Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4295 Statistical inference

Lecturer Fall 2023: *Gunnar Taraldsen*

Scribe: *Venja Reitan Fjøsne* and *Halvard Emil Sand-Larsen*

Lecture 1 in week 34: A general overview of the course

Contents

1 The NTNU pendulum

Wikipedia states that the length of the pendulum at NTNU is 25m, while a previous student measured the length to be 25.2965m. These statements are both incomplete because they lack uncertainty, which is an important part of this course. To get the uncertainty we need more data. For instance 20 measurements: $25.2965, ..., 25.1413$. We assume that the data is a random sample from $\mathrm{N}(\mu, \sigma^2)$ where $\mu = \lambda =$ true length and $n = 20$. We denote X_1 as random quantity while x_1 represents a realization. Note that we use quantity and not variable since the values have a unit, and variable only applies for real numbers. Furthermore, $x_1, ..., x_n$ and $X_1, ..., X_n$ are both random samples, which implies independence between the different elements. We define the random quantity as follows:

$$X_i : \Omega \to \mathbb{R} \cdot \mathrm{m} \quad , \quad X : \Omega \to \mathbb{R}^n \cdot \mathrm{m}$$

The cumulative function is defined as

$$F(X \leq x) = P(X \leq x) = P\{\omega | X(\omega) \leq x\}, \omega \in \Omega$$

In the formula above both ω and Ω are not known, and the convention is that Greek letters are unknown.

Definition 1.1 (Statistical model) A statistical model is a specification of a distribution for the data for each model $\theta \in \Omega_\theta$

Remark: Here Ω_θ is the model space, and the θ for the pendulum would be $\theta = (\mu, \sigma) \in H = \{(\mu, \sigma) | \mu \in \mathbb{R}, \sigma > 0\} = \Omega_\theta$

The data space for the pendulum is $\Omega_x = \mathbb{R}^n \cdot \mathrm{m}$.

2 Fundamental concepts

Definition 2.1 (Statistic) A statistic is a function of data.

Some examples of statistics: $\overline{x}, s, \hat{\sigma}, F_N, L, \ldots$ The likelihood function L is defined as $L(\theta) = f(x|\theta), X \sim f$. Note that it depends on the data.
When we have independence, the joint density function is equal to the likelihood function: $L(\theta) = \Pi_{i=1}^n f(x_i|\theta)$. In this case, since there is a $1 - 1$ correspondence, $(\hat{\sigma}, \hat{\mu})$ is a minimal sufficient statistic.

Definition 2.2 (Parameter) A parameter is a function of the model.

Some examples of parameters: $\theta, \mu, \sigma, f(\cdot|\theta), \ldots$ A general parameter is given by $\tau = \tau(\theta)|\tau(\cdot)$. The parameters may be a function, projection or a set of functions.

3 Data reduction

The expected value of the loss function equals the risk function. But it is not easy to choose the loss function since it needs to be unchanged under scaling and shifts. One solution is by the Fisher information $I_{ij} = g_{ij}$ gives the hyperbolic geometry to find the shortest distance between θ_1 and θ_2 as shown in figure (1). Also deducing from figure (1) the distribution, the length between two points, is found by $ln|\frac{AB}{ab}| = d$.

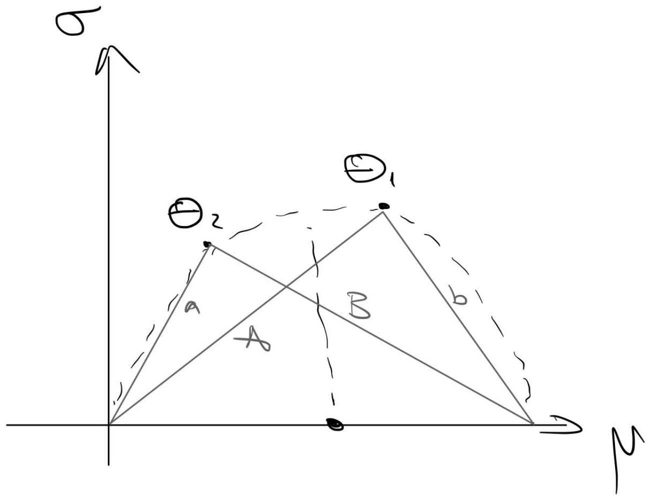

Figure 1: Distance between θ_1 and θ_2 in a hyperbolic geometry.

Furthermore, group theory grants the following results, where g is a combination of linear shift and scaling and a is only linear shifts.

$$\hat{\mu}(gx) = g\hat{\mu}(x)$$

$$\overline{(x+g)} = g + \overline{x}$$

$$\overline{gx} = g \cdot \overline{x}$$

$$\hat{\sigma}(ax) = a\hat{\sigma}(x)$$

Definitions

Index