

Mathematical Statistics Recitation 5

TA: Sonia Reilly

February 27, 2026

Lecture Review

Lecture 9

Parameter estimation: X_1, \dots, X_n i.i.d. from some distribution, but parameters are unknown.

Two methods of estimating parameters:

- Method of moments (MOM) — sometimes easier to apply
- Maximum likelihood estimation (MLE) — more mathematically rigorous

(Notation comment: θ denotes true parameter, $\hat{\theta}$ or $\hat{\theta}_n$ denote estimate. Used interchangeably.)

Method of Moments (MOM)

- $\hat{\theta}_{\text{MOM}}$ is the solution of the joint system of equations

$$\mathbb{E}[X^k] = \frac{1}{n} \sum_{i=1}^n X_i^k$$

up to $k = \dim(\theta)$, dimension/total number of parameters. [Left hand side is a function of θ .]

- MOM estimator is consistent:

$$\hat{\theta}_{\text{MOM}} \xrightarrow{p} \theta,$$

not always unbiased.

- MOM estimator is asymptotically normal (\approx normal for large n).

Two important theorems used for the proofs:

- Continuous Mapping Theorem:

$$\mu_n \xrightarrow{p} \mu \implies g(\mu_n) \xrightarrow{p} g(\mu) \quad \text{if } g \text{ continuous}$$

- Delta Method: If g differentiable,

$$\begin{aligned} \sqrt{n}(\mu_n - \mu) &\xrightarrow{d} N(0, \sigma^2) \\ \implies \sqrt{n}(g(\mu_n) - g(\mu)) &\rightarrow N(0, [g'(\mu)]^2 \sigma^2) \end{aligned}$$

Extensions of MOM

- Generalized MOM: for some function $g(X)$, solve

$$\mathbb{E}[g(X)] = \frac{1}{n} \sum_{i=1}^n g(X_i).$$

- Z-estimation: Define an $h(X, \theta)$ such that

$$\mathbb{E}[h(X, \theta)] = 0 \quad \text{for any } \theta$$

and solve

$$\frac{1}{n} \sum_{i=1}^n h(X_i, \hat{\theta}_n) = 0$$

for estimate $\hat{\theta}_n$. (generalization of generalized MOM)

- Example: median is a Z-estimate for certain pdfs (see lecture notes). Benefit of deriving it this way is that we then know that the median is consistent and asymptotically normal for those problems.

Lecture 10

Maximum Likelihood Estimator (MLE)

Likelihood is the joint pdf of the data viewed as a function of θ :

$$\mathcal{L}_n(\theta) = f_\theta(X_1, \dots, X_n).$$

MLE estimate finds the θ under which the probability of the data is highest:

$$\hat{\theta}_{\text{MLE}} = \arg \max_{\theta} \mathcal{L}_n(\theta) = \arg \max_{\theta} \ell_n(\theta),$$

where $\ell_n(\theta) = \log \mathcal{L}_n(\theta)$ is the log likelihood.

For X_1, \dots, X_n iid, can split joint pdf into independent terms:

$$\hat{\theta}_{\text{MLE}} = \arg \max_{\theta} \sum_{i=1}^n \log f_\theta(X_i)$$

Usual procedure: take log, take derivatives, set to 0, solve. If not easily differentiable, think about where max occurs.

Problems

1. [8.5] Suppose that X is a discrete random variable with $P(X = 1) = \theta$ and $P(X = 2) = 1 - \theta$. Three independent observations are made: $X_1 = 1$, $X_2 = 2$, and $X_3 = 2$.
 - (a) Find the method of moments estimator of θ .
 - (b) What is the likelihood function?
 - (c) What is the maximum likelihood estimator of θ ?

Solution:

(a) θ is 1-dimensional, so only need to look at first moment, and solve

$$\mathbb{E}[X] = \frac{1}{n} \sum_{i=1}^n X_i.$$

We compute

$$\mathbb{E}[X] = 1 \cdot \theta + 2 \cdot (1 - \theta) = 2 - \theta$$

and

$$\frac{1}{n} \sum_{i=1}^n X_i = \frac{5}{3},$$

so

$$2 - \hat{\theta}_{\text{MOM}} = \frac{5}{3} \implies \boxed{\hat{\theta}_{\text{MOM}} = \frac{1}{3}}.$$

(b)

$$\mathcal{L}_n(\theta) = \prod_{i=1}^n f_{\theta}(X_i) = \theta(1 - \theta)^2$$

because

$$f_{\theta}(x_i) = \begin{cases} \theta & \text{if } x_i = 1 \\ 1 - \theta & \text{if } x_i = 2. \end{cases}$$

(c)

$$\begin{aligned} \hat{\theta}_{\text{MLE}} &= \arg \max_{\theta} \theta(1 - \theta)^2 \\ &= \arg \max_{\theta} [\log \theta + 2 \log(1 - \theta)] \end{aligned}$$

$$\frac{\partial}{\partial \theta} (\log \theta + 2 \log(1 - \theta)) = \frac{1}{\theta} - \frac{2}{1 - \theta} = 0$$

$$\implies 1(1 - \theta) = 2\theta \implies \hat{\theta}_{\text{MLE}} = \frac{1}{3}$$

2. True or False?

- (a) MLE estimators are unbiased.
- (b) The square of the MLE estimator of a parameter θ converges to θ^2 in probability, for any parameter and any observations X_1, \dots, X_n .
- (c) The MOM estimator always exists.
- (d) The likelihood function integrates to 1, i.e.,

$$\int_{-\infty}^{\infty} \mathcal{L}_n(\theta) d\theta = 1.$$

Solution:

(a) False. In lecture we saw that when $X_1, \dots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$,

$$\hat{\sigma}_{\text{MLE}}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

which is a biased estimator.

(b) True. $\hat{\theta}_{\text{MLE}} \xrightarrow{p} \theta$ since the MLE is consistent, and so by the continuous mapping theorem,

$$\hat{\theta}_{\text{MLE}}^2 \xrightarrow{p} \theta^2,$$

since x^2 is continuous.

(c) False. For some distributions, $\mathbb{E}[X^k]$ is not defined since the integral is infinite.

(d) False. $\mathcal{L}_n(\theta) = f_\theta(X)$ is a probability distribution, but in X , not θ , so

$$\int_{-\infty}^{\infty} \mathcal{L}_n(\theta) dX = 1.$$

We can't say anything about its integral in θ .

3. [8.21] Suppose that X_1, X_2, \dots, X_n are i.i.d. with density

$$f(x) = e^{-(x-\theta)}, \quad x \geq \theta,$$

and $f(x) = 0$ otherwise.

(a) Find the method of moments estimate of θ .

(b) Find the mle of θ . (Hint: Be careful, and don't differentiate before thinking. For what values of θ is the likelihood positive?)

Solution:

(a) First compute the mean. Let $Y = X - \theta$. Then $Y \sim \text{Exp}(1)$, so

$$\mathbb{E}[Y] = 1.$$

Hence

$$\mathbb{E}[X] = \theta + 1.$$

Equating the sample mean \bar{X} to the population mean,

$$\bar{X} = \theta + 1,$$

so the method of moments estimator is

$$\boxed{\hat{\theta}_{\text{MOM}} = \bar{X} - 1.}$$

(b) The likelihood is

$$L(\theta) = \prod_{i=1}^n e^{-(X_i - \theta)} \mathbf{1}\{\theta \leq X_i\}$$

where $\mathbf{1}\{\theta \leq X_i\}$ is the indicator function

$$\mathbf{1}\{\theta \leq X_i\} = \begin{cases} 1 & \text{if } \theta \leq X_i \\ 0 & \text{if } \theta > X_i, \end{cases}$$

since the pdf of X_i is 0 if $X_i < \theta$. Thus

$$L(\theta) = \begin{cases} \exp(-\sum_{i=1}^n X_i + n\theta), & \theta \leq \min_i X_i, \\ 0, & \text{otherwise.} \end{cases}$$

For $\theta \leq \min_i X_i$, the likelihood is proportional to $e^{n\theta}$, which is increasing in θ . Therefore it is maximized at the largest allowable value of θ , namely

$$\hat{\theta}_{\text{MLE}} = \min\{X_1, \dots, X_n\}.$$

4. [8.16, expanded] Suppose X_1, \dots, X_n are i.i.d. with density

$$f(x | \sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right), \quad \sigma > 0,$$

which is the Laplace distribution centered at 0 with scale parameter σ . [Hint: you may use the fact that $\int_0^\infty x^n e^{-ax} dx = \frac{n!}{a^{n+1}}$.]

- Find the regular method of moments estimator of σ . (How many moments do you need to calculate?)
- Find the generalized method of moments estimator of σ using $g(X) = |X|$.
- Find the mle of σ .

Solution:

- The first moment $\mathbb{E}[X]$ is 0, since we are told the distribution is centered at 0, and we can see that it is symmetric. Since the first moment does not depend on θ , we cannot use it for a method of moments calculation, so we compute the second moment instead.

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx.$$

By symmetry,

$$\mathbb{E}[X^2] = 2 \int_0^\infty x^2 \frac{1}{2\sigma} e^{-x/\sigma} dx = \frac{1}{\sigma} \int_0^\infty x^2 e^{-x/\sigma} dx.$$

Using the hint with $n = 2$ and $a = 1/\sigma$,

$$\int_0^\infty x^2 e^{-x/\sigma} dx = \frac{2!}{(1/\sigma)^3} = 2\sigma^3.$$

Thus

$$\mathbb{E}[X^2] = \frac{1}{\sigma} (2\sigma^3) = 2\sigma^2.$$

So we solve

$$\frac{1}{n} \sum_{i=1}^n X_i^2 = 2\sigma^2$$

and get

$$\hat{\sigma}_{\text{MOM}} = \sqrt{\frac{1}{2n} \sum_{i=1}^n X_i^2}.$$

(b) By symmetry,

$$\mathbb{E}[|X|] = 2 \int_0^\infty x \frac{1}{2\sigma} e^{-x/\sigma} dx = \frac{1}{\sigma} \int_0^\infty x e^{-x/\sigma} dx.$$

Using the hint with $n = 1$ and $a = 1/\sigma$,

$$\int_0^\infty x e^{-x/\sigma} dx = \frac{1!}{(1/\sigma)^2} = \sigma^2,$$

so

$$\mathbb{E}[|X|] = \sigma.$$

Equating the sample moment to the population moment,

$$\hat{\sigma}_{\text{GMOM}} = \frac{1}{n} \sum_{i=1}^n |X_i|.$$

(c) The likelihood is

$$L(\sigma) = \prod_{i=1}^n \frac{1}{2\sigma} \exp\left(-\frac{|X_i|}{\sigma}\right) = (2\sigma)^{-n} \exp\left(-\frac{1}{\sigma} \sum_{i=1}^n |X_i|\right).$$

The log-likelihood is

$$\ell(\sigma) = -n \log(2\sigma) - \frac{1}{\sigma} \sum_{i=1}^n |X_i|.$$

Differentiate with respect to σ and set to 0:

$$\frac{d\ell}{d\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^2} \sum_{i=1}^n |X_i| = 0$$

$$\implies \hat{\sigma}_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^n |X_i|.$$