Today Lecture

- We will consider examples of using the ML technique
  - Building the likelihood function
  - Hypothesis testing
  - Comparison of two hypotheses
  - Comparison of two hypotheses when there are unknown parameters

- Other methods: MLS
Example

- We are searching for a resonance that has a known width but unknown mass and the cross-section on top of a known background
- The distribution being analyzed is the number of events per bin of invariant mass
  - Thus Poisson statistics

Signatures:
- Signal: a gaussian bump of unknown size in an unknown point in m
- Background: \( B/m \), where \( B \) is known, \( m \) is the mass

Shapes:
- Everything known except \( m_0 \) and normalization \( A \)
  - \( b(m) = 1/m \)
  - Normalization is known (=B)
Single Hypothesis

- **Single Hypothesis:**
  
  \[ L = \prod_{i=1}^{N_{\text{bin}}} P(N_i, \nu_i) \]
  
  - Where \( P(n, \nu) = \frac{\nu^n e^{-\nu}}{n!} \)
  
  - If S+B: \( \nu_i = S \times s_i(m) + B \times b_i(m) \) defined for specific S and \( m_0 \)
  
  - If B only: \( \nu_i = B \times b_i(m) \)

- Testing: make pseudo-data following background only model or (s+b) model and determine how often you get likelihoods smaller than the one observed in data

- Can test any specific hypothesis with fixed parameters, but can’t distinguish between the “signal with unknown parameters” and “background only”

- Note that in this definition when you test for background only hypothesis, you are asking a question of “how likely I am to observe a distribution as discrepant as the one I see in data”

  - It automatically accounts for bumps showing up at any mass
Two Hypotheses

• Hypothesis comparison:
  • Define a different test statistic:
    • \[ \lambda = \prod_{i=1}^{N_{\text{bin}}} P(N_i, \nu^S_i (S, m_0)) / \prod_{i=1}^{N_{\text{bin}}} P(N_i, \nu^B_i) \]
    • Well defined for any fixed S and m_0 and well suitable for comparisons of two specific hypotheses but not yet for any
  • In this case you ask a question of “assuming that it’s one (signal w/ fixed parameters) or the other (background only), what is the probability that what you see is a background fluctuation versus you seeing a signal with these parameters?”
Generalized Hypotheses

• Hypothesis comparison:
  • If the question you ask is how likely that this is signal of some mass and strength versus background, need to do more:
    • Define a different test statistic:
      • $\lambda = \max_{all \ S \ and \ m_0} \left( \prod_{i=1}^{N_{bin}} p(N_i, \nu^s_i(S, m_0)) \right) / \prod_{i=1}^{N_{bin}} p(N_i, \nu^b_i)$
      • You are now watching for “worst possible” signal-like fluctuation anywhere in S and m₀ space
      • If max is hard to imagine, think of lambda as a function of S and m₀, if it’s an analytical function, you can always find max or minimum
    • Can run now pseudo-experiments using background only hypothesis and see how often you get lambda even less than for the distribution in data
    • Tricks of taking maximum usually referred as “profiling” as in “profile likelihood”
Example We Looked at Before

- The p-value calculated here is almost what we just defined.
- One exception: the statistic was not max’ed over all possible $m_0$, but separately calculated for each $m_0$.
  - This is “local p-value” versus “global” in our definition.
  - Still scan over all possible signal strengths.
- You get p-value versus mass - nice, but the drawback is the look elsewhere effect.
  - If we were to maximize over $m_0$ as well, we would get a single number.
  - For a given data distribution, the “global” p-value we get will be higher.
Example We Looked at Before

- **Black line** – the **background only hypothesis**:
  - Make pseudoexperiments using background only model, run the machinery and calculate the p-value
- **It shows bumps at three masses** – this is where the fluctuations have happened in data
  - Local p-value is the probability of this being background rather than signal plus background assuming signal has mass $m_0$
  - Once you allow all possible masses to get global p-value, it degrades from 2.3 to 0.8 sigma
    - Ballpark p is increased by ~ ratio of the x-range divided by signal width in x
Example We Looked at Before

- **Blue line – signal+background hypothesis:**
  - Make pseudoexperiments using signal+background model, where signal strength is what SM says it should be
    - But estimate the p-value by doing pseudoexperiments with no signal
  - Scan over all possible masses run the machinery and calculate the median value of the p-value at the “correct” mass

- **Red line:**
  - Pick a single value of mass, generate pseudoexperiments following this hypothesis, calculate p-value versus mass
  - Essentially what you would expect on average if the signal was there
Other Ways for Hypothesis Testing

• The ML technique is powerful, but the machinery involved is complex and slow
  • In simple cases you may not need it, e.g. if you are comparing data and a fit, there are simpler ways to establish if the two are consistent with each other or not

• The Method of Least Squares (chi-square):
  \[
  \chi^2(\theta) = -2 \ln L(\theta) + \text{constant} = \sum_{i=1}^{N} \frac{(y_i - F(x_i; \theta))^2}{\sigma_i^2}
  \]

• Estimators for parameters are values that maximize the measured value of c2
  • As you see, it is related to ML method
  • In Poisson case:
    • Compare the number of observed events to the expected rate, variance in Poisson case is the mean rate
Distribution for $\chi^2$

- P-value versus $\chi^2$
  - $N$ – the number of degrees of freedom
    - The number of bins for testing a fixed hypothesis
  - Smaller $\chi^2$ – larger p-value
Distribution for $\chi^2$

- Rule of thumb: the reduced $\chi^2$ (divided by the number of degrees of freedom) should be about 1 (that’s where the p-value is about 50%)
- Too large $\chi^2$ indicates small p-value and the small probability that the data and hypothesis are compatible
- Too small value usually indicates that you made a mistake somewhere
Next Time

• We will focus on interpreting the likelihood functions to extract parameters
  • Parameter estimation and conventions for reporting measurements
  • Limit setting procedures
  • Some related topics: binned likelihood versus un-binned likelihood

• Combining datasets
  • Importance of likelihood functions