TMA4295 Statistical inference

Lecturer Fall 2023: *Gunnar Taraldsen*

Scribe: *Arkadiusz Mikuła* and *Julia Soufflet*

Norwegian University of Science and Technology
Department of Mathematical Sciences

Lecture 1 in week 36:'Birnbaum's theorem'

Contents

1 Fundamental concepts

I feel compelled to write this

Definition 1.1 (Statistic) A statistic is a measurable function of data.

Definition 1.2 (Parameter) A parameter is a function of model.

Definition 1.3 (Event) An event is a set BUT not every set is an event.

2 Birnbaum's theorem

Theorem 2.1 (Birnbaum's theorem) The likelihood principle equals to the sufficiency principle and the conditionality principle.

2.1 The likelihood

The likelihood statistic is always a minimal sufficient statistic.

The likelihood principle, which we did last week, says that all evidence from the data and the model is contained in the likelihood. This means that for two different experiments, with a common model space, if the likelihood is the same then we have the exact same information on both experiment. However, this principle relies on the fact that we have a likelihood.

We do not always have a likelihood, but we can always talk about conditional probability.

2.2 Conditional principle

All evidence from the data is given by the experiment actually performed.

Definition 2.1 (Ancillary statistic) A statistic is ancillary if it's distribution is known.

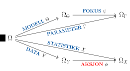
Example 2.1 Let (b,x) be the data. $b \sim$ Bernoulli $\left(\frac{1}{2}\right)$. So it is like tossing a coin. We have :

$x \mid b = 0 \sim \mathcal{N}_0(\mu, \sigma_0^2)$

and

$x \mid b = 1 \sim \mathcal{N}_1(\mu, \sigma_1^2)$

The model parameters are : μ, σ_0, σ_1. Here b is an ancillary statistic
If we get $b = 1$ then the evidence is as if we only did the experiment 1.
There is three experiments one can do here : toss the coin, then experience 0 which is draw a sample from $\mathcal{N}_0$ and then experience 1 which is draw a sample from $\mathcal{N}_1$.



3 Complete statistics

Definition 3.1 (Complete statistic) A statistic T is complete if :

$$\mathbb{E}(\Phi(T)) = 0 \implies \Phi(T) = 0.$$

Example 3.1 Let $(X_1, X_2, \ldots, X_n) \sim \mathcal{B}(p)$. $\Omega_\theta = \mathbb{R}$ and $R(\theta) = \{0,1\}$.

The model parameter $\theta = p$ and the empirical mean $t = \bar{X}$ is a complete and sufficient statistic.

Now let $Y = X_1 + X_2 + \ldots + X_n \sim \mathcal{B}_{in}(n, p)$ (also complete and sufficient).

Completeness : $\mathbb{E}(\Phi(Y)) = 0 \implies \Phi(Y) = 0$?

$$\mathbb{E}(\Phi(Y)) = \int \Phi(Y(\omega))\, P(d\omega) \tag{1}$$

$$= \sum_{y=0}^{n} \Phi(y) \binom{n}{y} p^y\, (1-p)^{n-y} \tag{2}$$

$$\tag{3}$$

This implies that $\Phi(Y) = 0$ for any p.

4 Basu's theorem

Theorem 4.1 (Basu's theorem) A complete sufficient statistic T is independent of any ancillary statistic A.

Example 4.1 If you take a random sample from a normal distribution we know that the empirical mean is independent from the empirical variance. The former is a complete sufficient statistic and the latter is an ancillary statistic.

Proof. $\mathbb{E}(\mathbb{E}(\Phi(A) \mid T) - \mathbb{E}(\Phi(A))) = 0$

$\mathbb{E}(\Phi(A))$ is an ancillary. $\mathbb{E}(\Phi(A) \mid T) - \mathbb{E}(\Phi(A))$ is a function of T.

T is sufficient so the expression is independent from θ , so it is a function of T. T is complete so the expression equals to 0.

We now know that A and T are independent. But what does it mean? Intuition : if you produce an event from A, and from T, they must be independent.

Let's take two independent events C and D. If A is independant from T then $(A \in C)$ is independent from $(T \in D)$ Reminder : $A \in C = \{\omega \mid A(\omega) \in C\}$

In Casella-Berger, you can read the proof of Basu's theorem for discrete distributions only on page 287. □

Theorem 4.2 (Natural statistic) The natural statistic in an exponential family is minimal sufficient and complete when the model parameter is the natural parameter.

Recall the definition of exponential family.

$$f(x) = g(x) \cdot e^{\eta t - \gamma} \tag{4}$$

Bernoulli distribution:

$$f(x) = p^x \cdot (1-p)^{1-x}, \;\; x = 0, 1 \tag{5}$$

We can represent above in the following way.

$$e^{xlnp} \cdot e^{(1-x)ln(1-p)} = e^{xln(\frac{p}{1-p})-(-ln(1-p))} \tag{6}$$

$$f(x_1, \ldots, x_n) = e^{\sum x_i \cdot ln(\frac{p}{1-p})-(-nln(1-p))} \tag{7}$$

$$f(x_1, \ldots, x_n) = e^{\sum x_i \cdot ln(\frac{p}{1-p})-(-nln(1-p))}, \;\; (0,1) \in p \iff \theta \in \mathbb{R} \tag{8}$$

Proof. We can skip gamma in this case.

$$0 = E(\varphi(T)) = \int g(x) \cdot e^{\eta \cdot t(x) - \gamma} \, dx, \;\; \forall \eta \tag{9}$$

$$0 = \int g(x) \cdot e^{\eta \cdot t(x)} \cdot \varphi(t(x)) \, \mu(dx), \;\; \Rightarrow \varphi = 0 \tag{10}$$

Definitions

Theorems

Examples

Index