General exam in "Statistical Methods" 17.1.2020

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) the Kolmogrov-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 an ML estimator and its variance are determined. Give the advantages and disadvantages of the ML method. [6p]

3. 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) In a time-of-flight (TOF) detector, the time used by a charged particle to traverse the detector is measured. These measurements are then used for separating different types of particles having the same momentum using their rest mass differences (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 1.5 for both hypotheses. Assume a kaon selection with a $w \geq 1.5$ requirement.
   (i) What is the kaon selection efficiency when requiring $w \geq 1.5$? [1p]
   (ii) The probability for a pion to be tagged as a kaon (i.e. $w \geq 1.5$)? [1p]
   (iii) What is the per centage of kaons in the subsample selected by $w \geq 1.5$ from a random particle sample consisting 95 % of pions and 5 % of kaons?[1p]
   (iv) What w requirement on the same sample would select ~50 % kaons?[1p]

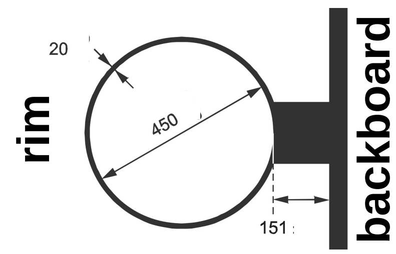
4. In a classic experiment on radioactivity, Rutherford and Geiger counted the number of alpha (α) decays. The table below gives the number of times n_m that they observed m decays of α's in a fixed time interval of 7.5 s. Assuming the source to consist of very many radioactive atoms and the probability 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]

TURN PAGE →

5. (i) Your professor is a basketball fan but not such a good basketball shooter. Determine analytically his scoring efficiency assuming that he throws the ball randomly and uniformly within a 40 cm distance of the centre of the rim (the "ring" which the ball has to go through to score in basketball). Hint: The uniform distribution is for the ball's geometrical centre. Take care that the whole ball fits into the rim. The circumference of a basketball is 749 mm and the inner diameter of the rim 450 mm. Assume both the ball and the rim to be perfect circles (a two-dimensional problem). [2.5p]
(ii) Repeat (i) taking a Gaussian-distributed and rim-centered throwing probability with a 60 cm width in both projections. Describe in detail how to determine the scoring efficiency using Monte Carlo (MC) methods [2.5p]
(iii) Players often use the backboard to score. Repeat (ii) allowing the ball to be reflected from a backboard if thrown too far according to a function $x \to -2x$, where $x =$ distance the ball was thrown behind the backboard. Assume he shoots perpendicular w.r.t. the backboard. Take also the rim width, 20 mm, and the rim-to-backboard distance, 151 mm, into account (see the figure to the left, all distances in mm). Describe how to modify the answer of (ii) to determine the scoring efficiency using MC in this case. [1p]



6. A pendulum is used to determine the gravitational acceleration g. If the pendulum swings in a small angle, the time of one swing (the period) T depends only on the pendulum length L and g, $T = 2\pi\sqrt{\frac{L}{g}}$. Some students did measurements with 5 different pendulums. The results are shown below:

Length (cm)	85.0±1.0	75.0±0.8	70.0±0.7	60.0±0.6	50.0±0.5
Period (s)	1.850±0.011	1.737±0.011	1.678±0.011	1.554±0.011	1.419±0.011

(i) Determine the gravitational acceleration g and its uncertainty at the location where the experiment was made, neglecting the uncertainties of the pendulum length. What is the χ^2_{min} value? [2p]

(ii) Calculate the variance of g using the RCF-bound assuming no bias. [2p]

(iii) From (i) and (ii), do you suspect that the students altered the measurements (or overestimated the uncertainties) to improve the result? Motivate.[2p]
NB! The RCF bound: $(1 + (\partial b/\partial g))^2 \,/\, \mathrm{E}\left[(\partial^2\chi^2/\partial g^2)/2\right]$, where $b =$ bias.

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