Practical Statistics for Particle Physicists
Lecture 1

Harrison B. Prosper

Florida State University

European School of High-Energy Physics

Parádfürdő, Hungary

5 – 18 June, 2013

Outline

- Lecture 1
  - Descriptive Statistics
  - Probability & Likelihood
- Lecture 2
  - The Frequentist Approach
  - The Bayesian Approach
- Lecture 3
  - Analysis Example

Descriptive Statistics

Descriptive Statistics – 1

Definition: A **statistic** is any function of the data ***X***.

Given a sample $\boldsymbol{X} = x_1, x_2, \ldots x_N$, it is often of interest to compute statistics such as

the **sample average** $$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

and the **sample variance** $$S^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})^2$$

In any analysis, it is good practice to study **ensemble averages**, denoted by < … >, of relevant statistics

Descriptive Statistics – 2

Ensemble Average	$< x >$
Mean	μ
Error	$\varepsilon = x - \mu$
Bias	$b = < x > - \mu$
Variance	$V = < (x - < x >)^2 >$
Mean Square Error	$\mathrm{MSE} = < (x - \mu)^2 >$

Descriptive Statistics – 3

$$\text{MSE} = \langle (x - \mu)^2 \rangle$$
$$= V + b^2$$

Exercise 1:
Show this

The **MSE** is the most widely used measure of **closeness** of an ensemble of statistics {**x**} to the **true value** μ

The **root mean square** (RMS) is

$$\text{RMS} = \sqrt{\text{MSE}}$$

Descriptive Statistics – 4

Consider the *ensemble* average of the *sample* **variance**

$$< S^2 > = < \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 >$$

$$= < \frac{1}{N}\sum_{i=1}^{N}x_i^2 - \frac{2}{N}\sum_{i=1}^{N}x_i\bar{x} + \frac{1}{N}\sum_{i=1}^{N}\bar{x}^2 >$$

$$= \frac{1}{N}\sum_{i=1}^{N} < x_i^2 > - < \bar{x}^2 >$$

$$= < x^2 > - < \bar{x}^2 >$$

Descriptive Statistics – 5

The ensemble average of the sample variance

$$< S^2 > = < x^2 > - < \bar{x}^2 >$$

$$= < x^2 > - \frac{< x^2 >}{N} - \left(\frac{N-1}{N} \right) < x >^2$$

$$= V - \frac{V}{N}$$

has a negative bias of $-V / N$

Exercise 2: Show this

Descriptive Statistics – Summary

The **sample average**
is an unbiased estimate
of the ensemble average

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

The **sample variance**
is a biased estimate
of the ensemble variance

$$S^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})^2$$

Probability

Probability – 1

Basic Rules

1. $P(A) \geq 0$
2. $P(A) = 1$ if A is true
3. $P(A) = 0$ if A is false

Sum Rule

4. $P(A+B) = P(A) + P(B)$ if AB is false *

Product Rule

5. $P(AB) = P(A|B)\ P(B)$ *

***A+B = A or B, AB = A and B, A|B = A given that B is true**

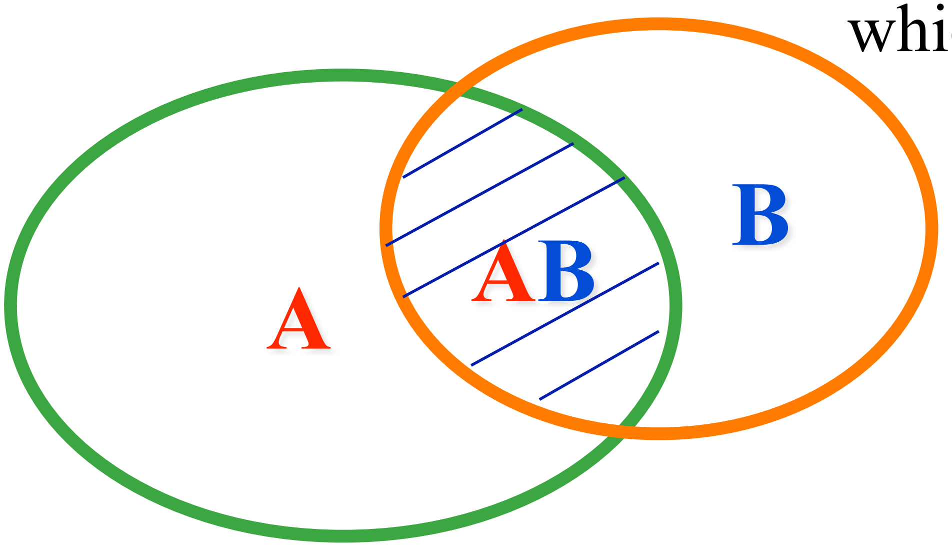
Probability – 2

By definition, the **conditional probability** of A *given* B is

$$P(A \mid B) = \frac{P(AB)}{P(B)}$$

$P(A)$ is the probability of A *without restriction*.

$P(A|B)$ is the probability of A when we *restrict* to the conditions under which B is true.



$$P(B \mid A) = \frac{P(AB)}{P(A)}$$

Probability – 3

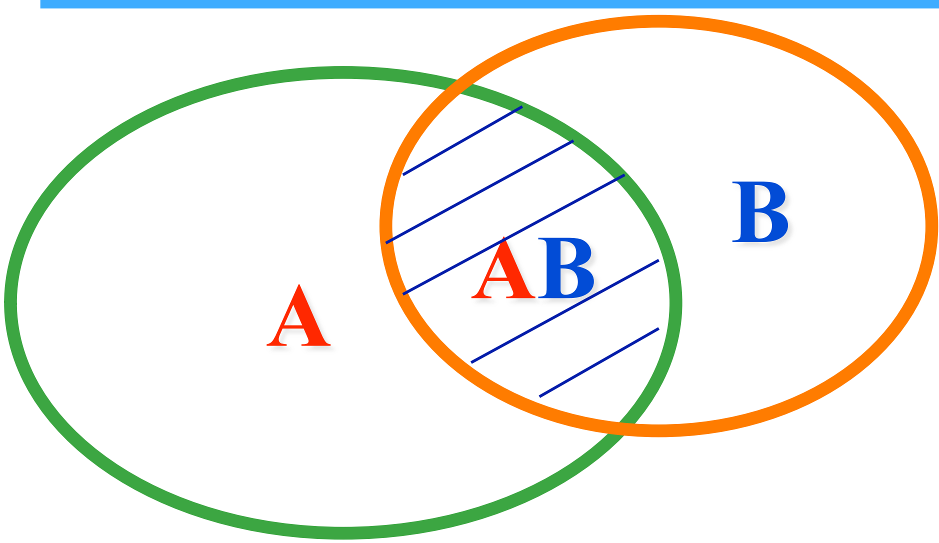
From

$$P(AB) = P(B \mid A)P(A)$$
$$= P(A \mid B)P(B)$$

we deduce

***Bayes' Theorem*:**

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$



Probability – 4

A and B are mutually exclusive if

$$P(AB) = 0$$

A and B are exhaustive if

$$P(A) + P(B) = 1$$

Theorem

$$P(A+B) = P(A) + P(B) - P(AB)$$

Exercise 3: Prove theorem

Probability

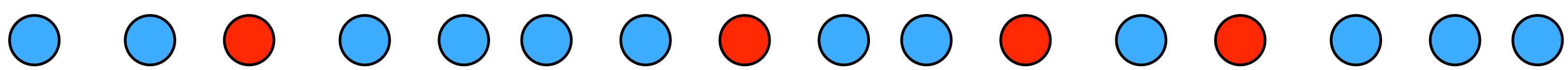
Binomial & Poisson Distributions

Binomial & Poisson Distributions – 1

A **Bernoulli** trial has two outcomes:

S = success or ***F*** = failure.

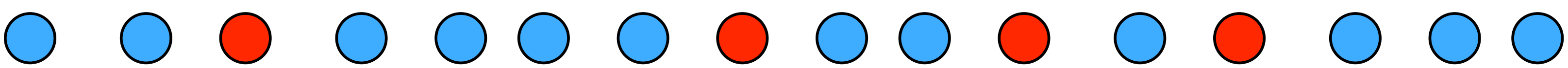
Example: Each collision between protons at the LHC is a Bernoulli trial in which something interesting happens (***S***) or does not (***F***).



Binomial & Poisson Distributions – 2

Let ***p*** be the probability of a success, which is assumed to be the ***same at each trial***. Since ***S*** and ***F*** are ***exhaustive***, the probability of a failure is **1 – *p***. For a given order ***O*** of ***n*** trails, the probability $\Pr(k,O|n)$ of ***exactly k*** successes and ***n – k*** failures is

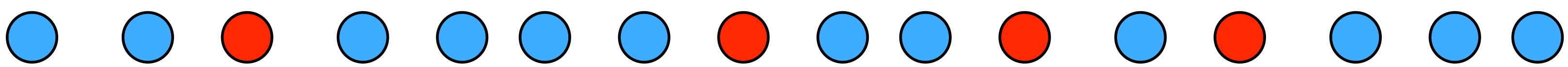
$$\Pr(k,O,n) = p^k (1 - p)^{n-k}$$



Binomial & Poisson Distributions – 3

If the order ***O*** of successes and failures is irrelevant, we can eliminate the order from the problem integrating over all possible orders

$$\Pr(k,n) = \sum_{O} \Pr(k,O,n) = \sum_{O} p^{k}(1-p)^{n-k}$$



This yields the **binomial distribution**

$$\text{Binomial}(k,n,p) \equiv \binom{n}{k} p^{k}(1-p)^{n-k}$$

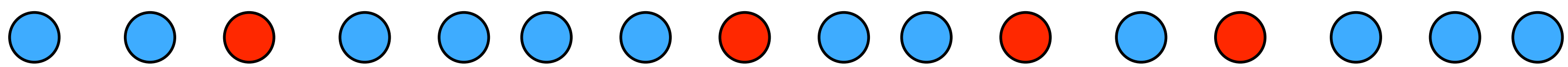
which is sometimes written as $k \sim \text{Binomial}(n,p)$

Binomial & Poisson Distributions – 3

We can prove that the mean number of successes ***a*** is

a = ***p n***. **Exercise 4**: Prove it

Suppose that the probability, ***p***, of a success is very small,



then, in the limit ***p*** → 0 and ***n*** → ∞, such that ***a*** is *constant*,

Binomial(k, n, p) → **Poisson**(k, a).

The Poisson distribution is generally regarded as a good model for a **counting experiment**

Exercise 5: Show that Binomial(k, n, p) → Poisson(k, a)

Common Distributions and Densities

$\mathrm{Uniform}(x,a)$	$1/a$
$\mathrm{Gaussian}(x,\mu,\sigma)$	$\exp[-(x-\mu)^2/(2\sigma^2)]/(\sigma\sqrt{2\pi})$
$\mathrm{Chisq}(x,n)$	$x^{n/2-1}\exp(-x/2)/[2^{n/2}\Gamma(n/2)]$
$\mathrm{Gamma}(x,a,b)$	$x^{b-1}a^b\exp(-ax)/\Gamma(b)$
$\mathrm{Exp}(x,a)$	$a\exp(-ax)$
$\mathrm{Binomial}(k,n,p)$	$\binom{n}{k}p^k(1-p)^{n-k}$
$\mathrm{Poisson}(k,a)$	$a^k\exp(-a)/k!$
$\mathrm{Multinomial}(k,n,p)$	$\frac{n!}{k_1!\cdots k_K!}\prod_{i=1}^{K}p_i^{k_i},\ \sum_{i=1}^{K}p_i=1,\ \sum_{i=1}^{K}k_i=n$

Probability – What is it Exactly?

There are at least two *interpretations* of probability:

1. **Degree of belief** in, or plausibility of, a proposition

   Example:

   It will snow in Geneva on Friday

2. **Relative frequency** of outcomes in an *infinite* sequence of *identically repeated* trials

   Example:

   trials: proton-proton collisions at the LHC

   outcome: the creation of a Higgs boson

Likelihood

Likelihood – 1

The ***likelihood function*** is simply the probability, or probability density function (**pdf**), ***evaluated at the observed data***.

Example 1: Top quark discovery (D0, 1995)

$p(D|\,d) = \mathrm{Poisson}(D\,|d)$ ***probability*** to get a count D

$p(\mathbf{17}|d) = \mathrm{Poisson}(\mathbf{17}|d)$ ***likelihood*** of observation $D = 17$

$\mathrm{Poisson}(D|d) = \exp(-d)\, d^D / D!$

Likelihood – 2

Example 2:

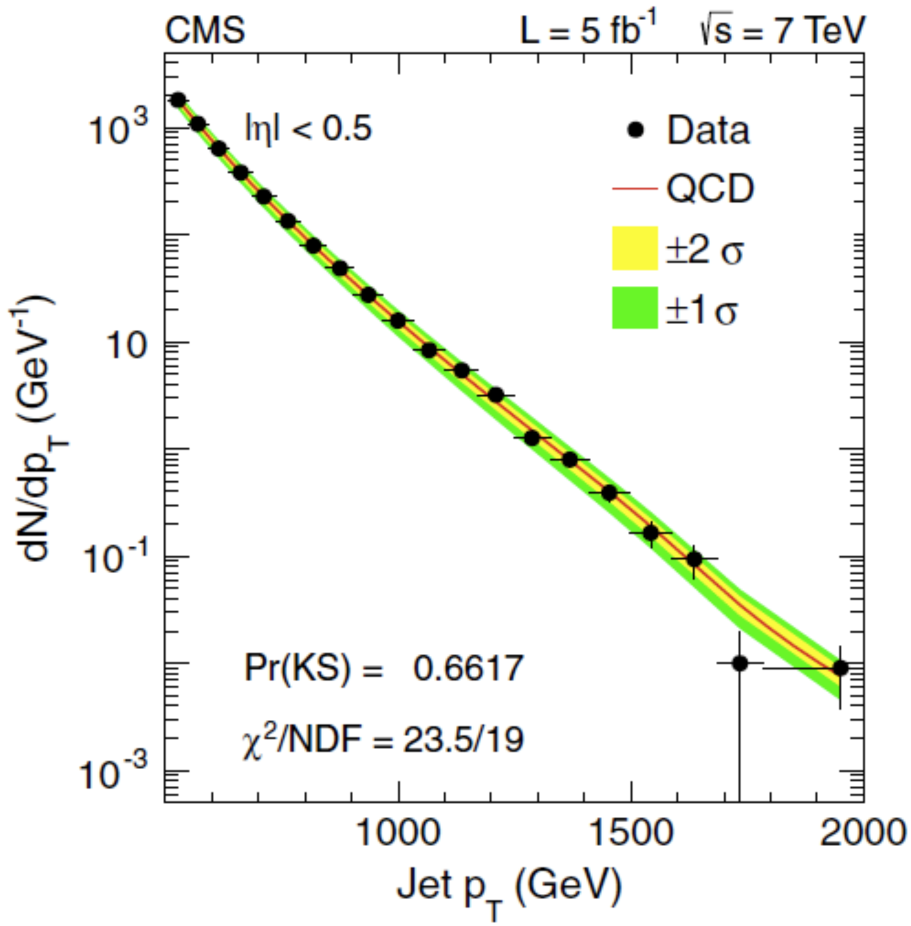
Multiple counts D_i with a fixed total count N

$$p(D \mid p) = \text{Multinomial}(D, N, p)$$

$$D = D_1, \cdots, D_K, \quad p = p_1, \cdots, p_K$$

$$\sum_{i=1}^{K} D_i = N$$

This is an example of a multi-*binned* likelihood



PHYSICAL REVIEW D **87,** 052017 (2013)

Search for contact interactions using the inclusive jet p_T spectrum in pp collisions at $\sqrt{s} = 7$ TeV

S. Chatrchyan *et al.**
(CMS Collaboration)

(Received 21 January 2013; published 26 March 2013)

Likelihood – 3

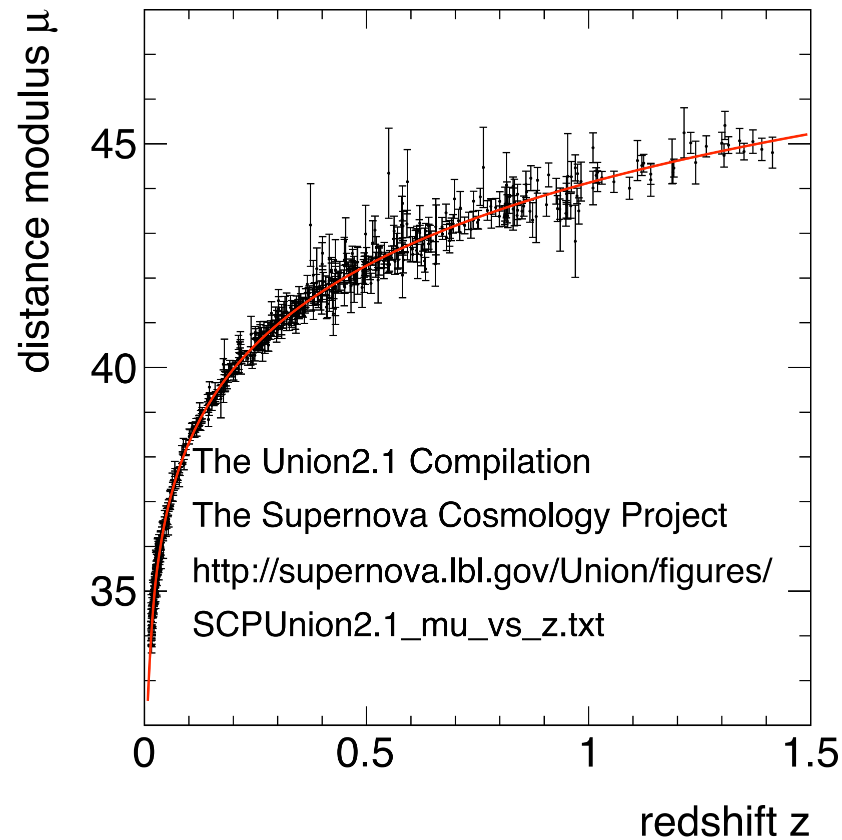
Example 3:

Red shift and distance modulus measurements of $N = 580$ Type Ia supernovae

$$p(D \mid \Omega_M, \Omega_\Delta, Q) =$$

$$\prod_{i=1}^{N} \text{Gaussian}(x_i, \mu(z_i, \Omega_M, \Omega_\Delta, Q), \sigma_i)$$

$$D = z_i, x_i \pm \sigma_i$$



This is an example of an *un-binned* likelihood

Likelihood – 4

Example 4: Higgs to γγ

The discovery of the neutral Higgs boson in the di-photon final state made use of an an *un-binned* likelihood,

$$p(x \mid s, m, w, b) = \exp[-(s+b)] \prod_{i=1}^{N} \left[s\, f_s(x_i \mid m, w) + b\, f_b(x_i) \right]$$

where x = di-photon masses

m = mass of new particle

w = width of resonance

s = expected signal

b = expected background

f_s = signal model

f_b = background model

Exercise 6: Show that a binned multi-Poisson likelihood yields an un-binned likelihood of this form as the bin widths go to zero

Likelihood – 5

Given the likelihood function we can answer questions such as:

1. How do I estimate a parameter?
2. How do I quantify its accuracy?
3. How do I test an hypothesis?
4. How do I quantify the significance of a result?

Writing down the likelihood function requires:

1. Identifying all that is *known*, e.g., the observations
2. Identifying all that is *unknown*, e.g., the parameters
3. Constructing a probability model *for both*

Likelihood – Top Quark Discovery

Example 1: Top Quark Discovery (1995), D0 Results

knowns:

D = 17 events

B = 3.8 ± 0.6 background events

unknowns:

b	expected background count
s	expected signal count
$d = b + s$	expected event count

Note: we are uncertain about *unknowns*, so **17 ± 4.1** is a statement about d, *not about the observed count* **17**!

Likelihood – Top Quark Discovery

Probability:

$$p(D \mid s, b) = \text{Poisson}(D, s+b)\ \text{Poisson}(Q, bk)$$

$$= \frac{(s+b)^D e^{-(s+b)}}{D!} \frac{(bk)^Q e^{-bk}}{\Gamma(Q+1)}$$

Likelihood:

$$p(17 \mid s, b)$$

where

$$B = Q / k \qquad Q = (B / \delta B)^2 = (3.8 / 0.6)^2 = 41.11$$

$$\delta B = \sqrt{Q} / k \qquad k = B / \delta B^2 = 3.8 / 0.6^2 = 10.56$$

Likelihood – Higgs to γγ (CMS)

Example 4: Higgs to γγ

background model

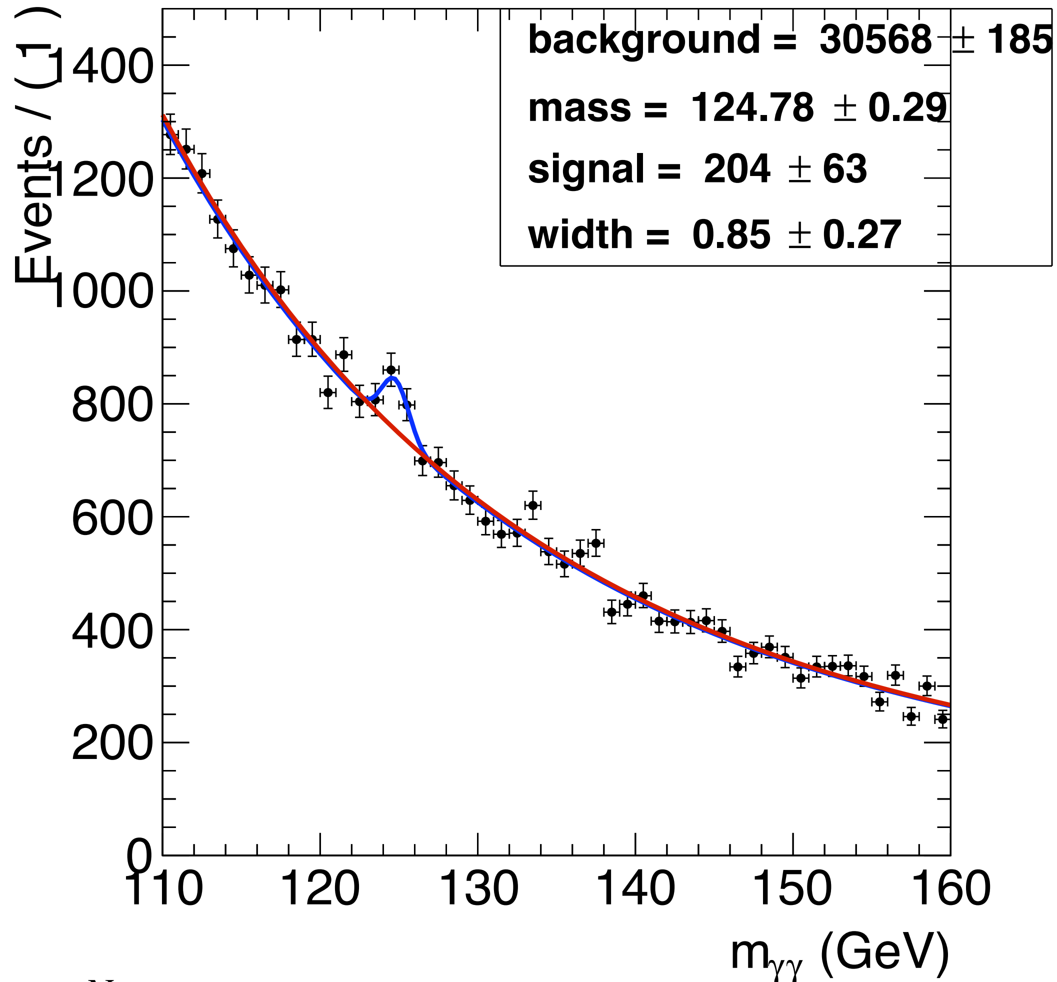
$x = m_{\gamma\gamma}$

$f_b(x) = A\exp[-(bx + ax^2)]$

signal model

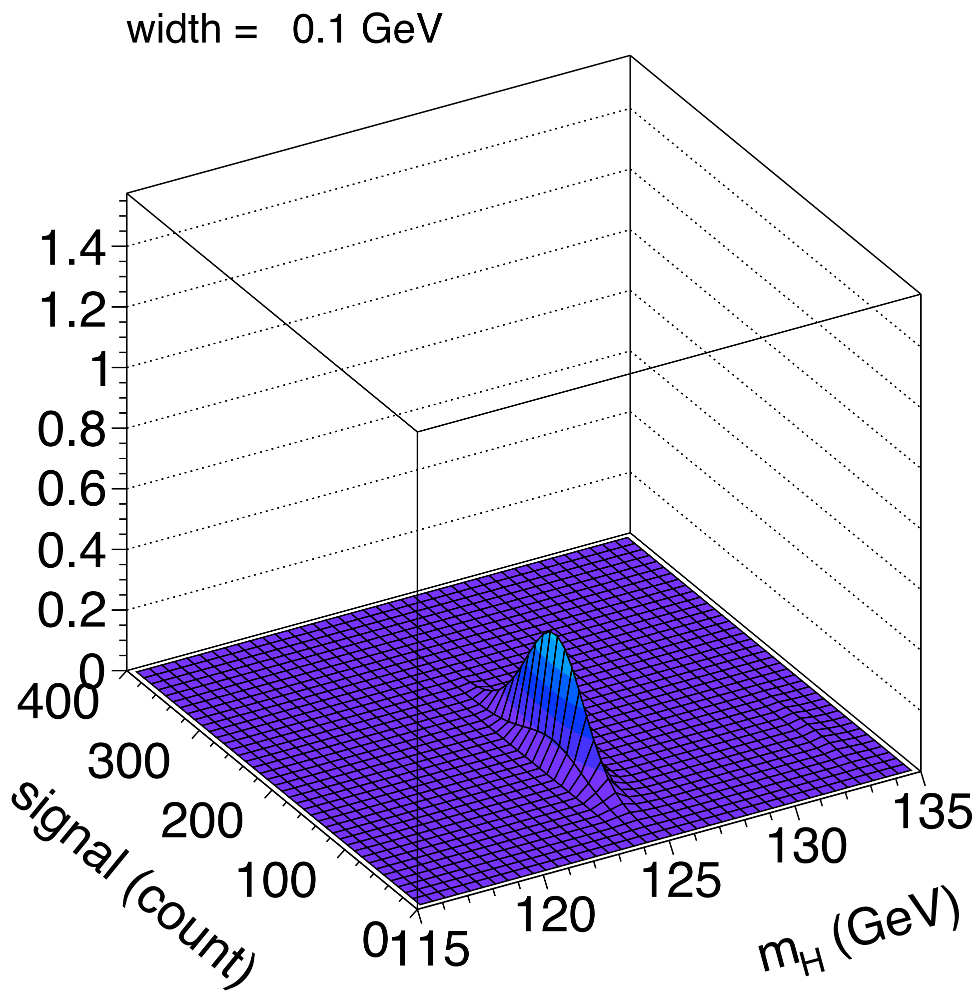
$f_s(x \mid m, w)$

$= \text{Gaussian}(x, m, w)$

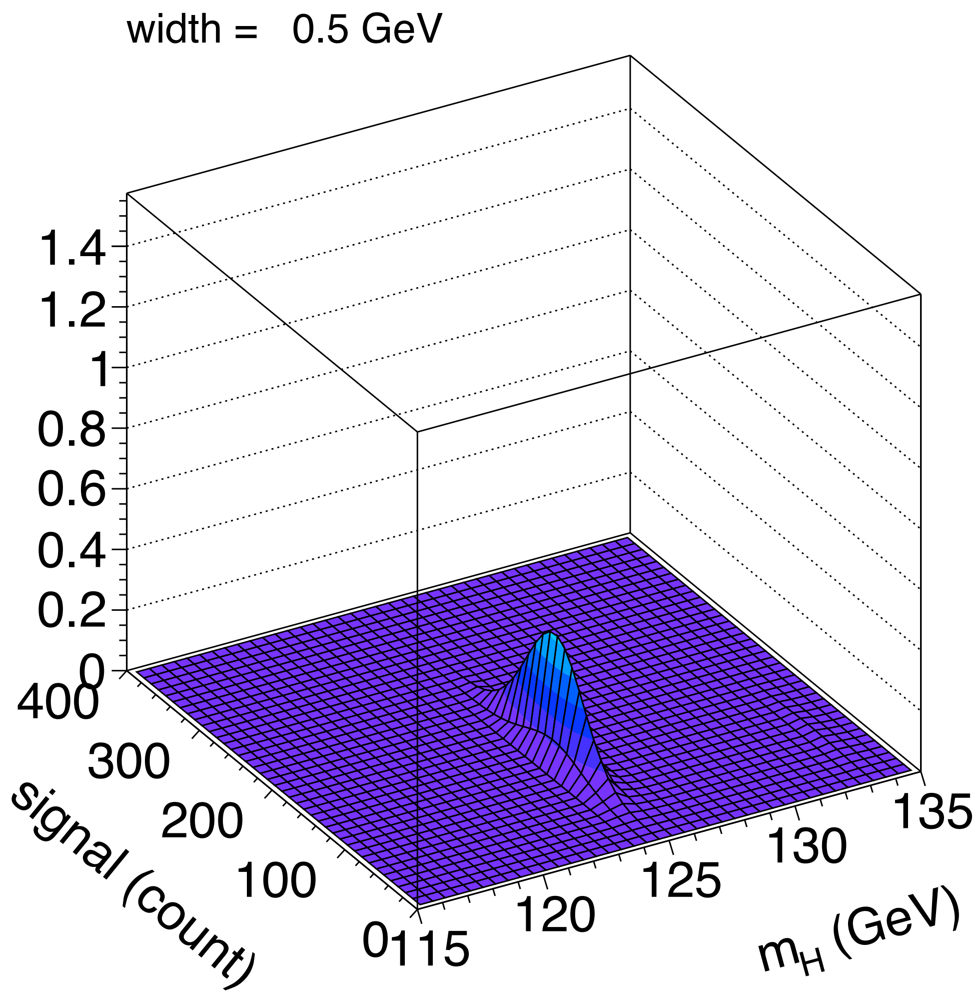


$$p(x \mid s, m, w, b) = \exp[-(s+b)] \prod_{i=1}^{N} \left[s\, f_s(x_i \mid m, w) + b\, f_b(x_i) \right]$$

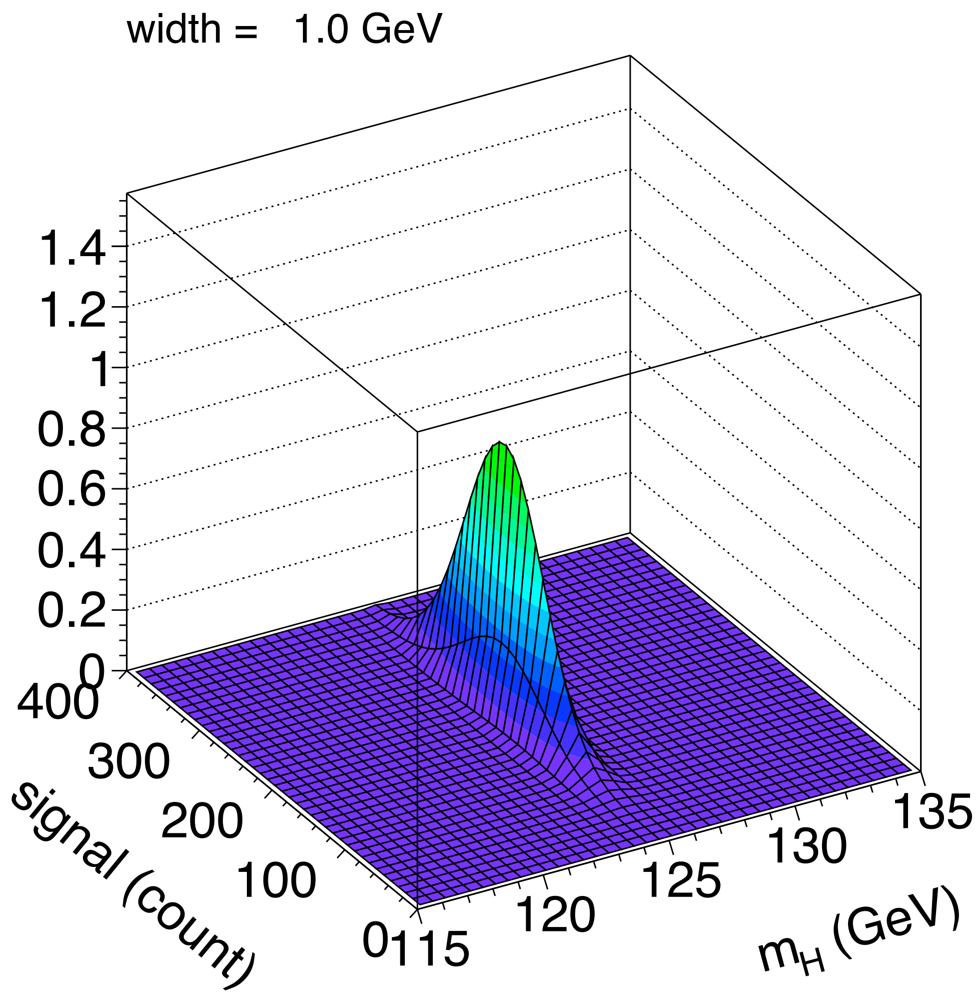
Likelihood – Higgs to γγ (CMS)



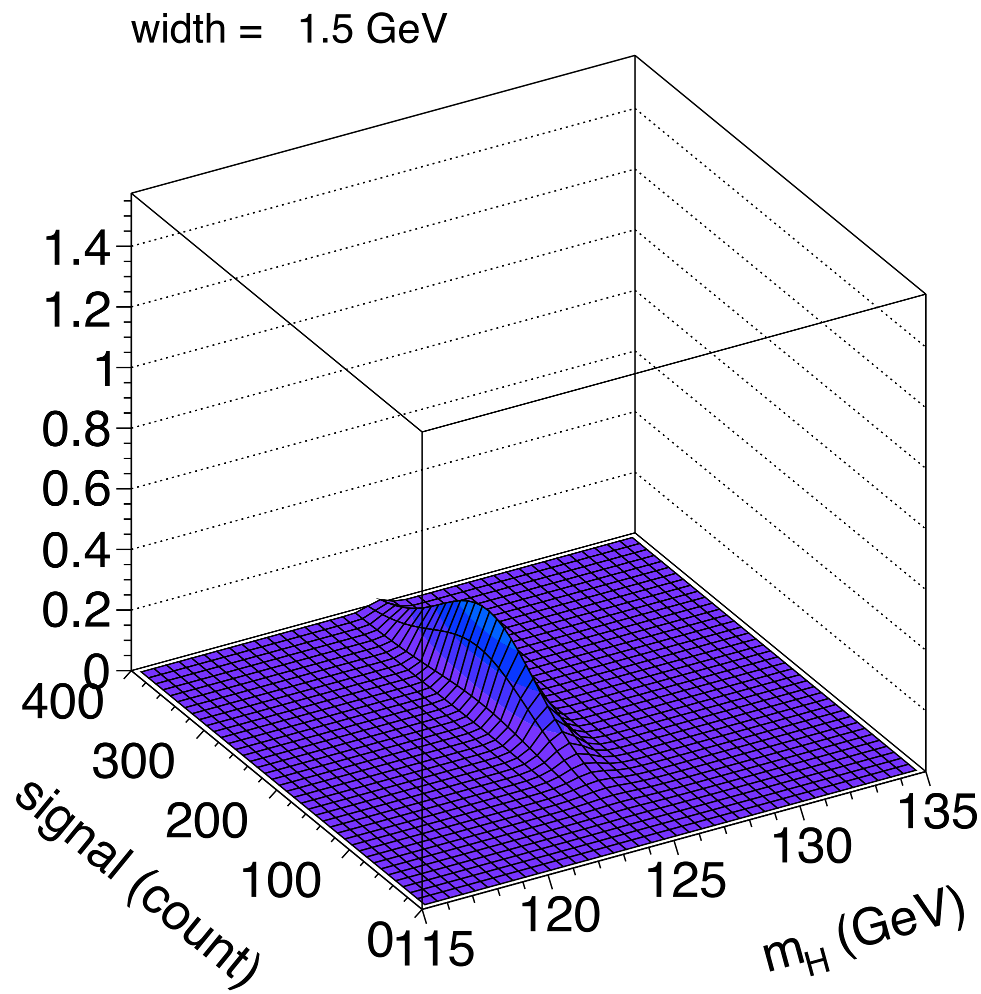
Likelihood – Higgs to γγ (CMS)



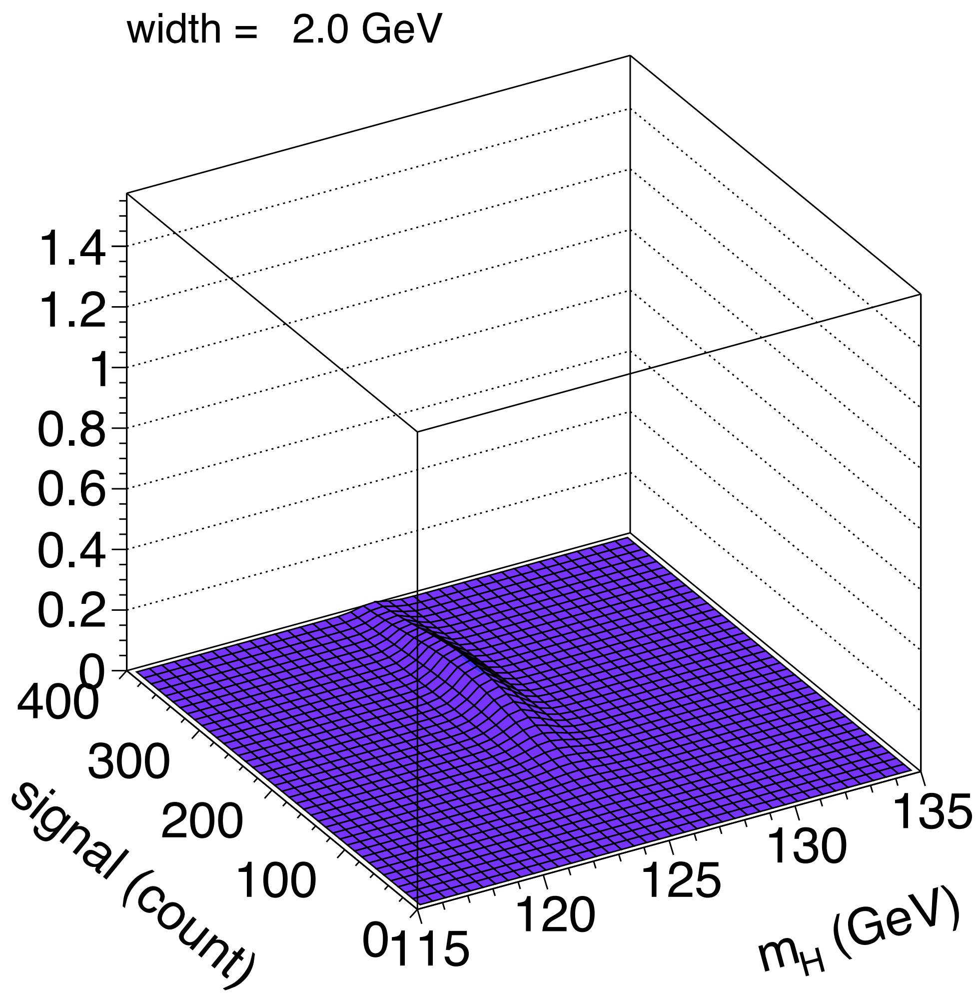
Likelihood – Higgs to γγ (CMS)



Likelihood – Higgs to γγ (CMS)



Likelihood – Higgs to γγ (CMS)



Summary

Statistic

A statistic is *any* function of potential observations

Probability

Probability is an *abstraction* that must be interpreted

Likelihood

The likelihood is the probability (or probability density) of potential observations *evaluated at the observed data*

Tutorials

Location:

http://www.hep.fsu.edu/~harry/ESHEP13

Download

tutorials.tar.gz

and unpack

tar zxvf tutorials.tar.gz

Need:

Recent version of Root linked with RooFit and TMVA