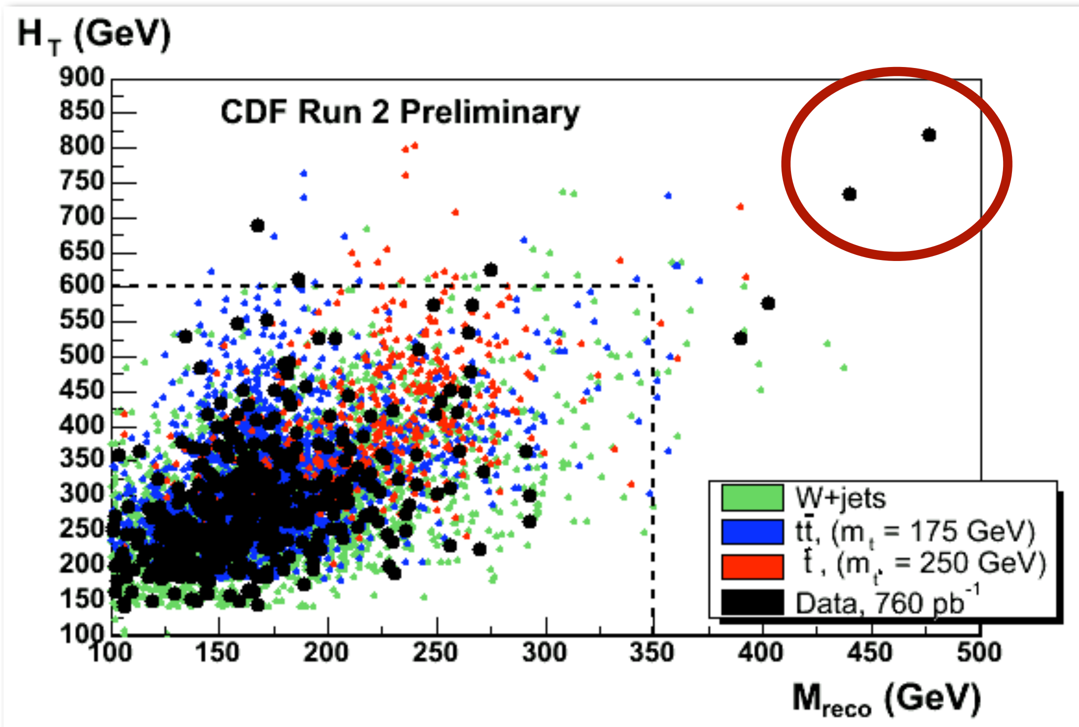
likelihood versus cross section including systematics

significance in the CDF $H \to \tau\tau$ search

- likelihood show that 150 GeV or so is the most significant mass at around 2.5σ (or p = 0.6%)
- generated many background-only pseudoexperiments
- fit each pseudoexperiment to each mass in previous slide
- recorded most significant likelihood ratio
- asked "how often do I get at least a LR = 25" ?
  - answer: 1.8%, or about 2.1σ
  - But what values for the nuisance parameters should be used to generate pseudoexperiments? Central values or varied? Do we risk double counting? (varying nuisance parameters: 1.8σ)

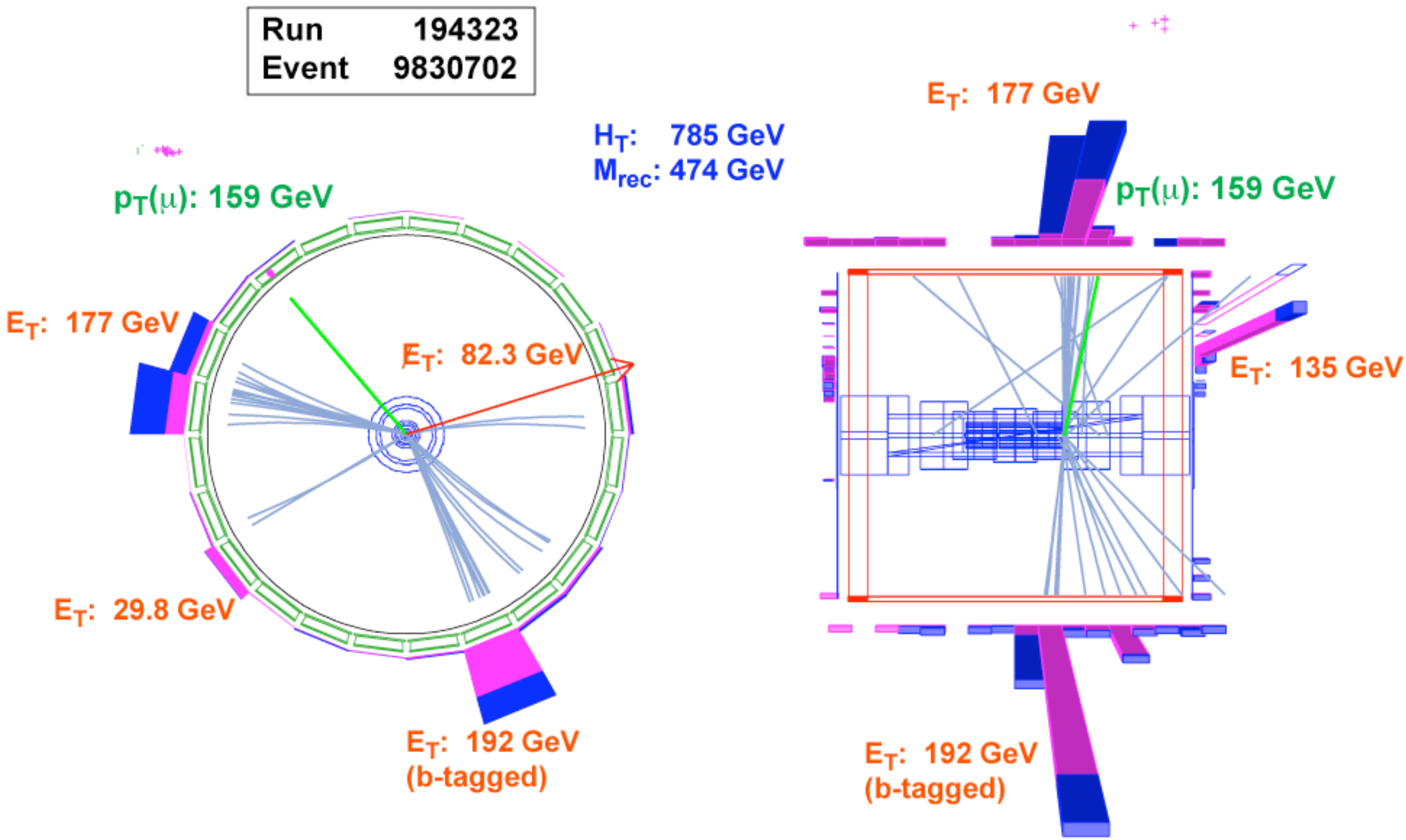
t′ search

- made a priori hypothesis in t′ search: count events with H_T > 600 GeV and M_{reco} > 350 GeV
- saw 7, expected 6.8 from background...nada

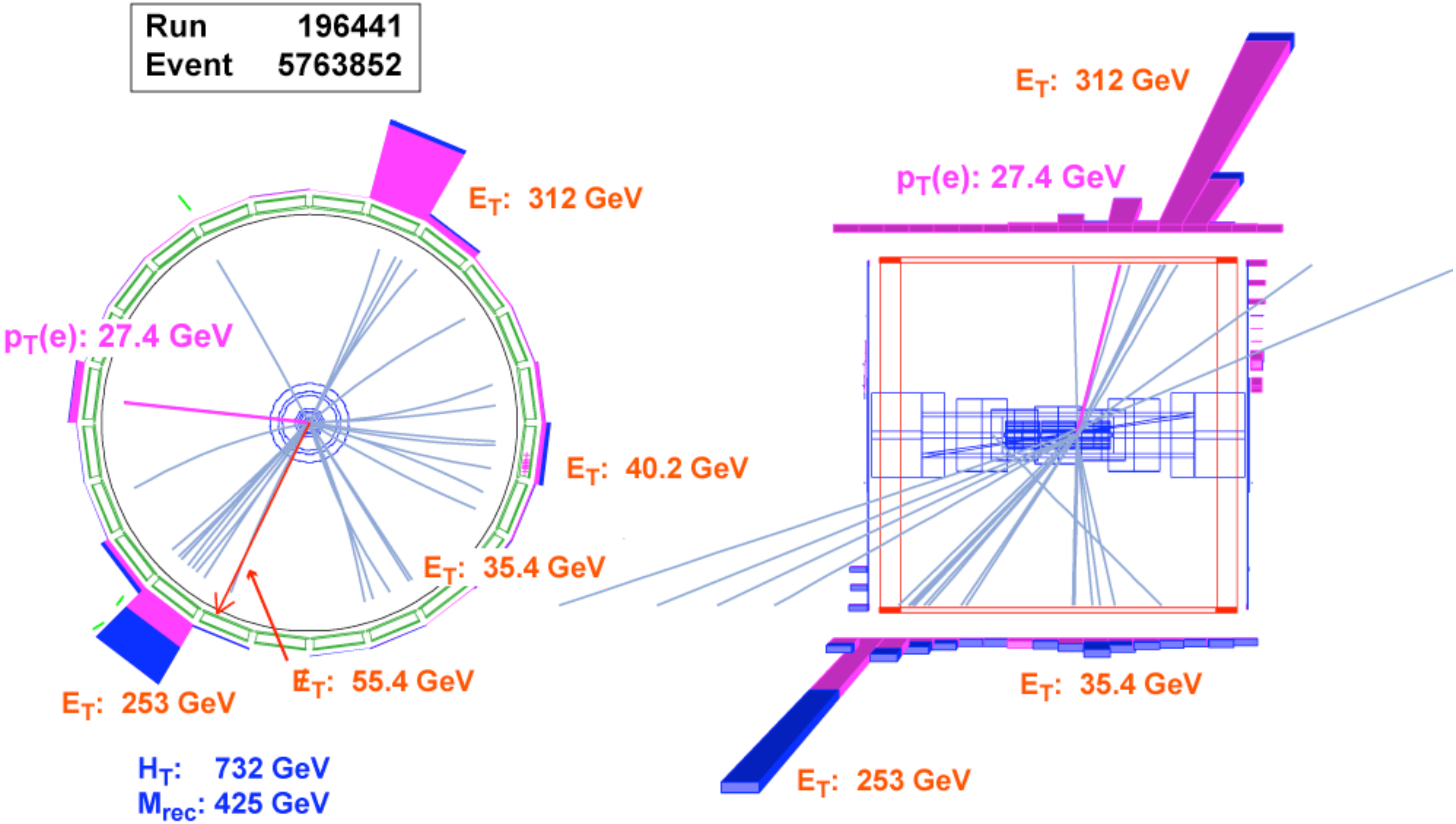


two of the events, though, were extraordinary...

t′ search


Run 194323
Event 9830702
H_T: 785 GeV
M_{rec}: 474 GeV
$p_T(\mu)$: 159 GeV
E_T: 177 GeV
$\not{E}_T$: 82.3 GeV
E_T: 29.8 GeV
E_T: 192 GeV
(b-tagged)
E_T: 177 GeV
$p_T(\mu)$: 159 GeV
E_T: 135 GeV
E_T: 192 GeV
(b-tagged)

t′ search



Run 196441
Event 5763852
E_T: 312 GeV
$p_T(e)$: 27.4 GeV
E_T: 40.2 GeV
E_T: 35.4 GeV
$\not{E}_T$: 55.4 GeV
E_T: 253 GeV
H_T: 732 GeV
M_{rec}: 425 GeV
E_T: 312 GeV
$p_T(e)$: 27.4 GeV
E_T: 35.4 GeV
E_T: 253 GeV

t' search

- a posteriori we cannot say much, statistically
- we know what to look for with 3x more data!
- these do not look much like top pair events, and not W+jets either
- but they could be background - right?
- we need a "discovery test" that could appropriately weight the presence of new physics events by how far out on the tails they are... (Andersen-Darling test?)

we can try our best to be objective, but we can't stop being physicists, or human...