NTNU
Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for
Exercise 4 in TMA4295 Statistical inference

Academic contact during examination: Professor Gunnar Taraldsen

Phone: see Wikipage

Examination date: Wednesday in week 36 to Monday in week 38, 2023

Examination time (from–to): Wed 21:00 – Mon 12:00

Permitted examination support material: C

— Tabeller og formler i statistikk, Akademika
– Mathematische Formelsamlung (Matematisk formelsamling) by K. Rottmann
— Stamped yellow A5 sheet with your own handwritten notes
— A specific basic calculator.

Other information:

You may write in English or Norwegian.

All answers must be justified. The answers must include enough details to see how they have been obtained. You must, as always, formulate necessary assumptions as part of the proof of a claim.

All 14 sub-problems carry the same weight for grading.

Language: English

Number of pages: 3

Number of pages enclosed: 0

Checked by:

Informasjon om trykking av eksamensoppgave	
Originalen er:	
1-sidig ☐	**2-sidig** ☒
sort/hvit ☐	**farger** ☒
skal ha flervalgskjema ☐	

Date Signature

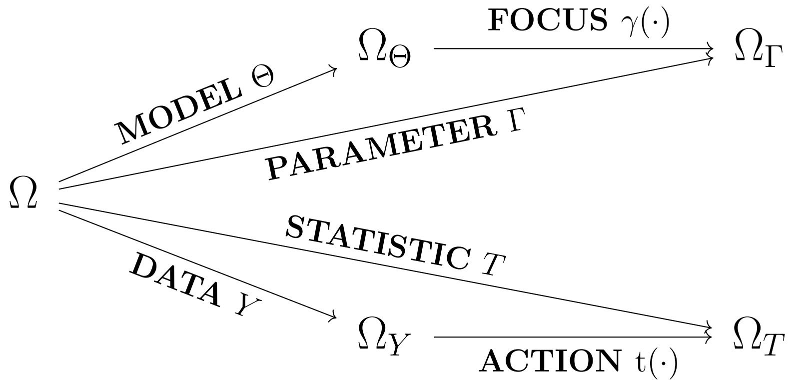

Figure 1: Functions in statistical inference.

Figure 1 can be useful for the more theoretical part of this exercis. The problem description and tasks are given on the next two pages.

Figure 2: A coiled paper dartboard with 10 beds defined by concentric rings.

Problem 1 Foundations of Statistics

a) Prove linearity of the expected value of simple random variables by considering: (i) Two Bernoulli variables. (ii) A simple variable and a Bernoulli variable. (iii) Two simple variables.

b) Prove the change-of-variables theorem $\mathrm{E}\,\phi(W) = \mathrm{E}_W\,\phi$ for a simple random variable W. Use this and the function $\phi(x, y) = \alpha x + \beta y$ to provide an alternative proof of linearity.

c) Show that the level sets of the data are included in the level sets of any given statistic. What is the relation between the level sets of a minimal sufficient statistic and the likelihood statistic? Use level sets and resulting equivalence relations to formulate the sufficiency principle.

Problem 2 Statistics and Darts

a) The dartboard shown in Figure 2 indicates a partition of the set

$$\Omega_D = \{d \mid d = (x, y), x, y \in \mathbb{R} \cdot \text{meter}\} \tag{1}$$

into 11 sets $A_0, A_1, \ldots, A_9, A_{20}$ where the subscript indicates the score s. Introduce notation, coordinates, and assumptions to define the sets precisely.

b) Let $p_s = \mathrm{P}(D \in A_s)$ where $D = (X, Y)$ is a random point given by throwing a dart at the dartboard. Show how these probabilities determine the probability of 2^{11} different events. Verify that this family $\mathcal{E}_D$ of events obeys the Kolmogorov axioms for a family of events. Explain that this defines Ω_D as a sample space and as a probability space. Is every subset of Ω_D an event?

c) The previous can be used to define a statistical model where the random point D is the data and the distribution P_D is unknown. Explain that $\Omega_\Theta = \mathbb{R}^{11}$ can be used as a model space. What is the range $\mathrm{R}(\Theta)$ of possible model points $\theta = (p_0, \ldots, p_9, p_{20})$?

d) Explain that the score S is a statistic with level sets given by the dartboard partition. Prove that the expected score $\mu = \mathrm{E}(S)$ is a parameter. How is this related to the change-of-variables theorem? Formulate the law of large numbers to give an interpretation of $\mathrm{E}(S)$ and $\mathrm{P}(S \in A)$.

e) A second statistical model is given by assuming that the data d is a random sample $d_1, \ldots, d_n$ from the distribution of a single dart throw as described above. The model space Ω_Θ for the corresponding experiment is unchanged, but what is the data space Ω_D? What experiment? The new family $\mathcal{E}_D$ of events is still finite, but with how many members? Is $\{d\}$ an event?

f) Explain that

$$m_s = (d_1 \in A_s) + \cdots + (d_n \in A_s) \tag{2}$$

defines a random variable M_s and a random vector $M = (M_0, \ldots, M_9, M_{20})$. Is M a minimal and complete sufficient statistic? Is the distribution of M from an exponential family? What is the name of the family of distributions M belongs to?

g) Find formulas for the maximum likelihood estimators $\hat{\theta}$ and $\hat{\mu}$ for the model θ and the expected score μ. Are the maximum likelihood estimators unbiased?

h) A third statistical model is given by assuming that the data $d_1, \ldots, d_n$ is a random sample $(x_1, y_1), \ldots, (x_n, y_n)$ from a bivariate normal distribution with mean $(0, 0)$m, independent components, and common unknown standard deviation σ. What is now the data space Ω_D and the family of events $\mathcal{E}_D$? Find a formula for the likelihood function. Show that the maximum likelihood estimator $\hat{\sigma}$ is a complete minimal sufficient statistic.

i) The estimate $\hat{\mu}(d)$ does not obey the sufficiency principle given the bivariate model, but the new maximum likelihood estimate $\mu(\hat{\sigma})$ does. Explain this. A third estimate is given by

$$\mu^* = \mathrm{E}(\hat{\mu}(D) \mid \hat{\sigma}(D) = \hat{\sigma}) \tag{3}$$

Show that μ^* is unbiased and why it has smaller standard uncertainty than any other unbiased estimate. How would you calculate the standard uncertainty? (You need not do the actual calculation!)

j) A group G of transformations is defined on the data space by $g(d) = gd$ where g is a positive real number. The third statistical model is a group family with respect to this group. Explain what this means. Is the estimate $\hat{\sigma}$ equivariant?

k) Determine the distribution of $\hat{\sigma}$. How can this be used to obtain confidence intervals for σ and μ? Use the Basu theorem to show that $\hat{\sigma}$ and $a = (x, y)/(\overline{x} + \overline{y})$ are independent statistics. Is a a maximal invariant statistic?