Applied Statistics

Exam in applied statistics 2013

The following problem set is the take-home exam for the course applied statistics. It will be distributed Thursday the 31st of October 2013, and a solution in writing (preferably sent by email) must be handed in by noon Friday the 1st of November. Working in groups is **not** allowed. The use of computers is both allowed and recommended.

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

I – Distributions and probabilities:

1.1 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?

1.2 2008 had 406 deaths in traffic, as did 2007. What is the chance of these being the same?

1.3 Let x be a uniformly distributed PDF in the interval $[\alpha, \beta]$, where $0 < \alpha < \beta$.

- What is the expectation value (i.e. mean) and the variance of $1/x$?
- Compare the expectation value $E[1/x]$ with $1/E[x]$. In what limit do they converge?

1.4 Elementary particles fall in two catagories: Fermions and bosons. In the last century six fermions and four bosons have been discovered in USA and Europe, respectively.

- Which distribution does the number of elementry particles discovered in USA follow?
- If the probability is the same of discovering particles in USA and Europe, what is the probability for all fermions to have been discovered in USA and all bosons in Europe?

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 $m = 1.27 \pm 0.15$ 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 was 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 Ptolemy [90-168 AD] estimated the distance to the Moon by triangulation from two points on Earth at distance l, as $d_{\text{Moon}} = l\sin(\alpha)\sin(\beta)/sin(\alpha+\beta)$. Ptolemy measured $\alpha = \beta = (89.0 \pm 0.1)^\circ$ and $l = 12900 \pm 700$ km.

- What value and uncertainty for d_{Moon} does Ptolemy's measurements yield?
- Is his estimate consistent with the currently accepted value $d_{\text{Moon}} = 384399$ km

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?

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

- Which method would you use to generate numbers according to $f(x)$? Explain.
- Produce such an algorithm and generate 1000 numbers following $f(x)$. Calculate the mean of these numbers, and compare it to the analytic value.

IV – Fitting data and optimization:

4.1 The file www.nbi.dk/∼petersen/data_Exam2013.txt contains 100000 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 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 15 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?

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]