

# Applied Statistics

## Problem set on curriculum for applied statistics

The following is a problem set on the curriculum lectured on so far. It will be handed out Tuesday the 13th of October 2009, and a solution in writing should be handed in Monday the 19th of October at the latest. Working in groups is allowed, but separate solutions are required. 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, in particular problem 4, where the dataset has to be obtained from the WWW.

Good Luck, Troels

---

### I – Distributions and probabilities:

**1.1** Little Peter rolls a normal die 13 times and obtains 7 sixes. What is the probability of obtaining this result, or something more extreme? Did he cheat?

**1.2** A beam of particles contains  $10^{-4}$  parts elektrons the rest being photons. The particles pass through a detector with two layers, which gives the signals 0, 1 or 2 depending on how many layers the particle is detected in. The probability for these signals are for electrons ( $e$ ) and photons ( $\gamma$ ) as follows:

$$\begin{array}{ll} P(0|e) = 0.001 & \text{og} & P(0|\gamma) = 0.99899 \\ P(1|e) = 0.01 & & P(1|\gamma) = 0.001 \\ P(2|e) = 0.989 & & P(2|\gamma) = 0.00001 \end{array}$$

- What is the probability for a particle to be a photon, if it is detected in one layer?
- What is the probability for a particle to be an electron, given signal in two layers?

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

- What is the expectation value and the variance of  $1/x$ ?
- Compare the expectation value  $E[1/x]$  with  $1/E[x]$ .

**1.4** Calculate the mean and the width of the following distributions:

- $f(x) = \frac{1}{2} \sin(x)$ ,  $x \in [0, \pi]$ .
- $f(x) = \frac{1}{2} e^{-|x-2|}$ ,  $x \in [-\infty, \infty]$ .

### II – Error propagation:

**2.1** Tennis balls are officially required to weight between 56.0 and 59.4 grams. For which mean and uncertainty will 90% of tennis balls lie within these requirements?

If a tennis serve has a speed of 73 m/s with an uncertainty of 2%, what is then the kinetic energy of the ball, if there are no correlations? And with a correlation of  $\rho = -0.7$ ?

**2.2** If  $\theta = 0.54 \pm 0.02$ , what is then 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:

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

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

**3.1** Let  $f(x) = \frac{1}{2\sqrt{x}}$  be a PDF for  $x \in [0, 1]$ .

- Which method should be used to generate this distribution? Why?
- Make an algorithm, which from a uniform distribution of random numbers in the interval  $[0, 1]$ , generates 1000 numbers following the PDF  $f(x)$ . Calculate the average on these numbers and the uncertainty on the average. Compare this value with the analytically calculated average.

### Estimators:

**4.1** Consider the dataset on Tibetan skull sizes, which can be found at:

[http://www.nbi.dk/~petersen/Teaching/Stat2009/Data\\_TibetanSkulls.txt](http://www.nbi.dk/~petersen/Teaching/Stat2009/Data_TibetanSkulls.txt)

- Hvor stor er middelværdien og spredningen for kranier af type A og type B for hver af de fem variable? Og hvor stor er separationen (i.e.  $(\mu_A - \mu_B)/\sqrt{\sigma_A^2 + \sigma_B^2}$ ) mellem type A og type B for hver af de fem variable?
- Hvor stor er den lineære korrelation mellem de forskellige variable for hver type kranie?
- Hvor godt kan du separere de to typer af kranier ved hjælp af en Fisher discriminant, henholdsvis uden og med korrelationer inkluderet?

**4.2** Consider the dataset on leaders lifetime after inauguration, election or coronation, which can be found at:

[http://www.nbi.dk/~petersen/Teaching/Stat2009/Data\\_LifetimeLeaders.txt](http://www.nbi.dk/~petersen/Teaching/Stat2009/Data_LifetimeLeaders.txt)

- Is the lifetime equally long for each of the three categories?
- Does the lifetime change with time for any of the categories?

### Fitting data:

**5.1** An experiment gave the following result, where the uncertainty on  $y$ ,  $\sigma_y$ , has been estimated to be the square root of the number plus a systematic uncertainty of 2.0 to be added in quadrature:

x	y	x	y	x	y	x	y
0.05	13.8	0.45	9.9	0.85	12.8	1.25	12.1
0.15	7.4	0.55	19.6	0.95	8.8	1.35	9.8
0.25	6.5	0.65	23.3	1.05	7.9	1.45	3.4
0.35	7.1	0.75	19.2	1.15	14.6	1.55	11.1

- Assume a linear relation between  $x$  og  $y$ , and make a  $\chi^2$ -fit to data. Is the fit good?
- If one expects a signal somewhere in the region 0.6 to 0.8, what would you then fit with? Discuss the validity of the fit and the significance of the signal.

---

*Those who are good at archery learnt from the bow and not from Yi the Archer.  
Those who know how to manage boats learnt from boats and not from Wo [the legendary boatman].  
Those who can think learnt for themselves and not from the sages.*

[Guan Yin, 8th Century]