METHODS OF ESTIMATION

Section 5.5

Statistics for Data Scientists An Introduction to Probability, Statistics, and Data Analysis

October 03, 2023

SECTION 5.5: ESTIMATION

- It would be much nicer if we could construct procedures that would allow us to directly **estimate the parameters of population distributions**, as opposed to estimating only characteristics of the distributions.
- Two such approaches:
  - ✓Method of moments estimation (MME)
  - ✓Maximum likelihood estimation (MLE)
- How to solve in R?

METHOD OF MOMENTS

- Idea: Use the sample moments to estimate the population moments.
- Sample moments:

$$M_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^r \,, r = 2, \dots ,$$

where $\bar{X}$ is the sample mean.

- Population moments:

$$\mu_r(f_\theta) = E\big(X - \mu\,(f_\theta)\big)^r = \int_{-\infty}^{\infty}\big(x - \mu\,(f_\theta)\big)^r f_\theta(x)dx \,, r = 2, \dots ,$$

where $\mu\,(f_\theta) = E(X)$.

- Match sample moments and population moments:
  - ✓ $\bar{X} = E(X)$
  - ✓ $M_r = \mu_r(f_\theta) = E\big(X - \mu\,(f_\theta)\big)^r, r = 2,3, \dots .$

EXAMPLE: LOG-NORMAL DISTRIBUTION

- $X_1, X_2, \dots, X_n \sim LN(\mu, \sigma)$
- Unknown parameters are (μ, σ)
- Match sample moments and population moments:

$$\bar{X} = \mu(f_L) = \exp\left(\mu + 0.5\sigma^2\right)$$
$$M_2 = \sigma^2(f_L) = \exp\left(2\mu + \sigma^2\right)\left(\exp\left(\sigma^2\right) - 1\right)$$

- After matching, the results are

$$\tilde{\sigma}^2 = \log\left(1 + \frac{M_2}{\bar{X}^2}\right) = \log\left(\bar{X}^2 + M_2\right) - 2\log\left(\bar{X}\right).$$

$$\begin{aligned}\tilde{\mu} &= \log\left(\bar{X}\right) - 0.5\left[\log\left(\bar{X}^2 + M_2\right) - 2\log\left(\bar{X}\right)\right] \\ &= 2\log\left(\bar{X}\right) - 0.5\log\left(\bar{X}^2 + M_2\right).\end{aligned}$$

EXAMPLE: LOG-NORMAL DISTRIBUTION

- If $X_1, X_2, \dots, X_n \sim LN(\mu, \sigma)$, then $\log(X_1), \log(X_2), \dots, \log(X_n) \sim LN(\mu, \sigma)$,
- Unknown parameters are (μ, σ)
- Match sample moments and population moments:

$$\tilde{\mu} = \frac{1}{n} \sum_{i=1}^{n} \log x_i \,,$$

$$\tilde{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\log x_i - \tilde{\mu})^2 \,.$$

HOMEWORK

In Homework 1,

2. Given two datasets,

   (a) Please provide the histograms of two datasets.

   (b) For each dataset, add the probability density functions of the given distributions (c)-(g) in **Question 1** to the figures in **Question 2(a)**. Tying to select more suitable distributions to the data based on your opinion.

a) Use the MME to estimate the parameters for the given distribution (c)-(g) in Question 1.

b) Add the estimated probability density functions to the histogram using different colors.

c) Add the estimated cumulative distribution functions to the empirical plots.

d) Use the test for testing the suitable estimated distributions. Please provide the *p*-value and the corresponding conclusion.