Applied Statistics

Exam in applied statistics 2011

The following problem set is the take-home exam for the course applied statistics. It will be distributed Thursday the 3rd of November 2011, and a solution in writing (and possibly by Email) should be handed in by noon Friday the 4th to HEP secretary Anette Uhl. Working in groups is **not** allowed.
The use of computers is both allowed and recommended along with modifications of the programs you've worked with. For some of the problems, the use of computers will be necessary.

Good luck and thanks for all your hard work, Troels & Sascha

I – Distributions and probabilities:

1.1 Orcs are attacking Minas Tirith with catapults. Gandalf is (by using magic) capable of destroying 98% of the stones shot at the city.

- If the orcs shoot 100 rocks, what is the chance that the city will be unharmed?
- How many stones do the Orcs need to shot at Minas Tirith in order to be 95% confident of at least one stone hitting the city?

1.2 The mean radii of the dwarf planets Pluto and Eris are 1153 ± 10 km and 1163 ± 6 km, respectively.

- Calculate analytically the probability that Eris is largest?
- Simulate their sizes 1000 times, and calculate the fraction of times that Eris comes out to be the largest of the two. How well does this result compare with the previous one?

1.3 A traffic consultant is optimizing the time between the arrival of a bus at a train station and the departure of the train, Δt. Both are running every 20 minutes, but while the train departure is accurate, the bus arrival has an uncertainty of 90 seconds. The time to get from the bus to the train is 15 seconds.

- For what value of Δt is the waiting minimized?

II – Error propergation:

2.1 Students were asked to measure the length of the table in Auditorium A. The results from ten of the measurements (in cm) were 334.0, 333.9, 338.2, 336.6, 336.5, 335.2, 336.1, 335.6, 337.8, and 337.3.

- What is the average length of the table and the uncertainty on that estimate?
- Given that the width has been measured with the same ruler to be 80.8 ± 0.6 cm, thus yielding a linear correlation of 80%, what is the length of the diagonal?

2.2 Eratosthenes of Cyrene [c. 276 – c. 195 BC] determined the circumference of the Earth by measuring the difference in angle to the sun in Syene and Alexandria $\Delta\theta = (7.2 \pm 0.7)^\circ$ along with the distance between them $\Delta d = 925 \pm 350$ km.

- Estimate the circumference of the Earth and the uncertainty on it.
- How consistant is this value with the known meridional circumference of Earth?

III – Monte Carlo:

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

- In order for the PDF to be normalized, what should the value of C be?
- 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 random numbers following the PDF $f(x)$. Plot the generated random numbers.
- Calculate the average of these numbers and the uncertainty on the average. Compare this to the analytical value for the average.

IV – Fitting data:

4.1 Robert Millikan's famous oil drop experiment yielded effective charges (EC) for oil drops in an electric field. The actual charge on the oil drops was unknown, but was clearly integer, and could be used to determine the unit charge, e, through the equation (based on the setup of the apparatus) $e = EC_{q=1} \times 0.3239 \times 10^{-19} C$. Assume that the relative uncertainties on the effective charges is 0.3%.

n	1	2	3	4	5	6	7	8	9
Charge (EC)	–	–	–	19.66	24.60	29.62	34.47	39.38	44.42
n	10	11	12	13	14	15	16	17	18
Charge (EC)	49.41	53.91	59.12	63.68	68.65	–	78.34	83.22	–

- Fit this data with a linear function and draw it. What is the χ^2 probability? Is the fit good?
- How consistant is the intercept (i.e. value of the fit function for n=0) with zero? Would it be reasonable to set it to zero? If so, refit with just a constant of proportionality and plot again.
- How consistant is the final result with the modern value of $e = (1.602176487 \pm 0.000000040) \times 10^{-19} C$?

V – Statistical tests:

5.1 Bode's Law stated that the distance from the Sun to our planets (including dwarf planet Ceres) follows the rule $d = 0.4 + 0.3 \times 2^k$ Astronomical Unit (AU). The distances for the eight innermost planets are given below:

Planet number (k)	-∞	0	1	2	3	4	5	6
Distance (in AU)	0.39	0.72	1.00	1.52	2.77	5.20	9.54	19.2

- Assuming that the relative uncertainty on the distances is 2% (except for Earth, which by definition is at 1 AU), calculate for Bode's Law the χ^2 and the probability of obtaining it.
- With the discovery of Neptune ($d = 30.06$ AU) in 1846, the rule could be further tested. How well does it fit?
- **Bonus problem:** Using the data for the eight innermost planets, determine the best parameters for a and b, such that Bode's Law $d = a + b \times 2^k$ is most accurate.

Don't worry too much about statistics! Just tell us what you do, and do what you tell us.

[Roger Barlow, ICHEP conference 2006, Moscow]