Exercises for the lecture „Moderne Methoden der Datenanalyse”

Prof. Dr. S. Hansmann-Menzemer, M. Schiller
Physikalisches Institut der Universität Heidelberg

May 18th 2009

Exercise 3: Central Limit Theorem

Measurements acquire errors form many different sources. Usually, it is assumed that the errors follow a Gaussian distribution. This is justified by the *Central Limit Theorem*. It states that the sum of N independent variables taken from the same distribution converges to a Gaussian when $N \to \infty$. Note that the variance has to be defined for the single distribution.

(Exercises with * can be skipped if no time is left.)

- **Exercise 3.1**
  Let us study the CLT using a sample of uniformly distributed random numbers. Write a macro that
  - generates uniform random numbers on the interval $[0, 5]$,
  - plots the distribution of the average of k random numbers with $2 \leq k \leq 20$ (e.g. 10,000 times),
  - fits a Gauss to each distribution
  - writes the results to a ROOT file
  - and plots the mean and the sigma versus k.

  For how many averaged random numbers does the distribution look like a genuine Gaussian?

- **Exercise 3.2**
  Consider a random variable x distributed according to the Cauchy (or Breit-Wigner) p.d.f.

  $$f(x) = \frac{1}{\pi} \frac{1}{1 + x^2}$$

  - Show that if r is uniformly distributed in [0,1], then

    $$x(r) = \tan\left[\pi(r - 1/2)\right]$$

    follows the Cauchy distribution.
  - Using this result write a macro to generate e.g. 10,000 Cauchy-distributed random numbers and plot them in a histogram.

– Modify your macro from before to generate repeated experiments each consisting of n independent Cauchy distributed values for e.g. $n = 10$. For each sample, compute the sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Compare a histogram of $\bar{x}$ with the original histogram of x. Does it agree with what you would expect from the Central Limit Theorem?

- **Exercise 3.3***

  The ultimate test for a particle physics tracking detector is to put it in a testbeam. Here, just like in a physics experiment, particles are shot through the detector. Before and after the detector under test (DUT) other detectors are placed. These detectors are known as the telescope. Using the hits in the telescope one predicts where the particle traversed the DUT. One of the most important measurements for the DUT is to measure the position resolution. Here, one measures the distance between the point where the particle should have gone through the detector according to the measurement in the telescope and the point where it went according to the measurement in the DUT itself.

  In principle, the particle goes in a straight line. Since detectors are all made from material, the charged particles scatter. This leads to an error in the predicted position. In this exercise, we will measure the error on the predicted position from a simulation.

  Assume we have five detectors in a row. The third one is the DUT. The tracks arrive horizontally at the first detector. At each plane the particle scatters. Multiple scattering is a statistical process. The angles are described by a Gaussian with a mean of zero and a width Θ of

  $$\Theta = \frac{13.6\text{MeV}}{\beta c p} z \sqrt{\frac{x}{X_0}} \left(1 + 0.038 \ln \frac{x}{X_0}\right)$$

  where p, β and z are the momemtum, velocity and charge number of the incident particle and x/X_0 is the thickness of the material in radiation lengths. Assume for now that $\Theta = 0.005$. The detectors are placed 25 mm apart. For the telescope, the hit resolution is 10 μm.
  Propagate 1,000 tracks through the telescope. Let the tracks scatter at each plane (i.e. detector) and fold the detector resolution for the telescope detectors in. Make a straight line fit using the four detectors of the telescope and make a histogram of the difference between the actual track position and the position predicted by the track fit at the position of the DUT. What is the error on the predicted position?

  (Hint: Assume that the beam is parallel to the x-axis. The beam is scattered in y. Use the following procedure:

  1. Propagate the track and remember the y-values.
  2. Smear the y-coordinates with the detector resolution and fit a straight line (using `TGraphErrors(npoints, xvalues, yvalues, xerrors, yerrors)` and the fit options `"pol1"` and `"Q"`).
  3. Compare track position and actual position in the DUT. How does your result compare to a Gaussian?)