

# SI 140A-02 Probability & Statistics for EECS, Fall 2024

## Homework 11

Name: **Wenye Xiong**

Student ID: 2023533141

Due on Dec. 24, 2024, 11:59 UTC+8

Read all the instructions below carefully before you start working on the assignment, and before you make a submission.

- You are required to write down all the major steps towards making your conclusions; otherwise you may obtain limited points of the problem.
- Write your homework in English; otherwise you will get no points of this homework.
- Any form of plagiarism will lead to 0 point of this homework.

## Problem 1

A coin with probability  $p$  of Heads is flipped repeatedly, where  $p$  is a known constant, with  $0 < p < 1$ .

- (a) What is the expected number of flips until the pattern  $HT$  is observed?
- (b) What is the expected number of flips until the pattern  $HH$  is observed?

### Solution

- (a) Let  $X_{HT}$  be the number of flips until the pattern  $HT$  is observed. We can write  $X_{HT}$  as the sum of two random variables  $X_1$  and  $X_2$ , where  $X_1$  is the number of flips until the first head is observed, and  $X_2$  is the number of flips until the first tail is observed after the first head. We have  $X_{HT} = X_1 + X_2$ . Since the coin is flipped repeatedly,  $X_1$  follows a geometric distribution with parameter  $p$ , and  $X_2$  follows a geometric distribution with parameter  $1 - p$ . Therefore, we have

$$E[X_1] = \frac{1}{p},$$

$$E[X_2] = \frac{1}{1-p}.$$

By the linearity of expectation, we have

$$E[X_{HT}] = E[X_1] + E[X_2] = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}.$$

Therefore, the expected number of flips until the pattern  $HT$  is observed is  $\frac{1}{p(1-p)}$ .

- (b) Let  $X_{HH}$  be the number of flips until the pattern  $HH$  is observed. We consider the result of the first flip:

$$E[X_{HH}] = E(X_{HH} | \text{first flip is H})P(\text{first flip is H}) + E(X_{HH} | \text{first flip is T})P(\text{first flip is T})$$

If the first flip is H, then the expected number of flips until the pattern  $HH$  is observed is  $2 \cdot p + (2 + E[X_{HH}]) \cdot (1 - p)$ . If the first flip is T, then the expected number of flips until the pattern  $HH$  is observed is  $1 + E[X_{HH}]$ . Therefore, we have

$$E[X_{HH}] = (2p + (2 + E[X_{HH}]) \cdot (1 - p)) \cdot p + (1 + E[X_{HH}]) \cdot (1 - p)$$

So we can solve the equation to get  $E[X_{HH}] = \frac{p+1}{p^2}$ . And the expected number of flips until the pattern  $HH$  is observed is  $\frac{p+1}{p^2}$ .

## Problem 2

Given two random variables  $X$  and  $Y$ , the corresponding joint PDF is

$$f_{X,Y}(x,y) = \begin{cases} x+y & \text{if } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{Otherwise} \end{cases}$$

Find  $E[Y | X]$  and  $L[Y | X]$ .

### Solution

The marginal PDF of  $X$  is

$$f_X(x) = \int_0^1 f_{X,Y}(x,y)dy = \int_0^1 (x+y)dy = x + \frac{1}{2}$$

The conditional PDF of  $Y$  given  $X$  is

$$f_{Y|X}(y | x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{x+y}{x + \frac{1}{2}}$$

The conditional expectation of  $Y$  given  $X$  is

$$E[Y | X = x] = \int_0^1 y f_{Y|X}(y | x) dy = \int_0^1 y \frac{x+y}{x + \frac{1}{2}} dy = \frac{3x+2}{6x+3}$$

Therefore, the conditional expectation of  $Y$  given  $X$  is  $E[Y | X] = \frac{3X+2}{6X+3}$ .

The linear MMSE estimator of  $Y$  given  $X$  is

$$L[Y | X] = E[Y] + \frac{\text{Cov}(X, Y)}{\text{Var}(X)}(X - E[X])$$

The covariance of  $X$  and  $Y$  is

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = \int_0^1 \int_0^1 xy(x+y)dxdy - \frac{49}{144} = -\frac{1}{144}$$

The variance of  $X$  is

$$\text{Var}(X) = E[X^2] - E[X]^2 = \int_0^1 x^2(x + \frac{1}{2})dx - \left(\int_0^1 x(x + \frac{1}{2})dx\right)^2 = \frac{11}{144}$$

Therefore, the linear MMSE estimator of  $Y$  given  $X$  is

$$L[Y | X] = \frac{7}{12} + \frac{-\frac{1}{144}}{\frac{11}{144}}(X - \frac{7}{12}) = \frac{7}{12} - \frac{1}{11}(X - \frac{7}{12}) = -\frac{1}{11}X + \frac{7}{11}$$

Therefore, the linear MMSE estimator of  $Y$  given  $X$  is  $L[Y | X] = -\frac{1}{11}X + \frac{7}{11}$ .

### Problem 3

Let  $X$  be the height of a randomly chosen adult man, and  $Y$  be his father's height, where  $X$  and  $Y$  have been standardized to have mean 0 and standard deviation 1. Suppose that  $(X, Y)$  is Bivariate Normal, with  $X, Y \sim \mathcal{N}(0, 1)$  and  $\text{Corr}(X, Y) = \rho$ .

- Let  $y = ax + b$  be the equation of the best line for predicting  $Y$  from  $X$  (in the sense of minimizing the mean squared error), e.g., if we were to observe  $X = 1.3$  then we would predict that  $Y$  is  $1.3a + b$ . Now suppose that we want to use  $Y$  to predict  $X$ , rather than using  $X$  to predict  $Y$ . Give and explain an intuitive guess for what the slope is of the best line for predicting  $X$  from  $Y$ .
- Find a constant  $c$  (in terms of  $\rho$ ) and an r.v.  $V$  such that  $Y = cX + V$ , with  $V$  independent of  $X$ .  
Hint: Start by finding  $c$  such that  $\text{Cov}(X, Y - cX) = 0$ .
- Find a constant  $d$  (in terms of  $\rho$ ) and an r.v.  $W$  such that  $X = dY + W$ , with  $W$  independent of  $Y$ .
- Find  $E(Y | X)$  and  $E(X | Y)$ .
- Reconcile (a) and (d), giving a clear and correct intuitive explanation.

### Solution

(a) Since the parameter  $\rho$  tells us what is the rate of change of second variable respective to the first one, we can assume that  $\rho$  is the slope of the line, i.e.  $a = \rho$ . Now, in order to predict  $X$  from  $Y$ , we just have to consider the line that is inverse to the original line. From the basic algebra, we know that inverse has slope one over the original slope. Thus, the required slope is  $\frac{1}{\rho}$ .

(b) Since we have to find  $V = Y - cX$  such that is independent from  $X$ , using the given hint, we have that

$$0 = \text{Cov}(X, Y - cX) = \text{Cov}(X, Y) - c \text{Var}(X) = \rho - c.$$

Hence, let's define  $c = \rho$  and it is the only candidate for the constant  $c$ . Let's check that  $X$  and  $V$  are independent. Observe that  $Y - \rho X$  is also Normal (as the linear combination of two Bivariate Normals). So, the fact that two Normals that construct Bivariate Normal are independent is equivalent to the fact that they are uncorrelated. Since we have the last information, we have found the required.

(c) With the same calculation and discussion as in part (b), we have that the answer is also  $d = \rho$ .

(d) Using the definition of conditional density function, we have that

$$\begin{aligned} f_{Y|X}(y | x) &= \frac{f(x, y)}{f(x)} = \frac{\frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 + y^2 - 2xy\rho)\right)}{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)} \\ &= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{x^2 + y^2 - 2xy\rho}{2(1-\rho^2)} + \frac{x^2}{2}\right) \\ &= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{(y - \rho x)^2}{2(1-\rho^2)}\right). \end{aligned}$$

Now, we see that

$$Y | X = x \sim \mathcal{N}(\rho x, 1 - \rho^2).$$

Hence,  $E(Y | X) = \rho X$ . Because of the symmetry, we also have that  $E(X | Y) = \rho Y$ .

(e) Since we know that means of  $X$  and  $Y$  are zero, we have that

$$X = \alpha Y$$

for some  $\alpha$ . Applying the conditional expectation  $E(\cdot | X)$  to the both sides, we have that

$$X = \alpha E(Y | X) = \alpha \rho X$$

Because of the fact that  $X \neq 0$  almost certainly, we can conclude that  $\alpha = \frac{1}{\rho}$ . Hence, we have proved the claimed.

## Problem 4

Show the following orthogonality properties of MMSE:

- (a) For any function  $\phi(\cdot)$ , one has

$$E[(Y - E[Y | X])\phi(X)] = 0$$

- (b) If the function  $g(X)$  satisfied

$$E[(Y - g(X))\phi(X)] = 0, \forall \phi(\cdot)$$

then  $g(X) = E(Y | X)$ .

## Solution

- (a) According to the linearity of expectation, we have

$$E((Y - E[Y|X])\phi(X)) = E(Y\phi(X)) - E(E[Y|X]\phi(X))$$

And by the product rule of conditional expectation and law of total expectation, we have

$$E(E[Y|X]\phi(X)) = E(E[Y\phi(X)|X]) = E(Y\phi(X))$$

Therefore, we have

$$E((Y - E[Y|X])\phi(X)) = E(Y\phi(X)) - E(Y\phi(X)) = 0$$

- (b) Since  $E[(Y - g(X))\phi(X)] = 0, \forall \phi(\cdot)$ , we have

$$E[(Y - g(X))\phi(X)] = E[Y\phi(X)] - E[g(X)\phi(X)] = 0$$

Therefore, we have

$$E[Y\phi(X)] = E[g(X)\phi(X)]$$

Since  $\phi(X)$  is arbitrary, we can choose  $\phi(X) = 1$ , then we have

$$E[Y] = E[g(X)]$$

Therefore, we have

$$E[Y] = E[g(X)] = E[E[Y|X]]$$

Therefore, we have  $g(X) = E(Y | X)$ .

## Problem 5

Given a coin with the probability  $p$  of landing heads.  $p$  is unknown and we need to estimate its value through data. In our data collection model, we have  $n$  independent tosses, result of each toss is either Head or Tail. Let  $X$  denote the number of heads in the total  $n$  tosses. Now we conduct experiments to collect data and find  $X = k$ . Then we need to find  $\hat{p}$ , the estimation of  $p$ .

- Assume  $p$  is an unknown constant. Find  $\hat{p}$  through the MLE (Maximum Likelihood Estimation) rule.
- Assume  $p$  is a random variable with a prior distribution  $p \sim \text{Beta}(a, b)$ , where  $a$  and  $b$  are known constants. Find  $\hat{p}$  through the MAP (Maximum a Posterior Probability) rule.
- Assume  $p$  is a random variable with a prior distribution  $p \sim \text{Beta}(a, b)$ , where  $a$  and  $b$  are known constants. Find  $\hat{p}$  through the MMSE (Minimal Mean Squared Error) rule.

## Solution

- The likelihood function is

$$L(p) = \binom{n}{k} p^k (1-p)^{n-k}$$

The log-likelihood function is

$$\log L(p) = k \log p + (n-k) \log(1-p) + \text{constant}$$

Taking the derivative of the log-likelihood function with respect to  $p$  and setting it to zero, we have

$$\frac{k}{p} - \frac{n-k}{1-p} = 0$$

Therefore, the MLE of  $p$  is

$$\hat{p} = \frac{k}{n}$$

- The posterior distribution of  $p$  is

$$f(p | X = k) \propto f(X = k | p) f(p) = \binom{n}{k} p^k (1-p)^{n-k} \frac{1}{B(a, b)} p^{a-1} (1-p)^{b-1} = \frac{\binom{n}{k}}{B(a, b)} p^{k+a-1} (1-p)^{n-k+b-1}$$

The MAP of  $p$  is the mode of the posterior distribution, which is the value of  $p$  that maximizes the posterior distribution. To find the mode more easily, we can take the log of the posterior distribution and find the maximum of the log-posterior distribution. The log-posterior distribution is

$$\log f(p | X = k) = \log \left( \frac{\binom{n}{k}}{B(a, b)} \right) + (k+a-1) \log p + (n-k+b-1) \log(1-p)$$

Taking the derivative of the log-posterior distribution with respect to  $p$  and setting it to zero, we have

$$\frac{k+a-1}{p} - \frac{n-k+b-1}{1-p} = 0$$

Therefore, the MAP of  $p$  is

$$\hat{p} = \frac{k+a-1}{n+a+b-2}$$

- The MMSE of  $p$  is

$$\hat{p} = E[p | X = k] = \int_0^1 p f(p | X = k) dp = \frac{k+a}{n+a+b}$$

**Problem 6**

**(Optional Challenging Problem)** Use two different methods to show that if  $X$  and  $Y$  are jointly Normal random variables, then

$$E[Y \mid X] = L[Y \mid X] = E(Y) + \frac{\text{Cov}(X, Y)}{\text{Var}(X)}(X - E(X))$$