Applied Statistics

Problems in fundamental concepts of statistics

The following problem set is a review of the fundamental concepts of statistics, which we have covered till now. It will be handed out Tuesday the 21st of September 2012, and a written solution is to be handed in Tuesday the 2nd of October at the latest. Problem solving in groups is allowed, but separate solutions are required.

Good luck, Troels

Those who are good at archery learnt from the bow and not from Yi the Archer. Those who know how to manage boats learnt from boats and not from Wo [the legendary boatman]. Those who can think learnt for themselves and not from the sages.

[Guan Yin, 8th Century]

I – Distributions and probabilities:

1.1 Let x be distributed according to the PDF $f(x) = x^2$ in the interval $[0, C]$.

- For which value of C is the PDF $f(x)$ normalized?
- What is the mean and width of x?

1.2 Little Peter goes to the casino and puts money on one number at a time ($p = 1/37$). If he is not cheating, what are the chances that he will win more than 3 times in 100 games?

1.3 Calculate the mean and width of the following PDFs:

- $f(x) = \frac{1}{2}\sin(x)$, $x \in [0, \pi]$.
- $f(x) = \ln(x)$, $x \in [1, e]$.

II – Error propagation:

2.1 If $\theta = 0.58 \pm 0.02$, what is then the uncertainty on $\cos\theta$, $\sin\theta$, and $\tan\theta$? What if $\theta = 1.58 \pm 0.02$?

2.2 Snell's Law states that $n_1 \sin\theta_1 = n_2 \sin\theta_2$. Find n_2 and its error from the following measurements:

$$\theta_1 = (22.03 \pm 0.2)^\circ \qquad \theta_2 = (14.45 \pm 0.2)^\circ \qquad n_1 = 1.0000$$

2.3 Let $x = 2yz + z^2$. Given that $y = 1.1$, $\sigma_y = 0.2$, $z = 3.5$, $\sigma_z = 0.3$, and that y and z are 60% linearly correlated, what is the uncertainty in x?

2.4 The initial activity N_0 and lifetime τ of a radioactive source is known with a relative uncertainty of 1%. When estimating the activity $N = N_0 e^{-t/\tau}$ the uncertainty will initially be dominated by the uncertainty in N_0 and later by the uncertainty in τ. For what value of t/τ will the to uncertainties contribute equally to the uncertainty on N?

III – Monte Carlo: (For this part the use of computers is adviced. Plots can be enclosed in the solution).

3.1 Let $f(x) = e^{-x^3+2x^2} - 1$ be proportional to a PDF for $x \in [0, 2]$.

- Which method should be used to generate numbers according to this distribution? Explain?
- Make an algorithm, which from a uniform distribtion in the interval $[0, 1]$ generates numbers following the PDF $f(x)$.
- Determine $\int_0^2 f(x)dx$ and its uncertainty by using this algorithm, such that you can normalize $f(x)$.

IV – Estimators:

4.1 In the past years several groups of students have been measuring the lifetime of the muon in the basement at NBI. Their results and estimated uncertainties are listed below:

Group	1	2	3	4	5	6	7	8	9	10
Result (μs)	1.82	1.95	1.46	2.12	2.09	1.70	1.93	1.87	2.25	2.16
Uncertainty (μs)	0.06	0.09	0.12	0.13	0.24	0.11	0.07	0.10	0.21	0.14

- Calculate the average and spread of the measurements along with the χ^2 and the probability of obtaining such a χ^2 value or something more extreme both in an unweighted and a weighted calculation.
- Is there a measurement, which does not fit very well in? Why?
- Repeat the previous calculation excluding the least probable measurement.
- How well do the results (unweighted and weighted) match the true value of τ_μ?

4.2 Consider the classic 1910 dataset on Polonium 210 decays by Rutherford and Geiger, showing the number of decays in a 72s period for 2608 periods:

Counts	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Frequency	57	203	383	525	532	408	273	139	45	27	10	4	0	1	1

- Argue what distribution these counts should follow and test if they do indeed. What is the average level of activity?
- Calculate the χ^2 and $-2\ln(likelihood)$ for a range of activity levels around the average value. Plot these and determine the (possibly asymmetric) statistical uncertainty on the average level both using the χ^2 and the likelihood.
- Given the lifetime of Polonium 210 and the time used for the experiment, which systematic uncertainty would you ascribe this result? Should Rutherford and Geiger actually have conducted a shorter experiment? Or reported the data differently?

V – Fitting data:

5.1 An experiment has yielded the following results, where the uncertainty on y, σ_y has been estimated to be 0.06:

x	y	x	y	x	y	x	y	x	y	x	y
-2.0	-.29	0.0	0.04	2.0	0.33	4.0	0.63	6.0	0.80	8.0	1.04
-1.0	-.19	1.0	0.06	3.0	0.57	5.0	0.89	7.0	0.81	9.0	0.94

- Assume a linear relation between x and y and make a χ^2-fit to data.
- Calculate from this χ^2 and the number of degrees of freedom the probability of obtaining such a χ^2 value or something more extreme. Is it a good fit?
- Try other hypothesis for the relation between x og y, and discuss their validity.