Applied Statistics

Problem Set in applied statistics 2014

The following problem for the course applied statistics will be distributed Friday the 19th of September 2014, and a solution in writing (preferably sent by mail) must be handed in by Wednesday the 1st of October. Discussing the problems with others is allowed, though each person should produce both the solutions (i.e. programs and calculations) and writeup by themselves.

Good luck and thanks for all your hard work so far, Troels, Florian, Arvad & Lars

---

I – Distributions and probabilities:

1.1 Let x be distributed according to the PDF $f(x) = x\exp(-x)$ in the interval $[0, \infty]$. What is the mean, mode, median, and RMS of this distribution? Possibly verify with simulation!

1.2 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 shoot at Minas Tirith in order to be 95% confident of at least one stone hitting the city?

1.3 In 2011 220 persons were killed in traffic and 4039 were injured. In 2012 the numbers were 167 and 3611, respectively. What is the percentage drop in number of deaths? And injuries? How significant are each of these drops?

II – Error propergation:

2.1 In a repeated experiment the velocity of a ball v is measured seven times.

Velocity (m/s)	94.1	86.3	93.9	89.8	101.2	97.5	118.3

- What is the average velocity and its uncertainty?
- If the mass of the ball is $m = 0.27 \pm 0.03$ kg, what is the kinetic energy $E_{\text{kin}} = \frac{1}{2}mv^2$ of the ball and its uncertainty?
- If, for some reason, there were a (linear) correlation between the velocity and the mass of the ball of $\rho_{vm} = -0.6$, what would the above answer then be?
- Do you find any of the measurements to be suspecious? Quantify your answer.

2.2 If $\theta = 0.54 \pm 0.02$, what is the uncertainty on $\cos\theta$, $\sin\theta$, and $\tan\theta$? What if $\theta = 1.54 \pm 0.02$?

2.3 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:

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

2.4 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:

3.1 Let $f(x) = ax^2$ be proportional to a PDF for $x \in [-1, 2]$.

- In order for this PDF to be normalized, what value should a have?
- What is the mean and width of $f(x)$?
- By which method would you generate random numbers according to this PDF?
- Produce an algorithm, which generates random numbers according to $f(x)$. Let t be a sum of twenty random values from $f(x)$, and generate 1000 values of t.
- Generate 1000 Gaussianly distributed numbers according to the mean and width of t (calculated analytically). Is this distribution the same as the one above?

IV – Fitting data:

4.1 The file [www.nbi.dk/∼petersen/data_HiggsGG.txt] contains 43742 measurements of the invariant mass between two photons in the range 100-160 GeV at the ATLAS experiment. (This problem is inspired by the Higgs search and discovery in 2012, which triggered the 2013 Nobel prize in physics).

- Read the file, plot the measurements with a reasonable binning and fit the distribution with a suitable smooth function. How well do you manage to describe this distribution?
- Do a Wald-Wolfowitz runs test on your fit residuals and comment on the result.
- Suppose that you are searching for a Gaussian peak of width 1.2 GeV in this spectrum. What is the significance of the largest peak found?

4.2 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 best combined measurements with uncertainty along with the χ^2 and the probability of obtaining such a χ^2 value or something more extreme.
- Is there a measurement, which does not fit very well in? And if so, why?
- How well do the results (unweighted and weighted) match the true value of τ_μ?

V – Statistical tests:

5.1 A magnet drops a ball through six timing gates providing t at various distances d to measure the acceleration due to gravity g. The distances are accurately know, while t has an uncertainty of 0.01 s.

Distance (m)	0.200	0.500	1.000	1.500	2.000	2.500
Time (s)	0.150	0.265	0.383	0.503	0.582	0.652

- Assume the magnet drops the ball at $t = 0$, and fit the data to obtain g. Comment on the fit quality.
- Drop the above assumption and measure the time of the magnet release t_0. How certain are you, that t_0 is not consistent with zero?

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]