Physics 736

Experimental Methods in Nuclear-, Particle-, and Astrophysics

- Statistics and Error Analysis -

Karsten Heeger
heeger@wisc.edu
Schedule

• No lecture next Monday, April 8, 2013.

• Walter will be running review session on error analysis and statistics.

• He will available to answer questions about homework #7.

• *Make-up lecture on Fri after colloquium?*
Statistics & Error Analysis

Topics

• introduction to statistics and error analysis
• probability distributions
• treatment of experimental data
• measurement process and errors
• maximum likelihood
• parameter estimation
• method of least squares
• Bayesian and frequentist approach
• hypothesis testing and significance
• confidence intervals and limits
Measurements and Limits
Confidence Levels
Measurement is a random process described by a probability distribution. How confident are we in our measurement? The best estimate is the central value of the distribution, and the standard deviation indicates the spread or variability of the measurements around the best estimate.
Gaussian distribution

$1\sigma = 68\%$

$2\sigma = 95\%$

$3\sigma = 99\%$
Data with Error Bars

For ±1σ,

1/3 of data should be outside fit
One-parameter Confidence Level

\[ \text{Area} = \frac{1 - C}{2} \]

\[ \text{Area} = \frac{1 - C}{2} \]
Multiparameter Confidence Levels

allow all parameters to vary

hold some parameters fixed, vary only subset
Confidence Intervals: Measurements and Limits

rate or flux or # of events

confidence interval (CL=68.3%)

confidence interval (CL=99%)
Issue of Coverage

- Confidence intervals 
  undercover
- Measurement 
  pretends to be 
  more accurate 
  than it actually is

- Correct 
  coverage

- Confidence intervals 
  overcover (i.e. 
  are too 
  conservative)
- Reduced power 
  to reject wrong 
  hypotheses

Proper coverage can be tested by Monte Carlo simulations
Sampling and Parameter Estimation
Maximum Likelihood
Method of Least Squares
Sampling and Parameter Estimation

best value minimizes variance between estimate and true value

• method of estimation
  – 1) determine best estimate
  – 2) determine uncertainty on best estimate
Maximum Likelihood Estimation

• general method of parameter estimation when functional form of parent distribution is known
• for large samples the ML estimators are normally distributed, and hence the variances are easy to determine
• for small samples, ML estimators possess most good properties

\[ L(x_1, x_2, ..., | \theta) = \prod f(x_i, \theta) \]

estimate \( \hat{\theta} \) is the value which maximizes \( L \)
Maximum Likelihood Estimation

\[ L(x_1, x_2, ..., | \theta) = \prod f(x_i, \theta) \]

Since \( L \) and \( \ln L \) attain their maximum values at the same point one usually uses \( \ln L \) since sums are easier to work with than products:

\[ \ln L = \sum_i \ln(f(x_i | \theta)) \]

Normally point of maximum likelihood is found numerically.
Maximum Likelihood Estimation

• Properties of ML estimators

  – invariant
    • under parameter transformation
  – consistent
    • estimators converge on true parameter
  – unbiased
    • sometimes biased for finite samples.
      \( \hat{\theta} \) may be unbiased but \( u(\hat{\theta}) \) may be biased
  – efficient
    • if a sufficient estimator exists, the ML method will produce it
Examples of Likelihood Distributions

central values and 1\(\sigma\) intervals

\[
\ln L(\hat{\theta} + \sigma(\hat{\theta})) = \ln L_{\text{max}} - 0.5
\]

uncertainty is deduced from position where \(\ln L\) is reduced by 1/2

even applies for non-Gaussian likelihood
Examples of Likelihood Distributions

central values and 1σ intervals

asymmetric errors

e.g. $4.0^{+2.5}_{-1.25}$
Likelihood for Two Parameters

Given $L(x|\theta_1, \theta_2)$ plot contours of constant likelihood in the $\theta_1, \theta_2$ plane.

To find the uncertainty, plot contour with

$$\ln L(\hat{\theta} + \sigma(\hat{\theta})) = \ln L_{\text{max}} - 0.5$$

and look at the projection of the contour on the two axes.

Using correct method, uncertainties do not depend on correlation of variables.
Likelihood Function and Binned Data

- If the sample is very large and \( f(x|\theta) \) is complex, computation can be reduced by grouping the sample into bins and write \( L \) as product of probability of finding \( n \) entries in each bin.

\[
L(n_1, n_2, \ldots | \theta) = n! \prod (n_i !)^{-1} p_i^{n_i}
\]

\( p_i = \) probability for bin \( i \)

\[
lnL = \sum n_i ln p_i(\theta)
\]

- There will be some **loss of information** by binning the data, but as long as the variation in \( f \) across each bin is small there should be no great loss in precision \( \hat{\theta} \).
Method of Least Squares

• relate data and model
• frequently used method for parameter estimation but no general optimal properties to recommend it
• if parameter dependence is linear, method of least squares (LS) produces unbiased estimators of minimum variance

\[ S = \sum_i \left( \frac{y_i - f(x_i|\theta_j)}{\sigma_i} \right)^2 \]

if data are Gaussian distributed then LS is equivalent to ML method

if in addition, observables are linear functions of the parameters, the S will follow \( \chi^2 \) distribution
Method of Least Squares

• Quality of Fit: How good is the function that we determined?
  – The value of S is a measure of the agreement between the fitted quantities and the measurements.

• Degrees of Freedom
  – If data is Gaussian distributed then S follows a $\chi^2$ distribution with N degrees of freedom.
    – N data points
      • in general n degrees of freedom

  – N data points, m parameters of linear model
    • N-m degrees of freedom
Degrees of Freedom

Figure 32.2: The ‘reduced’ $\chi^2$, equal to $\chi^2/n$, for $n$ degrees of freedom. The curves show as a function of $n$ the $\chi^2/n$ that corresponds to a given $p$-value.
Data with Error Bars

For ±1σ,

1/3 of data should be outside fit
Method of Least Squares

Binned data

two common choices
  - equal width
  - equal probability

Must not choose binning to make $S$ as small as possible. In this case it would no longer follow $\chi^2$ distribution.

Necessary to have several entries in bin to approximate Gaussian statistics (e.g. more than 5 entries).
Method of Least Squares

Goodness of Fit

- Least squares is a measure of the agreement between the fitted quantities and the measurements.

Since $X_{\text{min}}$ follows a known ($\chi^2$) distribution (for a linear model with Gaussian distributed observables), the value of $X_{\text{min}}^2$ obtained in a particular case is a measure of the agreement between the fitted quantities $\hat{\eta}$ and the measurements $y$.

A larger $X_{\text{min}}^2$ corresponds to a poorer agreement. The probability of obtaining a value of $X_{\text{min}}^2$ or larger is

$$P_{X_{\text{min}}^2} = \int_{X_{\text{min}}^2}^{\infty} f(\chi^2|N) \, d\chi^2 = 1 - F(X_{\text{min}}^2|N) = \alpha$$

where $F$ is the cumulative distribution.
Method of Least Squares

Example of $X^2$ Test

A simulated data sample is shown with distribution function that was not used to generate the data.

There are 20 bins. Distribution function was normalized to match number of events in data sample.

Does the model fit the data?

The value of $\chi^2$ for this distribution is 25.2 for 19 d.o.f. resulting in $P_{\chi^2} = 0.16$
Goodness of Fit Tests

Example of $X^2$ Test

A simulated data sample is shown with distribution function that was not used to generate the data.

There are 20 bins. Distribution function was normalized to match number of events in data sample.

Does the model fit the data?

The value of $\chi^2$ for this distribution is 25.2 for 19 d.o.f. resulting in $P_{\chi^2} = 0.16$

**BUT .... does this look right to you?**

*We will get back to this question.*
Example: CDF Results

Invariant Mass Distribution of Jet PairsProduced in Association with a $W$ boson in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV

T. Aaltonen,$^{21}$ B. Álvarez González$^w,^9$ S. Amerio,$^{41}$ D. Amidei,$^{32}$ A. Anastassov,$^{36}$ A. Annovi,$^{17}$ J. Antos,$^{12}$ G. Apollinari,$^{15}$ J.A. Appel,$^{15}$ A. Apresyan,$^{46}$ T. Arisawa,$^{56}$ A. Artikov,$^{13}$ J. Asaadi,$^{51}$ W. Ashmanskas,$^{15}$ B. Auerbach,$^{59}$ A. Aurisano,$^{51}$ F. Azfar,$^{40}$ W. Badgett,$^{15}$ A. Barbaro-Galtieri,$^{26}$ V. E. Barnes,$^{46}$ B.A. Barnett,$^{23}$ P. Barria$^{dd,44}$ P. Bartos,$^{12}$ M. Bause$^{bb,41}$ G. Bauer,$^{30}$ F. Bedeschi,$^{44}$ D. Beecher,$^{28}$ S. Behari,$^{23}$ G. Bellettini$^{cc,44}$

Karsten Heeger, Univ. of Wisconsin  Physics 736, Spring 2013
FIG. 1: The dijet invariant mass distribution. The sum of electron and muon events is plotted. In the left plots we show the fits for known processes only (a) and with the addition of a hypothetical Gaussian component (c). On the right plots we show, by subtraction, only the resonant contribution to $M_{jj}$ including $WW$ and $WZ$ production (b) and the hypothesized narrow Gaussian contribution (d). In plot (b) and (d) data points differ because the normalization of the background changes between the two fits. The band in the subtracted plots represents the sum of all background shape systematic uncertainties described in the text. The distributions are shown with a 8 GeV/$c^2$ binning while the actual fit is performed using a 4 GeV/$c^2$ bin size.

Example: CDF Results
Confidence Intervals and Limits
Hypothesis Testing
Goodness of Fit
Confidence Intervals and Limits
Confidence Intervals and Limits

Gaussian distribution

1σ = 68%
2σ = 95%
3σ = 99%
Confidence Intervals: Measurements and Limits

What if some measurements are in a non-physical region?
Frequentist vs Bayesian Approach

Two philosophies

- **Bayesian approach**
  - probability = degree of belief that something will happen or that a parameter will have a given value

- **Frequentist approach**
  - probability = relative frequency of something happening
  - one can define frequentist probability for observing data (which are random) but not for the true value of a parameter
  - independent of observer
Frequentist vs Bayesian Approach

Two philosophies

– Bayesian approach
  • require as input the prior beliefs of the physicist doing the analysis. necessarily subjective, and not allowed in frequentist method.

– Frequentist approach
  • require as input probabilities of observing all data, including both the data actually observed and that which could have been observed (the Monte Carlo). not allowed in Bayesian method.