Probability & Statistics for EECS: Homework #12

Due on Dec 2, 2023 at 23:59

Name:
Student ID:

Problem 1

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.

(a) Assume p is an unknown constant. Find $\hat{p}$ through the MLE (Maximum Likelihood Estimation) rule.

(b) 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.

(c) 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

(a) Let X_i be the outcome of ith toss. Then $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} \text{Bern}(p)$, where p is an unknown constant. The PMF of X_i can be formulated as

$$P_{X_i}(x_i; p) = p^{x_i}(1-p)^{1-x_i}$$

since

$$p^{x_i}(1-p)^{1-x_i} = \begin{cases} p, & \text{if } x_i = 1, \\ 1-p, & \text{if } x_i = 0. \end{cases}$$

The likelihood function is

$$P_X(x; p) = \prod_{i=1}^{n} P_{X_i}(x_i; p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^k(1-p)^{n-k}$$

So the corresponding log-likelihood function is

$$g(p) = \log P_X(x; p) = \log p^{S_n}(1-p)^{n-S_n} = S_n \log p + (n - S_n)\log(1-p)$$

Now we try to find $\widehat{p}_{\text{MLE}}$ such that $g(\widehat{p}_{\text{MLE}})$ is the maximum of $g(p)$. We have

$$\begin{aligned} g'(p) &= \frac{k}{p} - \frac{n-k}{1-p}, \\ g''(p) &= -\frac{k}{p^2} - \frac{n-k}{(1-p)^2} \leq 0 \end{aligned}$$

Let $g'(p) = 0$, we can get $p = \frac{k}{n}$. Since $g''(p) \leq 0$, then we know that

$$\widehat{p}_{\text{MLE}} = \frac{k}{n}$$

is the MLE of p.

(b) We know the posterior distribution

$$f_{p|X=k} \propto p^{a+k-1}(1-p)^{b+n-k-1},\ p \in (0,1)$$

by Beta-Binomial conjugacy. Then the MAP estimator

$$\widehat{p}_{\text{MAP}} = \arg\max_p f_{\theta|X=k} = \arg\max_p \log(f_{p|X=k})$$

Problem 1 continued on next page...

since logarithmic function is monotonically increasing. Let

$$g(p) = \log(f_{p|X=k}) = (a+k-1)\log p + (b+n-k-1)\log(1-p),$$

where we don't consider the proportional constant. Our goal is to find p^* such that $g(p^*)$ is maximum of $g(p)$. We have

$$\begin{aligned} g'(p) &= \frac{a+k-1}{p} - \frac{b+n-k-1}{1-p}, \\ g''(p) &= -\frac{a+k+1}{p^2} - \frac{b+n-k-1}{(1-p)^2} < 0. \end{aligned}$$

Let $g'(p^*) = 0$. We have $p^* = \frac{a+k-1}{a+b+n-2}$, and $g(p^*)$ is maximum of $g(p)$ since $g''(p) < 0$.

Then we can get the MAP estimate

$$\widehat{p}_{\text{MAP}} = \arg\max_p f_{p|X=k} = \arg\max_p \log(f_{\theta|X=k}) = p^* = \frac{a+k-1}{a+b+n-2}.$$

(c) Since the prior distribution is $p \sim \text{Beta}(a,b)$ and the conditional distribution of X given p is $X|p \sim \text{Bin}(n,p)$, we can get the posterior distribution

$$\Theta|X = k \sim \text{Beta}(a+k, b+n-k)$$

by Beta-Binomial conjugacy. It follows that

$$E(p|X=k) = \frac{a+k}{a+b+n},$$

so the MMSE estimation of Θ is

$$\widehat{p}_{\text{MMSE}} = E(p|X=k) = \frac{a+k}{a+b+n}.$$

Problem 2

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 $\text{Corr}(X,Y) = \rho$.

(a) 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.

(b) Find a constant c (in terms of ρ) and an r.v. V such that $Y = cX + V$, with V independent of X. Hint: Start by finding c such that $\text{Cov}(X, Y - cX) = 0$.

(c) Find a constant d (in terms of ρ) and an r.v. W such that $X = dY + W$, with W independent of Y.

(d) Find $E(Y \mid X)$ and $E(X \mid Y)$.

(e) Reconcile (a) and (d), giving a clear and correct intuitive explanation.

Solution

(a) Since the parameter ρ tells us what is the rate of change of second variable respective to the first one, we can assume that ρ 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

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

Now, we see that

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

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

Problem 2 continued on next page...

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

$$X = \alpha Y$$

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

$$X = \alpha E(Y \mid X) = \alpha \rho X.$$

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 3

Two chess players, Vishy and Magnus, play a series of games. Given p, the game results are i.i.d. with probability p of Vishy winning, and probability $q = 1-p$ of Magnus winning (assume that each game ends in a win for one of the two players). But p is unknown, so we will treat it as an r.v. To reflect our uncertainty about p, we use the prior $p \sim \text{Beta}(a, b)$, where a and b are known positive integers and $a \geq 2$.

(a) Find the expected number of games needed in order for Vishy to win a game (including the win). Simplify fully; your final answer should not use factorials or Γ.

(b) Explain in terms of independence vs. conditional independence the direction of the inequality between the answer to (a) and $1 + E(G)$ for $G \sim \text{Geom}\left(\frac{a}{a+b}\right)$.

(c) Find the conditional distribution of p given that Vishy wins exactly 7 out of the first 10 games.

Solution

(a) Denote N as the number of games needed for Vishy to win the game for one time, then there is $N|p \sim \text{FS}(p)$. Via Adam's law, we have:

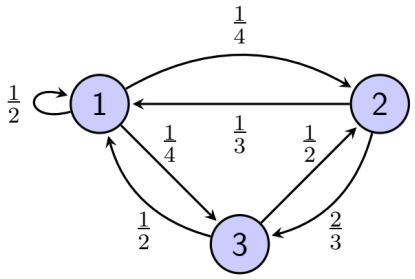
$$\begin{aligned}
\text{E}(N) &= \text{E}(\text{E}(N|p)) \\
&= \text{E}\left(\frac{1}{p}\right) \\
&= \int_0^1 \frac{1}{\beta(a,b)} \frac{1}{p} p^{a-1} (1-p)^{b-1} \, dp \\
&= \frac{\beta(a-1,b)}{\beta(a,b)} \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a-1)\Gamma(b)}{\Gamma(a+b-1)} \\
&= \frac{a+b-1}{a-1}
\end{aligned}$$

(b) $1 + E(G) = \frac{a+b}{a} < \frac{a+b-1}{a-1} = E(N)$ means that the games are conditionally independent given p while not dependent between each other. Since $1 + E(G)$ can been seen as the expectation of the number of trials for Vishy to win the first game given $p = \frac{a}{a+b}$. While p is unknown, and each time of Vishy's loss will decrease the probability of wining the game. Thus the expectation estimated by conditional probability is larger than the expectation given by prior distribution. Therefore $1 + E(G) < E(N)$.

(c) Via Beta-Binomial conjugacy, the conditional distribution of p give that Vish wins exactly 7 out out of the first 10 games is $\text{Beta}(a + 7, b + 3)$.

Problem 4

Given a Markov chain with state-transition diagram shown as follows:



(a) Is this chain irreducible?

(b) Is this chain aperiodic?

(c) Find the stationary distribution of this chain.

(d) Is this chain reversible?

Solution

The transition matrix of the Markov chain is

$$Q = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

(a) Yes, because the elements in the matrix are $Q_{1,2}, Q_{2,1}, Q_{1,3}, Q_{3,1}, Q_{2,3}, Q_{3,2}$ are all non-zero.

(b) Yes, the diagram we know that 1 is a possible return time for state 1, thus $d(1) = 1$ since both $2, 3$ are possible for state 2 and 3, $d(2) = d(3) = 1$ because the chain is irreducible and $d(1) = d(2) = d(3) = 1$. Therefore the chain is aperiodic.

(c) Denote π as the stationary distribution for the chain, then there is $\pi Q = \pi$. Then we solve the problem $\pi(Q - I) = 0$, as follows:

$$\begin{bmatrix} \pi_1 & \pi_2 & \pi_3 \end{bmatrix} \begin{bmatrix} -\frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & -1 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & -1 \end{bmatrix} = 0$$

where $\sum_{i=1}^{3} \pi_i$. The solution is $\pi = (\frac{16}{35}, \frac{9}{35}, \frac{2}{7})$.

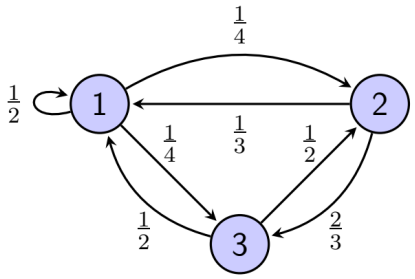
(d) No. If the chain is reversible, there exists a distribution π which satisfy:

$$\begin{cases} \pi_1 \cdot \frac{1}{4} = \pi_2 \cdot \frac{1}{3} \\ \pi_1 \cdot \frac{1}{4} = \pi_3 \cdot \frac{1}{2} \\ \pi_2 \cdot \frac{2}{3} = \pi_3 \cdot \frac{1}{2} \end{cases}$$

The solution of the above problem is $\pi_1 = \pi_2 = \pi_3 = 0$, which cannot satisfy the constraint $\sum \pi = 1$.

Problem 5

Given a Markov chain with state-transition diagram shown as follows:



(a) Find $P(X_3 = 3 \mid X_2 = 2)$ and $P(X_4 = 1 \mid X_3 = 2)$.

(b) If $P(X_0 = 2) = \frac{2}{5}$, find $P(X_0 = 2, X_1 = 3, X_2 = 1)$.

(c) Find $P(X_2 = 1 \mid X_0 = 2), P(X_2 = 2 \mid X_0 = 2)$, and $P(X_2 = 3 \mid X_0 = 2)$.

(d) Find $E(X_2 \mid X_0 = 2)$.

Solution

The transition matrix of the Markov chain is

$$Q = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

(a) From the state-transition diagram, we have

$$\begin{aligned} P(X_3 = 3|X_2 = 2) &= \frac{2}{3} \\ P(X_4 = 1|X_3 = 2) &= \frac{1}{3} \end{aligned} \tag{1}$$

(b)

$$\begin{aligned} P(X_0 = 2, X_1 = 3, X_2 = 1) &= P(X_1 = 3, X_2 = 1|X_0 = 2)P(X_0 = 2) \\ &= P(X_2 = 1|X_1 = 3, X_0 = 2)P(X_1 = 3|X_0 = 2)P(X_0 = 2) \\ &= P(X_2 = 1|X_1 = 3)P(X_1 = 3|X_0 = 2)P(X_0 = 2) \\ &= \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{2}{5} = \frac{2}{15} \end{aligned} \tag{2}$$

Problem 5 continued on next page...

(c)

$$
\begin{aligned}
&P(X_2 = 1|X_0 = 2) = \sum_{i=1}^{3} P(X_2 = 1|X_1 = i, X_0 = 2)P(X_1 = i|X_0 = 2) \\
&= \sum_{i=1}^{3} P(X_2 = 1|X_1 = i)P(X_1 = i|X_0 = 2) \\
&= \frac{1}{6} + 0 + \frac{2}{6} = \frac{1}{2} \\
&P(X_2 = 2|X_0 = 2) = \sum_{i=1}^{3} P\big(X_2 = 2\big|X_1 = i, X_0 = 2\big)P\big(X_1 = i\big|X_0 = 2\big) \\
&= \sum_{i=1}^{3} P(X_2 = 2|X_1 = i)P(X_1 = i|X_0 = 2) \\
&= \frac{1}{12} + 0 + \frac{2}{6} = \frac{5}{12} \\
&P(X_2 = 3|X_0 = 2) = \sum_{i=1}^{3} P(X_2 = 3|X_1 = i, X_0 = 2)P(X_1 = i|X_0 = 2) \\
&= \sum_{i=1}^{3} P(X_2 = 3|X_1 = i)P(X_1 = i|X_0 = 2) \\
&= \frac{1}{12} + 0 + 0 = \frac{1}{12}.
\end{aligned}
\tag{3}
$$

(d) The expectation is

$$
\mathrm{E}(X_2|X_0 = 2) = \sum_{i=1}^{3} iP(X_2 = i|X_0 = 2) = \frac{1}{2} + \frac{10}{12} + \frac{3}{12} = \frac{19}{12}.
$$

Problem 6

A fair coin is flipped repeatedly. We use H to denote "Head appeared" and T to denote the "Tail appeared".

(a) What is the expected number of flips until the pattern HTHT is observed?

(b) What is the expected number of flips until the pattern THTT is observed?

(c) What is the probability that pattern HTHT is observed earlier than THTT?

Solution

(a) Denote $t(\cdot)$ as the time for transferring from the current state to the ending state. The state space is $\{H, T, HT, HTH, HTHT\}$, and the transition relationship between is can be demonstrated as follows:

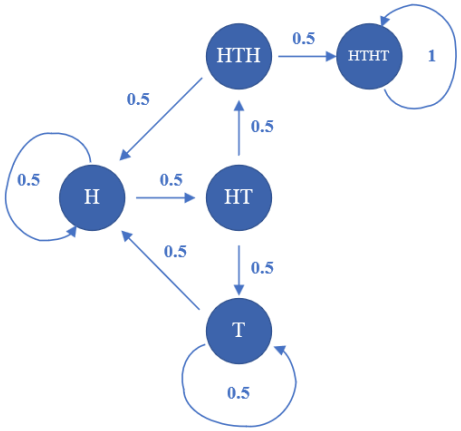


Figure 1: 6(1)

Then the expectation of step numbers for transferring from one state to the ending state can be listed as follows:

$$\begin{cases} E(t(H)) = \frac{1}{2}E(t(H)) + \frac{1}{2}E(t(HT)) + 1 \\ E(t(T)) = \frac{1}{2}E(t(T)) + \frac{1}{2}E(t(H)) + 1 \\ E(t(HT)) = \frac{1}{2}E(t(T)) + \frac{1}{2}E(t(HTH)) + 1 \\ E(t(HTH)) = \frac{1}{2}E(t(HTHT)) + \frac{1}{2}E(t(H)) + 1 \\ E(t(HTHT)) = 0 \end{cases}$$

Then we can obtain that $E(t(H)) = 18$ and $E(t(T)) = 20$. Therefore, the expected numbers of flips from starting is $\frac{1}{2}E(H) + \frac{1}{2}E(T) + 1 = 20$.

(b) Similarly, The state space is $\{H, T, HT, HTH, HTHT\}$, and the transition relationship between is can be demonstrated as follows: Then the expectation of step numbers for transferring from one state to the ending state can be listed as follows:

$$\begin{cases} E(t(T)) = \frac{1}{2}E(t(T)) + \frac{1}{2}E(t(TH)) + 1 \\ E(t(TH)) = \frac{1}{2}E(t(H)) + \frac{1}{2}E(t(THT)) + 1 \\ E(t(H)) = \frac{1}{2}E(t(H)) + \frac{1}{2}E(t(T)) + 1 \\ E(t(THT)) = \frac{1}{2}E(t(TH)) + \frac{1}{2}E(t(THTT)) + 1 \\ E(t(THTT)) = 0 \end{cases} \tag{4}$$

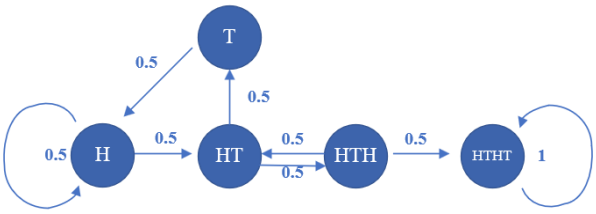


Figure 2: 6(2)

Then we can obtain that $E(t(H)) = 18$ and $E(t(T)) = 16$. Therefore, the expected numbers of flips from starting is $\frac{1}{2}E(H) + \frac{1}{2}E(T) + 1 = 18$.

(c) The state space in this problem can be conclude by $\{H, HT, HTH, HTHT, T, TH, THT, THTT\}$, and the relationship between each state can be demonstrated as follows:

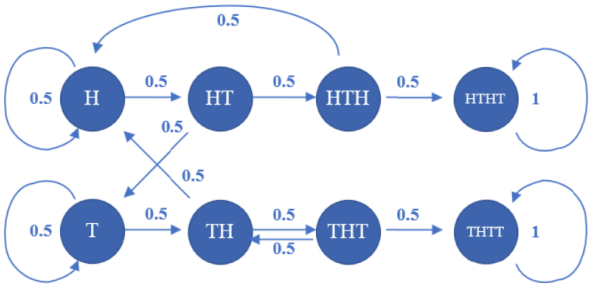


Figure 3: 6(3)

The relationship between the probability of each state finally ends up in $HTHT$ can be listed as follows:

$$\begin{cases} P(HTHT) = 1 \\ P(THTT) = 0 \\ P(HTH) = \frac{1}{2}P(HTHT) + \frac{1}{2}P(H) \\ P(THT) = \frac{1}{2}P(THTT) + \frac{1}{2}P(TH) \\ P(HT) = \frac{1}{2}P(HTH) + \frac{1}{2}P(T) \\ P(TH) = \frac{1}{2}P(THT) + \frac{1}{2}P(H) \\ P(H) = \frac{1}{2}P(H) + \frac{1}{2}P(HT) \\ P(T) = \frac{1}{2}P(T) + \frac{1}{2}P(TH). \end{cases} \tag{5}$$

Finally we can get that $P(H) = \frac{5}{7}$ and $P(T) = \frac{4}{7}$. Thus the probability of pattern $HTHT$ observed earlier than $THTT$ is $\frac{1}{2}P(H) + \frac{1}{2}P(T) = \frac{9}{14}$ by assuming the same initial state of entering states H and T.