General exam in "Statistical Methods" 18.1.2019

NB! Students are allowed to use a calculator as help when completing the exam!

1. Define following concepts (briefly):
   a) the central limit theorem (i.e. why are errors usually Gaussian) [2p]
   b) Kolmogorov-Smirnov test (application and test statistic) [2p]
   c) Fisher's linear discriminant (application and discriminating function) [2p]

2. Describe the maximum likelihood (ML) method ("suurimman uskottavuuden menetelmä"). Describe also how a ML estimator and its variance are determined. Give the advantages and disadvantages of the ML method. [6p]

3. a) In a time-of-flight (TOF) detector, the time used by a charged particle to traverse the detector is measured. These time measurements are then used for separating different types of particles having the same momentum using their rest mass differencies (special relativity applied). Suppose a test statistic w built from the TOF measurements follows a Gaussian distribution centered at $0.0(-3.0)$ for pions(kaons), with a standard deviation equal to one for both hypotheses. Assume a kaon selection with $w < -2.0$ requirement.
   (i) What is the kaon selection efficiency when requiring $w < -2.0$? [1p]
   (ii) What is the kaon tag probability for a true pion (i.e. $w < -2.0$)? [1p]
   (iii) Suppose a sample of particles consists 95 % pions and 5 % kaons. What is the purity of the kaon sample selected by $w < -2.0$? [1p]

   b) 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. Note $N(t) = N(t = 0) \cdot e^{-t \ln 2/\tau}$.
   (i) Given a current ratio $f = 0.00726 \pm 0.00005$, what is your estimate of the age of the Earth and the corresponding uncertainty? [2p]
   (ii) How does result and uncertainty change if $f(t = 0) = 0.30 \pm 0.10$? [1p]

4. a) Pekka's boss checks his workers twice per hour to make sure they work. Pekka is tired and needs a short nap. If he takes a 10 minute nap, what is the probability of getting away with it? How long could the nap maximally be for the probability of getting caught to be less than 1 in 10? [2p]

   b) A Breit-Wigner distribution describes the energy distribution of an unstable quantum state in e.g. atomic, molecular, nuclear or particle physics

$$dN/dE \quad \propto \quad \frac{(\Gamma_s)^2}{(E - E_s)^2 + (\Gamma_s/2)^2} \,,$$

   where E_s and Γ_s are the mean energy of the state and the uncertainty on the mean energy of the state. Latter is related to the Heisenberg uncertainty principle and inversely proportional to the lifetime of the state. Describe in detail how to generate the E_s distribution of any quantum state in your Monte Carlo code using the inverse transform method. [4p]

5. In a classic experiment on radioactivity, Rutherford and Geiger counted the number of alpha (α) decays. The table below gives the number of times n_m

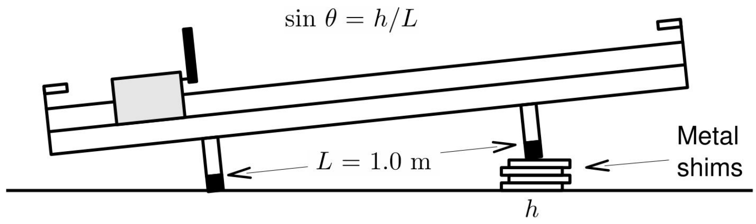
that they observed m decays of α's in a fixed time interval of 7.5 s. Assuming the source to consist of very many radioactive atoms and the probability for any one of them to emit an α in a short interval is small, one expects the number of decays in a time interval to follow a Poisson distribution. Deviations from this hypothesis would indicate that the decays were not independent e.g. that the emission of an α particle might cause neighboring atoms to decay, resulting in many decays in a short time interval.

m	n_m	m	n_m	m	n_m	m	n_m
0	57	4	532	8	45	12	0
1	203	5	408	9	27	13	1
2	383	6	273	10	10	14	1
3	525	7	139	11	4	> 14	0

(i) Verify with a test statistic whether data are Poisson distributed. [4p]
(ii) What is (approximately) the quality of the the Poisson hypothesis. Can this hypothesis be excluded? [2p]

6. A physics student studied carts sliding down an incline with a negliable friction (an "air track", see figure). The data are the measured times for a cart released from rest to travel various distances d along the track. Assume d to be precise and the uncertainty on t to be Gaussian with a σ of 0.05 s.

t [s]	d [cm]	t [s]	d [cm]	t [s]	d [cm]
0.75	1	2.45	16	4.15	49
6.10	100	7.85	144	9.10	225



(i) Consider the hypothesis $t = \alpha\sqrt{d}$. Find the least square estimator for α. What is the minimum χ^2 value? Can the hypothesis be excluded? [3p]
(ii) Estimate the uncertainty on α assuming variance of α to be equal to the RCF bound ($= (1 + (\partial b/\partial\alpha))^2 / E\left[-(\partial^2 \ln L/\partial\alpha^2)\right]$, where b = bias). [2p]
(iii) Assume no friction and $g = 9.81$ m/s^2. Estimate inclination angle θ. [1p]

Table of values of the confidence level for different values for the inverse of the cumulative gaussian distribution:

central		one-sided	
$\Phi^{-1}(1-\gamma/2)$	$1-\gamma$	$\Phi^{-1}(1-\alpha)$	$1-\alpha$
1	0.6827	1	0.8413
2	0.9544	2	0.9772
3	0.9973	3	0.9987