Humboldt-Universität zu Berlin
Institut für Physik

Winter Semester 2010/2011

Dr. U. Husemann, V. Ferrara

Statistical Methods of Data Analysis
Problem Set #7

Due Date: Thursday, January 20, 2010, during the lecture

Problem 1: (6 points)

A card player claims that he can predict which suit a randomly chosen play card has (hearts, diamonds, clubs or spades). Design a hypothesis test to check the player's clairvoyant abilities by drawing 10 random cards.

i) Write down the expected probability distribution of the null hypothesis (no clairvoyance).

ii) Use the expected probability distribution as a test statistic and assume a significance level of $\alpha = 0.01$ to determine how many correct predictions are required to reject the null hypothesis.

Problem 2: (10 points)

Consider the following experiment: you are determining the acceleration due to gravity g by switching off an electromagnet to release an object and measuring the time t it takes to fall a fixed distance d, so $d = \frac{1}{2}gt^2$. The distances are measured precisely, the time with an accuracy of 0.01 seconds. The results are:

Time (s)	0.16	0.40	0.58	0.72	0.97
Distance (m)	0.20	1.00	2.00	3.00	5.00

i) Determine the least squares estimator of the slope of a straight line passing through the origin and its variance (the corresponding results for a general straight line fit have been given in the lecture).

ii) Calculate the acceleration due to gravity g, with the appropriate uncertainty first assuming the object is released instantly and then that the field in the magnet takes an unknown but constant time to die away. Comment on the difference and on the χ^2 of the two fits.

(see reverse side)

Problem 3: (10 points)

If the underlying PDF of a dataset is unknown, empirical fit functions have to be employed. The most common empirical fit functions are n-th order polynomials with constant coefficients:

$$P_n(x) = \sum_{k=0}^{n} p_k\, x^k.$$

The fit results can usually be "stabilized" by using orthogonal polynomials

$$L_n(x) = \sum_{k=0}^{n} p_k\, l_k(x),$$

where $l_k(x)$ are Legendre polynomials, which can defined recursively by

$$l_0(x) = 1; \quad l_1(x) = x; \quad (k+1)\, l_{k+1}(x) = (2k+1)\, x\, l_k(x) - k\, l_{k-1}(x).$$

and fulfill the orthogonality relation

$$\int_{-1}^{1} \mathrm{d}x\, l_m(x)\, l_n(x) = \frac{2}{2n-1} \delta_{mn},$$

where δ_{mn} denotes the Kronecker delta. Use the ROOT class `TGraphErrors` to fit the data points given by the following pairs of x and y values to the polynomials $L_3(x)$, $L_5(x)$ and $L_7(x)$ and compare the result with fits to $P_3(x)$, $P_5(x)$ and $P_7(x)$:

```
Double_t x[20] = { -0.9, -0.8, -0.7, -0.6, -0.5,
                   -0.4, -0.3, -0.2, -0.1, 0.0,
                   0.1, 0.2, 0.3, 0.4, 0.5,
                   0.6, 0.7, 0.8, 0.9, 1.0 };
Double_t y[20] = { 5.0935, 2.1777, 0.2089, -2.3949, -2.4457,
                   -3.0430, -2.2731, -2.0706, -1.6231, -2.5605,
                   -0.7703, -0.3055, 1.6817, 1.8728, 3.6586,
                   3.2353, 4.2520, 5.2550, 3.8766, 4.2890 };
```

Assume that each data point has an uncertainty of $\sigma = 0.5$. Your solution should contain plots showing the data points and the fitted curves, as well as print-outs of the fitted coefficients p_k and their correlation matrices. The latter can be obtained in ROOT via `TFitResult::GetCorrelationMatrix()`. In which sense is the fit using orthogonal polynomials "more stable"?