General exam in "Statistical Methods" 22.1.2021

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

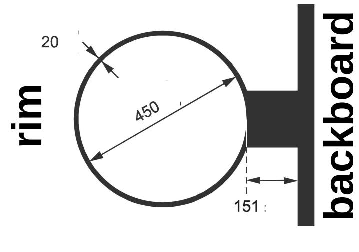
1. Define following concepts (briefly):
   a) the central limit theorem (i.e. why are errors usually Gaussian) [2p]
   b) Pearson's chi square test (application and test statistic) [2p]
   c) Fisher's linear discriminant (application and discriminating function) [2p]

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. (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. [3p]
   (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 methods [3p]



5. A group of physicists is trying to solve a longstanding debate between two famous physicists. Noble T. Menows model predicts a distribution in a measured variable x, labelled "Menow pred.", and Noble S. Mines model another distribution in x, labelled "Mine pred.". Table contains probability that x is contained in a certain bin for the two models. The group obtains an x distribution, labelled "data (events)", after background removal.

bin	1	2	3	4	5	6	7	8	9	10
bin starts at	-5.0	-4.0	-3.0	-2.0	-1.0	0.0	1.0	2.0	3.0	4.0
bin ends at	-4.0	-3.0	-2.0	-1.0	0.0	1.0	2.0	3.0	4.0	5.0
data (events)	1	2	9	35	49	57	28	15	4	0
Menow pred. (%)	0.15	0.75	4.5	15.8	28.8	28.8	15.8	4.5	0.75	0.15
Mine pred. (%)	0.5	2.5	8.0	15.0	24.0	24.0	15.0	8.0	2.5	0.5

(i) Plot the two models normalized to data together with the data. Can any of the two models be excluded based on the data? Which model is more consistent with the data? Assume measurement uncertainty in a bin to be square root of the number of observed events in that bin. Hint: P-value = 0.05 corresponds to $\chi^2 > 16.9\,(18.3)$ for nine (ten) degrees-of-freedom. [3p]
(ii) What statistics (= number of events) is needed (on average) to exclude both models if the true distribution lays in between, i.e. is:

bin	1	2	3	4	5	6	7	8	9	10
true distrib. (%)	0.3	1.2	6.3	15.3	26.9	26.9	15.3	6.3	1.2	0.3

Plot the corresponding "average measurement" with statistical uncertainties together with the two models. How many years of data taking does that correspond to (assume a yearly rate of 200 events)? [3p]

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	75.0	70.0	60.0	50.0
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. What is the $\chi^2_{\min}$ value? [3p]

(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.[1p]
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