Inference 2, 2023, lecture 1

Rolf Larsson

October 31, 2023

Outline of the course

- Book:
  Liero, Zwanzig: Introduction to the Theory of Statistical Inference, Chapman and Hall 2012.
- 12 theory lectures
- 3 problem solving lectures (e.g. old exams)
- Examination:
  - Three obligatory home assignments (see studium)
  - Written examination (not with book, but with hand-written formula sheet)

Outline of the lectures

Chapter	Topic	Lectures no.
1	Introduction	
2	Statistical Model	1
3	Inference Principles	2-5
	Problem solving	5
4	Estimation	6-9
5	Testing Hypotheses	10-13
	Problem solving	14
	Old exams	15

Today

Data and Statistical Model

Some notation:

- An observation vector $\mathbf{x} = (x_1, ..., x_n)$
- is a realization of the random vector $\mathbf{X} = (X_1, ..., X_n)$.
- We have $\mathbf{x} \in \mathcal{X}$, the sample space.
- On $\mathcal{X}$, we define a class of probability measures

$$\mathcal{P} = \{\mathrm{P}_\theta : \theta \in \Theta\},$$

  where the set Θ is called the **parameter space**.
- We call $\mathcal{P}$ the **statistical model**.
- Let the set $A \subseteq \mathcal{X}$. Then,
  - $\mathbf{P}_\theta(A) = \int_A f(\mathbf{x}; \theta) d\mathbf{x}$ if $\mathbf{X}$ is continuous,
  - $\mathbf{P}_\theta(A) = \sum_{\mathbf{x} \in A} \mathrm{P}_\theta(\mathbf{x})$ if $\mathbf{X}$ is discrete.

Data and Statistical Model

Definition (2.1)

A **sample** $\mathbf{X} = (X_1, ..., X_n)$ is a collection of **independent** random variables where X_i is distributed according to a distribution $\mathrm{P}_{i,\theta}$, $i = 1, 2, ..., n$, where n is the **sample size**.

Data and Statistical Model

Example 1: Exponential distribution

- Let $X_1, ..., X_n$ be independent and exponential with intensity β.
- Then, for X_i,

$$\mathbf{P}_{i,\beta}(A) = \int_A \beta e^{-\beta x} dx.$$

- Let $\mathbf{A} = A_1 \otimes ... \otimes A_n$.
- Then, (why?)

$$\mathbf{P}_\beta(\mathbf{A}) = \prod_{i=1}^{n} \mathbf{P}_{i,\beta}(A_i) = \beta^n \int_{\mathbf{x}\in\mathbf{A}} e^{-\beta \sum_i x_i} d\mathbf{x}.$$

Data and Statistical Model

Example 2: Poisson distribution

- Let $X_1, ..., X_n$ be independent and Poisson with intensity parameter λ.
- Then, for X_i,

$$\mathbf{P}_{i,\lambda}(A) = \sum_{x \in A} \frac{\lambda^x}{x!} e^{-\lambda}.$$

- Let $\mathbf{A} = A_1 \otimes ... \otimes A_n$. Write down $\mathbf{P}_\lambda(\mathbf{A})$ with notation as above.

Statistic

Definition (2.2)

A **statistic** T is a function of the sample

$$T : \mathbf{x} \in \mathcal{X} \to T(\mathbf{x}) = t \in \mathcal{T}$$

where $\mathcal{T}$ is a set.
With T as a random variable, its distribution is given by

$$\mathrm{P}_{\theta}^{T}(B) = \mathrm{P}_{\theta}(\{\mathbf{x} : T(\mathbf{x}) \in B\}).$$

Statistic

Example 2: Poisson distribution

Let $X_1, ..., X_n$ be independent and Poisson with intensity parameter λ. We observe $\mathbf{x} = (x_1, ..., x_n)$.

1. It $T(\mathbf{x}) = \sum_{i=1}^{n} x_i$ a statistic?
2. Is $T(\mathbf{x}) = x_1$ a statistic?
3. Is $T(\mathbf{x}) = 5$ a statistic?

The Exponential family

Definition (2.3)

A class of probability measures $\mathcal{P} = \{\mathrm{P}_\theta : \theta \in \Theta\}$ is called an **exponential family** if there exists

- a positive integer k,
- real-valued functions $\zeta_1, ..., \zeta_k$,
- real-valued statistics $T_1, ..., T_k$
- and a function h on $\mathcal{X}$
- such that the probability (density) function has the form

$$p(x;\theta) = A(\theta) \exp\left(\sum_{j=1}^{k} \zeta_j(\theta)\, T_j(x)\right) h(x).$$

The Exponential family

Example 1: Exponential distribution

- Let $X_1, \ldots, X_n$ be independent and exponential with intensity β.
- The density function is

$$f(x; \beta) = \beta \exp(-\beta x).$$

Does this distribution belong to the exponential family?

The Exponential family

Which of the following distributions belong to the exponential family?

1. Poisson with parameter λ.
2. $N(\mu, \sigma^2)$.
3. Uniform distribution on $[0, \theta]$.

The Exponential family

Definition

An exponential family with a minimal number k of statistics $T_1, \ldots, T_k$ is called a **strictly k-dimensional exponential family**.

Example 3: Let X be multinomial on $\mathcal{X} = \{1, 2, 3, 4\}$.

1. Does this distribution belong to the exponential family?
2. What is the minimal number k of statistics $T_1, \ldots, T_k$?

The Exponential family

Let $\mathcal{A} = \{x : p(x;\theta) > 0\}$.

Definition

The functions $T_1, ..., T_k$ is called $\mathcal{P}$*-affine independent* if for real constants $c_0, c_1, ..., c_k$,

$$\sum_{j=1}^{k} c_j T_j(x) = c_0 \quad \text{for all } x \in \mathcal{A}$$

$$\Rightarrow$$

$$c_j = 0 \quad \text{for } j = 0, 1, ..., k.$$

The Exponential family

Theorem (2.1)

Let $\mathcal{P}$ be an exponential family. Then

1. *The family $\mathcal{P}$ is strictly k-dimensional if the functions $1, \zeta_1, ..., \zeta_k$ are linearly independent and the statistics $T_1, ..., T_k$ are $\mathcal{P}$-affine independent.*
2. *The statistics $T_1, ..., T_k$ are $\mathcal{P}$-affine independent if the covariance matrix $\mathrm{Cov}_\theta T$ is positive definite for all $\theta \in \Theta$.*

cf example 3 (multinomial distribution on $\{1, 2, 3, 4\}$)

The Exponential family

Theorem (2.2)

Let $\mathcal{P}$ be an exponential family. Then

1. *If $X_1, \ldots, X_n$ is a sample of independent random variables with distributions belonging to the exponential family, then the joint distribution of the vector $\mathbf{X} = (X_1, \ldots, X_n)$ is an element of an exponential family.*
2. *If $X_1, \ldots, X_n$ is a sample of i.i.d. random variables with a distribution of the exponential family form with functions ζ_j, $j = 1, \ldots, k$, and $T = (T_1, \ldots, T_k)$, then the distribution of $\mathbf{X}$ belongs to an exponential family with functions ζ_j and $T(\mathbf{x}) = \sum_{i=1}^{n} T(x_i)$.*

The Exponential family

Example 4: Normal distribution

- Let $X_1, ..., X_n$ be independent $N(\mu, \sigma^2)$.
- Show that $T(\mathbf{x}) = (\sum_i x_i, \sum_i x_i^2)$.

News of today

- Sample
- Statistic
- The Exponential Family
  - Strictly k dimensional
  - The joint distribution of independent random variables which are in the exponential family is also in the exponential family.