Duke University
Department of Physics

Physics 391 Fall Term 2010

WUN2K FOR LECTURE 7

These are notes summarizing the main concepts you need to understand and be able to apply.

- The best fit parameters to some function $f(x,\mathbf{a})$, denoted $\hat{\mathbf{a}}$, from the maximum likelihood method come from maximizing $\mathcal{L}(\mathbf{a}) = \Pi_{i=1}^{N} P_i$ as a function of $\mathbf{a}$, where $P = P_i(x_i; a_1, ..., a_m)$ is the normalized probability density for the function. (Here the notation $\mathbf{a}$ means a column vector of m parameters, $\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}$. To find the uncertainties on the parameters $\hat{\mathbf{a}}$:
  - On can use a set of toy Monte Carlos to find the spread: *i.e.* assume the best fit parameters are the true ones, select random fake data from the probability distribution corresponding to the function, and fit each fake data set to find the spread of best-fit parameters.
  - If the toy MC method is computationally too expensive, one can often instead assume that the likelihood distribution approaches a Gaussian near its maximum (which is true for a large number N of measurements) and estimate (for one parameter a) $\sigma_{\hat{a}}^2 = \langle \frac{-1}{\left(\frac{\partial^2 l}{\partial a^2}\right)|_{a=\hat{a}}} \rangle$,where l is the log likelihood. This is equivalent to finding the values of $a_{\pm}$ such that $l = l(\hat{a}) - \frac{1}{2} = l_{\max} - \frac{1}{2}$.
  - For more than one parameter $\mathbf{a}$, this generalizes to an error matrix for the parameters $M(\hat{a})_{ij}^{-1} = -\left(\frac{\partial^2 l}{\partial a_i \partial a_j}\right)|_{\mathbf{a}=\hat{\mathbf{a}}}$.
- Least squares: here we looked at the case of fitting to a function which is linear in the *parameters*, i.e. one can write $f(x;\mathbf{a}) = \sum_{j=1}^{m} c_j(x) a_j$,

or in matrix notation $\mathbf{f} = C\mathbf{a}$, where $C_{ij} = C_j(x_i)$. For the case where the measurements are in general correlated, the χ^2 can be written $\chi^2 = (\mathbf{y} - C\mathbf{a})^T M^{-1}(\mathbf{y} - C\mathbf{a})$. The best fit is the solution to $\nabla\chi^2 = 0$, and comes out to a straightforward matrix equation giving the vector of parameters, $\hat{\mathbf{a}} = [(C^T M^{-1} C)^{-1} C^T M^{-1}]\mathbf{y}$. The error matrix for the best-fit parameters is given by $M(\hat{\mathbf{a}}) = [C^T M^{-1}(\mathbf{y}) C]^{-1}$. Finding least-squares fits and their errors by this method can be quite fast.

- In practice, one often just uses a software package to do the fitting and to spit out error matrices for parameters. But it's good to understand the underlying ideas, because sometimes these packages can be "leaky abstractions"!