A BRIEF DESCRIPTION OF THE MAXIMUM LIKELIHOOD (ML) ANALYSIS

GIORDANO CERIZZA

Data samples are a collection of N independent events. An event is the measurement of a set of $\mathcal{O}$ *observables* $\hat{x} = (x^1, ..., x^{\mathcal{O}})$ (energies, masses, spatial and angular variables...) recorded in a brief span of time by the detectors. The events can be classified in $\mathcal{S}$ different *species*, which are generally classified with *signal*, for the events of interest for their physics phenomena, and *backgrounds*, all the remaining. Each observable x^j is distributed for the given species $\mathcal{S}$ with a probability distribution function (PDF) $\mathcal{P}^j_{\mathcal{S}}(x^j; \hat{\theta}^j_{\mathcal{S}})$, where $\hat{\theta}^j_{\mathcal{S}}$ are the parameters of the PDF. Several data analysis techniques can be used to discriminate signal and background events, using particular observables which have different PDF distributions for signal and background events. The Maximum Likelihood (ML) fitting procedure is a popular statistical technique used to estimate the number of events belonging to each species and the parameters $\hat{\theta}_{\mathcal{S}} = (\hat{\theta}^1_{\mathcal{S}}, \ldots, \hat{\theta}^{\mathcal{O}}_{\mathcal{S}})$ of the PDFs, that can be related to the prediction obtained from physics models [1]. In the case in which the observables are uncorrelated to each other, the total PDF for the species $\mathcal{S}$ is expressed as

$$\mathcal{P}_{\mathcal{S}}(\hat{x}; \hat{\theta}_{\mathcal{S}}) = \prod_{j=1}^{\mathcal{O}} \mathcal{P}^j_{\mathcal{S}}(x^j; \hat{\theta}^j_{\mathcal{S}}) \tag{1}$$

The *extended likelihood function* is

$$\mathcal{L} = \frac{e^{-\sum\limits_{s=1}^{\mathcal{S}} n_s}}{N!} \prod_{i=1}^{N} \sum_{s=1}^{\mathcal{S}} n_s \mathcal{P}_{\mathcal{S}}(\hat{x}_i; \hat{\theta}_{\mathcal{S}}) \tag{2}$$

where n_s are the number of events belonging to each species. The ML technique allows to estimate the values of the parameters by maximizing this function with respect to the free parameters. Usually, it is used to minimize the equivalent function $-\ln(\mathcal{L})$, the negative log-likelihood (NLL). So the NLL to be minimized has the form [1]:

$$\mathrm{NLL} = \sum_{s=1}^{\mathcal{S}} n_s - \sum_{i=1}^{N} \left(\ln \sum_{s=1}^{\mathcal{S}} n_s \mathcal{P}_{\mathcal{S}}(\hat{x}_i; \hat{\theta}_{\mathcal{S}}) \right). \tag{3}$$

The function is a sum of logarithms and the search for its minimum can be carried out numerically. The whole procedure of minimization requires several evaluations of the NLL, and in turn requires the calculation of the corresponding PDFs for each variable and each event of the data sample. The algorithm for the evaluation of the NLL is implemented

[1] The $N!$ term in the expression is omitted because does not depend on the parameters.

in the ROOT/RooFit package [2]. It is based on the numerical minimization package MINUIT [3].

In the pedagogical example "fitter_sample.C", we defined four distinct categories: three signals, corresponding to the energy levels 1350, 1420, and 1850 keV and a continuum background. The probability density functions (PDFs) that describe the three peaks are parameterized as Gaussian functions, while the background as a second order Chebychev polynomial [2]. We generated from the full model 100k events with the default method "generate(variable,number of events)" and we fit to these sample to test the ML procedure. In the fit there are 12 parameters free to float, six mean and width values (sig_m and sig_w) corresponding to the energy peaks, two parameters that define the background shape (a0 and a1), and four yields for the four categories (nsig1, nsig2, nsig3, and nbkg). The relative strength of each peak can be easily evaluated over the total number of events generated.

The output plot is shown in Fig. 1.

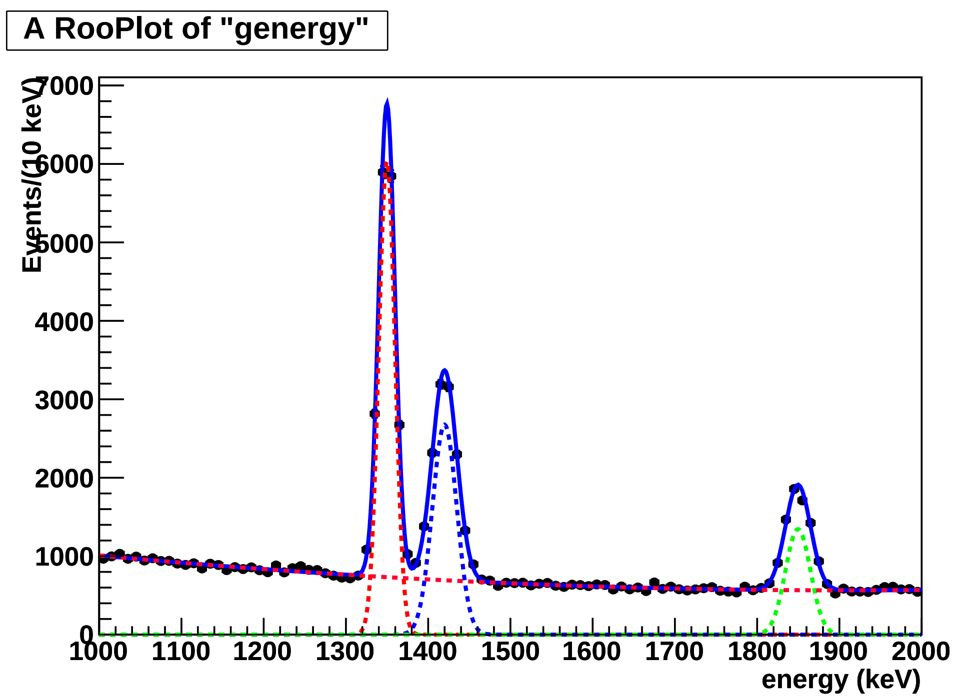

FIGURE 1. Resulting plot from the ML fit procedure on MonteCarlo generated events.

[2]This choice is only technical. RooFit internal integrator deals better with Cheby polynomials than simple polynomials when complex models are implemented.

```
**********
**    18 **HESSE          6000
**********
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=-383990 FROM HESSE     STATUS=OK           115 CALLS          442 TOTAL
                    EDM=2.08509e-05     STRATEGY= 1       ERROR MATRIX ACCURATE
 EXT PARAMETER                                   INTERNAL      INTERNAL
 NO.   NAME       VALUE            ERROR         STEP SIZE       VALUE
   1  a0          -3.04926e-01   6.96437e-03   1.18580e-04  -3.09861e-01
   2  a1           9.33051e-02   6.45392e-03   3.53614e-05   3.11067e-02
   3  nbkg         6.99615e+04   2.99049e+02   1.04830e-04   4.10678e-01
   4  nsig1        1.49319e+04   1.40322e+02   6.49464e-05  -7.77307e-01
   5  nsig2        9.96828e+03   1.31055e+02   3.42541e-04  -9.28353e-01
   6  nsig3        5.13845e+03   1.05581e+02   7.21900e-05  -1.11346e+00
   7  sig_m1       1.34979e+03   1.01329e-01   1.73320e-04  -4.25493e-03
   8  sig_m2       1.41997e+03   2.13004e-01   1.05178e-04   1.42546e-01
   9  sig_m3       1.84965e+03   3.37852e-01   2.88707e-04  -3.50263e-03
  10  sig_w1       9.87108e+00   8.80030e-02   1.45360e-04  -1.28925e-02
  11  sig_w2       1.48583e+01   1.98238e-01   3.60212e-04   5.07313e-01
  12  sig_w3       1.52361e+01   3.09874e-01   5.66417e-04   5.51080e-01
```

References

[1] Aldrich, J., *R.A. Fisher and the making of maximum likelihood 1912-1922* Statist Sci., 2 p. 162, 1997.

[2] I. Antcheva and others, *ROOT – A C++ framework for petabyte data storage, statistical analysis, and visualization* Comput. Phys. Commun., 180 p. 2499, 2009.

[3] James, F. and others, *MINUIT: a system for function minimization and analysis of the parameter errors and corrections* Comput. Phys. Commun. (CERN-DD-75-20), 10, p 343, 1975.