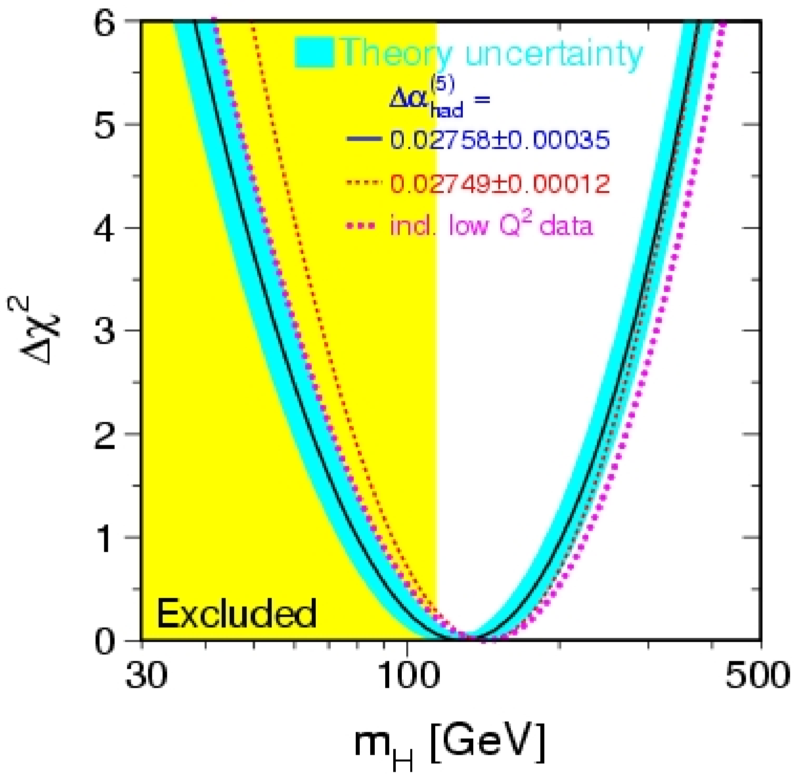
Last Lecture: Probability Distributions … Binomial

… Poisson

… Gaussian

: Central Limit Theorem … why Gaussians are "normal"

… why for N measurements, $\sigma(N) = \sigma(1)/\text{sqrt}(N)$



Today's lecture:

Extracting Information from Data:

… Estimation

… Fitting

… Likelihood

(See Barlow's text book for a more detailed discussion)

Case Study Reminder

- 10 minutes + 5 for questions is not long!

- Please focus on experimental and phenomenological aspects (i.e.things studied on this course) rather than giving too much theory detail

- Marks awarded (with equal weight) for:
    - Presentation
    - Scientific Content
    - Understanding Conveyed
    - Answers to questions

- Assessed by PRN and a couple of friendly Ph.D. students

- Aim to be (reasonably) fun, interesting and educational …

Case Study Schedule

1. Monday 1 December

Tony: Neutrinoless double beta decay experiments

Jack: The Totem experiment / forward physics @ LHC

2. Friday 5 December

Sukhbinder: Triggering events at the LHC

Pat: Heavy Ion experiments and the Quark-Gluon Plasma

Amelia: Evidence for neutrino mass from SuperK & elsewhere

3. Monday 8 December

Alex: The top quark discovery at the Tevatron

Rory: Searching for Supersymmetry at the LHC

Reminder: The Central Limit Theorem

When multiple measurements are added together:

$\langle x \rangle = \mu_1 + \mu_2 + \ldots \mu_N$ $V(x) = V_1 + V_2 + \ldots V_N$ $P(x) \rightarrow$ Gaussian

$$\sigma \sim \frac{1}{\sqrt{N}}$$

Examples of the CLT (i.e. Gaussian distributions) are everywhere!... eg people's heights are governed by many environmental & genetic factors, but overall distribution is Gaussian

There are counter-examples, eg people's weights ... because a single factor (food intake) dominates and gives a skew

Summary of Probability Distributions

Main points from last lecture …

- Binomial distribution (yes/no situations, eg effic's)
- Poisson distribution (numbers of events observed if we know the mean number or `expectation')
- Both look more and more Gaussian as stats improve

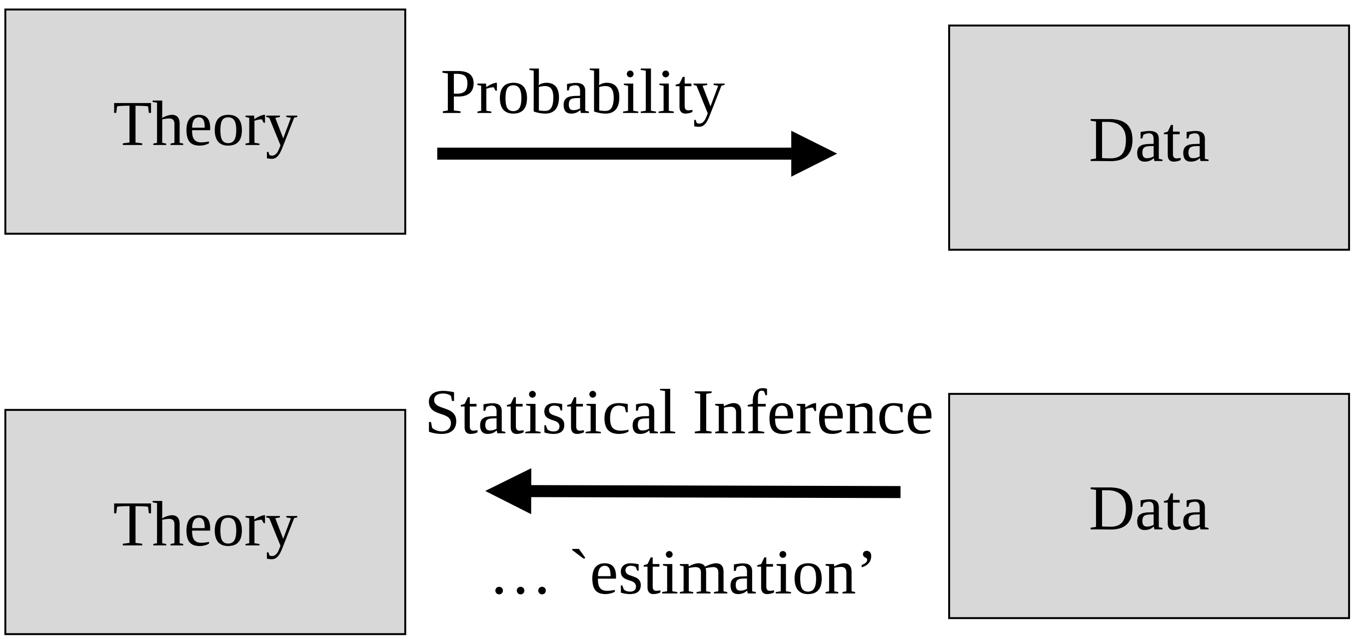
Distributions you need to know about in high energy physics ….

- Gaussian
- Poisson, Binomial
- Everything else (Top hat, Breit-Wigner, Landau, Gamma, Student's t, χ^2 …)

Extracting Information from Data: Estimation and Fitting

How to decide whether our data are well described by a model?

How to extract `best estimates' of model parameters from the data?



Rather than predicting data using theory …

… we want to extract best estimates of theory param's from data

What is an Estimator?

Formal Definition:

An estimator is a quantity extracted as a function of the data which gives a numerical value for a property of the parent distribution.

e.g. some possible estimators for the mean of a set of N points $\{x\} = x_1, x_2, x_3 \ldots x_N$:

$$\hat{\mu}(\{x\}) = x_1 \qquad \hat{\mu}(\{x\}) = \frac{x_{\max} + x_{\min}}{2} \qquad \hat{\mu}(\{x\}) = \frac{1}{N}\sum_i x_i$$

e.g. estimators for the variance of a set of N points $\{x\}$:

$$\hat{V}(\{x\}) = \frac{1}{N}\sum_i (x_i - \hat{\mu})^2 \qquad \hat{V}(\{x\}) = \frac{1}{N-1}\sum_i (x_i - \hat{\mu})^2$$

…. some estimators are better than others!

Desired Properties of Estimators

1. Consistency. An estimator is consistent if it tends towards the true value as $N \to \infty$.

2. Unbiased. An estimator is unbiased if it's expectation value is equal to the true value.

3. Efficient. An estimator is efficient if it's variance is small.

• Of examples on previous page, optimal choices are …

$$\hat{\mu}(\{x\}) = \frac{1}{N}\sum_i x_i$$

$$\hat{V}(\{x\}) = \frac{1}{N-1}\sum_i (x_i - \hat{\mu})^2$$

• Easy to see that these are consistent and unbiased.

• Efficiency is more complicated (see `Minimum Variance Bound' in text books)

Often, we want to "estimate" much more complicated things than means and variances – e.g. α_s from lots of different observables in jet data … need generalised estimation methods

The Likelihood Function

Suppose we have a set of data points $\{x_1, x_2, x_3 \dots x_N\}$ taken from a "*parent" probability distribution* characterised by a parameter *A*, to be estimated.

e.g. *A* might be the Higgs mass and $\{x_i\}$ might be LEP and TeVatron data which we believe obey the Standard Model

Define $P(x_1\,;\,a)$ = Probability of getting the result x_1 **if** *A*=a
Then the `likelihood' of *A*=a is just the combined probability that we get the set of points $\{x_i\}$ if *A*=a … i.e…

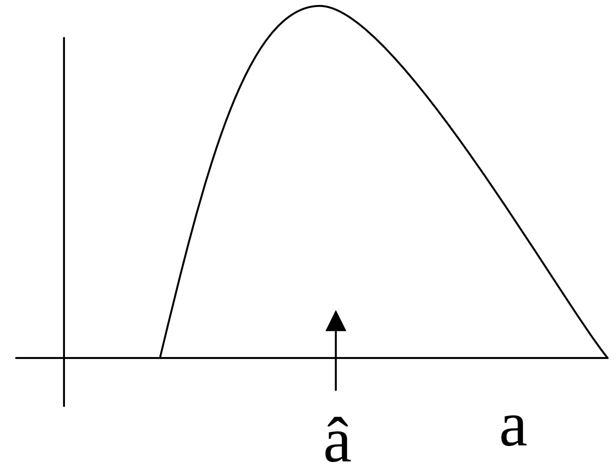
$$L(x_1, x_2, \dots x_N\,; a) = P(x_1; a)P(x_2; a)\dots .P(x_N\,; a) = \prod_{i=1}^{N} P(x_i\,; a)$$

Often more practical to work with `log likelihood'

$$\ln L(x_1 \dots x_N\,; a) = \sum_{i=1}^{N} \ln P(x_i\,; a)$$

Maximum Likelihood

To get best estimate $\hat{a}$ for a, just maximise the (log) likelihood of obtaining measured data with respect to a



… well defined general estimation method…

The maximum likelihood estimator $\hat{a}$ is the value of a which makes the probability of the observed results a large as possible.

$\hat{a}$ is obtained by the condition $\left.\dfrac{d \ln L}{da}\right|_{a=\hat{a}} = 0$

Maximum likelihood has lots of nice properties ….

…. consistent, unbiased and efficient for large N

…. Invariant: i.e, if we extract â and u is a function of a, û=u(â)

…. nice programs exist to do max / minimisation in real problems!

Numerical Max Likelihood Example

Suppose we have a sample $\{x_i\}$ = 0.89, 0.03, 0.50, 0.36, 0.49 …
… which we know comes from a parent distribution with
$P(x) \propto 1+a(x-0.5)$, but a is unknown.

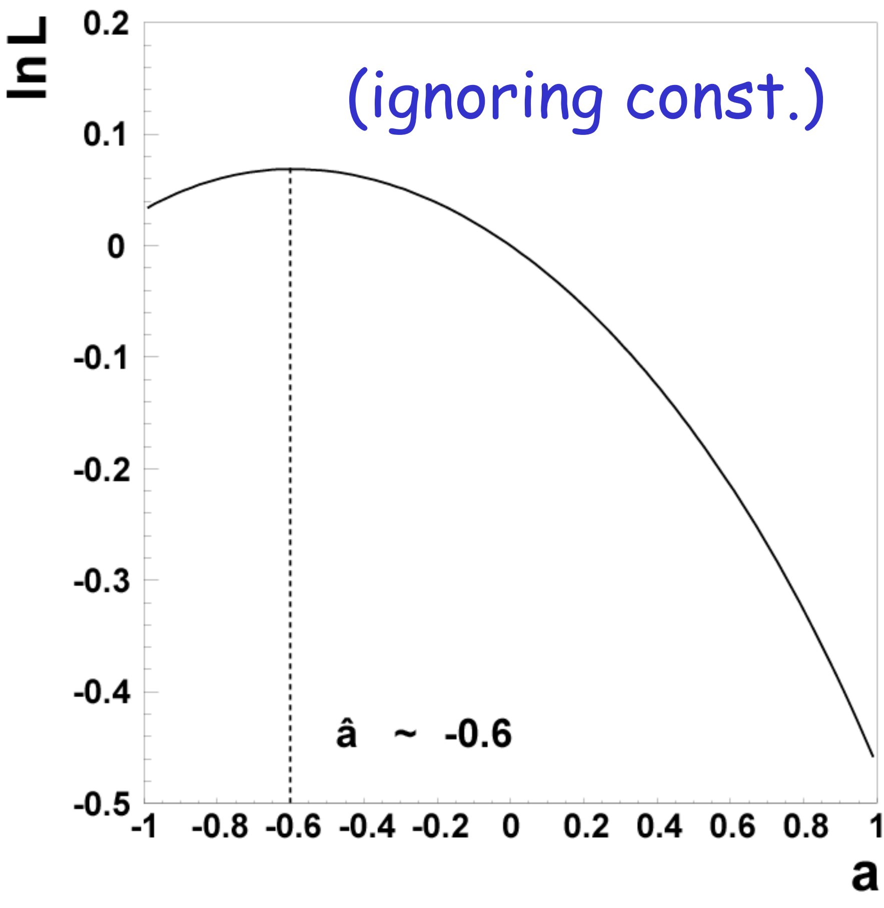
We have $P(0.89)=1+0.39a$ etc

Construct maximum likelihood:

$$\ln L=\ln(1+0.39a)+\ln(1-0.47a)+\ln(1)$$
$$\ln(1-0.14a)+\ln(1-0.01a)+\text{const.}$$

Plot this to see it has a maximum
Near a=0.6.

So log likelihood estimator is $\hat{a}$~-0.6



Important Example: Combining Measurements of the Same Quantity with Different Errors

What is our best estimate of a quantity, which has been measured several times (values x_i) with different errors σ_i ?

For each measurement, i,
$$P(x_i) = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-(x_i-\mu)^2/2\sigma_i^2}$$

So the log likelihood is:
$$\ln L = \sum_i -\ln \sigma_i \sqrt{2\pi} - \sum_i \frac{(x_i-\mu)^2}{2\sigma_i^2}$$

Differentiating:
$$\frac{\partial \ln L}{\partial \mu} = \sum_i \frac{x_i-\mu}{\sigma_i^2} = 0 \quad \text{at maximum}$$

Best estimate of mean is to weight the usual average by (variance)$^{-1}$ ...
$$\hat{\mu} = \frac{\sum_i (x_i/\sigma_i^2)}{\sum_i (1/\sigma_i^2)}$$

("the weighted mean")

Can also show that $\hat{\sigma}^2 = 1 \Big/ \sum_i (1/\sigma_i^2)$