Fakultät für Physik
Prof. Dr. M. Feindt, Prof. Dr. G. Quast
Dr. T. Kuhr
D. Piparo, F. Wick

20.11.2008

Praktikum zu
Moderne Methoden der Datenanalyse

Exercise 3: Fitting

The fitting of parametrised functions to measured data is important for the check of models and the determination of their parameters. However there are some pitfalls which can lead to wrong conclusions.

- **Exercise 3.1:**

  Plot the random numbers stored in the ntuple you created in exercise 2.3. Use a histogram with 100 bins from 0 to 10. Have a look at the online documentation of the `TTree::Draw` method for this task. (Or take the macro `getHisto.C` provided at `/home/staff/tkuhr/Praktikum/Exercise3`. This directory also contains a file with a random number ntuple.)

  The plotted numbers can be interpreted as the measurements of decay times of a radioactive material. Fit the following function to the distribution:

  $$f(t) = N \cdot \exp\left(-\frac{t - t_0}{\tau}\right)$$

  The fit parameters are the normalisation N, a possible offset t_0 and the lifetime τ. Use an object of type `TF1` to define the fit function. Fit the histogram using its `Fit` method with the default options (χ^2 fit). You can find some information about fitting of user-defined functions in the root howto's. Try out different start values for the three parameters. How does the result of the fit depend on the start values?

  Print the correlations between the fitted parameters. Consult the online documentation of `TH1::Fit` in order to find out how to obtain them. What can be learned from the correlations in this case?

  Improve the parametrisation of the fit function. How does it have to look like in order to have the number of measurements as a parameter? Repeat the fit with the improved parametrisation.

- **Exercise 3.2:**

  Make a histogram of the numbers from the ntuple with only 10 bins from 0 to 10. Fit the improved fit function to the histogram using once the default options and once the option ¨I¨. Compare the fitted parameters to the expected ones for both cases and explain the difference.

- **Exercise 3.3:**

  Make three different histograms with 10, 1000 and 100000 entries from the ntuple respectively. Use 1000 bins from 0 to 10. Fit the function to each histogram using the χ^2 method and the binned likelihood method. Compare the fitted parameters and the χ^2 value of both methods and try to explain the results.

- **Exercise 3.4:**

  Write an unbinned log likelihood fit for 10, 1000 and 100000 entries from the ntuple. Create an object of type `TMinuit` for this with the number of fit parameters as argument. Define a function, that calculates the log likelihood, and set it in the `TMinuit` object with `SetFCN`. The function has to have the following signature:

  ```
  void anyName(Int_t& npar, Double_t* gin, Double_t& logLikelihood,
               Double_t* param, Int_t flag)
  ```

  The first parameter of this function has to be the number of parameters, the third one the calculated log likelihood and the fourth one the array of fit parameters. The second and the last parameter can be ignored. In the function you have to loop over the ntuple entries and sum up the log likelihood values. The i-th entry in the ntuple can be accessed by first calling the method `GetEntry(i)` and then `GetArgs()[0]` of the ntuple. The pointer to the ntuple and the number of entries have to be defined outside the function in global variables.

  Set an initial parameter value and error with `TMinuit::DefineParameter`. The minimization is performed by `TMinuit::Migrad`.

  Plot a histogram from 0 to 10 with the entries used in the unbinned log likelihood fit together with the fitted function normalized to the number of entries. Display the log likelihood value as a function of the fit parameter τ from 0.5 to 5. What can be learned from this plot about the errors of the fitted parameter?

- **Exercise 3.5:**

  Fit a Gaussian function with a fixed mean of 0 to the histogram with 100 bins from 0 to 10 containing 100000 entries. What is a good parametrisation? Make a χ^2, a binned and an unbinned log likelihood fit. Does the result of the unbinned log likelihood fit describe the data? Invent a method to quantify the quality of the fit (without referring to the fit of the exponential function).