Applied Statistics

Exam in applied statistics

The following problem set is the take-home exam for the course applied statistics. It will be distributed Thursday the 29th of October 2009, and a solution in writing should be handed in by noon Friday the 30th. Working in groups is **not** allowed.

The use of computers is both allowed and recommended along with modifications of t-he programs you've worked with. For some of the problems, the use of computers will be necessary.

Good luck, Troels

I – Distributions and probabilities:

1.1 Little Peter has dropped gambling and is sitting on Blegdamsvej counting red cars. He has been told, that 3.1% of the cars passing are red, but he observes 11 red cars out of 100. What is the probability of this or something more extreme given the expected rate? Should Little Peter trust that number?

1.2 The members of a larger collaboration running the gigantic neutrino observatory Kamiokande in Japan, observed 11 neutrinos in their apparature on Monday the 23rd of February 1987. This coincided with the Supernova 1987A.

- What is the probability of observing 11 or more neutrinos in one day, if the expected average rate is 2.1 neutrinos per day?
- Actually the 11 neutrinos arrived within a period of 13 seconds. What is the probability of making such an observation, again given the expected rate?

II – Error propergation:

2.1 Five groups of students have determined the age of the Universe based on five different data samples. The results (in 10^9 years) were:

13.2 ± 0.4	15.6 ± 0.7	12.5 ± 0.6	14.2 ± 0.5	13.5 ± 0.9

- What is the mean age and the uncertainty on that mean?
- Do you find one of the measurements unlikely? Argue if this is the case and remove the measurement, if you find reason for it. Do the other measurements agree now?
- The star HE 1523-0901 has been measured to be $(13.2{\pm}0.1){\times}10^9$ years old by ESO's Very Large Telescope. What is the probability that the combined measurement of the students lies below the age of this star?

2.2 The period T of a pendulum is given by $T = 2\pi\sqrt{l/g}$, where l is the length of the pendulum and g is the gravitational acceleration. In an experiment the period is measured to be $T = (2.03 \pm 0.05)$s and the length of the pendulum $l = (0.998 \pm 0.003)$m.

- Assuming no correlations between the measurement of T and l, what is the result and the uncertainty on the measurement of g?
- What is the result, if T and l are 30% linearly correlated?

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

3.1 Let a Monte Carlo algorithm generate 1000 boxes with side lengths $a = 2.0 \pm 0.2$, $b = 3.0 \pm 0.15$, and $c = 4.0 \pm 0.1$, where the uncertainties are Gaussian and uncorrelated.

- Plot the distribution of box volumes, and determine the mean and the width.
- Compare the width with the result of analytically propagating the uncertainties.

3.2 Let $f(x) = \frac{1}{\pi}(1 + \sin(x))$ be a PDF for $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$.

- By which method should one generate random numbers according to this PDF?
- Produce an algorithm, which from a uniform distribution in the interval $[0, 1]$ generates 1000 randon numbers following the PDF $f(x)$. Calculate the average of these numbers and the uncertainty on the average. Compare this to the analytical value for the average.

IV – Statistical tests:

4.1 Charged particles passing through a gas produces ionization, the amount of which depends on the type of particle. Assume that a statistic, t, based on the signal from a detector is constructed such, that it is Gaussianly distributed around 2 for electrons and 0 for pions, both with a width of 1. A test is constructed which selects electrons by requiring that $t > 1.2$.

- What is the probability of accepting an electron and rejecting a pion?
- Assume that the fraction of pions make up 99% of the sample and the remaining 1% are electrons. What is the purity of the electrons selected by the criteria $t < 1.2$?
- If one requires an electron purity of 95%, which requirement on t most be demanded? What is the probability for an electron to pass this requirement?

V – Fitting data:

5.1 In an experiment wih a radioactive isotope, the number of decays in intervals of 1 second is measured to determine the lifetime of the isotope. The results are:

Time (s)	0.50	1.50	2.50	3.50	4.50	5.50	6.50	7.50	8.50	9.50
Counts	805	491	325	235	152	83	83	68	46	33

- Fit this data with an exponential function? What uncertainty do you ascribe each measurement? And is the fit good?
- It turns out that there is a constant background. Include this in your fit, and repeat it. Does this improve your fit significantly?
- A subsequent measurement shows that the background is 29.2. Given this knowledge, can you improve your lifetime measurement?

Coincidences, in general, are great stumblingblocks in the way of that class of thinkers who have been educated to know nothing of the theory of probabilities – that theory to which the most glorious objects of human research are endebted for the most glorious of illustration.

[Edgar Allan Poe (1809-1849), The murders in the Rue Morgue]