STA2002: Probability and Statistics II

Parameter Estimation II

Fangda Song and Ka Wai Tseng

School of Data Science, CUHK(SZ)

September, 2025

Introduction

In this lecture we will:

- Continue our discussion on the maximum likelihood estimation method
- Introduce the notion of **unbiasedness**
- Introduce another method of parameter estimation called the **method of moments (MoM)**
- **Suggested reading:** Chapter 6.4 in the book.

Parameter Estimation

Suppose that we repeat an experiment n times independently to observe the sample

$$X_1, X_2, \cdots, X_n$$

- How to estimate the unknown parameter $\theta \in \Omega$ based on the observations $x_1, x_2, \cdots, x_n$?
- As the estimators are not unique, how to evaluate different estimators?

Example: $N(\theta_1, \theta_2)$

Example 1

Suppose that $X_1, \ldots, X_n \sim N(\theta_1, \theta_2)$. The unknown parameter is two-dimensional $\theta = (\theta_1, \theta_2)$, and the pdf is given by

$$f(x; \theta) = \frac{1}{\sqrt{2\pi\theta_2}} \exp\left[-\frac{(x-\theta_1)^2}{2\theta_2}\right], \quad -\infty < x < \infty.$$

$\theta \in \Omega = \{(\theta_1, \theta_2) : -\infty < \theta_1 < \infty, 0 < \theta_2 < \infty\}$.

The likelihood function is

$$L(\theta_1, \theta_2) = \prod_{i=1}^{n} f(x_i; \theta) = \left(\frac{1}{\sqrt{2\pi\theta_2}}\right)^n \exp\left[-\frac{\sum_{i=1}^{n}(x_i - \theta_1)^2}{2\theta_2}\right].$$

The log-likelihood function is

$$\ell(\theta_1, \theta_2) = -\frac{n}{2}\ln(2\pi\theta_2) - \frac{\sum_{i=1}^{n}(x_i - \theta_1)^2}{2\theta_2}.$$

Example: $N(\theta_1, \theta_2)$

- Taking the partial derivatives of $\ell(\theta_1, \theta_2)$ with respect to θ_1 and θ_2 and setting them to zero give:

$$\frac{\partial}{\partial \theta_1} \ell(\theta_1, \theta_2) = \frac{1}{\theta_2} \sum_{i=1}^{n} (x_i - \theta_1) = 0$$

$$\frac{\partial}{\partial \theta_2} \ell(\theta_1, \theta_2) = -\frac{n}{2\theta_2} + \frac{1}{2\theta_2^2} \sum_{i=1}^{n} (x_i - \theta_1)^2 = 0$$

- The solutions are

$$\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}, \qquad \hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

Example: $N(\theta_1, \theta_2)$

- We check that this solution is indeed the global maximum by the second derivative test.
- As a result, the MLE are

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}, \qquad \hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

- X is the sample mean while $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is called the variance of the empirical distribution.

Unbiased Estimator

Definition

An estimator $u(X_1, X_2, \ldots, X_n)$ of θ is an **unbiased estimator** of θ if

$$\mathbb{E}(u(X_1, X_2, \ldots, X_n)) = \theta.$$

Otherwise, $u(X_1, X_2, \ldots, X_n)$ is called a **biased estimator**.

Example: $N(\theta_1, \theta_2)$ continued

Suppose $X_1, \ldots, X_n \sim N(\theta_1, \theta_2)$. The unknown parameter is two-dimensional $\theta = (\theta_1, \theta_2)$.
Task: To check that $\hat{\theta}_1$ is an unbiased estimator of θ_1, and $\hat{\theta}_2$ is a biased estimator of θ_2.

Example: $N(\theta_1, \theta_2)$ continued

- Recall the MLEs are,

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X},$$
$$\hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n-1}{n}\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = \frac{n-1}{n}S^2.$$

- Recall also that,

$$\bar{X} \sim N(\theta_1, \theta_2/n),$$
$$\frac{(n-1)S^2}{\theta_2} \sim \chi^2(n-1).$$

Example: $N(\theta_1, \theta_2)$ continued

- As a result,

$$\mathbb{E}(\hat{\theta}_1) = \mathbb{E}(\bar{X}) = \theta_1, \quad \mathbb{E}(S^2) = \theta_2 \cdot \frac{1}{n-1}\mathbb{E}\left(\frac{(n-1)S^2}{\theta_2}\right) = \theta_2.$$

- That is, $\hat{\theta}_1$ is an unbiased estimator of θ_1, and S^2 is an unbiased estimator of θ_2.
- However, for $\hat{\theta}_2$,

$$\mathbb{E}(\hat{\theta}_2) = \frac{n-1}{n}\mathbb{E}(S^2) = \frac{n-1}{n}\theta_2,$$

thus $\hat{\theta}_2$ is a biased estimator of θ_2.

Method of moments (MoM)

- Using the maximum likelihood estimation method requires maximizing the likelihood function $L(\theta)$.
- At times it can be hard to give an explicit formula for the maximum value, and so numerical optimization methods are required.
- Another method to derive a point estimator is called the **method of moments (MoM)**.

Method of moments (MoM)

- Suppose that the unknown parameter $\theta = (\theta_1, \ldots, \theta_k)$ is k-dimensional. For $1 \leq j \leq k$, denote the jth moment to be

$$\alpha_j(\theta) = \mathbb{E}_\theta(X_j).$$

- Denote the jth sample moment to be

$$\hat{\alpha}_j = \frac{1}{n} \sum_{i=1}^{n} X_i^j.$$

Method of moments (MoM)

Definition (method of moments estimator)

The method of moments estimator $\tilde{\theta}$ is defined to be the value of θ such that:

$$\begin{aligned}\alpha_1(\theta) &= \hat{\alpha}_1\\ \alpha_2(\theta) &= \hat{\alpha}_2\\ &\vdots\\ \alpha_k(\theta) &= \hat{\alpha}_k\end{aligned}$$

This gives a system of k equations with k unknowns.

Example: Gamma(θ_1, θ_2)

Example

Suppose $X_1, \ldots, X_n \sim \text{Gamma}(\theta_1, \theta_2)$. The unknown parameter is two-dimensional $\theta = (\theta_1, \theta_2)$, and its first two moments are

$$\alpha_1(\theta_1, \theta_2) = \mathbb{E}_\theta(X_1) = \theta_1 \theta_2,$$
$$\alpha_2(\theta_1, \theta_2) = \mathbb{E}_\theta(X_1^2) = \theta_1 \theta_2^2 + \theta_1^2 \theta_2^2.$$

The first two sample moments are

$$\hat{\alpha}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i,$$
$$\hat{\alpha}_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2.$$

Example: Gamma(θ_1, θ_2)

- To carry out the method of moments, set the first moment to equal the first sample moment, and the second moment to the second sample moment:

$$\theta_1\theta_2 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} \tag{1}$$

$$\theta_1\theta_2^2 + \theta_1^2\theta_2^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \tag{2}$$

- Let

$$V := \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 2\bar{X}X_i + \bar{X}^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

Example: Gamma(θ_1, θ_2)

- Plugging (1) into (2) yields

$$\bar{X}\theta_2 + \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

- Rearranging gives

$$\theta_2 = \frac{V}{\bar{X}}$$

- Substituting this θ_2 back into (1) gives

$$\theta_1 = \frac{\bar{X}^2}{V}$$

- Therefore, the method of moments estimators are

$$\tilde{\theta}_1 = \frac{\bar{X}^2}{V}, \qquad \tilde{\theta}_2 = \frac{V}{\bar{X}}.$$

Example: Poisson(λ)

Example

Suppose $X_1, \ldots, X_n$ independently follow a Poisson distribution with parameter λ.

- Using the method of moments, we set the first moment equal to the first sample moment, that is,

$$\tilde{\lambda}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X},$$

  which is the same as its MLE.

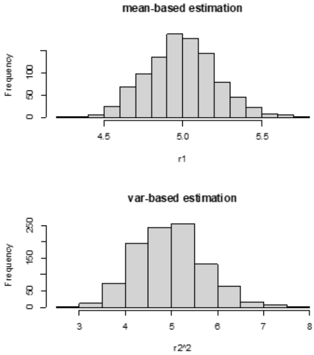
- MoM estimator for Poisson distribution is not unique: second-moment estimator

$$\mathbb{E}(X) = \lambda, \mathbb{E}(X^2) = \lambda + \lambda^2 \Rightarrow \lambda = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

  Therefore, $\tilde{\lambda}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X})^2$

Example: Poisson(λ)

- Simulation to compare two MoM estimators
  - Mean-based: $\tilde{\lambda}_1 = \bar{X}$
  - Variance-based: $\tilde{\lambda}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X})^2$
- Repeatedly generate 1,000 Poisson samples with sample size 20 and draw the histogram of MoM estimators



The mean-based estimator is preferred because of smaller variance.

Example: Uniform$(0, \theta)$

Example

Suppose $X_1, \ldots, X_n$ are uniformly distributed on the interval $(0, \theta)$. Recall that the maximum likelihood estimator is

$$\hat{\theta} = \max\{X_1, \ldots, X_n\}$$

- Using the method of moments, we set the first moment equal to the first sample moment, that is,

$$\frac{\theta}{2} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

Example: Uniform$(0, \theta)$

- Therefore, the MoM estimator is

$$\tilde{\theta} = 2\bar{X}$$

- Note that $\hat{\theta} \neq \tilde{\theta}$. Also, $\tilde{\theta}$ may not be a good estimator if

$$2\bar{X} < \max\{X_1, \ldots, X_n\}.$$

Summary

- Usually MLE is preferred if available
- When large amount of computation is needed, MoM can be useful due to speed.
- Sometimes MoM can give analytical solution while MLE cannot; in this case MLE can still use numerical solution (Ex: Gamma distribution)
- MoM estimators are not unique, so we prefer using the lowest-order moments possible to construct a more stable estimator
- Sometimes MoM gives unrealistic solution (Ex: uniform distribution)