Home exam in "Statistical Methods" 8–10.12.2020

(To be returned by 17.00 Helsinki time on Thu 10.12.2020 through Moodle course-page: `https://moodle.helsinki.fi/course/view.php?id=30207`)

1. (i) The weights of the eggs produced by a farmer's hens have a standard deviation of 15 g. He feeds some hens a vitamin supplement, which would be cost-effective if the weights of the eggs increase by at least 3 g. He measures 25 eggs from the vitamin-fed hens and the average weight has increased by 4 g. Is the increase significant? [1.5p]

   (ii) During a particular meteor shower, meteors fall at the rate of 6 per hour. What's the probability to observe none in 30 minutes? Estimate also the probability of observing at least one in 10 minutes? [1.5p]

   (iii) A particle beam consists of a tiny 10^{-3} fraction of electrons, the rest are photons. The beam passes through a double-layered particle detector giving signals in either zero, one or two layers. The probability for an electron to give signal in zero and one of the layers are 10^{-3} and 0.1, respectively. The probability for a photon to give signal in one and two of the layers is 0.01 and 10^{-4}, respectively. Give the probabilities:
   a) that a particle detected in one layer is a photon?
   b) that a particle detected in two layers is an electron? [1.5p]

   PS1 GC AAchen

   only? WHat is the probability for the particle to be a photon given a detected signal in one lay only

   (iv) The age of the Earth can be estimated from the ratio of current amounts of Uranium isotopes 235 ("U235") and 238 ("U238"), $f = N_{\mathrm{U235}}/N_{\mathrm{U238}}$. Assume equal amounts at Earth's creation (i.e. $f(t=0) = 1$). The half-lifes are $\tau_{\mathrm{U235}} = (0.703 \pm 0.003) \cdot 10^9$ y and $\tau_{\mathrm{U238}} = (4.47 \pm 0.02) \cdot 10^9$ y. Assume that τ and f uncertainties are uncorrelated.
   a) Given a current ratio $f = 0.00726 \pm 0.00005$, what is your estimate of the age of the Earth and the corresponding uncertainty?
   b) How much does the estimate and its uncertainty change if $f(t=0)$ is quite uncertain, e.g. $\mathbf{0.30 \pm 0.05}$? [2.5p]

2. Simulate the electron production at the last dynode in a photomultiplier tube (PMT). The number of electrons produced at the ith dynode for each incoming electron can be modeled as a Poisson variable n_i with mean ν_i. Suppose a PMT has N dynodes. The number of electrons n_{out} produced for a single incident photoelectron (PE) has an expectation value (= "gain"):

$$\mu_{out} = E[n_{out}] = \prod_{i=1}^{N} \nu_i \, .$$

   (i) Determine the n_{out} distribution from a single initial PE after $N = 6$ dynodes with $\nu = 3.3$ for each dynode. Estimate the expectation value μ_{out} and its standard deviation $\sigma_{\mu_{out}}$. Make sure that the relative precision of your estimates for these variables are better than 1 %. [4p]

(ii) Calculate from your simulations the fraction of single initial PE's producing no electron signal at all. Compare the result with an analytic estimate where you sum all the largest contributions to this fraction using directly their Poisson probability distribution estimates. [1.5p]

3. In an analysis, the measured events need to be classified into two different categories: a and b. For each event two variables (x_1, x_2) are measured.

   (i) Build a new variable, x_3, as a linear (alternatively non-linear) combination of x_1 and x_2 for linear (alternatively non-linear) optimal discrimination between the two event categories. In the same folder as the exam paper, you find two training samples, `exam2020_cata.txt` and `exam2020_catb.txt` that contain the x_1 (1. column) and x_2 (2. column) values of events from only category a and only category b, respectively. Plot the x_3 value for the events of category a and b. [2p]

   (ii) What is the requirement on x_3 to have a 90 % rejection of events from category a? What is the efficiency for accepting events of category b with this requirement? Plot the data points of the two training samples in the $x_1 x_2$-plane using different colours. Add the x_3 requirement giving 90 % rejection of category a into the same plot. [2p]

   (iii) Assume the observation of an event at $x_1 = 0.30$ and $x_2 = 0.30$. Calculate the probability that the event belongs to category a and that the event belongs to category b. Assume the category-dependent probability distributions for both variables (x_1 and x_2) to be independent gaussians and that the event belongs to one of the two categories. [1.5p]

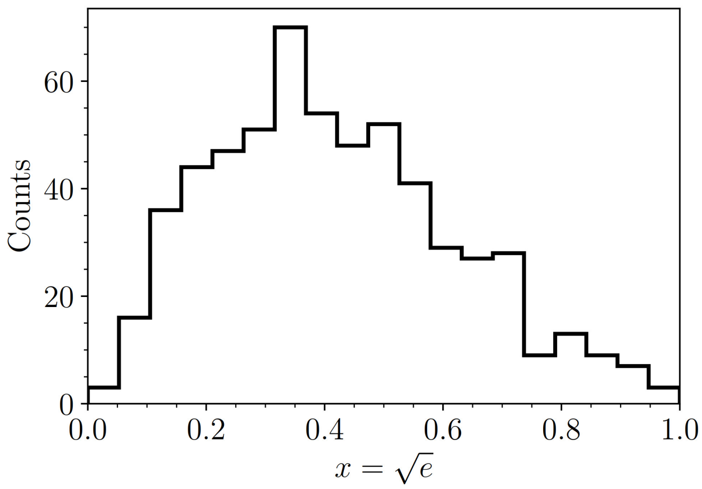
4. A precise gamma ray instrument measure the decay time of nuclei. Data sets have been produced from two samples, `nucl_sample1.txt` and `nucl_sample2.txt`. The decay times are given in arbitrary units.

   (i) Compare the two data sets: Are they compatible to originate from same decay time distribution i.e. decay of same nucleus or nuclei? [2p]

   (ii) Assume that the data sets correspond to exponential lifetime distributions, estimate the lifetime and its uncertainty for each set. [3.5p]

5. Lately accurate space telescopes have been able to measure the orbits of "exoplanets", planets outside our solar system. The orbits of planets bound to a star are elliptical and can be described by their eccentricity, e, defined as $(d_{\mathrm{far}} - d_{\mathrm{close}})/(d_{\mathrm{far}} + d_{\mathrm{close}})$, where d_{far} (d_{close}) is the furthest (closest) distance from the star the planet is during its orbit. For bound orbits e is $0 \leq e < 1$. In the same folder as the exam paper, you find the

file `exoplanet_ecc.txt` containing the eccentricities of 587 exoplanet orbits taken from the Exoplanet Orbit Database 29.12.2017.



The figure above shows the distribution of $x = \sqrt{e}$ for these exoplanets (the bin width is 0.05). The histogram indicates that the distribution of x is approximately triangular, i.e., increasing linearly from zero at $x = 0$ to a maximum at some point $x = c$, and then decreasing linearly from the maximum to zero at $x = 1$. This triangular shape defines a continuous probability density functions $f(x|c)$ on the $0 < x < 1$ interval depending on the single parameter (one can assume $0 < c < 1$).

(i) Define the probability distribution $f(x|c)$ using the above info. [1.5p]

(ii) Determine the parameter c and its uncertainty. Assume that the uncertainties of the eccentricity measurements can be neglected. Is the triangular shape a good assumption? What is the P-value? [3p]

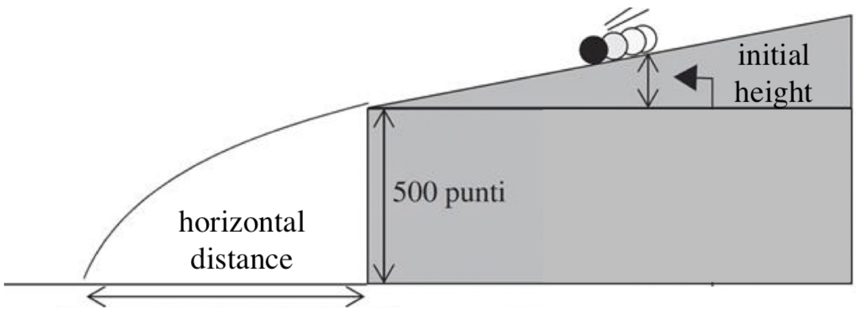
(iii) Estimate the uncertainty of c also using Monte Carlo methods making sufficient number of "pseudo-experiments" with the same statistics as Figure 1 so that the relative statistical precision is 2 % or better. Describe in detail how to generate random values x with the probability distribution $f(x|c)$ as defined in (i) using the estimated c of (ii) as "true" c-value and add any computer code used to your answer. [2p]

6. Galileo Galilei was one of the first to study the trajectory of a freely falling body. In 1609, he proved mathematically the relation between the horizontal velocity component and the initial height. His discovery of this result, which preceded the mathematical proof by a year, was

the result of empirical findings from studies with a ball and an inclined ramp. The ramp was placed on a table near the table edge as shown by the Figure below. The ball first rolls down the ramp from an initial height, h, over the table. Hitting the table, the balls' motion is smoothly turned horizontal, then it rolls over the edge of the table to hit the floor at a horizontal distance, d, from the table edge.

(i) Galilei's data can be found in the table below. Heights and distances are given in punti, one punti is slightly less than a mm.

d	h	d	h	d	h	d	h
253	100	337	200	395	300	438	400
469	500	495	600	520	700	534	800
557	900	573	1000	609	1200	643	1400



Let's assume the heights h are known with negligible uncertainties ($\sigma_h \approx 0$), and that the horizontal distances d can be regarded as Gaussian random variables with standard deviation $\sigma_d = 13$ punti. Test the following hypotheses with Galilei's data:

a) linear: $d = a \cdot h + b$

b) quadratic: $d = a \cdot h + b \cdot h^2 + c$

c) square root: $d = a \cdot \sqrt{h} + b$,

where a, b and c are constants to be determined from data. Determine the constants and their uncertainties as well as the goodness of fit for all three hypotheses. Based on the tests, can you exclude some hypothesis (assuming the assumption on the values of σ_h and σ_d are correct). What is the physics conclusion on the horizontal velocity assuming the friction is minimised in Galilei's experiment. [3p]

(ii) Assume also that the heights h have Gaussian uncertainties with a standard deviation of $\sigma_h = 4$ punti. Determine a, b and c and their uncertainties as well as the goodness of fit for the same three hypotheses as (i). Any change in the results and in the physics conclusions? [3p]