A. Giovanidis 2021

2. Estimation (Classical)

Network Data Analysis – NDA 2021-2022
Anastasios Giovanidis

Sorbonne-LIP6

Septembre 22, 2021

Bibliography

A. Giovanidis 2021

B.1 H. Pishro-Nik, "Introduction to probability, statistics, and random processes", available at https://www.probabilitycourse.com, Kappa Research LLC, 2014.

☞ Chapter 8

B.2 I. Goodfellow, Y. Bengio, and A. Courville, "Deep learning", MIT Press, 2017.

☞ Chapter 5.4, 5.5

A. Giovanidis 2021

Intro

Statistical Inference is a collection of methods that deal with drawing conclusions from data that are prone to random variation.

Examples

- Predict the outcome of an election: Use a random sample to poll part of the population about their potential vote.
  Randomness from sampling and the uncertainty of vote.
- Decoding in wireless communications: A message is transmitted to a receiver, but the received message is corrupted with noise (channel and thermal).

We work with real data!

A. Giovanidis 2021

Data Analysis

> Data analysis is very much related - but also very different from - probability models.

Let X be a normal random variable with mean $\mu = 100$ and variance $\sigma^2 = 15$. Find the probability that $X > 110$. But, in reality:

- We do not know what is the distribution of X.
- Even if we knew it, we do not know the values of μ and σ.

☞ We need to collect real data, to check if the Central Limit Theorem works, and estimate the values for μ and σ.

Statistical Inference

A. Giovanidis 2021

General Setup: There is an unknown quantity θ that we want to estimate. We get some data. From the data we estimate the desired quantity.

There are 2 main approaches:

- Frequentist (classical) Inference.
- Bayesian Inference.

Classical...

A. Giovanidis 2021

- Frequentist (classical) Inference: The unknown quantity θ is assumed to be **fixed and deterministic**.
  Using data collected, we estimate it by $\hat{\Theta}$, a random variable.

E.g. [Polling] If n is the sample size, let $Y(n)$ be the number of users in the sample, who vote for a certain candidate.
The estimate of the real unknown percentage θ, is $\hat{\Theta} = Y(n)/n$.

...vs Bayesian

A. Giovanidis 2021

- Bayesian Inference: The unknown quantity Θ is assumed to be a **random variable**, and we assume an initial guess about its distribution.
  After observing the data we update the distribution using Bayes' Rule.

E.g. We want to know if information of 1 *bit* was transmitted. The received signal is $\Theta = 1$ *bit* with probability p, or $\Theta = 0$ *bit* with probability $1 - p$.

Then $\Theta \sim \mathrm{Bernoulli}(p)$. We receive a noisy version X of Θ. To estimate Θ we use X and the prior knowledge over the distribution to update p'.

Point Estimator

A. Giovanidis 2021

Frequentist approach:
A point estimator is a function of the random sample $\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$ that estimates the unknown quantity θ.

Any function can do? Yes!

But, a good estimator should be close to the true underlying θ that generated the data.

E.g. To estimate the average file-size we can define the point estimator:

$$\hat{\Theta} = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

A. Giovanidis 2021

Sample Mean

The sample mean is a point estimator for the quantity $\theta = \mathbb{E}[X]$,

$$\overline{X} \quad = \quad \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Properties:

- $\mathbb{E}[\overline{X}] = \mu$.
- $Var(\overline{X}) = \frac{\sigma^2}{n}$.
- Central Limit Theorem: The random variable (r.v.)

$$Z_n = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$$

converges in distribution to the standard normal r.v. as $n \to \infty$.

Properties

A. Giovanidis 2021

- $\mathbb{E}\left[\overline{X}\right] = \mathbb{E}\left[\frac{X_1+\ldots+X_n}{n}\right] = \frac{1}{n} n\mathbb{E}[X_1] = \mu.$

- $Var[\overline{X}] = \mathbb{E}\left[(\frac{X_1+\ldots+X_n}{n} - \mu)^2\right] = \mathbb{E}\left[(\frac{(X_1-\mu)+\ldots+(X_n-\mu)}{n})^2\right] =$
  $\frac{1}{n^2}\mathbb{E}\left[((X_1-\mu)+\ldots+(X_n-\mu))^2\right] =$
  $\frac{1}{n^2}\mathbb{E}\left[(X_1-\mu)^2+\ldots+(X_n-\mu)^2+\sum_{i=1,j>i}^{n} 2(X_i-\mu)(X_j-\mu)\right] =$
  $\frac{n}{n^2}\mathbb{E}\left[(X_1-\mu)^2\right] = \frac{1}{n} Var(X_1) = \frac{\sigma^2}{n}.$

A. Giovanidis 2021

Order Statistics

We can be interested in the largest, smallest, or middle sample value.

Consider a random sample of size n: $X_1, X_2, \ldots, X_n$ from a continuous distribution with CDF $F_X(x)$. We order the X_i's from the smallest to the largest and get the resulting sequence:

$$X_{(1st)}, X_{(2nd)}, \ldots, X_{(nth)}$$

Thus we have

$$\begin{aligned} X_{(1st)} &= \min\left(X_1, X_2, \ldots, X_n\right), \\ X_{(nth)} &= \max\left(X_1, X_2, \ldots, X_n\right) \end{aligned}$$

☞ We call $X_{(1st)}, X_{(2nd)}, \ldots, X_{(nth)}$ the order statistics of the random sample, and are interested in their PDFs and CDFs.

A. Giovanidis 2021

Point Estimator Properties

Some Point Estimator Properties

A. Giovanidis 2021

Point Estimator Properties:

A. Bias
B. Variance
C. Mean Squared Error (MSE)
D. Consistency*

A. Estimator Bias

A. Giovanidis 2021

The bias of point estimator $\hat{\Theta} = h(X_1, \ldots, X_n)$ is defined by

$$B(\hat{\Theta}) \quad = \quad \mathbb{E}[\hat{\Theta}] - \theta.$$

The bias tells us how far is the estimator from the real value.
☞ We say that $\hat{\Theta}$ is an **unbiased** estimator of θ if

$$B(\hat{\Theta}) = 0, \qquad \textit{for all possible values of } \theta.$$

Note: An unbiased estimator is not necessarily a "good" one!

✎ Exercise: Show that $\hat{\Theta}_1 = X_n$ is an unbiased estimator of $\theta = \mathbb{E}[X]$. Same for the sample mean $\hat{\Theta}_2 = \overline{X}$. Which one is better?

Solve Exercise Bias

A. Giovanidis 2021

Solution: Both the single independent sample $\hat{\Theta}_1$ and the sample mean $\hat{\Theta}_2$ are unbiased:

$$\begin{aligned} B(\hat{\Theta}_1) &= \mathbb{E}[\hat{\Theta}_1] - \mu = \mathbb{E}[X_n] - \mu \\ &= 0 \\ &= \mathbb{E}[\hat{\Theta}_2] - \mu = \mathbb{E}[\overline{X}] - \mu = B(\hat{\Theta}_2). \end{aligned}$$

☞ We suspect that the sample mean is a better estimator than the single random sample. How can we show this?

B. Estimator Variance

A. Giovanidis 2021

Let $X_1, \ldots, X_n$ be a random sample, and let θ be an unknown parameter of the distribution that generated it (e.g. the mean).

- The estimator $\hat{\Theta} = h(X_1, \ldots, X_n)$ is a random variable.
- The variance of the estimator

$$Var(\hat{\Theta}) = \mathbb{E}[(\hat{\Theta} - \mathbb{E}[\hat{\Theta}])^2]$$

  strongly depends on the variance of the individual X_is.
- The Standard Error (SE) is given by

$$SE(\hat{\Theta}) = \sqrt{Var(\hat{\Theta})}$$

Exercise Estimator Variance

A. Giovanidis 2021

✎ Exercise: Given a random sample $X_1, \ldots, X_n$ generated by a $\mathrm{Bernoulli}(\theta)$, use $\hat{\Theta}$ to estimate θ: For (a) $\hat{\Theta}_1 = X_n$ and (b) $\hat{\Theta}_2 = \overline{X}$

- calculate the bias of the estimator $B(\hat{\Theta})$.
- calculate the variance of the estimator $Var(\hat{\Theta})$.
- calculate the Standard Error $SE(\hat{\Theta})$.

Solve Exercise

A. Giovanidis 2021

- For the sample mean $\hat{\Theta}$,

$$\hat{\Theta}_1 = X_n, \qquad \hat{\Theta}_2 = \overline{X} = \frac{X_1 + \ldots + X_n}{n}$$

- For the bias of the estimator,

$$B(\hat{\Theta}_1) = \left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i]\right) - \theta = \quad 0 \quad = \mathbb{E}[X_n] - \theta = B(\hat{\Theta}_2)$$

- For the variance of the estimator,

$$\begin{aligned} Var(\hat{\Theta}_1) &= Var(X_i) = \theta(1-\theta), \\ Var(\hat{\Theta}_2) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\theta(1-\theta)}{n}. \end{aligned}$$

Solve Exercise cont'd

A. Giovanidis 2021

- For the Standard Error of the estimator,

$$SE(\hat{\Theta}_1) = \sqrt{\theta(1-\theta)}, \qquad SE(\hat{\Theta}_2) = \sqrt{\frac{\theta(1-\theta)}{n}}.$$

☞ The sample size can determine the accuracy of $\hat{\Theta}$. As $n \to \infty$, we see that $SE(\hat{\Theta}_2) \to 0$, but not $SE(\hat{\Theta}_1)$.

C. Mean Squared Error

A. Giovanidis 2021

The mean squared error (MSE) of point estimator $\hat{\Theta}$ is defined as

$$MSE(\hat{\Theta}) \quad = \quad \mathbb{E}[(\hat{\Theta} - \theta)^2].$$

(Note the difference in definition compared to the $Var(\hat{\Theta})$)

Again, this is a measure of distance (error) between the estimator and the real value. A smaller MSE is indicative of a better estimator.

✎ Exercise: Let $X_1, \ldots, X_n$ be a random sample from an original distribution with mean $\mathbb{E}[X_i] = \theta$ and variance $Var(X_i) = \sigma^2$. Consider the following two estimators again: $\hat{\Theta}_1 = X_n$ and $\hat{\Theta}_2 = \overline{X}$.

Which one is better? *Hint:* $MSE(\hat{\Theta}_1) > MSE(\hat{\Theta}_2)$.

Solve Exercise MSE

A. Giovanidis 2021

Solution: Both the single random sample $\hat{\Theta}_1$ and the sample mean $\hat{\Theta}_2$ are unbiased:

$$MSE(\hat{\Theta}_1) = \mathbb{E}[(\hat{\Theta}_1 - \mu)^2] = \mathbb{E}[(X_n - \mu)^2] = Var(X_n) = \sigma^2.$$

$$MSE(\hat{\Theta}_2) = \mathbb{E}[(\hat{\Theta}_2 - \mu)^2] = \mathbb{E}[(\overline{X} - \mu)^2] = Var(\overline{X}) = \frac{\sigma^2}{n}.$$

☞ With respect to the MSE, the sample mean is indeed a better estimator than the single random sample, because $MSE(\hat{\Theta}_2) < MSE(\hat{\Theta}_1)$, for $n > 1$.

MSE, Bias and Variance

A. Giovanidis 2021

If $\hat{\Theta}$ is a point estimator for θ,

$$\begin{aligned} MSE(\hat{\Theta}) &= \mathbb{E}[(\hat{\Theta}-\theta)^2] = \mathbb{E}[(\hat{\Theta} - \mathbb{E}[\hat{\Theta}] + \mathbb{E}[\hat{\Theta}] - \theta)^2] \\ &= Var(\hat{\Theta}) + (\mathbb{E}[\hat{\Theta}]-\theta)^2 + 2\mathbb{E}[(\hat{\Theta}-\mathbb{E}[\hat{\Theta}])(\mathbb{E}[\hat{\Theta}]-\theta))] \\ &= Var(\hat{\Theta}) + B(\hat{\Theta})^2, \end{aligned}$$

where $B(\hat{\Theta}) = \mathbb{E}[\hat{\Theta}] - \theta$ is the bias of $\hat{\Theta}$.

MSE contains a part due to estimator variance and a part due to bias.

• Bias measures the expected deviation from the true value θ.
• Variance measures the deviation from the expected estimator, due to the particular sample.

☞ Often, these two cannot be minimised simultaneously by the choice of estimator, and there is a trade-off.

A. Giovanidis 2021

Variance Estimator

Point Estimator for Variance

A. Giovanidis 2021

✔ We saw that $\overline{X}$ is a reasonable point estimator for the mean.

☞ Suppose we want a point estimator for the variance σ^2. By definition:

$$\sigma^2 \quad = \quad \mathbb{E}\left[(X-\mu)^2\right].$$

The reasonable estimator (similar to the mean) is just:

$$\hat{\sigma}^2 \quad = \quad \frac{1}{n}\sum_{k=1}^{n}(X_k - \mu)^2.$$

This is an unbiased and consistent estimator of σ^2. However it assumes a known value for the mean μ. In practice this value is unknown and estimated by $\overline{X}$.

✎ Exercise: Show that the following estimator has strictly negative bias:

$$\overline{S}^2 \quad = \quad \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^2 = \frac{1}{n}\left(-n\overline{X}^2 + \sum_{k=1}^{n} X_k^2\right).$$

Solve Exercise Variance Estimator

A. Giovanidis 2021

Solution: We want to calculate the bias

$$\begin{aligned}
B(\overline{S}^2) &= \mathbb{E}[\overline{S}^2] - \sigma^2 \\
&= \frac{1}{n}\left(-n\mathbb{E}[\overline{X}^2] + \sum_{k=1}^{n}\mathbb{E}[X_k^2]\right) - \sigma^2 \\
&= \frac{1}{n}\left(-n(\frac{\sigma^2}{n} + \mu^2) + \sum_{k=1}^{n}(\sigma^2 + \mu^2)\right) - \sigma^2 \\
&= \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n} < 0.
\end{aligned}$$

☞ The bias can be corrected by multiplying $\frac{n}{n-1}\ \overline{S}^2$.

Sample Variance

A. Giovanidis 2021

Let $X_1, \ldots, X_n$ be a random sample with mean $\mathbb{E}[X_i] = \mu < \infty$ and variance $0 < Var(X_i) = \sigma^2 < \infty$.
The sample variance of this random sample is defined as:

$$S^2 = \frac{n}{n-1}\overline{S}^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2 = \frac{1}{n-1}\left(-n\overline{X}^2 + \sum_{k=1}^{n} X_k^2\right).$$

The sample variance is an **unbiased estimator** of σ^2.

The sample standard deviation is defined as

$$S \quad = \quad \sqrt{S^2},$$

and is commonly used as an estimator for σ, but is a **biased** one.

Solve Exercise

A. Giovanidis 2021

✎ Exercise: Calculate the sample mean and sample variance for the sample $\{18, 21, 17, 16, 24, 20\}$.

A. Giovanidis 2021

Solve Exercise

✎ Exercise: Calculate the sample mean and sample variance for the sample $\{18, 21, 17, 16, 24, 20\}$.

- Sample mean

$$\overline{T} = \frac{18 + 21 + 17 + 16 + 24 + 20}{6} = 19,33$$

- Sample variance

$$S^2 = \frac{\sum_{i=1}^{6}(T_i - 19,33)^2}{6 - 1} = 8,67$$

- Sample standard deviation

$$S = \sqrt{S^2} = \sqrt{8,67} = 2,94$$

A. Giovanidis 2021

Maximum Likelihood

Estimation: A systematic way

A. Giovanidis 2021

- Our method to propose estimators for mean and variance have been somewhat ad hoc.
- Is there a systematic way of parameter estimation? Yes!
- Introducing the Maximum Likelihood Estimation (MLE).

An example for MLE 1

A. Giovanidis 2021

Example: A bag contains 3 balls, some Red some Blue. We do not know their exact number; the number of Blue balls is the unknown parameter θ. Possible values for θ are $0, 1, 2$ or 3.

☞ I will choose 4 balls from the bag, using random sampling **with replacement** (I will pick one see the colour and put it back in the bag). The colours in these draws are the r.v.s X_1, X_2, X_3, X_4, where

$$X_i = \begin{cases} 1 & \text{,if the } i\text{th chosen ball is Blue} \\ 0 & \text{,if the } i\text{th ball is Red} \end{cases} .$$

Then X_is are i.i.d. and $X_i \sim \mathrm{Bernoulli}(\frac{\theta}{3})$.

After doing the experiment, the following values are observed: $x_1 = 1, x_2 = 0, x_3 = 1, x_4 = 1$. (3 balls Blue and 1 ball Red).

An example for MLE 2

A. Giovanidis 2021

Questions:

a) Find the probability of the observed sample,
$(x_1, x_2, x_3, x_4) = (1, 0, 1, 1)$.

b) For which value of θ is the observed probability the largest?

An example for MLE 2

A. Giovanidis 2021

Questions:
a) Find the probability of the observed sample, $(x_1, x_2, x_3, x_4) = (1, 0, 1, 1)$.
b) For which value of θ is the observed probability the largest?

Answers:
a) Sample values are i.i.d. $X_i \sim \mathrm{Bernoulli}(\frac{\theta}{3})$

$$P_{X_1 X_2 X_3 X_4}(1,0,1,1) \quad = \quad \left(\frac{\theta}{3}\right)^3 (1 - \frac{\theta}{3}).$$

b) The possible values for θ are $0, 1, 2, 3$ and the highest probability for the sample is obtained for $\theta^* = 2$.

The Likelihood function

A. Giovanidis 2021

Let $X_1, \ldots, X_n$ be a random sample from a distribution with a parameter θ. We have observed $X_1 = x_1$, ..., $X_n = x_n$.

- If X_is are discrete, then the likelihood function is defined as

$$L(x_1, \ldots, x_n; \theta) \quad = \quad P_{X_1 \ldots, X_n}(x_1, \ldots, x_n; \theta).$$

- If X_is are jointly continuous, then the likelihood function is

$$L(x_1, \ldots, x_n; \theta) \quad = \quad f_{X_1 \ldots, X_n}(x_1, \ldots, x_n; \theta).$$

Often, it is easier to work with the log-likelihood function given by

$$\ln L(x_1, \ldots, x_n; \theta),$$

because if θ^* maximises $L(\bullet; \theta)$ it also maximises $\log L(\bullet; \theta)$.

Maximum Likelihood Estimator

A. Giovanidis 2021

A maximum likelihood estimate of θ, shown by $\hat{\theta}_{ML}$, is a value of θ that maximises the likelihood function for the sample values $(x_1, \ldots, x_n)$

$$\hat{\theta}_{ML} \quad = \quad \arg\max L(x_1, \ldots, x_n; \theta).$$

The maximum likelihood estimator (MLE) of θ, denoted by $\hat{\Theta}_{ML}$ is a random variable $\hat{\Theta}_{ML}(X_1, \ldots, X_n)$, whose value when $(X_1, \ldots, X_n) = (x_1, \ldots, x_n)$ is given by $\hat{\theta}_{ML}$.

MLE alternative

A. Giovanidis 2021

If X_is are drawn i.i.d. (random sample) we can write:

$$
\begin{aligned}
\hat{\Theta}_{ML}(x_1, \ldots, x_n) &= \arg\max_{\theta} \frac{1}{n} \sum_{i=1}^{n} \log P_X(x_i; \theta) \\
&= \arg\max_{\theta} \mathbb{E}_{empirical} \left[\log P_X(x_i; \theta) \right].
\end{aligned}
$$

MLE exercise A

A. Giovanidis 2021

Exercise: Find the maximum likelihood estimate(s) for

1. $X_i \sim \mathrm{Binomial}(3, \theta)$ with observations $(x_1, x_2, x_3, x_4) = (1, 3, 2, 2)$.
2. $X_i \sim \mathrm{Exponential}(\theta)$ with observations $(1.23, 3.32, 1.98, 2.12)$.

Solve exercise A

A. Giovanidis 2021

1. $\text{Binomial}(3,\theta)$:

$$\begin{aligned} L(x_1,x_2,x_3,x_4;\theta) &= \begin{pmatrix} 3 \\ x_1 \end{pmatrix}\begin{pmatrix} 3 \\ x_2 \end{pmatrix}\begin{pmatrix} 3 \\ x_3 \end{pmatrix}\begin{pmatrix} 3 \\ x_4 \end{pmatrix}\theta^{x_1+x_2+x_3+x_4}\cdot \\ &\quad \cdot(1-\theta)^{12-x_1-x_2-x_3-x_4} \\ &= \begin{pmatrix} 3 \\ 1 \end{pmatrix}\begin{pmatrix} 3 \\ 3 \end{pmatrix}\begin{pmatrix} 3 \\ 2 \end{pmatrix}\begin{pmatrix} 3 \\ 2 \end{pmatrix}\theta^{8}\cdot(1-\theta)^{12-8} \\ &= 27\cdot\theta^8\cdot(1-\theta)^4 \end{aligned}$$

For the maximum likelihood estimate:

$$\frac{dL(1,3,2,2;\theta)}{d\theta} = 27[8\theta^7(1-\theta)^4 - 4\theta^8(1-\theta)^3] \Rightarrow \hat{\theta}_{ML} = \frac{2}{3}.$$

Solve exercise A

A. Giovanidis 2021

2. Exponential(θ) : $f_X(x) = \theta \exp(-\theta x) u(x)$

$$L(x_1, x_2, x_3, x_4; \theta) = \theta^4 \exp(-\theta(x_1 + x_2 + x_3 + x_4))$$

For the maximum likelihood estimate:

$$\frac{d}{d\theta} \ln L(1.23, 3.32, 1.98, 3.12; \theta) = \frac{d}{d\theta}\left(4 \log\theta - \theta \sum_{i=1}^{4} x_i\right) \Rightarrow$$

$$\hat{\theta}_{ML} = \frac{4}{1.23 + 3.32 + 1.98 + 2.12} = 0.46.$$

MLE exercise B

A. Giovanidis 2021

Exercise: Find the Maximum Likelihood Estimator (MLE) of θ

1. $X_i \sim \mathrm{Binomial}(\mathrm{m}, \theta)$ with observations $(x_1, \ldots, x_n)$.
2. $X_i \sim \mathrm{Exponential}(\theta)$ with observations $(x_1, \ldots, x_n)$.
3. $X_i \sim \mathcal{N}(\theta_1, \theta_2)$ with observations $(x_1, \ldots, x_n)$.

Solve exercise B

A. Giovanidis 2021

$$1.\ \mathrm{Binomial}(\mathrm{m},\theta):\ f_X(x;\theta) = \binom{m}{x}\theta^x(1-\theta)^{m-x}$$

$$L(x_1,\ldots,x_n;\theta) = \left[\prod_{i=1}^{n}\binom{m}{x}\right]\theta^{x_1+\ldots+x_n}(1-\theta)^{mn-(x_1+\ldots+x_n)}$$

For the maximum likelihood estimate: (let $s = x_1 + \ldots + x_n$)

$$\frac{d}{d\theta}L(x_1,\ldots,x_n;\theta) = \frac{d}{d\theta}\mathbf{C}\theta^s(1-\theta)^{mn-s} \Rightarrow \hat{\theta}_{ML} = \frac{s}{nm}.$$

☞ Hence $\hat{\Theta}_{ML} = \frac{1}{mn}\sum_{i=1}^{n} X_i$.

Solve exercise B

A. Giovanidis 2021

2. Exponential(θ) : $f_X(x;\theta) = \theta \exp(-\theta x) u(x)$

$$L(x_1, \ldots, x_n; \theta) \quad = \quad \theta^n \exp(-\theta(x_1 + \ldots + x_n))$$

For the maximum likelihood estimate: (let $s = x_1 + \ldots + x_n$)

$$\frac{d}{d\theta} \ln L(x_1, \ldots, x_n; \theta) = \frac{d}{d\theta} (n \ln(\theta) - \theta s) \Rightarrow \hat{\theta}_{ML} = \frac{n}{s}.$$

☞ Hence $\hat{\Theta}_{ML} = \frac{n}{\sum_{i=1}^{n} X_i}$.

A. Giovanidis 2021

Solve exercise B

2. Normal $\mathcal{N}(\theta_1, \theta_2)$: $f_X(x; \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi\theta_2}} \exp\left(-\frac{(x-\theta_1)^2}{2\theta_2}\right)$

$$L(x_1, \ldots, x_n; \theta) \quad = \quad \frac{1}{(2\pi)^{n/2}\theta_2^{n/2}} \exp\left(-\frac{(x_1-\theta_1)^2 + \ldots + (x_n-\theta_1)^2}{2\theta_2}\right)$$

It is better to work with log-likelihood:

$$\ln L(x_1, \ldots, x_n; \theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\theta_2) - \frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i - \theta_1)^2$$

☞ Hence we can maximise over θ_1 and θ_2 separately.

$\hat{\Theta}_{ML,1} = \arg\max \sum_{i=1}^{n}(x_i - \theta_1)^2$. ((Just the MSE!))

$\hat{\Theta}_{ML,1} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\hat{\Theta}_{ML,2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\Theta}_{1,ML})^2$,
or $\hat{\Theta}_{ML,1} = \overline{X}$, and $\hat{\Theta}_{ML,2} = \overline{S}^2$.

Asymptotic Properties of the MLE

A. Giovanidis 2021

When the sample size becomes large:

- $\hat{\Theta}_{ML}$ is asymptotically consistent, i.e.

$$\lim_{n\to\infty} P\left(|\hat{\Theta}_{ML} - \theta| > \epsilon\right) \quad = \quad 0.$$

- $\hat{\Theta}_{ML}$ is asymptotically unbiased, i.e.

$$\lim_{n\to\infty} \mathbb{E}[\hat{\Theta}_{ML}] \quad = \quad \theta.$$

- As n becomes large, $\hat{\Theta}_{ML}$ is approximately a normal random variable. More precisely the r.v.

$$\frac{\hat{\Theta}_{ML} - \theta}{\sqrt{Var(\hat{\Theta}_{ML})}}$$

converges in distribution to $\mathcal{N}(0,1)$. See the CLT.

A. Giovanidis 2021

Summary ML Estimators

☞ Let n be the sample size.

- Binomial distribution estimator $\mathrm{Binomial}(m, \theta)$

$$\hat{\Theta}_{ML} = \frac{1}{mn}\sum_{i=1}^{n} X_i$$

- Exponential distribution estimator $\mathrm{Exponential}(\theta)$

$$\hat{\Theta}_{ML} = \frac{n}{\sum_{i=1}^{n} X_i}$$

- Gaussian distribution estimator $\mathrm{Normal}(\theta_1, \theta_2)$

$$\hat{\Theta}_{ML,1} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

$$\hat{\Theta}_{ML,2} = \frac{1}{n}\sum_{i=1}^{n} \left(X_i - \hat{\Theta}_{ML,1}\right)^2 = \overline{S}^2 \tag{1}$$

A. Giovanidis 2021

END