Probability & Statistics for EECS: Homework #11

Due on Dec 2, 2023 at 23:59

Name:

Student ID:

Problem 1

Show the following orthogonality properties of MMSE:

(a) For any function $\phi(\cdot)$, one has
$$E[(Y - E[Y \mid X])\phi(X)] = 0.$$

(b) If the function $g(X)$ satisfied
$$E[(Y - g(X))\phi(X)] = 0, \forall \phi(\cdot).$$
then $g(X) = E(Y \mid X)$.

Solution

(a) Define the conditional expectation $E[Y|X]$ as $\hat{Y}$:
$$\hat{Y} = E[Y|X]$$
We want to show that for any function $\phi(X)$:
$$E[(Y - \hat{Y})\phi(X)] = 0$$
First, recall that the expectation of a function of Y given X is $\hat{Y}$:
$$\hat{Y} = E[Y|X]$$
This means that $Y - \hat{Y}$ is the difference between Y and its conditional expectation given X. Consider the expectation $E[(Y - \hat{Y})\phi(X)]$:
$$E[(Y - \hat{Y})\phi(X)] = E[E[(Y - \hat{Y})\phi(X)|X]] = \phi(X)E[(Y - \hat{Y})|X]$$
But since $\hat{Y} = E[Y|X]$, we have:
$$E[(Y - \hat{Y})|X] = E[Y|X] - E[Y|X] = 0$$
Therefore, we have:
$$E[(Y - \hat{Y})\phi(X)] = 0$$

(b) Let's start with the assumption:
$$E[(Y - g(X))\phi(X)] = 0$$
for all functions $\phi(X)$.

Since this holds for all functions $\phi(X)$, it must hold in particular for $\phi(X) = h(X)$ where $h(X)$ is any arbitrary function of X. Let's choose $\phi(X) = (g(X) - E[Y|X])$. Then:
$$E[(Y - g(X))(g(X) - E[Y|X])] = 0$$
Expanding this expectation, we have:
$$E[Yg(X) - YE[Y|X] - g(X)g(X) + g(X)E[Y|X]] = 0$$
Which simplifies to:
$$E[Yg(X)] - E[YE[Y|X]] - E[g(X)^2] + E[g(X)E[Y|X]] = 0$$
Since $E[YE[Y|X]] = E[E[Y|X]E[Y|X]]$ and $E[g(X)E[Y|X]] = E[E[Y|X]g(X)]$, we have
$$E[Yg(X)] - E[E[Y|X]E[Y|X]] - E[g(X)^2] + E[E[Y|X]g(X)] = 0$$
Given that $E[Y|X]$ is the best (in the mean-square error sense) predictor of Y given X, $g(X)$ must equal $E[Y|X]$ to minimize the mean squared error. Therefore:
$$g(X) = E[Y|X]$$

Problem 2

Let X and Y be independent, positive r.v.s. with finite expected values.

(a) Give an example where $E(\frac{X}{X+Y}) \neq \frac{E(X)}{E(X+Y)}$, computing both sides exactly. Hint: Start by thinking about the simplest examples you can think of!

(b) If X and Y are i.i.d., then is it necessarily true that $E(\frac{X}{X+Y}) = \frac{E(X)}{E(X+Y)}$?

(c) Now let $X \sim \text{Gamma}(a, \lambda)$ and $Y \sim \text{Gamma}(b, \lambda)$. Show *without using calculus* that

$$E\left(\frac{X^c}{(X+Y)^c}\right) = \frac{E(X^c)}{E((X+Y)^c)},$$

for every real $c > 0$.

Solution

(a) Suppose that $X = 1$ almost certainly, *i.e.*, $X = 1$ with the probability of 1.

Meanwhile, suppose that Y is a random variable that has following distribution

$$Y = \begin{cases} 5 & \text{w.p. } 1/2 \\ 10 & \text{w.p. } 1/2. \end{cases}$$

Suppose that X and Y are independent. We have that

$$\frac{X}{X+Y} = \begin{cases} \frac{1}{6} & \text{w.p. } 1/2 \\ \frac{1}{11} & \text{w.p. } 1/2. \end{cases}$$

Hence, we obtain

$$E\left(\frac{X}{X+Y}\right) = \frac{1}{6} \cdot \frac{1}{2} + \frac{1}{11} \cdot \frac{1}{2} = \frac{17}{132}.$$

But on the other hand we get that

$$\frac{E(X)}{E(X+Y)} = \frac{E(X)}{E(X)+E(Y)} = \frac{1}{1+7.5} = \frac{2}{27},$$

which are not equal.

(b) Yes, this claim is true. Observe that by symmetry, we have that

$$E\left(\frac{X}{X+Y}\right) = E\left(\frac{Y}{X+Y}\right),$$

and because of

$$E\left(\frac{X}{X+Y} + \frac{Y}{X+Y}\right) = 1,$$

we obtain

$$E\left(\frac{X}{X+Y}\right) = E\left(\frac{Y}{X+Y}\right) = \frac{1}{2}$$

On the other hand, since X and Y are *i.i.d.*, we have $E(X) = E(Y)$. Then we obtain

$$\frac{E(X)}{E(X+Y)} = \frac{E(X)}{E(X)+E(Y)} = \frac{E(X)}{E(X)+E(X)} = \frac{1}{2}.$$

Hence we have proved the claimed.

Problem 2 continued on next page...

(c) By the bank-post office story, we know that $\frac{X}{X+Y}$ and $X+Y$ are independent. So, we can conclude that $\frac{X^c}{(X+Y)^c}$ and $(X+Y)^c$ are independent as functions of two independent variables. Thus $\frac{X^c}{(X+Y)^c}$ and $(X+Y)^c$ are uncorrelated.

In this way, we have

$$\begin{aligned}\mathrm{Cov}\left(\frac{X^c}{(X+Y)^c}, (X+Y)^c\right) &= E\left(\frac{X^c}{(X+Y)^c}\cdot(X+Y)^c\right) - E\left(\frac{X^c}{(X+Y)^c}\right)E((X+Y)^c) \\ &= E(X^c) - E\left(\frac{X^c}{(X+Y)^c}\right)E((X+Y)^c) = 0.\end{aligned}$$

Hence, we have

$$E\left(\frac{X^c}{(X+Y)^c}\right) = \frac{E(X^c)}{E((X+Y)^c)},$$

for every real $c > 0$.

Problem 3

A DNA sequence can be represented as a sequence of letters, where the "alphabet" has 4 letters: A,C,T,G. Suppose such a sequence is generated randomly, where the letters are independent and the probabilities of A,C,T,G are p_1, p_2, p_3, p_4 respectively.

(a) In a DNA sequence of length 115, what is the expected number of occurrences of the expression "CATCAT" (in terms of the p_j)? (Note that, for example, the expression "CATCATCAT" counts as 2 occurrences.)

(b) What is the probability that the first A appears earlier than the first C appears, as letters are generated one by one (in terms of the p_j)?

(c) For this part, assume that the p_j are unknown. Suppose we treat p_2 as a $\text{Unif}(0, 1)$ r.v. before observing any data, and that then the first 3 letters observed are "CAT". Given this information, what is the probability that the next letter is C?

Solution

(a) Let I_j be 1 if the j^{th} and the next 5 positions have the pattern "CATCAT", zero otherwise, where $j \in \{1, 2, \ldots, 110\}$. Then $\mathrm{E}(I_j) = (p_2 p_1 p_3)^2$. So the expected number of occurrences is

$$\mathrm{E}\left(\sum_{j=1}^{110} I_j\right) = \sum_{j=1}^{110} \mathrm{E}(I_j) = 110(p_1 p_2 p_3)^2.$$

(b) What we concern is "A" and "C", ignoring others, we have

$$\Pr\{\text{A appears earlier than C}\} = \Pr\{\text{A appears}|\text{A or C appears}\} = \frac{p_1}{p_1 + p_2}.$$

(c) Let X be the number of C's in the first 3 letters (so $X = 1$ is observed here). The prior is $p_2 \sim \text{Beta}(1, 1)$, so the posterior is $p_2 | X = 1 \sim \text{Beta}(2, 3)$ (by the connection between Beta and Binomial, or by Bayes Rule). Given p_2, the indicator of the next letter being C is $\text{Bern}(p_2)$. So given X (but not given p_2), the probability of the next letter being C is:

$$\begin{aligned} &\Pr\{\text{next letter is C}|X\} = \int_0^1 \Pr\{\text{next letter is C}|p_2 = \theta, X\} f_{p_2|X}(\theta)\, d\theta \\ &= \int_0^1 \theta f_{p_2|X}(\theta)\, d\theta = \mathrm{E}(p_2|X) = \frac{2}{5}. \end{aligned} \tag{1}$$

Sanity check: It makes sense that the answer should be strictly in between 1/2 (the mean of the prior distribution) and 1/3 (the observed frequency of C's in the data).

Problem 4

A coin with probability p of Heads is flipped repeatedly. For (a) and (b), suppose that 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?

(c) Now suppose that p is unknown, and that we use a Beta(a, b) prior to reflect our uncertainty about p (where a and b are known constants and are greater than 2). In terms of a and b, find the corresponding answers to (a) and (b) in this setting.

Solution

(a) This can be thought of as "Wait for Heads, then wait for the first Tails after the first Heads," so the expected value is $\frac{1}{p} + \frac{1}{q}$, with $q = 1 - p$.

(b) Let X be the waiting time for HH and condition on the first toss, writing H for the event that the first toss is Heads and T for the complement of H:

$$E(X) = E(X|H)p + E(X|T)q = E(X|H)p + (1 + E(X))q.$$

To find $E(X|H)$, condition on the second toss:

$$E(X|H) = E(X|HH)p + E(X|HT)q = 2p + (2 + E(X))q.$$

Solving for $E(X)$, we have

$$E(X) = \frac{1}{p} + \frac{1}{p^2}.$$

(c) Let X and Y be the number of flips until HH and until HT, respectively. By (a), $E(Y|p) = \frac{1}{p} + \frac{1}{1-p}$. So $E(Y) = E(E(Y|p)) = E\left(\frac{1}{p}\right) + E\left(\frac{1}{1-p}\right)$. Likewise, by (b), $E(X) = E(E(X|p)) = E\left(\frac{1}{p}\right) + E\left(\frac{1}{p^2}\right)$. By LOTUS,

$$\begin{aligned}
E\left(\frac{1}{p}\right) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 p^{a-2}(1-p)^{b-1}dp = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a-1)\Gamma(b)}{\Gamma(a+b-1)} = \frac{a+b-1}{a-1}, \\
E\left(\frac{1}{1-p}\right) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 p^{a-1}(1-p)^{b-2}dp = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a)\Gamma(b-1)}{\Gamma(a+b-1)} = \frac{a+b-1}{b-1}, \\
E\left(\frac{1}{p^2}\right) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 p^{a-3}(1-p)^{b-1}dp = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a-2)\Gamma(b)}{\Gamma(a+b-2)} = \frac{(a+b-1)(a+b-2)}{(a-1)(a-2)}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
E(Y) &= \frac{a+b-1}{a-1} + \frac{a+b-1}{b-1}, \\
E(X) &= \frac{a+b-1}{a-1} + \frac{(a+b-1)(a+b-2)}{(a-1)(a-2)}.
\end{aligned}$$

Problem 5

(a) Let $p \sim \text{Beta}(a, b)$, where a and b are positive real numbers. Find $E(p^2(1-p)^2)$, fully simplified (Γ should not appear in your final answer).

Two teams, A and B, have an upcoming match. They will play five games and the winner will be declared to be the team that wins the majority of games. Given p, the outcomes of games are independent, with probability p of team A winning and $(1-p)$ of team B winning. But you don't know p, so you decide to model it as an r.v., with $p \sim \text{Unif}(0, 1)$ a priori (before you have observed any data).
To learn more about p, you look through the historical records of previous games between these two teams, and find that the previous outcomes were, in chronological order, $AAABBAABAB$. (Assume that the true value of p has not been changing over time and will be the same for the match, though your beliefs about p may change over time.)

(b) Does your posterior distribution for p, given the historical record of games between A and B, depend on the specific order of outcomes or only on the fact that A won exactly 6 of the 10 games on record? Explain.

(c) Find the posterior distribution for p, given the historical data.

The posterior distribution for p from (c) becomes your new prior distribution, and the match is about to begin!

(d) Conditional on p, is the indicator of A winning the first game of the match positively correlated with, uncorrelated with, or negatively correlated of the indicator of A winning the second game of the match? What about if we only condition on the historical data?

(e) Given the historical data, what is the expected value for the probability that the match is not yet decided when going into the fifth game (viewing this probability as an r.v. rather than a number, to reflect our uncertainty about it)?

Solution

(a) Since the PDF of r.v. p is $f_p(x) = \frac{1}{\beta(a,b)} x^{a-1}(1-x)^{b-1}$ for $p \in (0, 1)$, we have

$$\begin{aligned} \mathrm{E}\left(p^2(1-p)^2\right) &= \int_0^1 x^2(1-x)^2 \cdot \frac{1}{\beta(a,b)} x^{a-1}(1-x)^{b-1}\, dx \\ &= \frac{1}{\beta(a,b)} \int_0^1 x^{a+2-1}(1-x)^{b+2-1}\, dx. \end{aligned}$$

Because of the equation $\int_0^1 \frac{1}{\beta(a+2,b+2)} x^{a+2-1}(1-x)^{b+2-1}\, dx = 1$, we further have

$$\mathrm{E}\left(p^2(1-p)^2\right) = \frac{1}{\beta(a,b)} \cdot \beta(a+2, b+2).$$

Moreover, by the fact that $\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ we can write the result as

$$\mathrm{E}\left(p^2(1-p)^2\right) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+2)\Gamma(b+2)}{\Gamma(a+b+4)}.$$

Substitute $\Gamma(n) = (n-1)!$, we get

$$\mathrm{E}\left(p^2(1-p)^2\right) = \frac{(a+b-1)!}{(a-1)!(b-1)!} \cdot \frac{(a+1)!(b+1)!}{(a+b+3)!} = \frac{ab(a+1)(b+1)}{(a+b)(a+b+1)(a+b+2)(a+b+3)}.$$

Problem 5 continued on next page...

(b) The posterior distribution for p does not depend on the specific order of outcomes. The reason is that the probability of all possible orders are the same when p is given, *i.e.*, orders don't influence our belief about p.

(c) The prior distribution of p is $\mathrm{Unif}(0,1)$, which is equivalent to $\mathrm{Beta}(1,1)$. By Beta-Binomial conjugacy, the posterior distribution of p is $\mathrm{Beta}(1+6,1+4)$, *i.e.*, $\mathrm{Beta}(7,5)$.

(d) Let I_1 be the indicator of A winning the first game of the match, and I_2 be the indicator of A winning the second game of the match. Conditional on p, I_1 and I_2 are uncorrelated since they are i.i.d. Bernoulli r.v.s when p is given.

If we only condition on the historical data, *i.e.*, $p \sim \mathrm{Beta}(7,5)$, we have

$$\mathrm{E}\left(I_1\right) = \Pr\left(I_1 = 1\right) = \int_0^1 \Pr\left(I_1 = 1 | p = x\right) f_p\left(x\right) dx = \int_0^1 x f_p\left(x\right) dx = \mathrm{E}\left(p\right),$$

where the third equation is because $I_1|p \sim \mathrm{Bern}(p)$. Similarly we also have $\mathrm{E}(I_2) = \mathrm{E}(p)$. On the other hand,

$$\begin{aligned}\mathrm{E}\left(I_1 I_2\right) &= \Pr\left(I_1 I_2 = 1\right) = \Pr\left(I_1 = 1, I_2 = 1\right) \\ &= \int_0^1 \Pr\left(I_1 = 1, I_2 = 1 | p = x\right) f_p\left(x\right) dx = \int_0^1 x^2 f_p\left(x\right) dx = \mathrm{E}\left(p^2\right)\end{aligned}$$

Since I_1 and I_2 are independent Bernoulli r.v.s with distribution $\mathrm{Bern}(p)$ given p. Therefore

$$\mathrm{Cov}\left(I_1, I_2\right) = \mathrm{E}\left(I_1 I_2\right) - \mathrm{E}\left(I_1\right)\mathrm{E}\left(I_2\right) = \mathrm{E}\left(p^2\right) - \mathrm{E}^2\left(p\right) = \mathrm{Var}\left(p\right) > 0,$$

which implies that I_1 and I_2 are positively correlated.

(e) Let X be the number of games that A win in the first 4 games of the match, then we have $X|p \sim \mathrm{Bin}(4,p)$. The probability that the match is not yet decided when going into the fifth game given p is

$$\Pr\left(X = 2 | p\right) = \binom{4}{2} p^2 \left(1-p\right)^2.$$

Given the historical data, *i.e.*, $p \sim \mathrm{Beta}(7,5)$, the expected value of probability that the match is not yet decided when going into the fifth game is

$$\mathrm{E}\left(\binom{4}{2} p^2 \left(1-p\right)^2\right) = \binom{4}{2} \mathrm{E}\left(p^2\left(1-p\right)^2\right).$$

From (a) we know that when $p \sim \mathrm{Beta}(7,5)$, $\mathrm{E}\left(p^2\left(1-p\right)^2\right) = \frac{2}{39}$. Thus the expected value of probability is $\binom{4}{2} \times \frac{2}{39} = \frac{4}{13}$.

Problem 6

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

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

Solution

(a) Let $\mu_X = \mathrm{E}(X), \mu_Y = \mathrm{E}(Y), \sigma_X^2 = \mathrm{Var}(X), \sigma_Y^2 = \mathrm{Var}(Y), \rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$,
$X = \sigma_X W + \mu_X$, $Y = \sigma_Y \left(\rho W + \sqrt{1-\rho^2} Z\right) + \mu_Y$, where W and Z are independent

$$\begin{aligned}
\mathrm{E}[Y|X] &= \mathrm{E}\left[\sigma_Y \left(\rho W + \sqrt{1-\rho^2} Z\right) + \mu_Y | \sigma_X W + \mu_X\right] \\
&= \frac{\rho \sigma_Y}{\sigma_X}(X - \mu_X) + \mu_Y \\
&= \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}(X - \mathrm{E}(X)) + \mathrm{E}(Y)
\end{aligned}$$

(b) For jointly normal random variables (X, Y) with mean vector (μ_X, μ_Y) and covariance matrix:

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \sigma_Y^2 \end{pmatrix}$$

The conditional distribution $Y|X$ is also normally distributed with mean and variance given by:

$$E[Y|X] = \mu_Y + \frac{\mathrm{Cov}(X,Y)}{\sigma_X^2}(X - \mu_X)$$

$$\mathrm{Var}(Y|X) = \sigma_Y^2 - \frac{\mathrm{Cov}(X,Y)^2}{\sigma_X^2}$$

Thus, the conditional expectation $E[Y|X]$ is:

$$E[Y|X] = \mu_Y + \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}(X - \mu_X)$$

This again confirms the given expression:

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

In both methods, we have derived that if X and Y are jointly normal random variables, then the conditional expectation of Y given X is:

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