Some Comments on Likelihood Functions

The likelihood $\mathcal{L}$ is a function of the parameters of a statistical model. It is used to estimate the values of and uncertainties on those parameters for a given set of measurements. For an ensemble of $n$ measurements, the likelihood is defined as

$$\mathcal{L}(x; \theta) = \prod_{i=1}^{n} \mathcal{L}_i(x; \theta) = \prod_{i=1}^{n} f(x; \theta)$$

where $f(x; \theta)$ is the probability density function for the statistical model of interest. The best value of the parameters $\theta$ can be determined by maximizing the likelihood function, or equivalently, the log of the likelihood function.

$$\frac{\partial \ln \mathcal{L}}{\partial \theta} = \frac{\partial}{\partial \theta} \ln \prod_{i=1}^{n} \mathcal{L}_i$$

$$= \frac{\partial}{\partial \theta} \sum_{i=1}^{n} \ln \mathcal{L}_i$$

$$= 0$$

Often this procedure is described instead as minimizing $-\ln \mathcal{L}$ (minimizing the minus log likelihood).

The log likelihood can be Taylor expanded about its minimum. Since $\partial \mathcal{L}/\partial \theta|_{\theta=\theta_{\text{min}}} = 0$: $\ln \mathcal{L} = \ln \mathcal{L}_{\text{min}} + \frac{1}{2} \frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} |_{\theta=\theta_{\text{min}}} (\theta - \theta_{\text{min}})^2$.

$$2 (\ln \mathcal{L} - \ln \mathcal{L}_{\text{min}}) = \frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} |_{\theta=\theta_{\text{min}}} (\theta - \theta_{\text{min}})^2$$

In the limit of large $n$, the distribution $\mathcal{L}$ (due to the central limit theorem) becomes Gaussian. Since for a Gaussian distribution a change in $2 \ln \mathcal{L}$ of one unit corresponds to a $1-\sigma$ variation in the parameter $\theta$, the uncertainty on $\theta$ is given by:

$$\sigma_{\theta}^2 \equiv \left\langle (\theta - \theta_{\text{min}})^2 \right\rangle = -\frac{1}{\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2}}$$

Alternatively, the uncertainty on the estimated values of the parameters $\theta$ can be obtained by calculating the value of $\Delta \theta$ at which $-2 \ln \mathcal{L}$ increases by 1.0. In cases where $\ln \mathcal{L}$ is not parabolic, the uncertainties can be asymmetric.
The definitions above depend on the fact that the probability density function \( f(x; \theta) \) is normalized over the region of \( x \) where measurements can occur

\[
\int_{x_{\text{min}}}^{x_{\text{max}}} f(x; \theta) \, dx = 1
\]

For this reason, likelihood fits are not sensitive to the value of \( n \). It is possible to add a Poisson term to the likelihood function to include the number of events in the likelihood fit. When a fit includes such a term, it is called an extended likelihood fit.

**An example**

Suppose a set of measurements \( x_i \) are made in an experimental setup where the number of events as a function of \( x \) follows the distribution

\[
N(x) = A + Bx \quad \text{for } 0 < x < 10
\]

We would like to use the likelihood method to estimate the value of \( \kappa \equiv A/B \). Let’s see how to setup this problem.

The total number of events \( N_{\text{Tot}} \) can be determined

\[
N_{\text{Tot}} = \int_0^{10} A + Bx = \left[ Ax + \frac{1}{2} Bx^2 \right]_0^{10} = 10A + \frac{1}{2} (100B) = 10A + 50B
\]

and normalized probability density function \( f(x; \theta) \) is

\[
f(x; A, B) = \frac{1}{N_{\text{Tot}}} (A + Bx) = \frac{1}{10A + 50B} (A + Bx) = \frac{A}{10A + 50B} + \frac{Bx}{10A + 50B}
\]

\[
= 0.1 \left( \frac{\kappa}{\kappa + 5} + \frac{x}{\kappa + 5} \right)
\]
The overall likelihood function is therefore

\[ L(x; \kappa) = \prod_{i=1}^{n} 0.1 \left( \frac{\kappa}{\kappa + 5} + \frac{x_i}{\kappa + 5} \right) \]

and the log likelihood is

\[
\ln (L(x; \kappa)) = \sum_{i=1}^{n} \ln \left( 0.1 \left( \frac{\kappa}{\kappa + 5} + \frac{x_i}{\kappa + 5} \right) \right) \\
= \sum_{i=1}^{n} \ln \left( \left( \frac{\kappa}{\kappa + 5} + \frac{x_i}{\kappa + 5} \right) \right) + n \ln(0.1)
\]

If the values \( x_i \) are known, then \(-\ln (L(x; \kappa))\) can be minimized with respect to \( \kappa \). Because the last term in independent of \( x_i \), it merely adds a constant term to the log likelihood and is irrelevant for the minimization.

There are many programs available to do such minimization (Root for example, has a good interface to the Minuit minimization package). However, you can also find the minimum by by seeing how \( \ln (L(x; \kappa)) \) changes when \( \kappa \) is varied. An example root macro to generate fake data for the case \( A = 1, B = 2 \) and to use these data to determine \( \kappa \) can be found here:

http://physics.lbl.gov/shapiro/Physics226/myLikelihoodFit.C

The output of this macro is provided on the next page:
Determination of $\kappa$ from 1,000 MC events

True $\kappa=0.5$

Estimate of kappa: $0.53^{+0.24}_{-0.21}$

$\Delta(-2\ln(L))=1.0$