

# 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  $p_X(x; \theta)$  (if  $X$  discrete) or density  $f_X(x; \theta)$  (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
  - (c) Set it to 0 to find a candidate for the MLE,  $\hat{\theta}$ . (note: at this step, we change from the  $\theta$  to the  $\hat{\theta}$  because in this step we are solving for the *maximum* likelihood estimator for  $\theta$ )
  - (d) 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. Content Review

- (a) **True or False:** The Log-Likelihood gives a slightly different estimate, but because it is close enough and easier to compute we use it for our estimate of  $\theta$ .

**Solution:**

**False:** 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.

(b) **True or False:**  $\hat{\theta}$  is the true parameter and  $\theta$  is our estimate. **Solution:**

**False:** It is the other way around. Remember to switch to  $\hat{\theta}$  when you set your equation to zero!

(c) **True or False:** An estimator is unbiased if  $\text{Bias}(\hat{\theta}, \theta) = \mathbb{E}[\hat{\theta}] - \theta = 0$  or equivalently  $\mathbb{E}[\hat{\theta}] = \theta$  **Solution:**

**True** by definition of

(d) You flip a coin 10 times and observe HHHTHHTHHH (8 heads, 2 tails). What is the MLE of  $\theta$ , where  $\theta$  is the true probability of seeing tails?

- $\hat{\theta} = .2$
- $\hat{\theta} = .25$
- $\hat{\theta} = .8$
- $\hat{\theta} = .3$

**Solution:**

Option 1:  $\hat{\theta} = .2$

## 2. 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  $\hat{\theta}$ , the maximum likelihood estimator for  $\theta$ .

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

### 3. 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 for  $\theta$ .

**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  $\hat{\theta}$ .

$$\begin{aligned} L(x_1, \dots, x_n; \theta) &= \prod_{i=1}^n e^{-\theta} \frac{\theta^{x_i}}{x_i!} \\ \ln L(x_1, \dots, x_n; \theta) &= \sum_{i=1}^n [-\theta - \ln(x_i!) + x_i \ln(\theta)] \\ \frac{\partial}{\partial \theta} \ln L(x_1, \dots, x_n; \theta) &= \sum_{i=1}^n \left[-1 + \frac{x_i}{\theta}\right] \\ -n + \frac{\sum_{i=1}^n x_i}{\hat{\theta}} &= 0 \\ \hat{\theta} &= \frac{\sum_{i=1}^n x_i}{n}. \end{aligned}$$

Notice that the  $-\ln(x_i!)$  term disappears since it is a constant relative to  $\theta$ , of which we take the derivative.

### 4. 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.

## 5. 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  $\hat{\theta}$ .

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

## 6. Bird Watching

You are an ornithologist studying a rare species of birds in a nature reserve. Over a period of 50 days, you record the number of sightings of this bird (you see  $x_1, x_2, \dots, x_{50}$  birds on each day). Your research has shown that the number of sightings on this species depends on the average number of monkeys in the reserve,  $\theta_1$ , and the average number of eagles in the reserve,  $\theta_2$ . After years of studying this rare species in other environments, you've found the number of birds observed on a particular day follows the following distribution:

$$p_X(k) = \frac{1}{k!} (\theta_1^k \cdot e^{-\theta_1} \cdot \theta_2^k \cdot e^{-3\theta_2})$$

Find the MLE for  $\theta_1$  and  $\theta_2$  (i.e., find  $\hat{\theta}_1$  and  $\hat{\theta}_2$ ).

(a) What is the likelihood function? **Solution:**

Once again, the likelihood of seeing the above samples is just their PDFs multiplied together.

$$L(x; \theta_1, \theta_2) = \prod_{i=1}^{50} \left( \frac{1}{x_i!} (\theta_1^{x_i} e^{-\theta_1} \cdot \theta_2^{x_i} e^{-3\theta_2}) \right)$$

(b) What is the log-likelihood function? **Solution:**

We take the log of the above and simplify:

$$\ln(L(x; \theta_1, \theta_2)) = \ln\left(\prod_{i=1}^{50} \left(\frac{1}{x_i!} (\theta_1^{x_i} \cdot e^{-\theta_1} \cdot \theta_2^{x_i} \cdot e^{-3\theta_2})\right)\right) \quad (1)$$

$$= \sum_{i=1}^{50} \left(\ln\left(\frac{1}{x_i!}\right) + \ln(\theta_1^{x_i}) + \ln(e^{-\theta_1}) + \ln(\theta_2^{x_i}) + \ln(e^{-3\theta_2})\right) \quad (2)$$

$$= \sum_{i=1}^{50} \left(\ln\left(\frac{1}{x_i!}\right) + x_i \ln(\theta_1) - \theta_1 + x_i \ln(\theta_2) - 3\theta_2\right) \quad (3)$$

(c) We want to find values of  $\theta_1$  and  $\theta_2$  that maximize the likelihood function. To do this, we will take the partial derivative with respect to each of these parameters and solve for the values that make them both zero.

First, take the partial derivative of the likelihood function with respect to  $\theta_1$ . **Solution:**

When taking the partial derivative with respect to a certain variable, we take the derivative as usual, but treat other variables as a constant! So here, we treat  $\theta_2$  as a constant and derivate with respect to  $\theta_1$ .

$$\frac{\partial}{\partial \theta_1}(\ln(L(x; \theta_1, \theta_2))) = \sum_{i=1}^{50} \left( \frac{x_i}{\theta_1} - 1 \right) \quad (4)$$

$$= \sum_{i=1}^{50} \left( \frac{x_i}{\theta_1} \right) - 50 \quad (5)$$

(d) Now, take the partial derivative with respect to  $\theta_2$ . **Solution:**

Now, we treat  $\theta_1$  as a constant and derivate with respect to  $\theta_2$ .

$$\frac{\partial}{\partial \theta_2}(\ln(L(x; \theta_1, \theta_2))) = \sum_{i=1}^{50} \left( \frac{x_i}{\theta_2} - 3 \right) \quad (6)$$

$$= \sum_{i=1}^{50} \left( \frac{x_i}{\theta_2} \right) - 150 \quad (7)$$

(e) Set both these partial derivatives to 0, and solve for  $\hat{\theta}_1$  and  $\hat{\theta}_2$ . **Solution:**

We end up with the equations (notice we added the hats to the thetas at this point - since we set these derivatives to 0, we are now solving for the *maximum* likelihood estimator):

$$\sum_{i=1}^{50} \left( \frac{x_i}{\hat{\theta}_1} \right) - 50 = 0$$

$$\sum_{i=1}^{50} \left( \frac{x_i}{\hat{\theta}_2} \right) - 150 = 0$$

We now solve this system of equations for  $\hat{\theta}_1$  and  $\hat{\theta}_2$ . The first equation gives us:

$$\sum_{i=1}^{50} \left( \frac{x_i}{\hat{\theta}_1} \right) = 50$$

$$\left( \frac{\sum_{i=1}^{50} x_i}{\hat{\theta}_1} \right) = 50$$

$$\hat{\theta}_1 = \left( \frac{\sum_{i=1}^{50} x_i}{50} \right)$$

With similar steps, we get that:

$$\hat{\theta}_2 = \left( \frac{\sum_{i=1}^{50} x_i}{150} \right)$$

## 7. 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.

## 8. It Means Nothing

Suppose  $x_1, x_2, \dots, x_n$  are samples from a normal distribution whose mean is known to be  $\mu$  but the variance is unknown. How does the maximum likelihood estimator for the variance differ from the maximum likelihood estimator when both mean and variance are unknown? Which if any is unbiased? **Solution:**

Begin with the same derivation as before, however we now use  $\mu$  instead of the mean of 0, which gets us:

$$\begin{aligned}
 L(x_1, \dots, x_n; \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{-(x_i - \mu)^2}{2\sigma^2} \\
 \ln L(x_1, \dots, x_n; \sigma^2) &= \sum_{i=1}^n -\ln \sqrt{2\pi\sigma^2} - \frac{(x_i - \mu)^2}{2\sigma^2} \\
 &= \sum_{i=1}^n -\frac{1}{2} \ln 2\pi\sigma^2 - \frac{(x_i - \mu)^2}{2\sigma^2} \\
 &= \sum_{i=1}^n -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma^2 - \frac{(x_i - \mu)^2}{2\sigma^2} \\
 &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2} \\
 \frac{\partial}{\partial \sigma^2} \ln L(x_1, \dots, x_n; \sigma^2) &= -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^4} = 0 \\
 \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^4} &= \frac{n}{2\sigma^2} \\
 \sigma^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2
 \end{aligned}$$

Then, we do the same but with two parameters,  $\theta_1, \theta_2$ , the former being the mean, and the latter being the variance. We can take the derivative with respect to  $\theta_2$ , and do effectively the same as before.

$$\begin{aligned}
 L(x_1, \dots, x_n; \theta_1, \theta_2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta_2}} \exp \frac{-(x_i - \theta_1)^2}{2\theta_2} \\
 \ln L(x_1, \dots, x_n; \theta_1, \theta_2) &= \sum_{i=1}^n -\ln \sqrt{2\pi\theta_2} - \frac{(x_i - \theta_1)^2}{2\theta_2} \\
 &= \sum_{i=1}^n -\frac{1}{2} \ln 2\pi\theta_2 - \frac{(x_i - \theta_1)^2}{2\theta_2} \\
 &= \sum_{i=1}^n -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \theta_2 - \frac{(x_i - \theta_1)^2}{2\theta_2} \\
 &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \theta_2 - \frac{\sum_{i=1}^n (x_i - \theta_1)^2}{2\theta_2} \\
 \frac{\partial}{\partial \theta_2} \ln L(x_1, \dots, x_n; \theta_1, \theta_2) &= -\frac{n}{2\theta_2} + \frac{\sum_{i=1}^n (x_i - \theta_1)^2}{2\theta_2^2} = 0 \\
 \frac{\sum_{i=1}^n (x_i - \theta_1)^2}{2\theta_2^2} &= \frac{n}{2\theta_2} \\
 \hat{\theta}_2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\theta}_1)^2
 \end{aligned}$$

Now, we just need to find if both estimators are biased or unbiased. We do this by seeing if their expected value is equal to the original parameter or not. Let's start with the former. We move the expectation in with linearity of expectation, and then can identify that the remaining expectation is just the definition of variance (expected

deviation from the mean squared) and see that it is unbiased.

$$E[\hat{\sigma}^2] = \mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E} [(x_i - \mu)^2] = \frac{1}{n} \sum_{i=1}^n \sigma^2 = \frac{1}{n} n \sigma^2 = \sigma^2$$

We do the same with the other estimator, and find that is biased, since it does not reduce down to the true parameter  $\theta_2$ :

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

$$\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

vs.

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

(The former turns out to be unbiased, the latter biased.)

## 9. Covariance Connection

Let  $X$  be the network connection status, where  $X = 0$  represents a stable connection and  $X = 1$  represents an unstable connection. Let  $Y$  be the number of successes in data transmission, taking values in the set  $\{0, 1, 2\}$ . If  $X = 0$ ,  $Y$  follows a Binomial distribution  $\text{Bin}(2, 0.8)$ , and if  $X = 1$ ,  $Y$  follows a Binomial distribution  $\text{Bin}(2, 0.3)$ . The probabilities for  $X$  are given by  $P(X = 0) = 0.8$  and  $P(X = 1) = 0.2$ . Find  $\text{Cov}(X, Y)$ . (note that we don't know that  $X$  and  $Y$  are independent here!)

**Solution:**

To calculate the covariance  $\text{Cov}(X, Y)$ , we need to determine  $E[X]$ ,  $E[Y]$ , and  $E[XY]$ . The covariance is then given by the formula:

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

First, we calculate  $E[X]$ :  $E[X] = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = 0 \cdot 0.8 + 1 \cdot 0.2 = 0.2$

Next, we calculate  $E[Y]$ . First, we calculate  $E[Y | X = 0]$  and  $E[Y | X = 1]$ . Based on what's given in the problem and using the formula for expectation for a binomial:  $E[Y | X = 0] = 2 \cdot 0.8 = 1.6$  and  $E[Y | X = 1] = 2 \cdot 0.3 = 0.6$ . Using the law of total expectation:

$$E[Y] = E[Y | X = 0]P(X = 0) + E[Y | X = 1]P(X = 1) = 1.6 \cdot 0.8 + 0.6 \cdot 0.2 = 1.4$$

To compute  $E[XY]$ , we first construct the joint PMF for  $XY$  and then use the definition of expectation. The possible values for  $XY$  are 0, 1, and 2. Let's compute the probabilities for each value:

$$\begin{aligned} P(XY = 0) &= P(X = 0 \cup Y = 0) = P(X = 0) + P(Y = 0) - P(X = 0 \cap Y = 0) \\ &= P(X = 0) + P(Y = 0) - P(X = 0)P(Y = 0 | X = 0) = 0.8 + 0.13 - 0.8 \cdot 0.2^2 = 0.898 \end{aligned}$$

$$P(XY = 1) = P(X = 1 \cap Y = 1) = P(X = 1)P(Y = 1 | X = 1) = 0.2 \cdot (2 \cdot 0.3 \cdot 0.7) = 0.084$$

$$P(XY = 2) = P(X = 1 \cap Y = 2) = P(X = 1)P(Y = 2 | X = 1) = 0.2 \cdot (0.3^2) = 0.018$$

In the above calculations, we use that  $P(Y = 0) = P(Y = 0 | X = 0)P(X = 0) + P(Y = 0 | X = 1)P(X = 1) = 0.2^2 \cdot 0.8 + 0.7^2 \cdot 0.2 = 0.13$  Now, using the definition of expectation, we have:

$$E[XY] = 0 \cdot 0.898 + 1 \cdot 0.084 + 2 \cdot 0.018 = 0.12$$

Therefore,  $E[XY] = 0.12$ . Finally, we calculate the covariance  $\text{Cov}(X, Y)$ :

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0.12 - 0.2 \cdot 1.4 = -0.16$$

Therefore, the covariance  $\text{Cov}(X, Y)$  is  $-0.16$ . The negative covariance of  $-0.16$  between the network connection status  $X$  and the number of successes in data transmission  $Y$  indicates an inverse relationship, suggesting that as the network connection status becomes less stable (i.e., as  $X = 1$ ), the likelihood of success in data transmission decreases, and vice versa, as expected!