

Maximum Likelihood Estimation and Bayesian Statistics

Rebecca C. Steorts

Agenda

- ▶ Maximum Likelihood Estimation
- ▶ Unbiased Estimators
- ▶ Invariance Property of MLEs
- ▶ Mean Squared Error
- ▶ Practice Exercises

Traditional inference

You are given **data** X and there is an **unknown parameter** you wish to estimate θ

How would you estimate θ ?

- ▶ Find an unbiased estimator of θ .
- ▶ Find the maximum likelihood estimate (MLE) of θ by looking at the likelihood of the data.
- ▶ Suppose that $\hat{\theta}$ estimates θ .

Note: $\hat{\theta}$ may depend on the data $x_{1:n} = x_1, \dots, x_n$.

Unbiased Estimator

Recall that $\hat{\theta}$ is an **unbiased estimator** of θ if

$$E[\hat{\theta}] = \theta. \quad (1)$$

Maximum Likelihood Estimation

Assume sample points $x_{1:n}$.

Let $\hat{\theta}$ be a parameter value at which $p(x_{1:n} \mid \theta)$ attains its maximum as a function of θ , with $x_{1:n}$ held fixed.

A **maximum likelihood estimator** (MLE) of the parameter θ based on a sample $x_{1:n}$ is denoted by $\hat{\theta}$.

Finding the MLE

The solution to the MLE are the possible candidates (θ) that solve

$$\frac{\partial p(x_{1:n} \mid \theta)}{\partial \theta} = 0. \quad (2)$$

The solution to equation 2 are only **possible candidates** for the MLE.

Our job is to find a **global maximum**, and make sure that we have not found a **local maximum**.

MLE of Normal distribution

Consider

$$X_1, \dots, X_n \stackrel{iid}{\sim} \text{Normal}(\theta, 1).$$

Show that the MLE is $\hat{\theta} = \bar{x}$.

MLE of Normal distribution

Proof:

$$p(x_{1:n} \mid \theta) = (2\pi)^{-n/2} \times \exp\left\{\frac{-1}{2} \sum_{i=1}^n (x_i - \theta)^2\right\} \quad (3)$$

Consider

$$\log p(x_{1:n}) = -n/2 \log(2\pi) - \frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2 \quad (4)$$

MLE of Normal distribution

$$\frac{\partial \log p(x_{1:n} \mid \theta)}{\partial \theta} = \sum_{i=1}^n (x_i - \theta) \quad (5)$$

This implies that

$$\sum_i (x_i - \theta) = 0 \implies \hat{\theta} = \bar{x}.$$

MLE of Normal distribution

Consider

$$\frac{\partial^2 \log p(x_{1:n} \mid \theta)}{\partial \theta^2} = -n < 0.$$

Thus, our solution is unique (and a global solution).

Invariance property of MLE's

If $\hat{\theta}$ is the MLE of θ , then for any function $g(\theta)$, the MLE of $g(\theta)$ is the MLE of $g(\hat{\theta})$.

Proof: Theorem 7.2.10, Casella and Berger, page 318.

Evaluation of Estimators

How do we evaluate estimators? We often use the mean squared error.

$$\text{MSE}(\hat{\theta}) = E_{\theta}[(\hat{\theta} - \theta)^2].$$

Observe that

$$\text{MSE}(\hat{\theta}) = \text{Var}_{\theta}(\hat{\theta}) + E_{\theta}[(\hat{\theta} - \theta)^2] = \text{Var}_{\theta}(\hat{\theta}) + (\text{Bias}_{\theta}(\hat{\theta}))^2,$$

where the

$$\text{Bias}_{\theta}(\hat{\theta}) = E_{\theta}(\hat{\theta}) - \theta.$$

For a more in depth treatment of MSE and bias, see Section 7.3.1, Casella and Berger, p. 330 - 334.

Exercise 0

Show that $\text{MSE}(\hat{\theta}) = \text{Var}_{\theta}(\hat{\theta}) + (\text{Bias}_{\theta}(\hat{\theta}))^2$.

Proof:

$$\text{MSE}(\hat{\theta}) \tag{6}$$

$$= E_{\theta}[(\hat{\theta} - \theta)^2] \tag{7}$$

$$= E_{\theta}[(\hat{\theta} - E_{\theta}(\hat{\theta}) + E_{\theta}(\hat{\theta}) - \theta)^2] \tag{8}$$

$$= E_{\theta}[(\hat{\theta} - E_{\theta}(\hat{\theta}))^2] \tag{9}$$

$$+ 2(E_{\theta}(\hat{\theta}) - \theta) \times (E_{\theta}[\hat{\theta} - E_{\theta}(\hat{\theta})]) + E_{\theta}[(E_{\theta}(\hat{\theta}) - \theta)^2] \tag{10}$$

$$= E_{\theta}[(\hat{\theta} - E_{\theta}(\hat{\theta}))^2] + [E_{\theta}(\hat{\theta}) - \theta]^2 \tag{11}$$

$$= \text{Var}_{\theta}(\hat{\theta}) + [\text{Bias}_{\theta}(\hat{\theta})]^2. \tag{12}$$

Exercise 1

Show that

$$\hat{\theta} = \bar{x}$$

is an unbiased estimator for θ .

Solution to Exercise 1

Proof.

$$E[\hat{\theta}] = E[\bar{x}] = \frac{1}{n} \sum_i E[x_i] = \frac{1}{n} \sum_i \theta = \theta.$$

Thus, we have showed that the MLE is an unbiased estimator for θ .

Exercise 2

Consider

$$X_1, \dots, X_n \stackrel{iid}{\sim} \text{Normal}(\theta, 1) \quad (13)$$

$$\theta \stackrel{ind}{\sim} \text{Normal}(\mu, \tau^2) \quad (14)$$

Write the posterior mean as a function of the MLE and the prior mean μ .

Solution to Exercise 2

Let $\lambda = 1$ and $\lambda_o = 1/\tau^2$.

Recall that from module 3,

$$\theta \mid x_{1:n} \sim N(M, L^{-1}),$$

where

$$L = n\lambda + \lambda_o$$

and

$$M = \frac{n\lambda\bar{x} + \lambda_o\mu}{n\lambda + \lambda_o}.$$

Solution to Exercise 2

Observe that

$$M = \frac{n\lambda\bar{x} + \lambda_o\mu}{n\lambda + \lambda_o} = \frac{n\lambda\hat{\theta} + \lambda_o\mu}{n\lambda + \lambda_o} = \frac{n\lambda}{n\lambda + \lambda_o}\hat{\theta} + \frac{\lambda_o}{n\lambda + \lambda_o}\mu.$$

Thus, we can write the posterior mean as a function of the MLE and the prior mean μ .

Exercise 3

$$X_1, \dots, X_n \stackrel{iid}{\sim} \text{Bernoulli}(\theta). \quad (15)$$

$$\theta \sim \text{Beta}(a, b) \quad (16)$$

Observe that $Y = \sum_i X_i \sim \text{Binomial}(n, \theta)$.

It can be shown that the MLE for θ is $\bar{x} = y/n$.

Recall that

$$\theta \mid y \sim \text{Beta}(y + a, n - y + b).$$

Exercise 3

Show that the posterior mean can be written as

$$E[\theta \mid y] = \text{MLE} \times \frac{n}{a + b + n} + \text{priorMean} \times \frac{a + b}{a + b + n},$$

where $\text{MLE} = \bar{x}$ and $\text{priorMean} = \frac{a}{a+b}$.

Solution to Exercise 3

Proof:

$$\begin{aligned} E[\theta | y] &= \frac{y + a}{y + a + n - y + b} = \frac{y + a}{a + n + b} \\ &= \frac{y}{a + b + n} + \frac{a}{a + b + n} \\ &= \frac{y}{n} \times \frac{n}{a + b + n} + \frac{a}{a + b} \times \frac{a + b}{a + b + n} \\ &= \text{MLE} \times \frac{n}{a + b + n} + \text{priorMean} \times \frac{a + b}{a + b + n} \end{aligned}$$

Thus, we have written the posterior mean as a linear combination of the MLE and prior mean with weights being determined by a, b , and n .

Binomial MLE Exercise

Let

$$X_1, \dots, X_n \stackrel{iid}{\sim} \text{Bernoulli}(\theta).$$

Show that the MLE is $\hat{\theta} = \bar{x}$.

Proof: Casella and Berger, Example 7.2.7, page 317-318.

Normal-Normal model Exercise

Suppose that

$$X_1, \dots, X_n \stackrel{iid}{\sim} \text{Normal}(\theta, \sigma^2),$$

where θ, σ^2 are both unknown.

Show that $(\bar{x}, n^{-1} \sum_i (x_i - \bar{x})^2)$ are the MLE's for (θ, σ^2) .

Proof: Casella and Berger, Example 7.2.7, page 317-318.

Summary

- ▶ Maximum Likelihood Estimators (MLEs)
- ▶ Invariance of MLEs
- ▶ Mean squared errors
- ▶ Unbiased Estimator
- ▶ Practice exercises