

# Physics 222 UCSD/225b UCSB

## Lecture 8

A brief aside on statistics and data analysis.  
Beginning of Chapter 13 in H&M.

# Aside on Statistics

- Disclaimer:
  - I am by no means an expert on statistics. To me it's just a tool.
  - As all tools, the simpler the better !!!
  - ⇒ *The statistical interpretation of what I see in my ONE experiment is guided by what I expect to see if I repeated the experiment N times.*

# Aside on Philosophy

Frequentist statistics provides the usual tools for reporting objectively the outcome of an experiment without needing to incorporate prior beliefs concerning the parameter being measured or the theory being tested. As such they are used for reporting most measurements and their statistical uncertainties in High Energy Physics.

In Bayesian statistics, the interpretation of probability is more general and includes *degree of belief* (called subjective probability). One can then speak of a probability density function (p.d.f.) for a parameter, which expresses one's state of knowledge about where its true value lies. Bayesian methods allow for a natural way to input additional information, such as physical boundaries and subjective information; in fact they *require* as input the *prior* p.d.f. for the parameters, *i.e.*, the degree of belief about the parameters' values before carrying out the measurement. Using Bayes' theorem Eq. (31.4), the prior degree of belief is updated by the data from the experiment. Bayesian methods for interval estimation are discussed in Sections 32.3.1 and 32.3.2.6

*Ref. from PDG*

I guess I'm a frequentist,  
because it's simple,  
and I like my tools simple.

# Simple Classification of “Use Cases” for Statistics in HEP

- Parameter Estimation
  - I do an experiment to estimate a parameter and its error.
    - E.g. measuring a cross section, or a branching fraction, or deriving a theoretical parameter from a complex set of experimental measurements.
- Hypothesis Testing
  - I measure a distribution, and want to compare data against multiple hypotheses to determine which of them is the most likely.
    - Statistical significance in a new particle search.
      - Distinguish bkg only hypothesis from hypothesis that allows for signal.
    - Having observed a new particle, we then want to distinguish among a few possibilities for its spin, based on measurements of some angular distributions in its production and/or decay.

# Parameter vs Hypothesis

- Let's say we search for ZZ production at the Tevatron in the decay ZZ→4 leptons.

Channel	Observed	Background
without trackless e	2	$0.014^{+0.010}_{-0.007} \pm 0.003$
with trackless e	1	$0.082^{+0.089}_{-0.060} \pm 0.016$

The *error on the cross section is  $\sim 1/\sqrt{3} \sim 60\%$* .

However, the *probability for the background to fluctuate to the observed yield is  $\sim 10^{-5}$* .

We thus consider this an observation of ZZ production with a statistical significance of  $4.2\sigma$ .

# Meaning of $4.2\sigma \sim 10^{-5}$

If we draw a random number from a Gaussian distribution with mean=0 and unit width. The probability for picking a number larger than 4.2 is given as  $\sim 10^{-5}$ .

Having established the use of statistics in an example, let's now start over and define some of our terms.

I will follow a mix of PDG and Frodesen, Skjeggstad, and Tofte "Probability and Statistics in Particle Physics".

# Measurement, Random variable, and Probability Density Function

- A measurement  $x$  is generally viewed as randomly distributed according to some probability density function  $f(x)$ .
- To be a PDF the following must be true:
  - If the measurement of  $x$  is repeated many times, then the probability to find a value in the range of  $[x, x+dx]$  is given by  $f(x)dx$
  - The integral  $f(x)dx$  over all possible measurement outcomes,  $x$ , is 1. I.e.  $f(x)$  is normalized to 1 when integrated across the space of all possible values of  $x$ .

# Expectation Value

- Any function of  $u(x)$  is again a random variable, generally with a different pdf  $g(u)$  than  $f(x)$ .
- We refer to the “expectation value” of  $E(u)$  as:

$$E(u) = \int_{-\infty}^{+\infty} u(x) f(x) dx$$



# Marginal Distributions

- Let's assume we have two random variables  $x, y$  that have a joint PDF  $f(x, y)$  then we define the marginal distributions  $f_1(x)$  and  $f_2(y)$  as:

$$f_1(x) = \int_{-\infty}^{+\infty} f(x, y) dy$$

- The probability for  $x$  to be within  $[x, x+dx]$  if I couldn't care less about the value of  $y$  is given by  $f_1(x)dx$ .

# Conditional Distributions

- So what if I do care about  $y$ ?

$$f_4(x|y) = f(x, y) / f_2(y)$$

- The probability for  $x$  to be within  $[x, x+dx]$  under the condition of a fixed  $y$  is given by  $f_4(x|y) dx$  .
- And playing the same game the other way around:

$$f_3(y|x) = f(x, y) / f_1(x)$$

# Bayes Theorem:

$$f_4(x|y) = \frac{f_3(y|x)f_1(x)}{f_2(y)}$$

- Let's look at a trivial example, uncorrelated variables:  $f(x,y) = g(x) h(y)$

$$f_1(x) = \int_{-\infty}^{+\infty} f(x,y)dy = g(x) \int_{-\infty}^{+\infty} h(y)dy = g(x)$$

$$f_2(y) = \int_{-\infty}^{+\infty} f(x,y)dx = h(y) \int_{-\infty}^{+\infty} g(x)dx = h(y)$$

$$f_3(y|x) = \frac{f(x,y)}{f_1(x)} = h(y)$$

$$f_4(x|y) = \frac{f(x,y)}{f_2(y)} = g(x)$$

*Bayes Theorem for uncorrelated variables:  
 $g(x) = h(y)g(x)/h(y)$*

# Parameter Estimation

Here we review the frequentist approach to *point estimation* of parameters. An *estimator*  $\hat{\theta}$  (written with a hat) is a function of the data whose value, the *estimate*, is intended as a meaningful guess for the value of the parameter  $\theta$ .

There is no “a priori” right way of constructing the estimator. Instead, we define a set of “desirable” features we want from the estimation procedure.

- a) Consistency
- b) Lack of Bias
- c) Efficiency
- d) Robustness

# Consistency

- The estimator should converge to the true value as the amount of data used in the estimate increases to infinity.

## Lack of a bias

- For finite amounts of data, the expectation value of the estimator is equal to the true value.

# More on Bias

- In most cases, we are interested in estimating mean value and standard deviation simultaneously.
- In those cases we want both to be unbiased, i.e. we want the “pull distribution” to be normal (Gaussian with mean =0 and  $\sigma=1$ ).
  - Pull distribution is the pdf:  $f((x-\text{mean})/\sigma)$

# Efficiency

- It can be shown under very general conditions that the minimum variance of an estimator is given by the Cramer-Rao bound.
- An estimator is called efficient if its variance is minimal in the above sense, i.e. Cramer-Rao inequality becomes an equality.
- E.g.: You could use either the median or the mean as estimator of the peak of a Gaussian distribution. Both are consistent and unbiased. However, only the mean is efficient.

# In Practice

- We pick a procedure to estimate the physics quantity of interest.
- We use Monte Carlo methods to repeat our experiment many times, and thus study the properties of our estimator.
  - We plot the pull distribution for realistic sample sizes -> study bias.
  - We plot the pull distribution for “large” sample sizes -> study consistency.
  - Efficiency is studied less often. Some people, myself included, prefer Maximum Likelihood Method for parameter estimation because it leads to efficient estimators if they exist at all.



# Maximum Likelihood Method

- If the pdf is known a priori, and the different data points measured are mutually independent, then a likelihood function can be constructed by forming the joint probability of all measured data points:

$$L(\theta) = \prod_i f(x_i|\theta)$$

The ML method simply says that you obtain  $\theta$  by maximizing  $L(\theta)$  given a set of measurements  $x_i$ .

# In Practice

- ML fits are by far the most desirable “multi-variate” technique for deriving estimates of parameters of a physical theory.
- They require/allow you to develop a physical model of your experiment, parts of which you can often test via auxiliary measurements.
- It’s generally straightforward to build into your code that implements the ML fit the Monte Carlo methods to draw toy experiments from the distributions, and thus evaluate pull distributions.

# Cases when ML fit is impossible

- If the variables you measure have non-linear correlations that you do not understand a priori, then it is generally impossible to write down a sufficiently accurate pdf.
- In such cases you may have to either look for a different set of input variables, or resort to multivariate techniques with less well understood characteristics:
  - Neural Networks
  - Boosted decision trees
  - ...

# Hypothesis Testing

- Distinguish between competing physics hypotheses.
- Test consistency of different datasets taken at different times.
- Test consistency of data and Monte Carlo expectations.
- Establish the probability for a given signal to be consistent with a background fluctuation.

*Let's focus on the last, and discuss two simple examples.*

# Example: Yield in signal region

- Assume you chose a set of cuts to define a signal region.
- Assume you have a background expectation in the signal region  $bkg \pm \sigma$ .
- Draw toy experiments as follows:
  - Draw an expected bkg from a Gaussian with mean= $bkg$  and variance=  $\sigma^2$  .
  - Draw an actual number of bkg events from a poisson distribution with mean = expected background.
  - Record the actual bkg from a billion of such experiments.
  - Define the p-value as:  
(# of toy experiments with actual bkg  $\geq$  yield in data) / 1 billion.

# Example: Likelihood ratio

- Assume you have a signal likelihood  $L_s$  and a background likelihood  $L_{\text{bkg}}$  defined for your data. Define  $LR = L_s / L_{\text{bkg}}$  or  $LLR = \log LR$ .
- Draw 1 billion experiments from background only Monte Carlo, and record LR for each.
- Your p-value is defined as:

(# of toy experiments with  $LR \geq LR$  in data) / 1 billion

# Interpretation of p-value

- It is an arbitrary, but common, criteria to require  $> 5\sigma$  significant excess before you call it an “observation”.
- This means that the p-value has to be  $< 2.85 \cdot 10^{-7}$ , i.e. less than the area in the one-sided tail,  $5\sigma$  away from the mean of a Gaussian distribution.
- In some cases, where the interpretation of “success” may include fluctuations in both directions, a p-value  $< 5.7 \cdot 10^{-7}$  may be considered sufficient for an observation.

# An example pathology

- Let's assume your model for both signal and bkg are gaussian pdfs with unit variance, but means:
  - Signal = 0
  - Bkg = +1.3
- Let's assume you have 11 events in your sample, 7 of which are below -1, and 4 between 0.5 and 1.5.
- **What do you do?**



# Lesson learned from pathology

- You are allowed to be “a little lucky”, and still claim an observation.
- However, if your distribution in data is inconsistent with both signal and bkg hypothesis, then you’ve got to pause.
- Chances are the data you have is just plain junk that is not modeled by either your signal or your background hypothesis.

# Back to Physics

- Weak Neutral Currents
- Start of Chapter 13

# Weak Neutral Currents

- “*Observation of neutrino-like interactions without muon or electron in the gargamelle neutrino experiment*” Phys.Lett.B46:138-140,1973.
- This established weak neutral currents.

$$M = \frac{4G}{\sqrt{2}} 2\rho J_{\mu}^{NC} J^{NC\mu}$$

$\rho$  allows for different coupling from charged current.

$$J_{\mu}^{NC}(q) = \bar{u} \gamma_{\mu} \frac{1}{2} (c_V - c_A \gamma^5) u$$

$c_V = c_A = 1$  for neutrinos, but not for quarks.

*Experimentally: NC has small right handed component.*

# EWK Currents thus far

- Charged current is strictly left handed.
  - EM current has left and right handed component.
  - NC has left and right handed component.
- => Try to symmetrize the currents such that one  $SU(2)_L$  triplet that is strictly left-handed, and an  $SU(2)_L$  singlet.

# Starting with Charged Current

- Follow what we know from isospin, to form doublets:

$$\chi_L = \begin{pmatrix} \nu \\ e^- \end{pmatrix}_L; \tau_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}; \tau_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

$$J_\mu^\pm(x) = \bar{\chi}_L \gamma_\mu \tau_\pm \chi_L$$

$$J_\mu^3(x) = \bar{\chi}_L \gamma_\mu \frac{1}{2} \tau_3 \chi_L = \frac{1}{2} \bar{\nu}_L \gamma_\mu \nu_L - \frac{1}{2} \bar{e}_L \gamma_\mu e_L$$

***We thus have a triplet of left handed currents  $W^+, W^-, W^3$ .***

# Hypercharge, $T^3$ , and $Q$

- We next take the EM current, and decompose it such as to satisfy:

$$Q = T^3 + Y/2$$

$$j_{\mu}^{em} = J_{\mu}^3 + \frac{1}{2} j_{\mu}^Y$$

- The symmetry group is thus:  $SU(2)_L \times U(1)_Y$
- And the generator of  $Y$  must commute with the generators  $T^i$ ,  $i=1,2,3$  of  $SU(2)_L$ .
- All members of a weak isospin multiplet thus have the same eigenvalues for  $Y$ .

# Resulting Quantum Numbers

Lepton	T	T <sup>3</sup>	Q	Y
$\nu$	1/2	1/2	0	-1
$e^-_L$	1/2	-1/2	-1	-1
$e^-_R$	0	0	-1	-2

Quark	T	T <sup>3</sup>	Q	Y
$u_L$	1/2	1/2	2/3	1/3
$d_L$	1/2	-1/2	-1/3	1/3
$u_R$	0	0	2/3	4/3
$d_R$	0	0	-1/3	-2/3

You get to verify the quark quantum numbers in HW3.

# Now back to the currents

- Based on the group theory generators, we have a triplet of  $W$  currents for  $SU(2)_L$  and a singlet “B” neutral current for  $U(1)_Y$ .

*Basic EWK interaction:*

$$-ig(J^i)^\mu W_\mu^i - i\frac{g'}{2}(J^Y)^\mu B_\mu$$

- The two neutral currents B and  $W^3$  can, and do mix to give us the mass eigenstates of photon and Z boson.



# $W^3$ and B mixing

- The physical photon and Z are obtained as:

$$A_\mu = W_\mu^3 \sin \theta_W + B_\mu \cos \theta_W$$

$$Z_\mu = W_\mu^3 \cos \theta_W - B_\mu \sin \theta_W$$

- And the neutral interaction as a whole becomes:

$$\begin{aligned} & -ig(J^3)^\mu W_\mu^3 - i\frac{g'}{2}(J^Y)^\mu B_\mu = \\ & = -i\left(g \sin \theta_W J_\mu^3 + g' \cos \theta_W \frac{J_\mu^Y}{2}\right) A^\mu \\ & \quad - i\left(g \cos \theta_W J_\mu^3 - g' \sin \theta_W \frac{J_\mu^Y}{2}\right) Z^\mu \end{aligned}$$

# Constraints from EM

$$ej^{em} = e \left( J_\mu^3 + \frac{J_\mu^Y}{2} \right) = -i \left( g \sin \theta_W J_\mu^3 + g' \cos \theta_W \frac{J_\mu^Y}{2} \right)$$

$$\Rightarrow g \sin \theta_W = g' \cos \theta_W = e$$

$$\Rightarrow g' = \frac{\sin \theta_W}{\cos \theta_W} g$$

We now eliminate  $g'$  and write the weak NC interaction as:

$$-i \frac{g}{\cos \theta_W} \left( J_\mu^3 - \sin^2 \theta_W j_\mu^{em} \right) Z^\mu \equiv -i \frac{g}{\cos \theta_W} J_\mu^{NC} Z^\mu$$

# Summary on Neutral Currents

$$j_{\mu}^{em} = J_{\mu}^3 + \frac{1}{2} j_{\mu}^Y$$

$$J_{\mu}^{NC} = J_{\mu}^3 - \sin^2 \theta_W j_{\mu}^{em}$$

This thus re-expresses the “physical” currents for photon and Z in form of the “fundamental” symmetries.





