Probability & Statistics for EECS: Homework #14

Due on May 21, 2023 at 23:59

Name: **Wang Yunfei**
Student ID: 2021533135

Problem 1

Solution

(a) Assume the indicator r.v. I_j represents "CATCAT" sequence appearing starting from j to j+5, $j \in \{1, \cdots, 110\}$.
Then the total number of the occurence of the expression "CATCAT" is $N = \sum_{j=1}^{110} I_j$.

$$\begin{aligned} E[N] &= E[\sum_{j=1}^{110} I_j] \\ &= \sum_{j=1}^{110} E[I_j] \\ &= \sum_{j=1}^{110} P(I_j = 1) \\ &= \sum_{j=1}^{110} (p_2 p_1 p_3 p_2 p_1 p_3) \\ &= 110(p_1 p_2 p_3)^2 \end{aligned}$$

(b) First because we treat $p_2 \sim Unif(0, 1)$, then we can get the prior distribution of p_2, which is $p_2 \sim Beta(1, 1)$.
And then we can use the data X we observed, which is "CAT".
From the story we have learned about Beta-Binomial conjugacy in the lesson, so we can seem A and T as failures, adding the latter parameter in the prior distribution. At the meanwhile, seenming C as a success, adding the former parameter in the prior distribution.
Then we can get the posterior distribution which is denoted as $p_2|X$, that is $p_2|X \sim Beta(2, 3)$.
Finally we can easily get the probability of next time the letter C appearing, which is $\frac{2}{5}$. (Bayesian average in nature)

Problem 2

a Assume I_j is the original index of the r.v. X_j^*.
Then by Adam's law, we can get,

$$E[X_j^*] = E[E[X_j^*|I_j]] = E[[E[X_{I_j}]]] = E[\mu] = \mu$$

By Eve's law, we can get,

$$\begin{aligned} Var[X_j^*] &= Var[E[X_j^*|I_j]] + E[Var[X_j^*|I_j]] \\ &= Var[E[X_{I_j}]] + E[Var[X_{I_j}]] \\ &= Var[\mu] + E[\sigma^2] \\ &= \sigma^2 \end{aligned}$$

b (1)$E[\overline{X^*}|X_1, \cdots, X_n]$

$$\begin{aligned} E[\overline{X^*}|X_1, \cdots, X_n] &= E[\frac{1}{n}(X_1^* + \cdots + X_n^*)|X_1, \cdots, X_n] \\ &= \frac{1}{n}\sum_{j=1}^{n} E[X_j^*|X_1, \cdots, X_n] \\ &= E[X_1^*|X_1, \cdots, X_n] \\ &\textit{By the hint} \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{n}X_j \\ &= \overline{X} \end{aligned}$$

$E[\overline{X^*}|X_1, \cdots, X_n] = \overline{X}$
(2)$Var[\overline{X^*}|X_1, \cdots, X_n]$

$$\begin{aligned} Var[\overline{X^*}|X_1, \cdots, X_n] &= Var[\frac{1}{n}(X_1^* + \cdots + X_n^*)|X_1, \cdots, X_n] \\ &= \frac{1}{n^2}\sum_{j=1}^{n} Var[X_j^*|X_1, \cdots, X_n] \\ &= \frac{1}{n^2}\sum_{j=1}^{n}[E[(X_j^*)^2|X_1, \cdots, X_n] - E[X_j^*|X_1, \cdots, X_n]^2] \\ &= \frac{1}{n^2}\sum_{j=1}^{n}(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X_n^2}) \end{aligned}$$

c By Adam's law, we can get,

$$E[\overline{X^*}] = E[E[\overline{X^*}|X_1, \cdots, X_n]] = E[\overline{X}] = \mu$$

By Eve's law, we can get. And $Var(\overline{X}) = \sigma^2/n$, $E[\overline{X}^2] = \frac{\sigma^2}{n} + \mu^2$, Then we can get,

$$\begin{aligned} Var[\overline{X^*}] &= E[Var[\overline{X^*}|X_1, \cdots, X_n]] + Var[E[\overline{X^*}|X_1, \cdots, X_n]] \\ &= E[\frac{1}{n^2}\sum_{j=1}^{n}(X_j - \overline{X})^2] + Var[\overline{X}] \\ &= \frac{(n-1)\sigma^2}{n^2} + \frac{\sigma^2}{n} \end{aligned}$$

d $\overline{X^*}$ has bigger uncertainty than the original r.v.s, because $X_1^*, \cdots, X_n^*$ are formed from X_j choosing randomly for $j \in \{1, \cdots, n\}$.
Then we have $Var(\overline{X}) < Var(\overline{X^*})$.

Problem 3

Solution

(a) (1)HT

First assume w_1 is the number of tosses until the "H" appears for the first time.
Assume w_2 is the number of tosses until the "T" appears for the first time.
Assume w_{HT} is the number of tosses until the "HT" appears for the first time.
Then we can easily get $w_1 \sim Fs(p)$, $w_2 \sim Fs(1-p)$, and $w_{HT} = w_1 + w_2$.
From the property of Fs distribution, finally we get the answer,

$$E[w_{HT}] = E[w_1 + w_2] = E[w_1] + E[w_2] = \frac{1}{p} + \frac{1}{1-p}$$

(2)HH

Assume O_1 is the outcome of the first toss, O_2 is the outcome of the second toss.
Assume w_{HH} is the number of the tosses until the "HH" appears for the first time.
Next, we can use LOTE to get the result,

$$\begin{aligned}
E[w_{HH}] &= E[w_{HH}|O_1 = H]P(O_1 = H) + E[w_{HH}|0_1 = T]P(O_1 = T) \\
\textit{Because of memeryless} \\
&= E[w_{HH}|O_1 = H]p + (E[w_{HH}] + 1)(1-p) \\
E[w_{HH}|O_1 = H] &= E[w_{HH}|O_1 = H, O_2 = H]P(O_2 = H|O_1 = H) \\
&\quad + E[w_{HH}|O_1 = H, O_2 = T]P(O_2 = T|O_1 = H) \\
\textit{Because } O_1 \textit{ and } O_2 \textit{ are independent} \\
E[w_{HH}|O_1 = H] &= E[w_{HH}|O_1 = H, O_2 = H]P(O_2 = H) + E[w_{HH}|O_1 = H, O_2 = T]P(O_2 = T) \\
E[w_{HH}|O_1 = H] &= E[w_{HH}|O_1 = H, O_2 = H]p + E[w_{HH}|O_1 = H, O_2 = T](1-p) \\
E[w_{HH}|O_1 = H] &= E[w_{HH}|O_1 = H, O_2 = H]p + E[w_{HH}|O_1 = H, O_2 = T](1-p) \\
\textit{Because of memeryless} \\
E[w_{HH}|O_1 = H] &= 2p + (E[w_{HH}] + 2)(1-p)
\end{aligned}$$

Then we can get the final expression,

$$\begin{aligned}
E[w_{HH}] &= p(2p + (E[w_{HH}] + 2)(1-p)) + (1-p)(E[w_{HH}] + 1) \\
&= \frac{1}{p} + \frac{1}{p^2}
\end{aligned}$$

(b) From the question, we need to calculate three parameters to get the corresponding answers in (a). They are $E[\frac{1}{p}]$, $E[\frac{1}{1-p}]$, $E[\frac{1}{p^2}]$.
And we already have $p \sim Beta(a, b)$.
(1)$E[\frac{1}{p}]$

$$\begin{aligned}
E[\frac{1}{p}] &= = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \frac{1}{p} p^{a-1}(1-p)^{b-1}\, dp \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a-1)\Gamma(b)}{\Gamma(a+b-1)} \\
&= \frac{a+b-1}{a-1}
\end{aligned}$$

(2)$E[\frac{1}{p^2}]$

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

(3)$E[\frac{1}{(1-p)}]$

$$\begin{aligned} E[\frac{1}{1-p}] = &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \frac{1}{1-p} p^{a-1}(1-p)^{b-1}\, dp \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a)\Gamma(b-1)}{\Gamma(a+b-1)} \\ &= \frac{a+b-1}{b-1} \end{aligned}$$

Then we can subtitude the value we get into the expression in (a) and get the final answer.

$$\begin{aligned} E[w_{HH}] &= \frac{a+b-1}{a-1} + \frac{(a+b-1)(a+b-2)}{(a-1)(a-2)} \\ E[w_{HT}] &= \frac{a+b-1}{a-1} + \frac{a+b-1}{b-1} \end{aligned}$$

Problem 4

Solution

(a) Let N be the total number of rolls we needed, and X_j represents the j_{th} roll number. Then we can use LOTE to have $E[N] = E[N|X_1 = 1]\frac{1}{6} + E[N|X_1 \neq 1]\frac{5}{6}$. And then because memoryless property we can simlify it and get the following expression.

$$E[N] = E[N|X_1 = 1]\frac{1}{6} + (E[N] + 1)\frac{5}{6}$$

And then for $E[N|X_1 = 1]$, we can also use LOTE,

$$E[N|X_1 = 1] = \sum_{i=1}^{6} E[N|X_1 = 1, X_2 = k]P(X_2 = k|X_1 = 1) = \frac{2}{6} + \frac{E[N|X_1 = 1] + 1}{6} + \frac{4(E[N] + 2)}{6}$$

Then we subtitude $E[N|X_1 = 1]$ into the expression of $E[N]$, finally we can get $E[N] = 36$.

(b) For this question, the concrete process is similar with (a), and what we need to change is the expression of $E[N|X_1 = 1]$.

$$E[N|X_1 = 1] = \frac{2}{6} + \frac{5(E[N] + 2)}{6}$$

And then we subtitude $E[N|X_1 = 1]$ into the expression of $E[N]$, finally we can get $E[N] = 42$.

(c) Assume X_n is the number of tosses until there are n continuous appearance of the same number.
And then on the basis of X_n, next time we have $\frac{1}{6}$ probability to get X_{n+1}, and with $\frac{5}{6}$ probability we need to get another n+1 continuous same number tosses, which can be denoted as $X_n + X_{n+1}$.
Therefore we can get the expression as follows,

$$a_{n+1} = \frac{1}{6}(a_n + 1) + \frac{1}{6}(a_n + a_{n+1})$$

Finally we can get the answer $a_{n+1} = 6a_n + 1$ and $a_1 = 1$, $n >= 1$.

(d) Then we can use the equation in (c) to get the answer.
$a_1 = 1$, $a_2 = 1 + 6 = 7$, $a_3 = 6 * 7 + 1 = 43$, $a_4 = 43 * 6 + 1 = 259$, $a_5 = 259 * 6 + 1 = 1555$, $a_6 = 1555 * 6 + 1 = 9331$, $a_7 = 9331 * 6 + 1 = 55987$.

Problem 5

Solution

(a) Intuitively, from the expression $y = ax + b$, we can directly get $x = \frac{1}{a}(y - b)$. So at the first sight, we may think $\frac{1}{a}$ is the slope of the best line for predicting X form Y.

(b) Construct $Cov(X, Y - cX)$, and from the properties of covariance, then we can have,

$$
\begin{aligned}
Cov(X, Y - cX) &= Cov(X, Y) - cCov(X, X) \\
\textit{Because they are standard normal distribution} \\
&= \rho - c
\end{aligned}
$$

And then let $Cov(X, Y - cX) = 0$. Thereby $\rho = c$.
Then we can define $V = Y - cX$.
Because X, Y are Bivariate Normal, then we can get any linear combination of X, V are normal distribution. So they are Bivariate Normal too.
Then V, X are independent of each other since $Cov(X, Y - cX) = 0$ and they are uncorrelated.
Therefore $V = Y - \rho X$, $c = \rho$.

(c) So in this question, we can get the correct answer by the same way in (b).
And the answer is $W = X - \rho Y$, and $d = \rho$.

(d) (1)$E[X|Y]$

$$E[X|Y] = E[\rho Y + W|Y] = E[\rho Y|Y] + E[W|Y] = \rho Y + E[W] = \rho Y$$

(2)$E[Y|X]$

$$E[Y|X] = E[\rho X + V|X] = E[\rho X|X] + E[V|X] = \rho X + E[V] = \rho X$$

(e) Then from the two expressions we get in (d), we can obviously get the correct answer, that is ρ is the slope of the best line for predicting X from Y, not the reciprocal.
That is because there exists symmetry property in Correlation, that is $Corr(X, Y) = Corr(Y, X) = \rho$.
By the way, take $\rho = 0$ as example. When $\rho = 0$, X, Y are independent of each other. Then we can get X given Y, that is $E[X] = 0$. But if we want to get it by inverting ρ, obviously this is wrong.