Prof. Dr. K. Desch
Dr. P. Wienemann

Statistical methods of data analysis
Summer term 2010

Homework 5

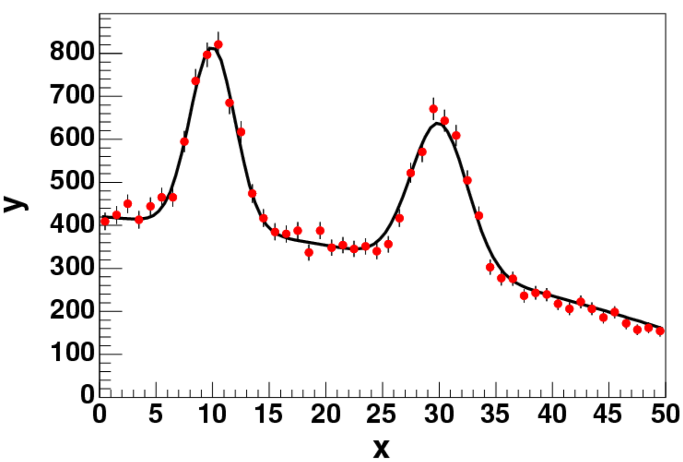
due on June 28, 2010 after the lecture

Exercise 15: (5 points)

A frequent problem in physics is the reconstruction of peaks in spectra above a certain background. An example is shown in the below plot. The shown histogram entries are listed in the table below the plot (available in electronic form on `http://pi.physik.uni-bonn.de/~wienemann/teaching/ss10/homework/data_sheet5.dat`). The uncertainty of each bin entry is approximately given by the root of the number of entries. To parameterise the spectrum, we make the following assumptions: The background can be described by a polynomial of degree two and the peaks have a Gaussian shape. In summary, the parametrisation has the following form:

$$y(x;\vec{p}) = p_1 + p_2 x + p_3 x^2 + p_4 \exp\left(-\frac{(x-p_5)^2}{2p_6^2}\right) + p_7 \exp\left(-\frac{(x-p_8)^2}{2p_9^2}\right)$$

Determine the parameters $p_1, \ldots, p_9$ including their uncertainties using a χ^2 fit. Check the quality of the parametrisation based on the minimal χ^2 value.



Bin nr.	Bin content	Bin nr.	Bin content	Bin nr	Bin content	Bin nr.	Bin content	Bin nr.	Bin content
1	409	11	821	21	348	31	643	41	218
2	424	12	685	22	354	32	609	42	206
3	450	13	617	23	345	33	505	43	222
4	413	14	474	24	351	34	423	44	206
5	444	15	417	25	340	35	303	45	186
6	465	16	385	26	356	36	277	46	198
7	465	17	380	27	417	37	276	47	172
8	595	18	388	28	522	38	236	48	157
9	736	19	337	29	571	39	243	49	162
10	797	20	388	30	671	40	239	50	154

Exercise 16: (5 points)

The method of least squares is often used for straight line fits

$$y(x) = mx + c$$

to data (x_i, y_i). Assume that all y values have the same (uncorrelated) statistical uncertainty σ and that they share a common (fully correlated) systematic uncertainty s.

a) How does the covariance matrix look like?

b) Determine estimators $\hat{m}$ and $\hat{c}$ for the slope m and the intercept c using the method of least squares.

c) What is the variance of $\hat{m}$ and $\hat{c}$?

Exercise 17: (5 points)

A counting experiment with 7 search channels has observed n_i events in channel i and expects on average b_i events for the "background only" hypothesis and $s_i + b_i$ events for the "signal+background" hypothesis. The corresponding numbers are summarised in the table below.

i	1	2	3	4	5	6	7
n_i	50	36	29	9	7	7	1
b_i	46	35	26	9	5	3	1
$s_i + b_i$	53	42	32	13	8	6	3

a) Create two histograms showing the probability density functions of the test statistic $-2\ln Q$ with

$$Q = \prod_i \frac{\mathcal{L}(\mu_i = s_i + b_i; n_i)}{\mathcal{L}(\mu_i = b_i; n_i)} \qquad \text{and} \qquad \mathcal{L}(\mu_i; n_i) = \frac{\mu_i^{n_i}}{n_i!} \exp(-\mu_i)$$

for the "background only" and the "signal+background" hypothesis. Do this by creating many toy data sets n_i' where n_i' is Poisson distributed around the "background only" and the "signal+background" expectation, respectively.

b) How large is CL_b, the p-value for the "background only" hypothesis?