Likelihood Fits

Craig Blocker
Brandeis
August 19, 2002

Outline

I. What is the question?
II. Likelihood Basics
III. Mathematical Properties
IV. Uncertainties on Parameters
V. Miscellaneous
VI. Goodness of Fit

What is the Question?

Often in HEP, we make we make a series of measurements and wish to deduce the value of a fundamental parameter.

For example, we may measure the mass of many $B \rightarrow J/\psi\ K_S$ decays and then wish to get the best estimate of the B mass.

Or, we might measure the efficiency for detecting such events as a function of momentum and then wish to derive a functional form.

The question is: What is the best way to do this?

Likelihood Method

$P(X|\alpha) \equiv$ Probability of getting a measurement X on a given event.

α is a parameter of set of parameters, which P depends on.

Suppose we make a series of measurements, yielding a set of X_i's. The likelihood function is defined as

$$L = \prod_{i=1}^{N} P(X_i \mid \alpha)$$

The value of α that maximizes L is known as the Maximum Likelihood Estimator (MLE) of α, which we will denote as α^*.

Note that we often work with $\ln(L)$.

Example

Suppose we measure a variable x, which we believe is Gaussian, and wish to get the best estimate of the mean and width.

$$\mathbf{P(x\,|\,\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

$$\mathbf{\ln \mathit{L} = \text{-}N \ln\left(\sqrt{2\pi}\sigma\right) - \sum_{i=1}^{N} \frac{\left(x_i - \mu\right)^2}{\sigma^2}}$$

Maximizing with respect to μ and σ gives

$$\mu^* = \frac{\sum \mathbf{x_i}}{\mathbf{N}} = \bar{\mathbf{x}}$$

$$\sigma^{*2} = \frac{\sum \left(\mathbf{x_i} - \mu^*\right)^2}{\mathbf{N}} = \overline{\mathbf{x}^2} - \bar{\mathbf{x}}^2$$

Bias, Consistency, and Efficiency

What does "best" mean?

We want the estimate of the parameter to be close to the true value.

Unbiased $\Rightarrow \quad \overline{\alpha^*} = \alpha_0$

Consistent $\Rightarrow$ **Unbiased for large N**

Efficient $\Rightarrow \quad \overline{\left(\alpha^* - \alpha_0\right)^2}$ **is minimal**

Maximum Likelihood Estimators are NOT unbiased but are consistent and efficient for large N.

Other Types of Fits

Chi-square:

If data is binned and uncertainties are Gaussian, then maximum likelihood is equivalent to a χ^2 fit.

Binned Likelihood:

If data is binned and not Gaussian, can still do a binned likelihood fit. Common case is when data are Poisson distributed.

$$P_i = \frac{e^{-\bar{\mu}_i}\left(\bar{\mu}\right)^{n_i}}{n_i!}$$

$$\ln L = \sum_{\text{bins}} \ln P_i$$

Comparison of Fits

Chi-square:

❶ Goodness of fit.
❷ Can plot function with binned data.
❸ Data must be Gaussian, in particular, χ^2 doesn't do well with bins with a small number of events.

Binned likelihood:

❶ Goodness of fit?
❷ Can plot function with binned data.
❸ Still need to be careful of bins with small number of events (don't add in too many zero bins.

Unbinned likelihood:

❶ Usually most powerful.
❷ Don't need to bin data.
❸ Works well for multi-dimensional data.
❹ No goodness of fit estimate.
❺ Can't plot fit with data.

Examples

Consider again, the mean and width of a Gaussian.

$$\overline{\mu^*} = \overline{\frac{1}{N}\sum x_i} = \frac{1}{N}\sum \overline{x_i} = \frac{1}{N}\sum \mu_0 = \mu_0$$

$$\overline{\sigma^{*2}} = \frac{N}{N-1}\sigma_0^2$$

Note that the MLE of the mean is unbiased, But for the width squared is not. However,

$$\tilde{\sigma}^2 = \frac{\sum \left(x_i - \mu^*\right)^2}{N-1}$$

is unbiased.

Uncertainty on Parameters

Just as important as getting an estimate of a parameter is knowing the uncertainty of that estimate.
The maximum likelihood method also provides an estimate of the uncertainty.
For one parameter, *L* become Gaussian for large N. Thus,

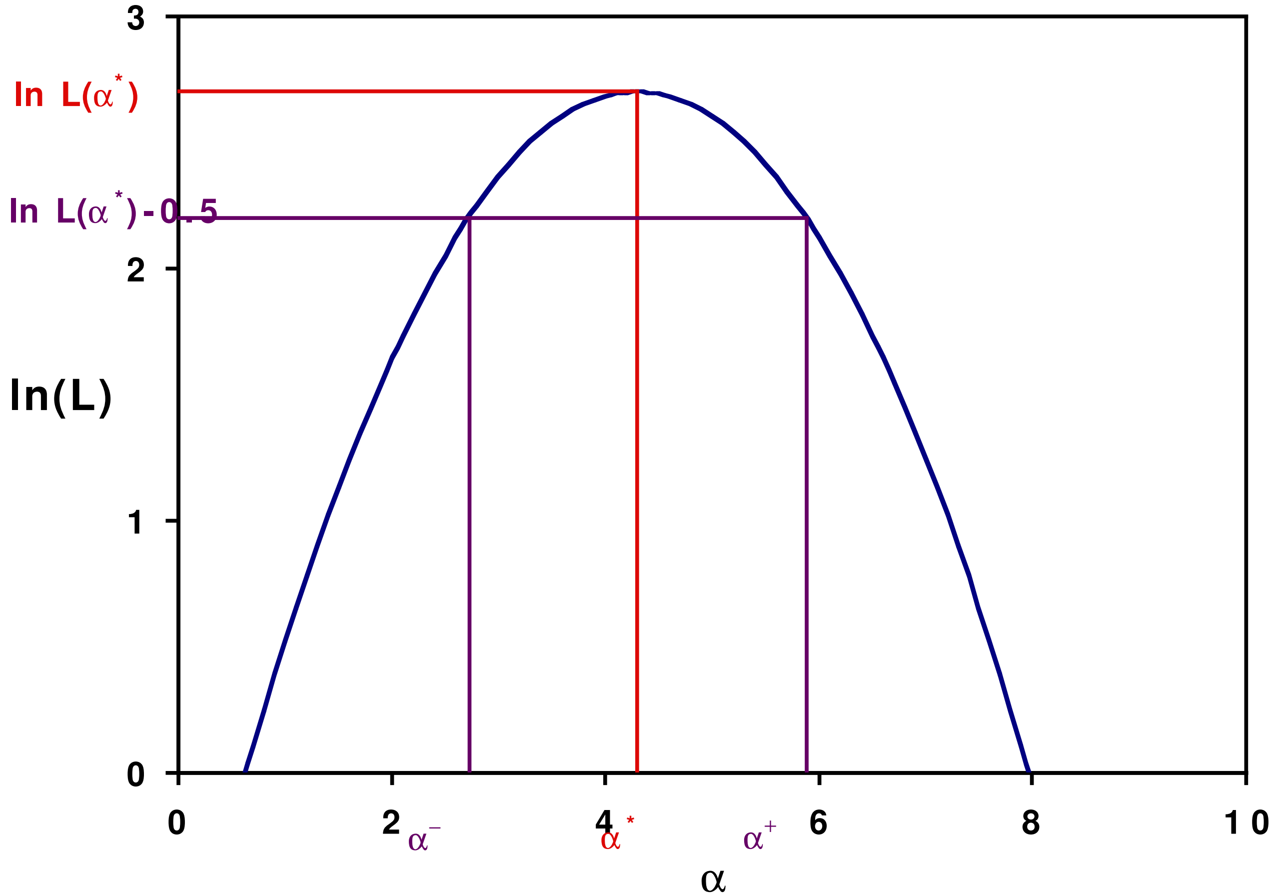
$$\ln L \cong \ln L^* + \frac{1}{2}\left.\frac{\partial^2 \ln L}{\partial \alpha^2}\right|_{\alpha=\alpha^*}\left(\alpha-\alpha^*\right)^2$$

$$\Rightarrow \Delta\alpha^2 \equiv \overline{\left(\alpha^* - \alpha_0\right)^2} = -\frac{1}{\left.\frac{\partial^2 \ln L}{\partial \alpha^2}\right|_{\alpha=\alpha^*}}$$

We usually write this as $\alpha = \alpha^* \pm \Delta\alpha$

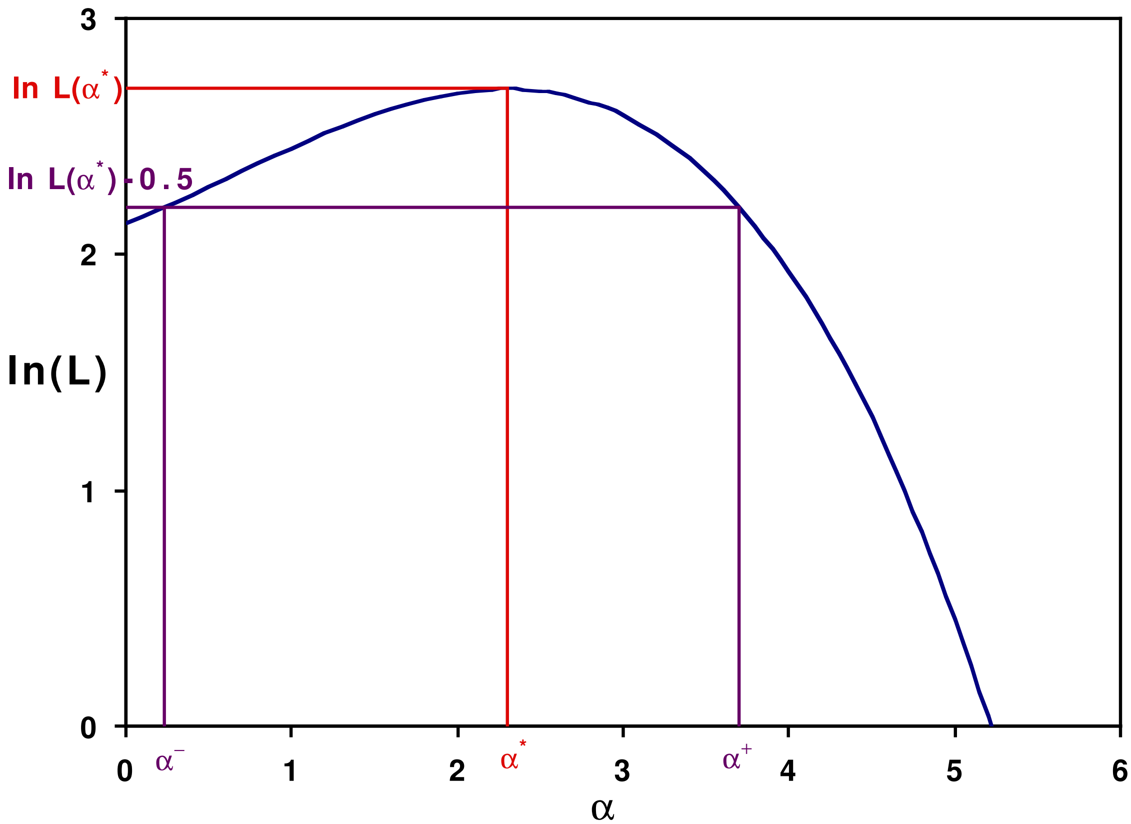
If $\alpha = \alpha^* \pm \Delta\alpha$, $\ln L = \ln L^* - \frac{1}{2}$

Likelihood Example



Asymmetric Uncertainties

Sometimes ln*L* may not be parabolic and there may be asymmetric uncertainties.

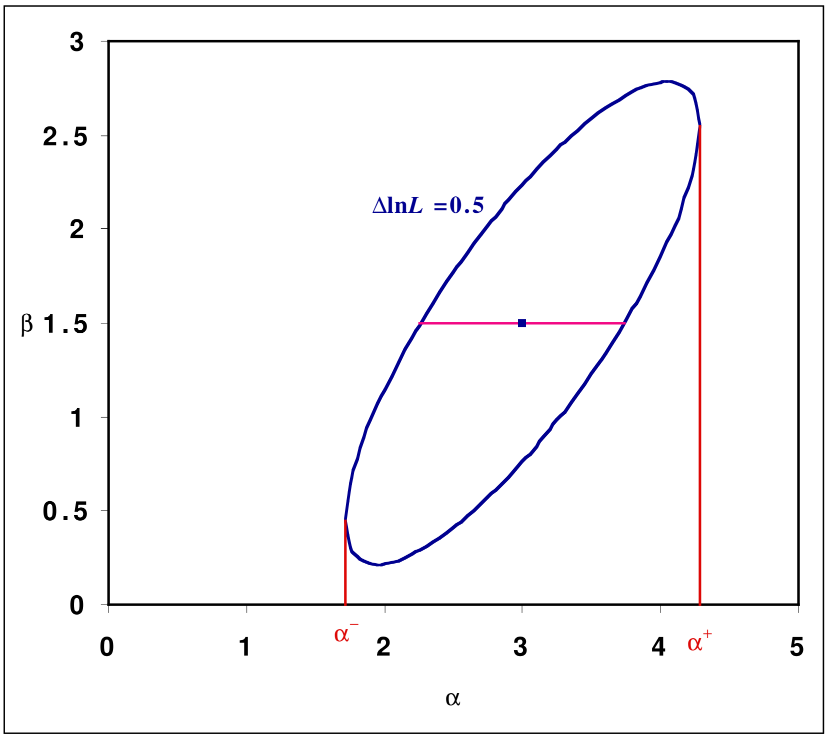


We write $\alpha = \alpha^{*}{}^{+\alpha_+}_{-\alpha_-}$

Note: the ΔlnL = 1/2 interval does NOT always give a 68% confidence interval.

Correlations

If there are multiple parameters, things are more complicated due to possible correlations.



Covariance matrix V is given by

$$\mathbf{V_{ij}} = \overline{(\alpha_i - \overline{\alpha_i})(\alpha_j - \overline{\alpha_j})}$$

V is equal to U^{-1}, where

$$\mathbf{U_{ij}} = -\frac{\partial^2 \ln L}{\partial \alpha_i \partial \alpha_j}$$

Normalization

Sometimes, people will say they don't need to normalize their probability distributions. This is sometimes true.

For the Gaussian example, if we omitted the normalization factor of $1/\sqrt{2\pi}\sigma$ **we get the mean correct but not the width. In general, if the normalization depends on any of the parameters of interest, it must be included.**

My advice is always normalize.

Extended Likelihood

Suppose we have a Gaussian mass distribution with a flat background and wish to determine the number of events in the Gaussian.

$$P = f_S \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(M-M_0)^2}{2\sigma^2}} + (1 - f_S)\frac{1}{\Delta M}$$

where f_S is the fraction of signal events and ΔM is the mass range of the fit.

We can fit for f_S and get Δf_S.
Nf_S is a good estimate of the number of events in the Gaussian, but $N\Delta f_S$ is not a good estimate of the uncertainty (because f_S is actually binominally distributed.

We can fix this by adding a Poisson term in the total number of events. This is called an Extended Likelihood fit.

Constrained Fits

Suppose there is a parameter in the likelihood that is somewhat known from elsewhere. This information can be incorporated in the fit.

For example, we we are fitting for the mass of a particle decay with resolution σ. Suppose the Particle Data Book lists the mass as $M_0 \pm \sigma_M$.

$$L = \frac{1}{\sqrt{2\pi}\sigma_M} e^{-\frac{(M-M_0)^2}{2\sigma_M^2}} \prod_i^N \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(m_i-M)^2}{2\sigma^2}} \right]$$

This is only useful of σ_M and the resolution of the fit ($\Delta\sigma$) are comparable.

Simple Monte Carlo Tests

It is possible to write simple, short, fast Monte Carlo programs that generate data for fitting.

❶ Tests likelihood function.
❷ Tests for bias.
❸ Tests that uncertainty from fit is correct.

This does NOT test the correctness of the model of the data. For example, if you think that some data is Gaussian distributed, but it is really Lorentzian, then the simple Monte Carlo test will not reveal this.

Goodness of Fit

Unfortunately, the likelihood method does not, in general, provide a measure of the goodness of fit (as a χ^2 fit does).

For example, consider fitting lifetime data to an exponential.

$$L = \prod_{i}^{N} \Gamma e^{-\Gamma t_i}$$

$$\Gamma^* = \frac{N}{\sum t_i}$$

$$L(\Gamma^*) = -N\left(1 + \ln \frac{\sum t_i}{N}\right)$$

Thus the value of *L* at the maximum depends only on the number of events and average value of the data.

Numerical Methods

Most likelihood fits have many parameters (perhaps scores) and can't be done analytically.

However, numerical methods are still very effective.

MINUIT is a powerful program from CERN for doing maximum likelihood fits.

Systematic Uncertainties

When fitting for one parameter, there often are other parameters that are imperfectly known.

It is tempting to estimate the systematic uncertainty due to these parameters by varying them and redoing the fit.

Because of statistical variations, this overestimates the systematic uncertainty (often called double counting).

Best way to estimate such systematics is probably with a high statistics Monte Carlo program.

Summary

Maximum Likelihood methods are a powerful tool for extracting measured parameters from data.

However, it is important to understand their proper use and avoid potential problems.