July 13, 2011

Statistical Methods in BaBar

Aaron Roodman

(lecture notes by M. Peskin)

In this lecture, I will describe the procedures for evaluating estimates and error intervals used by the BaBar experiment. Most of our results are based on the *maximum likelihood* method.

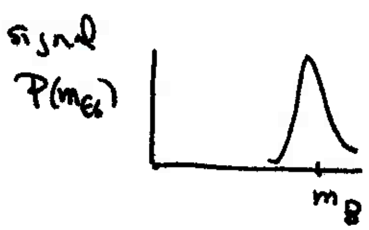
In particular, we typically use *unbinned maximum likelihood.* That is, we construct a likelihood function for the variables x_j that we actually measure, we evaluate the likelihood function at the measured values, and we maximize this function to obtain our estimates.

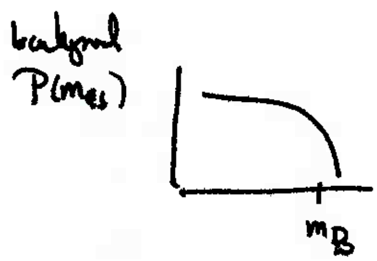
A typical BaBar analysis will be based on variables such as: m_{ES} and ΔE (variables used to kinematically identify $B\overline{B}$ events), NN (the output of a neural network classifier), Δt and $\sigma_{\Delta t}$ (the value and measurement error for the B lifetime), and information from a flavor tag. The PDF used in the likelihood function will also depend on some parameters α_k. These include underlying physics parameters and parameters such as the width of the Gaussian distribution of m_{ES} in signal events. The likelihood also depends on the number of events in a way that I will be more explicit about below.

For each event, the PDF is built up as a product of factors

$$P(x_j;\alpha_k) = P(m_{ES})\,P(\Delta E)\,P(NN)\cdots$$

For example, for m_{ES}, the distribution $P(m_{ES})$ has the shape





"Argus function"

respectively, for signal and background processes. Systematical errors are included in the PDF by convolving the underlying distributions with resolution functions, for example,

$$P(\Delta\tau) = \frac{1}{\tau} e^{-\Delta t/\tau} \otimes \frac{1}{\sqrt{2\pi\sigma_{\Delta t}^2}} e^{-\frac{(\Delta t - \Delta t_0)^2}{2\sigma_{\Delta t}^2}}$$

To combine these PDF's into a likelihood function, we must take account that events are generated by a variety of processes that fall into various catagories

$$\{ \text{signal}, \text{bkgd1}, \text{bkgd2}, \ldots \}$$

The kinematic distributions are different for events of each category. The likelihood function will depend on quantities n_i, the number of events from processes in the category i. Let

$$f_i = \frac{n_i}{\sum_i n_i}$$

The likelihood function is now constructed as follows: For each event, we assemble the product of PDF's for individual measured quantities, sum over the contributions of different categories of events, and then take the product over events:

$$\mathcal{L} = \prod_{\substack{\text{events} \\ j}} \sum_{\substack{\text{categories} \\ i}} f_i \, \mathcal{P}_i(x_j, \alpha_k)$$

We also include a Poisson factor for the total number of events; this gives the *extended likelihood function*

$$\mathcal{L} = \frac{1}{n_{obs}!} \left(\sum_i n_i\right)^{n_{obs.}} e^{-\sum n_i} \prod_{\substack{events\\ j=1}}^{n_{obs.}} \sum_{cat\ i} \frac{n_i}{(\sum n_i)} P_i(x_j; \alpha_k)$$

We now must maximize this function with respect to the n_i and the variables α_k, keeping the number of observed events (which is the sum of the n_i) fixed. Dropping factors that are constant in the maximization

$$\mathcal{L} = (\text{const}) \cdot e^{-\sum n_i} \prod_j \sum_{cat\ i} n_i P_i(x_j; \alpha_k)$$

It is easiest to work with the log(likelihood), for which the product becomes a sum.

To quote errors in σ's, it is useful to relate the results to those for a Gaussian distribution. Work through the simplest case of a Gaussian PDF, with one measured variable x and one unknown μ. The likelihood is

$$\mathcal{L} = P(x; \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Taking the logarithm,

$$-\log \mathcal{L} = \log \sqrt{2\pi\sigma^2} + \frac{(x-\mu)^2}{2\sigma^2}$$

Minimization gives

$$0 = \frac{\partial}{\partial \mu}(-\log \mathcal{L}) = -\frac{(x-\mu)}{\sigma^2}$$

or

$$\mu = x$$

If there are multiple measurements, possibly with different errors, the likelihood function would be

$$\mathcal{L} = \prod_j \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-(x_j-\mu)^2/2\sigma_j^2}$$

leading to the estimate

$$\mu = \frac{\sum_j x_j/\sigma_j^2}{\sum_j 1/\sigma_j^2}$$

The uncertainty on μ is given by

$$\sigma_\mu^2 = \mathrm{Var}(\mu) = \frac{1}{-\frac{\partial^2}{\partial\mu^2}(\log\mathcal{L})}$$

For multiple measurements, this evaluates to

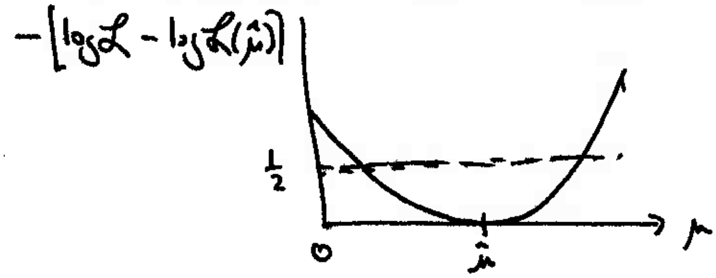
$$\sigma_\mu^2 = \left[\sum_j \frac{1}{\sigma_j^2}\right]^{-1}$$

which reduced to the familiar result σ^2/N for N measurements with the same error.

Another way to evaluate this uncertainty is to look at the shape of the likelihood distribution. We usually study the quantity

$$\mathbb{L} = -\left[\log\mathcal{L} - \log\mathcal{L}(max)\right]$$

which takes a minimum value of 0 at the best estimate of μ. Then



For a Gaussian likelihood function, the standard error interval is found by setting

$$\mathbb{L} = -[\log\mathcal{L} - \log\mathcal{L}(\hat{\mu})] = \frac{1}{2}$$

For a likelihood function that is not a Gaussian, we minimize the log likelihood function using the program MINUIT. This is a heavy-duty 1960's-era function minimizer that is the standard in high-energy. It is especially forgiving in finding a minimum from a wide range of starting values. (However, it is necessary to experiment with a number of starting values to be sure of finding the *global* minimum). One might describe MINUIT as old, but battle-tested. MINUIT contains associated routines for estimating the error matrix about the minimum.

To quote an error, we find the contour where

$$\mathbb{L} = \frac{1}{2}$$

Alternatively, the significance of a result $\mu \neq 0$ is estimated by

$$\sqrt{2\mathbb{L}(0)} = \text{Number of } \sigma \text{ from } \hat{\mu}$$

In computing the log likelihood function as a function of μ, we minimize, for each fixed μ, over the additional parameters α_k. To evaluate upper limits, we integrate the area under the exponential of this function,

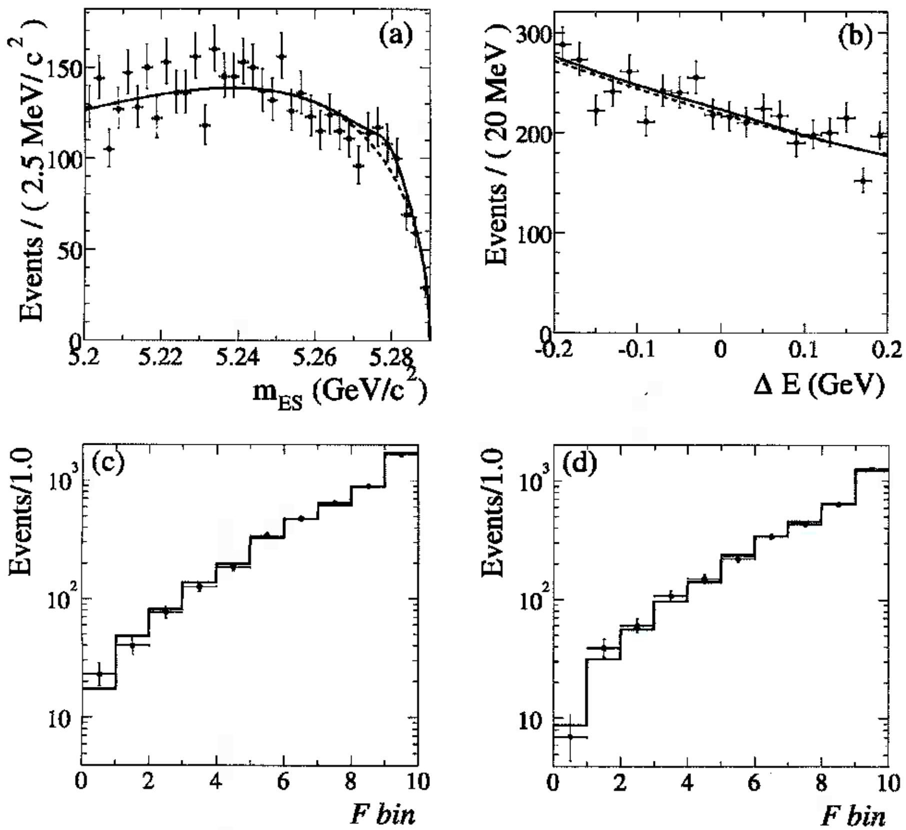
Many BaBar papers give as their result a plot of the log likelihood function. This is especially useful in cases where the likelihood is very different from a Gaussian, and especially when there are multiple minima.

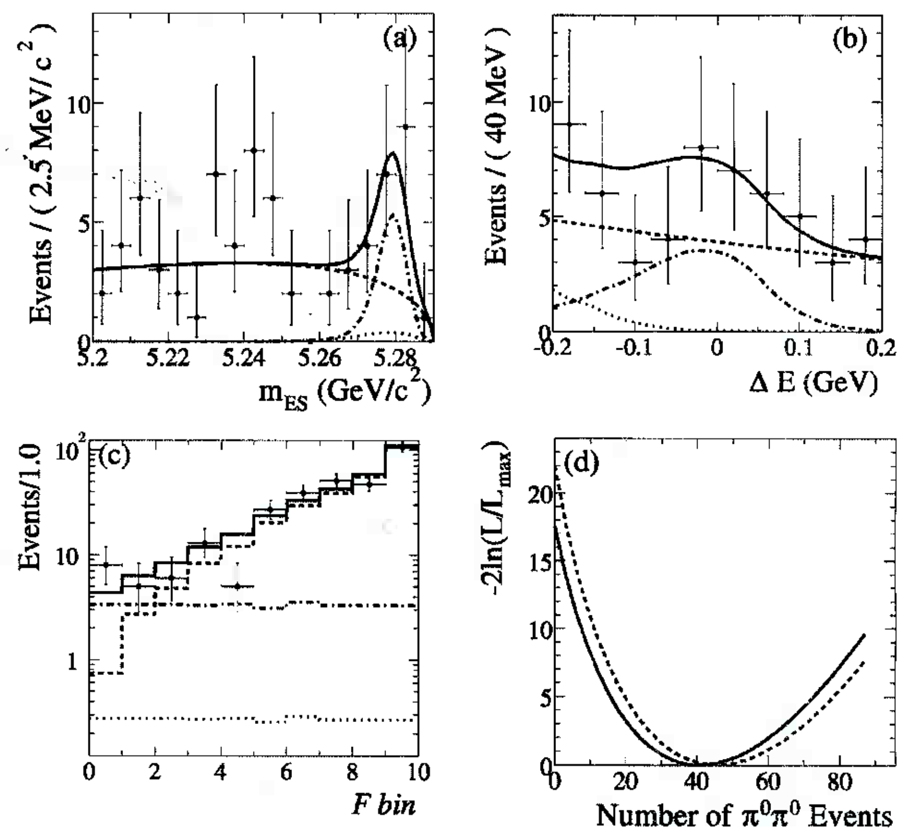
The figures show some examples of this approach in actual BaBar measurements. The first set of figures are taken from the paper "Observation of the decay $B^0 \to \pi^0\pi^0$", Phys. Rev. Lett. 91, 241801 (2003) [hep-ex/0308012]. The first set of figures shows the kinematic distributions in three variables in the original data sample. The second set of distributions shows the fit to these variables in the final data sample after selection cuts, and the final log likelihood function. The analysis gave a significance of 4.2 σ from the maximum likelihood fit. This can be compared to a significance of 2.7 σ from a parallel, purely cut-based analysis.

The second set of figures are taken from the paper "Search for Dimuon Decays of a Light Scalar Boson in Radiative Transitions $\Upsilon \to \gamma A^0$", Phys. Rev. Lett. 103, 081803 (2009) [arXiv:0905.4539 [hep-ex]]. This experiment involved searching for a photon of a definite energy that might be recoiling against a very weakly-coupled scalar boson. This was a search for a small signal on top of a large background. The data was fit to a smooth background plus a Gaussian with the known resolution in photon energy. Both positive and negative values of the coefficient of the signal Gaussian were allowed. The distribution of fitted coefficients for the signal is shown in the lower left. Some 3σ excesses were observed, but the number of these was consistent with the expectation from statistical fluctuations of the background. The curve in the upper right shows the log likelihood function for a particular value of the scalar boson mass. Such curves were used to set the upper limits on the branching fraction to this mode shown in the lower right.

The third set of figures are taken from the paper "Measurements of CP-violating asymmetries in the decay $B^0 \to K^+K^-K^0$", Phys. Rev. Lett. 99, 161802 (2007) [arXiv:0706.3885 [hep-ex]]. These show the likelihood function used in a measurement of the effect value of the CKM angle β contributing to the CP asymmetry in this decay. The fitted Δt distributions are shown in the lower right. The curve in the lower left is the final log likelihood distribution. The two local minima correspond to the two values of β giving the same value of $\sin 2\beta$. The likelihood curve quantified the level of confidence with which the lower of these solutions is selected.

Observation of a Significant Excess of $\pi^0\pi^0$ Events in B Meson Decays



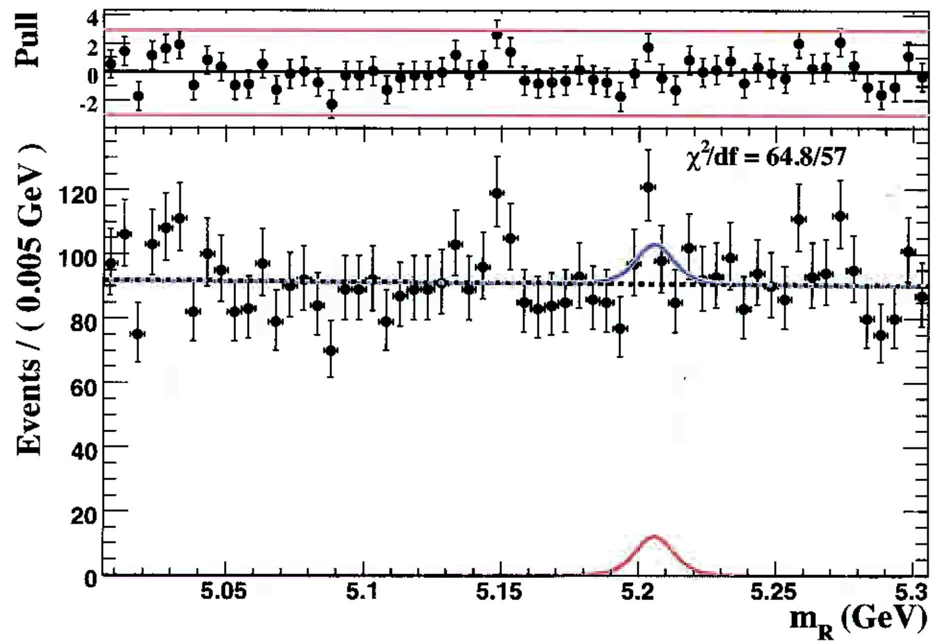


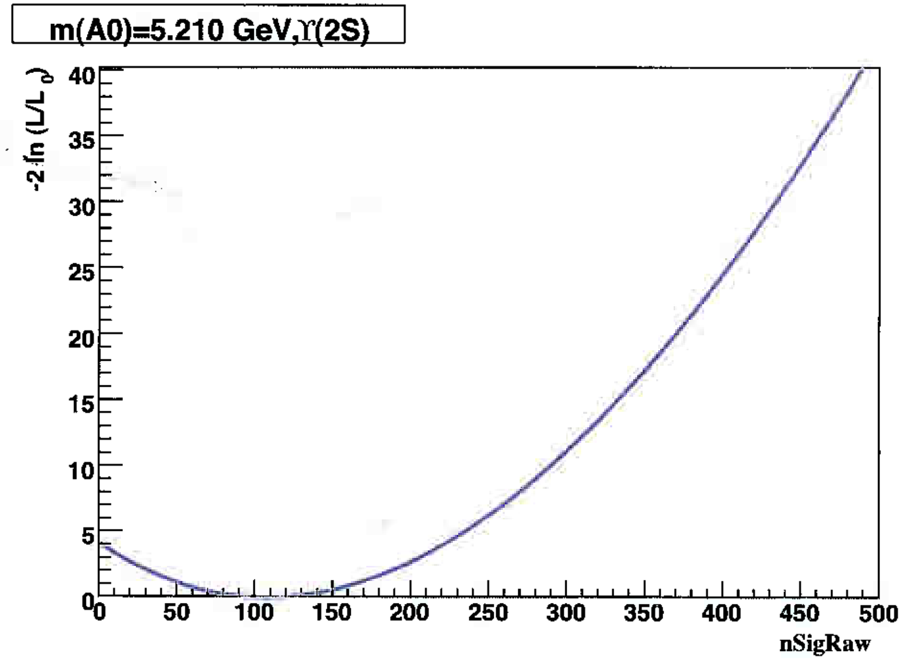
$$\mathcal{L} = \exp\left(-\sum_{i=1}^{3} n_i\right) \prod_{j=1}^{N}\left[\sum_{i=1}^{3} n_i \mathcal{P}_i(\vec{x}_j;\vec{\alpha}_i)\right].$$

$46 \pm 13 \pm 3 \quad B^0 \rightarrow \pi^0\pi^0$

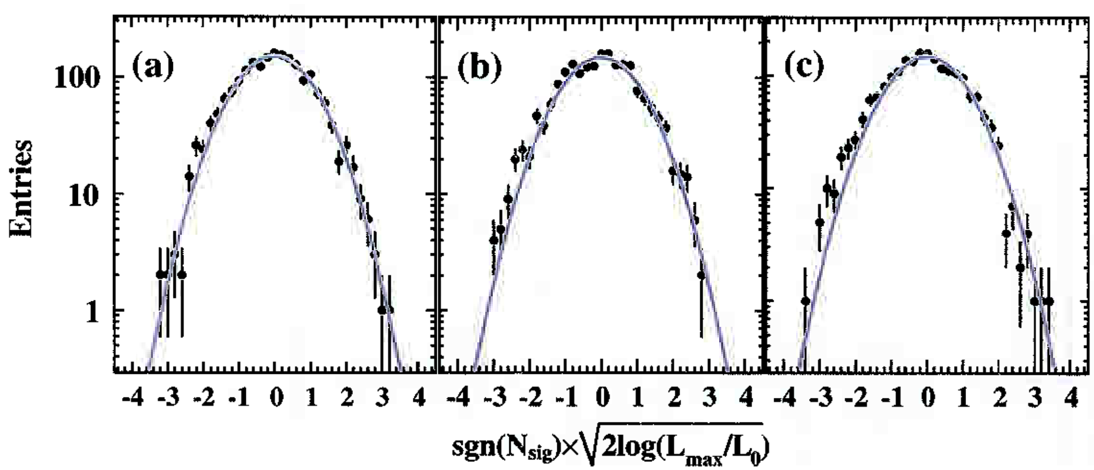
4.2 standard deviations

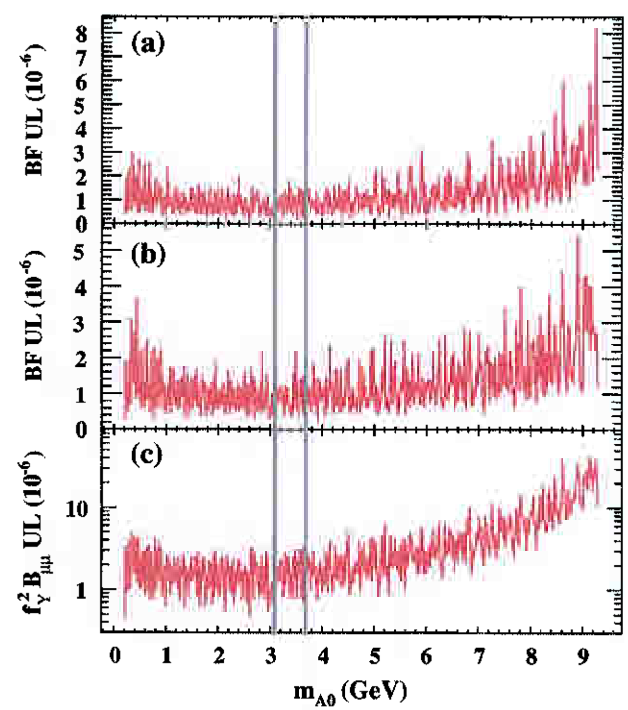
Search for Dimuon Decays of a Light Scalar Boson in Radiative Transitions $\Upsilon \rightarrow \gamma A^0$





$$S(m_{A^0}) = \operatorname{sgn}(N_{\rm sig})\sqrt{2\log(L_{\rm max}/L_0)},$$





Measurements of CP-Violating Asymmetries in the Decay $B^0 \rightarrow K^+K^-K^0$

$$\mathcal{P}_i \equiv \mathcal{P}(m_{\mathrm{ES}})\mathcal{P}(\Delta E)\mathcal{P}_{\mathrm{Low}}\mathcal{P}_{\mathrm{DP}}(m_{K^+K^-}, \cos\theta_H, \Delta t, q_{\mathrm{tag}}) \otimes \mathcal{R}(\Delta t, \sigma_{\Delta t}),$$

$$\mathcal{P}_{\mathrm{DP}} = d\Gamma \times \varepsilon(m_{K^+K^-}, \cos\theta_H) \times |J|,$$

$$\frac{d\Gamma}{d\Delta t} \propto \frac{e^{-|\Delta t|/\tau}}{2\tau}[|\mathcal{A}|^2 + |\bar{\mathcal{A}}|^2 + q_{\mathrm{tag}}2\mathrm{Im}(\xi\bar{\mathcal{A}}\mathcal{A}^*) \times \sin\Delta m_d\Delta t - q_{\mathrm{tag}}(|\mathcal{A}|^2 - |\bar{\mathcal{A}}|^2)\cos\Delta m_d\Delta t],$$

$$\overset{(-)}{\mathcal{A}}(m_{K^+K^-}, \cos\theta_H) = \sum_r c_r(1 \mp b_r)e^{i(\varphi_r \mp \delta_r)} \times f_r(m_{K^+K^-}, \cos\theta_H),$$