Applied Statistics

Problems in fundamental concepts of statistics

The following problem set is a review of the fundamental concepts of statistics. It will be handed out Friday the 20th of September 2013 at 9:00, and a written solution is to be handed in Wednesday the 2nd of October by midnight. The use of computers is recommended and plots (not code) should be enclosed in the solution. Problem solving in groups is allowed, but each person should hand in their own solution.

Good luck, Troels

I – Distributions and probabilities:

1.1 Let t be distributed according to the PDF $f(t) = Ce^{-t/\tau}$ in the interval $t \in [t_0, \infty]$.

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

1.2 Little Peter flips 20 coins. What is the chance of getting 14 or more heads?
Afterwards, little Peter flips 20 coins 100 times. What is the chance that he gets 18 or more heads at least once?

II – Error propagation:

2.1 Several student groups have through the Cavendish Experiment measured the gravitational constant G, as summarized in the table below (in $10^{11}\mathrm{m}^3\mathrm{kg}^{-1}\mathrm{s}^{-1}$).

Measurement	6.5 ± 0.5	7.5 ± 0.6	5.9 ± 0.3	6.2 ± 0.4	8.0 ± 0.8	7.1 ± 0.5

- Calculate the average of the measurements and its uncertainty along with the χ^2 and the probability of obtaining an equal or larger χ^2 value.
- How well do the result match the true value of G?

2.2 John Harrison's Gridiron pendulum was invented to minimize the impact of temperature uncertainties on time measurements through the period $T = 2\pi\sqrt{(}L/g)$. The coefficient of expantion for iron (brass) is 11.3 (18.7) $\times$ 10^{-6} m/mK. If the uncertainty on the average temperature is 2.5 K, what is the uncertainty in a time measurement,

- if the entire pendulum of length L is made of iron?
- if there is a brass "counter part" in the pendulum of length $l = 0.6L$?

2.3 Snell's Law states that $n_1 \sin\theta_1 = n_2 \sin\theta_2$. The index of refraction (n) of pure water is 1.3330, while sugar solutions have an index of refraction that increases linearly up to 1.4774 (75% solution).
Given the measurements $\theta_{\mathrm{air}} = (25.21 \pm 0.09)^\circ$ and $\theta_{\mathrm{solution}} = (17.91 \pm 0.08)^\circ$:

- Find n_2 and its error.
- Determine the percentage of sugar in the solution and its error.

III – Monte Carlo:

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

- For what value of C is $f(x)$ normalized?
- 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 $f(x)$, and show 1000 such points in a plot.

3.2 You roll 3 10-sided, 6 8-sided, and 10 6-sided (i.e. normal) dice.

- What is the mean and width of the sum of the eyes?
- Assuming the sum to follow a Gaussian distribution, what is the chance to roll a sum of 100 or more?
- Using a Monte Carlo, what do you estimate the chance of getting a sum of more than or equal to 100?

IV – Statistical tests:

4.1 The ATLAS Transition Radiation Tracker (TRT) is a particle detector, which can separate electrons from pions through the number of High Threshold (HT) signal. On a track which crosses 35 detector units (straws), the probability for each straw to yield a HT signal is 4.2% for pions and 22.6% for electrons.

- What distribution does the number of HT signals N_{HT} follow for electrons?
- Imagine a test, which requires $N_{\text{HT}} > 6$. What is the efficiency of this test for pions and electrons?
- If there are 100 times more pions than electrons, what is the minimum requirement on N_{HT} that gives an 80% pure electron sample? And what is the electron efficiency for such a selection?

V – Fitting data:

5.1 Counting the number of photons from an experiment yielded the following histogram:

bin	N	bin	N	bin	N	bin	N	bin	N	bin	N
$[-6, -5]$	1	$[-4, -3]$	43	$[-2, -1]$	86	$[0, 1]$	99	$[2, 3]$	76	$[4, 5]$	18
$[-5, -4]$	11	$[-3, -2]$	64	$[-1, 0]$	115	$[1, 2]$	99	$[3, 4]$	41	$[5, 6]$	1

- What is the mean of this data and is it consistent with zero?
- Assume a Gaussian distribution 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 alternative hypotheses with a maximum of 3 parameters to find a better match.

5.2 An experiment has yielded the following 14 measurements, where the uncertainty on y, σ_y has been estimated to be 0.02:

x	y	x	y	x	y	x	y	x	y	x	y	x	y
0.1	0.19	0.3	0.43	0.5	0.73	0.7	0.98	0.9	1.11	1.1	1.22	1.3	1.27
0.2	0.32	0.4	0.60	0.6	0.85	0.8	1.08	1.0	1.15	1.2	1.22	1.4	1.34

- Test if there is a linear relation between x and y.
- Assume the function to be piecewise linear, i.e. consisting of two linear functions joining continuously in one point (a kink). How would you write down such a function algebraically?
- Is the kinked fit good? And what is the position of the kink and its uncertainty?