Data Analysis Example

• Let’s do a simple example analysis together:
  • Start with all data your experiment has collected
  • Select events that have exactly two good energetic muons (momentum in the transverse plane $p_T > 20$ GeV/c)
  • Calculate the following quantity using measured momenta $p_1$ and $p_2$ of the two muons:
    $$m_{\mu\mu} = (p_1 + p_2)^2 - (\vec{p}_1 - \vec{p}_1)^2$$
  • Require $m_{\mu\mu} > 70$ GeV and make a plot of this quantity

• Simple? Yea, but we just discovered the Z boson
  – Too bad we are 40 years late, otherwise we would be in line for the Nobel prize
Measurements

• Let’s come back to our Z boson “discovery”
• Once you discover it, you will want to find these:
  • Mass $m_0$ of the new particle, production rate $S$, and width $\Gamma$ of the new particle
    • Lifetime $\tau \sim 1/\Gamma$
• How? Fit a function to the data to measure parameters
  • A simple function would do okay here:
    • $f(m_0, \sigma, S, B) = \frac{S}{\sigma \sqrt{2\pi}} e^{-(m-m_0)^2/2\sigma^2} + \frac{B}{m^2}$
    • Second term is added to describe backgrounds
• Method of least squares: obtain parameters by finding the minimum in 4-dimensional space $(m_0, \sigma, S, B)$ of the following
  $$\chi^2 = \sum_{i=1}^{N_{\text{bin}}} \frac{(N_i - f(m))^2}{\delta_i^2}$$
  $\delta_i = \sqrt{N_i}$ (assume poisson error in each bin assumed)
• Caveat: MLS does not work well if in some bins $N_i$ is small or zero
Probability and Statistics

• Important when looking for small signals when you have little opportunity to quickly increase your statistics
  • Unimportant when you do something that has large statistics
  • Unimportant when you do something that you can easily reproducible

• In the world of small numbers, everything boils down to probability of this versus that
  • A question of “Do I see signal or do I see background?” requires some sort of probabilistic qualification to be a well defined question
Probability

• Fairly intuitive term referring to “inexact” processes

• Poisson process example:
  • When something happens with a certain rate (frequency over time), how many of these occurrences will you get over an hour?
  • Slight simplification but close: you toss a coin every second and count how many times will get tails over an hour
    • If you repeat this experiment 1,000,000 times and plot the outcomes (the number of heads in each experiment) on the graph and then normalize.
    • You will get the Poisson distribution of getting N outcomes out of 1,000 when the expected rate is 500 (given that the probability is $p=0.5$, right?)
Poisson Probability

• Poisson process example:
  • Can figure this out by doing probability calculations explicitly: every outcome is independent and has a probability of 50% and you consider all possible outcomes, sum them up get the probability density function

\[ P(n, \lambda) = \frac{\lambda^n e^{-\lambda}}{n!} \]

• Note it starts looking more and more gaussian with more and more trials
Using Probability Distributions

• What can you do with this $P$?

- $P(n, \lambda) = \frac{\lambda^n e^{-\lambda}}{n!}$

• E.g. you can calculate the probability to “observe” 100 radioactive decays over an hour when you expect 10

• Or you can calculate the mean expected value (if you were to do this experiment million times and average out):
  - Integrate $n$ from zero to infinity with weight $P$
    - In this example probability is discrete, so you sum things up instead. The result is $<n>=\lambda$
  - Can also calculate second moment $\Sigma(n-\lambda)^2*P(n)$, which is the variance. The result is $\sigma^2=\lambda$
Probability in Statistics

• The usual “frequentist” definition of probability is:
  • If you had to repeat the same experiment many times, in the limit of infinite number of trials, what would be the frequency of something happening
    • This frequency is your probability
  • Assumes no prior knowledge or beliefs, just statistics in action

• One can also incorporate subjective belief
  • Could be prior knowledge like previous measurements of the same quantity or physical bounds (like a cross-section can’t be less than zero)
    • This is characteristic of Bayesian statistics
  • These are best discussed in examples
Other Statistical Distributions

• Gaussian or normal:
  \[ s(m) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(m - m_0)^2}{2\sigma^2}\right) \]

  • The mean is \( m_0 \), variance is \( \sigma^2 \)

• The central limit theorem:
  • For \( x_i \) following any distribution with finite mean and variance, \( y = \Sigma x_i \) follows gaussian distribution in the limit of large \( n \)
    • The sum of a large number of fluctuations \( x_i \) will be distributed as gaussian even if \( x_i \) themselves are not gaussian
Binomial

• When something has two outcomes (like a coin toss) and the probability of a specific outcome is $p$ (correspondingly the probability of the “other” outcome is $1-p$ in each trial)
  - The probability to observe $r$ successful outcomes out of $N$ trials given true probability of $p$:
    \[
    f(k; n, p) = \Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}
    \]
  - Where:
    \[
    \binom{n}{k} = \frac{n!}{k!(n-k)!}
    \]
Probability Density Function

• Consider Poisson distribution:
  • Can read it two ways:
    • The probability to observe particular values \( n \) for a fixed \( \lambda \)
      \[
P(n, \lambda) = \frac{\lambda^n e^{-\lambda}}{n!}
      \]
    • Or probability of particular values of \( \lambda \), given that the number of observed events is \( n \)
      • If you integrate Poisson distribution over all possible lambdas you will get 1 for any fixed \( n \), just as if you sum it up over all possible \( n \) for a fixed lambda
  • If this is allowable, you have a joint probability density function (p.d.f.), which you can view either as a function of the data \( (n) \) or the parameters \( (\lambda) \)
Measurements and Statistics

- A typical example is a fit, in which you want to measure some unknown parameters of the “theory”
  - However, when you start thinking of it, there is a lot of conditionality in that, e.g. you assume that at least the parameterization of the “theory” is true – how do you know it to be the case?
  - Therefore, it may well be insufficient to do a simple fit, you often may need to answer many other questions
- Asking proper questions in statistics is very important as the answer and whether it is meaningful depend on the question
  - Will get back to it later
Estimators

• A typical fit problem: we have a set of N measured quantities $x = (x_1, \ldots, x_N)$ described by a joint p.d.f. $f(x; \mu)$, where $\mu = (\mu_1, \ldots, \mu_n)$ is set of n parameters whose values are unknown.

• Any measurement is an attempt to estimate true parameters of the “theory” using observed data.
  • But you are only estimating and your estimates may have biases, so people usually talk about “estimators”.
    • There are various ways to build estimators, most (if not all usually used) estimators converge to the true value of the parameters.
    • But not all can be equal (the “efficiency”, which is how fast your estimator converges to the true value with more data, can matter).

• Various fitting techniques are effectively different ways of picking the estimators.
Common Estimators

• For a set of data $x_i$ following the same pdf with a common mean, the estimators:
  • The mean:
    $$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} x_i$$
  • The variance:
    $$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \hat{\mu})^2$$
  • Variance of the estimator of the mean: $\sigma^2 / N$
  • Variance of the estimator of the variance is more complex (relies on higher moments of the true distribution):
    $$V[\hat{\sigma}^2] = \frac{1}{N} \left( m_4 - \frac{N-3}{N-1} \sigma^4 \right)$$
Maximum Likelihood

- Likelihood for N measurements $x_i$:
  - It is essentially a joint p.d.f. seen as a function of the parameters $\theta$
    $$L(\theta) = \prod_{i=1}^{N} f(x_i; \theta)$$
  - While $x_i$ could be any measurements, it’s easy to visualize it as a histogram of $x$ with $N$ bins; in each bin you calculate how probable it is to see $x_i$ for a particular $\theta$ (or set of $\theta$’s)
  - In ML, the estimators are those values of $\theta$ that maximize the likelihood function
    - One could use MINUIT to find the position of minimum for $-\log(L)$
TO BE CONTINUED