General exam in "Statistical Methods" 14.2.2025

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) unfolding/deconvolution (of a measured distribution) [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) 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 $f(t=0) = 0.30 \pm 0.03$? [2p]

   b) 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]

4. During the 1987 Supernova (SN1987A), interactions from cosmic neutrinos were seen in two large underground experiments: IMB in USA and Kamiokande (KAM) in Japan. Most of the cosmic neutrino events are expected to produce recoil particles with an angular distribution:

   $$dN/d\cos\theta \propto 1 + \gamma\cos\theta\,,$$

   where $\gamma \approx 0.1$, w.r.t. direction towards the source, assumed to be SN1987A. The recoil angles θ of the 8 IMB events were 80, 44, 56, 65, 33, 52, 42 and 104 degree and of the 12 KAM events 18, 40, 108, 70, 135, 68, 32, 30, 38, 122, 49 and 91 degree. The uncertainty in θ determination can be neglected.
   (i) Test first whether the two results are compatible (i.e. that they originate from the same distribution). [3p]
   (ii) Test also whether the combined IMB and KAM data is compatible with the expected angular distribution (when treated as one data sample). [3p]

5. (a) Suppose a charged particle beam consists 90 % of pions ("π") and 10 % of kaons ("K"), i.e. 2 types of elementary particles, and we measure the beam with a particle identification device with following performance. If a K traverses, 82 % of time identified as a K, 5 % as a π and reminder no decision reported. If a π traverses, 70 % of time identified as a π, 10 % as a K and reminder no decision reported. Give probabilities after measurement that particle was a π and that particle was a K (2 probabilities per case) if:
   (i) Particle identified as a π. [1p]
   (ii) Particle identified as a K. [1p]

(iii) The particle identification device reports no decision. [1p]

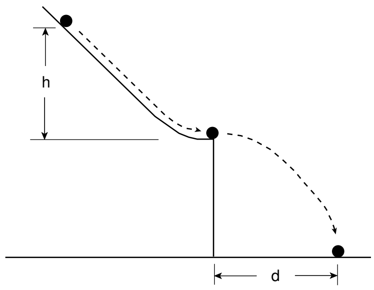
(b) In W gauge boson decay $W^- \to e^- \bar{\nu}_e$ at CERNs Large Hadron Collider, the distribution of the transverse momentum (p_T) of the electron (e^-) is:

$$f\left(p_T\right) \propto p_T / \sqrt{1-\left(p_T / p_{max}\right)^2},$$

where p_{max} is maximal momentum of e^-. Describe in detail how to generate p_T distribution of e^- with Monte Carlo using inverse transform method. [3p]

6. Galileo studied the motion of balls on an inclined ramp. The ball trajectory was made horizontal before it got over the edge (see figure). The horizontal distance d was measured for different values of the initial height h. Assume h to be precise and the uncertainty on d to be Gaussian with a σ of 15 mm.

height h [mm]	horizontal distance d [mm]
1000	1500
828	1340
800	1328
600	1172
300	800



(i) Consider the hypothesis $d = \alpha\sqrt{h}$. Find the least square estimator for α. What is the minimum χ^2 value? Discuss validity of the hypothesis. [3p]

(ii) Estimate the uncertainty on α. Assume the variance of α to be equal to the RCF bound ($= \left(1 + (\partial b/\partial\alpha)\right)^2 / \mathrm{E}\left[(\partial^2\chi^2/\partial\alpha^2)/2\right]$, where $b =$ bias). [3p]

Kolmogorov-Smirnov table:

Critical values, $d_{alpha};(n)^a$, of the maximum absolute difference between sample $F_n(x)$ and population $F(x)$ cumulative distribution.

Number of trials, n	Level of significance, α			
	0.10	0.05	0.02	0.01
1	0.95000	0.97500	0.99000	0.99500
2	0.77639	0.84189	0.90000	0.92929
3	0.63604	0.70760	0.78456	0.82900
4	0.56522	0.62394	0.68887	0.73424
5	0.50945	0.56328	0.62718	0.66853
6	0.46799	0.51926	0.57741	0.61661
7	0.43607	0.48342	0.53844	0.57581
8	0.40962	0.45427	0.50654	0.54179
9	0.38746	0.43001	0.47960	0.51332
10	0.36866	0.40925	0.45662	0.48893
11	0.35242	0.39122	0.43670	0.46770
12	0.33815	0.37543	0.41918	0.44905
13	0.32549	0.36143	0.40362	0.43247
14	0.31417	0.34890	0.38970	0.41762
15	0.30397	0.33760	0.37713	0.40420
16	0.29472	0.32733	0.36571	0.39201
17	0.28627	0.31796	0.35528	0.38086
18	0.27851	0.30936	0.34569	0.37062
19	0.27136	0.30143	0.33685	0.36117
20	0.26473	0.29408	0.32866	0.35241