

# Section 9: Solutions

---

## Review of Main Concepts

- **Realization/Sample:** A realization/sample  $x$  of a random variable  $X$  is the value that is actually observed.
- **Likelihood:** Let  $x_1, \dots, x_n$  be iid realizations from probability mass function  $p_X(\mathbf{x}; \theta)$  (if  $X$  discrete) or density  $f_X(\mathbf{x}; \theta)$  (if  $X$  continuous), where  $\theta$  is a parameter (or a vector of parameters). We define the likelihood function to be the probability of seeing the data.

If  $X$  is discrete:

$$L(x_1, \dots, x_n | \theta) = \prod_{i=1}^n p_X(x_i | \theta)$$

If  $X$  is continuous:

$$L(x_1, \dots, x_n | \theta) = \prod_{i=1}^n f_X(x_i | \theta)$$

- **Maximum Likelihood Estimator (MLE):** We denote the MLE of  $\theta$  as  $\hat{\theta}_{\text{MLE}}$  or simply  $\hat{\theta}$ , the parameter (or vector of parameters) that maximizes the likelihood function (probability of seeing the data).

$$\hat{\theta}_{\text{MLE}} = \arg \max_{\theta} L(x_1, \dots, x_n | \theta) = \arg \max_{\theta} \ln L(x_1, \dots, x_n | \theta)$$

- **Log-Likelihood:** We define the log-likelihood as the natural logarithm of the likelihood function. Since the logarithm is a strictly increasing function, the value of  $\theta$  that maximizes the likelihood will be exactly the same as the value that maximizes the log-likelihood.

If  $X$  is discrete:

$$\ln L(x_1, \dots, x_n | \theta) = \sum_{i=1}^n \ln p_X(x_i | \theta)$$

If  $X$  is continuous:

$$\ln L(x_1, \dots, x_n | \theta) = \sum_{i=1}^n \ln f_X(x_i | \theta)$$

- **Bias:** The bias of an estimator  $\hat{\theta}$  for a true parameter  $\theta$  is defined as  $\text{Bias}(\hat{\theta}, \theta) = \mathbb{E}[\hat{\theta}] - \theta$ . An estimator  $\hat{\theta}$  of  $\theta$  is unbiased iff  $\text{Bias}(\hat{\theta}, \theta) = 0$ , or equivalently  $\mathbb{E}[\hat{\theta}] = \theta$ .
- **Steps to find the maximum likelihood estimator,  $\hat{\theta}$ :**
  - (a) Find the likelihood and log-likelihood of the data.
  - (b) Take the derivative of the log-likelihood and set it to 0 to find a candidate for the MLE,  $\hat{\theta}$ .
  - (c) Take the second derivative and show that  $\hat{\theta}$  indeed is a maximizer, that  $\frac{\partial^2 L}{\partial \theta^2} < 0$  at  $\hat{\theta}$ . Also ensure that it is the global maximizer: check points of non-differentiability and boundary values.

## 1. Mystery Dish!

A fancy new restaurant has opened up which features only 4 dishes. The unique feature of dining here is that they will serve you any of the four dishes randomly according to the following probability distribution: give dish A with probability 0.5, dish B with probability  $\theta$ , dish C with probability  $2\theta$ , and dish D with probability  $0.5 - 3\theta$

Each diner is served a dish independently. Let  $x_A$  be the number of people who received dish A,  $x_B$  the number of people who received dish B, etc, where  $x_A + x_B + x_C + x_D = n$ . Find the MLE for  $\theta$ ,  $\hat{\theta}$ . **Solution:**

The data tells us, for each diner in the restaurant, what their dish was. We begin by computing the likelihood of seeing the given data given our parameter  $\theta$ . Because each diner is assigned a dish independently, the likelihood is equal to the product over diners of the chance they got the particular dish they got, which gives us:

$$L(x|\theta) = 0.5^{x_A} \theta^{x_B} (2\theta)^{x_C} (0.5 - 3\theta)^{x_D}$$

From there, we just use the MLE process to get the log-likelihood, take the first derivative, set it equal to 0, and solve for  $\hat{\theta}$ .

$$\ln L(x|\theta) = x_A \ln(0.5) + x_B \ln(\theta) + x_C \ln(2\theta) + x_D \ln(0.5 - 3\theta)$$

$$\frac{\partial}{\partial \theta} \ln L(x|\theta) = \frac{x_B}{\theta} + \frac{x_C}{\theta} - \frac{3x_D}{0.5 - 3\theta}$$

$$\frac{x_B}{\hat{\theta}} + \frac{x_C}{\hat{\theta}} - \frac{3x_D}{0.5 - 3\hat{\theta}} = 0$$

Solving yields  $\hat{\theta} = \frac{x_B + x_C}{6(x_B + x_C + x_D)}$ .

## 2. A Red Poisson

Suppose that  $x_1, \dots, x_n$  are i.i.d. samples from a  $\text{Poisson}(\theta)$  random variable, where  $\theta$  is unknown. Find the MLE of  $\theta$ . **Solution:**

Because each Poisson RV is i.i.d., the likelihood of seeing that data is just the PMF of the Poisson distribution multiplied together for every  $x_i$ . From there, take the log-likelihood, then the first derivative, set it equal to 0 and solve for  $\hat{\theta}$ .

$$L(x_1, \dots, x_n | \theta) = \prod_{i=1}^n e^{-\theta} \frac{\theta^{x_i}}{x_i!}$$

$$\ln L(x_1, \dots, x_n | \theta) = \sum_{i=1}^n [-\theta - \ln(x_i!) + x_i \ln(\theta)]$$

$$\frac{\partial}{\partial \theta} \ln L(x_1, \dots, x_n | \theta) = \sum_{i=1}^n \left[-1 + \frac{x_i}{\theta}\right]$$

$$-n + \frac{\sum_{i=1}^n x_i}{\hat{\theta}} = 0$$

$$\hat{\theta} = \frac{\sum_{i=1}^n x_i}{n}$$

## 3. Independent Shreds, You Say?

You are given 100 independent samples  $x_1, x_2, \dots, x_{100}$  from  $\text{Bernoulli}(\theta)$ , where  $\theta$  is unknown. (Each sample is either a 0 or a 1). These 100 samples sum to 30. You would like to estimate the distribution's parameter  $\theta$ . Give all answers to 3 significant digits.

(a) What is the maximum likelihood estimator  $\hat{\theta}$  of  $\theta$ ? **Solution:**

Note that  $\sum_{i \in [n]} x_i = 30$ , as given in the problem spec. Therefore, there are 30 1s and 70 0s. (Note that they come in some specific order.) Therefore, we can setup  $L$  as follows, because there is a  $\theta$  chance of getting a 1, and a  $(1 - \theta)$  chance of getting a 0 and they are each i.i.d. From there, take the log-likelihood,

then the first derivative, set it equal to 0 and solve for  $\hat{\theta}$ .

$$\begin{aligned}L(x_1, \dots, x_n | \theta) &= (1 - \theta)^{70} \theta^{30} \\ \ln L(x_1, \dots, x_n | \theta) &= 70 \ln(1 - \theta) + 30 \ln \theta \\ \frac{\partial}{\partial \theta} \ln L(x_1, \dots, x_n | \theta) &= -\frac{70}{1 - \theta} + \frac{30}{\theta} \\ -\frac{70}{1 - \hat{\theta}} + \frac{30}{\hat{\theta}} &= 0 \\ \frac{30}{\hat{\theta}} &= \frac{70}{1 - \hat{\theta}} \\ 30 - 30\hat{\theta} &= 70\hat{\theta} \\ \hat{\theta} &= \frac{30}{100}\end{aligned}$$

(b) Is  $\hat{\theta}$  an unbiased estimator of  $\theta$ ? **Solution:**

An estimator is unbiased if the expectation of the estimator is equal to the original parameter, i.e.:  $E[\hat{\theta}] = \theta$ . Setting up the expectation of our estimator and plugging it in for the generic case, we get the following, which we can then reduce with linearity of expectation:

$$\begin{aligned}\mathbb{E}[\hat{\theta}] &= \mathbb{E}\left[\frac{1}{100} \sum_{i=1}^{100} X_i\right] \\ &= \frac{1}{100} \sum_{i=1}^{100} \mathbb{E}[X_i] \\ &= \frac{1}{100} \cdot 100\theta = \theta.\end{aligned}$$

so it is unbiased.

## 4. Y Me?

Let  $y_1, y_2, \dots, y_n$  be i.i.d. samples of a random variable with density function

$$f_Y(y|\theta) = \frac{1}{2\theta} \exp\left(-\frac{|y|}{\theta}\right)$$

Find the MLE for  $\theta$  in terms of  $|y_i|$  and  $n$ . **Solution:**

Since the samples are i.i.d., the likelihood of seeing  $n$  samples of them is just their PDFs multiplied together.

From there, take the log-likelihood, then the first derivative, set it equal to 0 and solve for  $\hat{\theta}$ .

$$\begin{aligned}
 L(y_1, \dots, y_n | \theta) &= \prod_{i=1}^n \frac{1}{2\theta} \exp\left(-\frac{|y_i|}{\theta}\right) \\
 \ln L(y_1, \dots, y_n | \theta) &= \sum_{i=1}^n \left[ -\ln 2 - \ln \theta - \frac{|y_i|}{\theta} \right] \\
 \frac{\partial}{\partial \theta} \ln L(y_1, \dots, y_n | \theta) &= \sum_{i=1}^n \left[ -\frac{1}{\theta} + \frac{|y_i|}{\theta^2} \right] \\
 \sum_{i=1}^n \left[ -\frac{1}{\hat{\theta}} + \frac{|y_i|}{\hat{\theta}^2} \right] &= 0 \\
 -\frac{n}{\hat{\theta}} + \frac{\sum_{i=1}^n |y_i|}{\hat{\theta}^2} &= 0 \\
 \hat{\theta} &= \frac{\sum_{i=1}^n |y_i|}{n}
 \end{aligned}$$

## 5. A biased estimator

In class, we showed that the maximum likelihood estimate of the variance  $\theta_2$  of a normal distribution (when both the true mean  $\mu$  and true variance  $\sigma^2$  are unknown) is what's called the *population variance*. That is

$$\hat{\theta}_2 = \left( \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\theta}_1)^2 \right)$$

where  $\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^n x_i$  is the MLE of the mean. Is  $\hat{\theta}_2$  unbiased?

**Solution:**

Let  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ . Then

$$E(\hat{\theta}_2) = E\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right) = E\left(\frac{1}{n} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2)\right)$$

which by linearity of expectation (and distributing the sum) is

$$\begin{aligned}
 &= \frac{1}{n} \sum_{i=1}^n E(X_i^2) - E\left(\frac{2}{n} \bar{X} \sum_{i=1}^n X_i\right) + E(\bar{X}^2) \\
 &= \frac{1}{n} \sum_{i=1}^n E(X_i^2) - 2E(\bar{X}^2) + E(\bar{X}^2) \\
 &= \frac{1}{n} \sum_{i=1}^n E(X_i^2) - E(\bar{X}^2). \quad (**)
 \end{aligned}$$

We know that for any random variable  $Y$ , since  $Var(Y) = E(Y^2) - (E(Y))^2$  it holds that

$$E(Y^2) = Var(Y) + (E(Y))^2.$$

Also, we have  $E(X_i) = \mu$ ,  $Var(X_i) = \sigma^2 \forall i$  and  $E(\bar{X}) = \mu$ ,  $Var(\bar{X}) = \frac{\sigma^2}{n}$ . Combining these facts, we get

$$E(X_i^2) = \sigma^2 + \mu^2 \quad \forall i \quad \text{and} \quad E(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2.$$

Substituting these equations into (\*\*) we get

$$\begin{aligned} E\left(\frac{1}{n}\sum_{i=1}^n(X_i - \bar{X})^2\right) &= \frac{1}{n}\sum_{i=1}^n E(X_i^2) - E(\bar{X}^2) = \sigma^2 + \mu^2 - \left(\frac{\sigma^2}{n} + \mu^2\right) \\ &= \left(1 - \frac{1}{n}\right)\sigma^2. \end{aligned}$$

Thus  $\hat{\theta}_2$  is not unbiased.