Probability & Statistics for EECS: Homework #014

Due on May 21, 2023 at 23:59

Name: **Penghao Wang**
Student ID: 2021533138

Problem 1

(a) As for the expected number of occurrences of the expression "CATCAT" we have that the probability of "CAT" occurs is that

$$p_1p_2p_3,$$

so the probability of "CATCAT" occurs is that

$$(p_1p_2p_3)^2,$$

so the expected number of occurrences of the expression "CATCAT" is that

$$(p_1p_2p_3)^2 \cdot 115 = 115(p_1p_2p_3)^2.$$

(b) Denote that n is the number of C observed in the sequence, then we have that as p_2 is a Unif(0, 1), then

$$p_2 \sim Beta(1, 1),$$

then with the Beta-Binomial conjugate prior, we have that

$$p_2|n \sim Beta(1 + n, 1 + N - n),$$

where N is the total observation times, so we have

$$p_2|n \sim Beta(2, 3)$$

Then as we want to calculate the probability that the next letter is C, we have that the probability is that

$$E(p_2|n).$$

So with the beta distribution, we have that

$$E(p_2|n) = \frac{2}{5}.$$

So given this information, the probability that the next letter is C is $\frac{2}{5}$.

Problem 2

(a) According to the method of the bootstrap sample, the X_j^* is generated by sampling with replacement with equal probabilities, so in the X_j^*, each X_j has equal probabilities to be selected. Then we have that as $X_1, ..., X_n$ i.i.d. r.v.s with mean μ and variance σ^2

$$E(X_j^*) = \frac{X_1 + X_2 + ... + X_n}{n} = \frac{n\mu}{n} = \mu.$$

$$Var(X_j^*) = \frac{Var(X_1 + X_2 + ... + X_n)}{n^2} = \frac{n^2 Var(X_1)}{n^2} = Var(X_1) = \sigma^2.$$

Which is the same for all the X_j^*, so we have that

$$E(X_j^*) = \mu, Var(X_j^*) = \sigma^2 \quad \text{for each j.}$$

(b) As for the $E(\bar{X}^*|X_1, ..., X_n)$ and the $Var(\bar{X}^*|X_1, ..., X_n)$, we have that

$$E(\bar{X}^*|X_1, ..., X_n) = \frac{1}{n}(E(X_1^*|X_1, ..., X_n) + ... + E(X_n^*|X_2, ..., X_n)) = E(X_1^*|X_1, ..., X_n).$$

As we have that conditional on $X_1, ..., X_n$, the X_j^* , $\forall j \in 1, ..., n$ are independent, so we have that

$$E(X_1^*|X_1, ..., X_n) = \bar{X}.$$

As for the variance, we have that

$$Var(\bar{X}^*|X_1, ..., X_n) = \frac{n}{n^2} Var(X_1^*|X_1, ..., X_n) = \frac{\sum_{j=1}^n (X_j - \bar{X})^2}{n^2}.$$

(c) With Adam's law, we have that

$$E(\bar{X}^*) = E(E(\bar{X}^*|X_1, ..., X_n)) = E(\bar{X}) = \mu.$$

With Eve's law, we have that

$$\begin{aligned}
Var(\bar{X}^*) &= E(Var(\bar{X}^*|X_1, ..., X_n)) + Var(E(\bar{X}^*|X_1, ..., X_n)) \\
&= E(\frac{\sum_{j=1}^n (X_j - \bar{X})^2}{n^2}) + Var(\bar{X}) \quad \text{(from question b)} \\
&= \frac{(n-1)\sigma^2}{n^2} + \frac{\sigma^2}{n} \\
&= \frac{2n-1}{n^2}\sigma^2.
\end{aligned}$$

(d) To intuitively explain $Var(\bar{X}) < Var(\bar{X}^*)$, we have that this is due to the different sampling method, as $\bar{X}^*$ is the that we sample with replacement, which means that each X_j^* may be chosen more than once, while the $\bar{X}$ is just the mean of $X_1, ..., X_n$, so the X_j^* will have more complex situation which will make the variance greater. So the variance of $\bar{X}^*$ is larger than that of $\bar{X}$, that is $Var(\bar{X}^*) > Var(\bar{X})$.

Problem 3

(a) By solving the problems using conditional expectation. With LOTE and first step analysis, we denote that O_1 is the outcome of the first toss, O_2 is the outcome of the second toss. The probability of a toss head is that p, and the probability of a toss tail is that q, where $p+q=1$.

1. HT

$$E(W_{HT}) = E(W_{HT}|O_1 = H)P(O_1 = H) + E(W_{HT}|O_1 = T)P(O_1 = T) \quad \text{with LOTE.}$$

Then as for the $E(W_{HT}|O_1 = H)$, we have that

$$\begin{aligned} E(W_{HT}|O_1 = H) &= E(W_{HT}|O_1 = H, O_2 = H)P(O_2 = H|O_1 = H) \\ &\quad + E(W_{HT}|O_1 = H, O_2 = T)P(O_2 = T|O_1 = H) \\ &= (1 + E(W_{HT}|O_1 = H))p + 2q \quad \text{, as } O_1 \text{ and } O_2 \text{ are independent.} \end{aligned}$$

Then we get that $E(W_{HT}|O_1 = H) = \dfrac{p+2q}{1-p} = \dfrac{2-p}{1-p}$.
Then as for the $E(W_{HT}|O_1 = T)$, we have that

$$E(W_{HT}|O_1 = T) = 1 + E(W_{HT}).$$

So we have that

$$\begin{aligned} E(W_{HT}) &= E(W_{HT}|O_1 = H)P(O_1 = H) + E(W_{HT}|O_1 = T)P(O_1 = T) \\ &= (\frac{2-p}{1-p})p + (1 + E(W_{HT}))q \\ &= \frac{2p-p^2}{1-p} + q + qE(W_{HT}) \\ &= \frac{2p-p^2}{1-p} + 1 - p + (1-p)E(W_{HT}) \end{aligned}$$

So we get that

$$E(W_{HT}) = \frac{\frac{2p-p^2}{1-p} + 1 - p}{p} = \frac{1}{(1-p)p} = \frac{1}{p} + \frac{1}{1-p}.$$

2. HH

$$E(W_{HH}) = E(W_{HH}|O_1 = H)P(O_1 = H) + E(W_{HH}|O_1 = T)P(O_1 = T) \quad \text{with LOTE.}$$

Then as for the $E(W_{HH}|O_1 = H)$, we have that

$$\begin{aligned} E(W_{HH}|O_1 = H) &= E(W_{HH}|O_1 = H, O_2 = H)P(O_2 = H|O_1 = H) \\ &\quad + E(W_{HH}|O_1 = H, O_2 = T)P(O_2 = T|O_1 = H) \\ &= 2p + (2 + E(W_{HH}))q \quad \text{, as } O_1 \text{ and } O_2 \text{ are independent.} \end{aligned}$$

Then as for the $E(W_{HH}|O_1 = T)$, we have that

$$E(W_{HH}|O_1 = T) = 1 + E(W_{HH}).$$

So we have that

$$\begin{aligned} E(W_{HH}) &= E(W_{HH}|O_1 = H)P(O_1 = H) + E(W_{HH}|O_1 = T)P(O_1 = T) \\ &= (2p + (2 + E(W_{HH}))q)p + (1 + E(W_{HH}))q \\ &= 2p^2 + 2pq + pqE(W_{HH}) + q + qE(W_{HH}) \end{aligned}$$

So we get that

$$E(W_{HH}) = \frac{2p^2 + 2pq + q}{1 - pq - q} = \frac{p+1}{p^2} = \frac{1}{p} + \frac{1}{p^2}.$$

So we get that

$$E(W_{HT}) = \frac{1}{p} + \frac{1}{1-p} \quad \text{and that} \quad E(W_{HH}) = \frac{1}{p} + \frac{1}{p^2}.$$

(b) Suppose that p is unknown, with a Beta(a, b) prior, we have that

$$E(W_{HT}) = E(E(W_{HT}|p)) = E(\frac{1}{p}) + E(\frac{1}{1-p}).$$

$$E(W_{HH}) = E(E(W_{HH}|p)) = E(\frac{1}{p}) + E(\frac{1}{p^2}).$$

With LOTUS, as for the $E(\frac{1}{p})$, we have that

$$E(\frac{1}{p}) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 p^{a-2}(1-p)^{b-1}dp = \frac{\Gamma(a+b)\Gamma(a-1)\Gamma(b)}{\Gamma(a)\Gamma(b)\Gamma(a+b-1)} = \frac{a+b-1}{a-1}$$

$$E(\frac{1}{1-p}) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 p^{a-1}(1-p)^{b-2}dp = \frac{\Gamma(a+b)\Gamma(a)\Gamma(b-1)}{\Gamma(a)\Gamma(b)\Gamma(a+b-1)} = \frac{a+b-1}{b-1}$$

$$E(\frac{1}{p^2}) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 p^{a-3}(1-p)^{b-1}dp = \frac{\Gamma(a+b)\Gamma(a-2)\Gamma(b)}{\Gamma(a)\Gamma(b)\Gamma(a+b-2)} = \frac{(a+b-1)(a+b-2)}{(a-1)(a-2)}$$

Then we take them back, we have that

$$E(W_{HT}) = \frac{a+b-1}{a-1} + \frac{a+b-1}{b-1}$$

$$E(W_{HH}) = \frac{a+b-1}{a-1} + \frac{(a+b-1)(a+b-2)}{(a-1)(a-2)}$$

Problem 4

(a) Denote that N is the number of rolls, O_1 is the first roll values. Then we again use conditional expectation and first step analysis, we have that

$$E(N) = E(N|O_1 = 1)P(O_1 = 1) + E(N|O_1! = 1)P(O_1! = 1).$$

Then as for the $E(N|O_1 = 1)$, we have that

$$\begin{aligned} E(N|O_1 = 1) &= \sum_{i=1}^{6} E(N|O_1 = 1, O_2 = i)P(O_2 = i|O_1 = 1) \\ &= \frac{2}{6} + \frac{1 + E(N|O_1 = 1)}{6} + \frac{4(2 + E(N))}{6}. \end{aligned}$$

From where we get that

$$E(N|O_1 = 1) = \frac{2 + 1 + 8 + 4E(N)}{5} = \frac{11 + 4E(N)}{5}.$$

Then as for the $E(N|O_1! = 1)$, we have that

$$E(N|O_1! = 1) = 1 + E(N).$$

So we have that

$$\begin{aligned} E(N) &= E(N|O_1 = 1)P(O_1 = 1) + E(N|O_1! = 1)P(O_1! = 1) \\ &= \frac{11 + 4E(N)}{5}\frac{1}{6} + (1 + E(N))\frac{5}{6} \\ &= \frac{11}{30} + \frac{4}{30}E(N) + \frac{5}{6} + \frac{5}{6}E(N) \\ &= \frac{36}{30} + \frac{29}{30}E(N) \end{aligned}$$

we have that $E(N) = 36$.

(b) As for the consecutive 1's, it is the same with (a), we denote that N is the number of rolls, O_1 is the first roll values. Then we again use conditional expectation and first step analysis, we have that

$$E(N) = E(N|O_1 = 1)P(O_1 = 1) + E(N|O_1! = 1)P(O_1! = 1).$$

Then as for the $E(N|O_1 = 1)$, we have that

$$\begin{aligned} E(N|O_1 = 1) &= \sum_{i=1}^{6} E(N|O_1 = 1, O_2 = i)P(O_2 = i|O_1 = 1) \\ &= \frac{2}{6} + \frac{5(2 + E(N))}{6} \\ &= 2 + \frac{5E(N)}{6}. \end{aligned}$$

Then as for the $E(N|O_1! = 1)$, we have that

$$E(N|O_1! = 1) = 1 + E(N).$$

So we have that

$$\begin{aligned} E(N) &= E(N|O_1 = 1)P(O_1 = 1) + E(N|O_1! = 1)P(O_1! = 1) \\ &= (2 + \frac{5E(N)}{6})\frac{1}{6} + (1 + E(N))\frac{5}{6} \\ &= \frac{7}{6} + \frac{35}{36}E(N) \end{aligned}$$

we have that $E(N) = 42$.

(c) As for get the same values n times in a row, we have that as a_n is the expected number of rolls to get the same values n times in a row, we have that there are mainly 2 cases, that is the the next roll after a_n is the same with the previous one, or not.
If the next roll is still value j, this cases is of probability $\frac{1}{6}$,
If the next roll is not value j, the roll need to re-roll, and this cases is of probability $\frac{5}{6}$.
We have that
$$a_{n+1} = \frac{a_n + 1}{6} + \frac{5}{6}(a_n + a_{n+1}),$$
which can be simplified to
$$a_{n+1} = 6a_n + 1.$$

(d) As for the formula for a_n for all $n \geq 1$, we could use the recursive formula from (c). We have that
$$a_1 = 1, \quad a_2 = 1 + 6, \quad a_3 = 1 + 6 + 6^2, ...$$
So we have that
$$a_n = 1 + 6 + 6^2 + ... + 6^{n-1} = \frac{1 - 6^n}{1 - 6} = \frac{6^n - 1}{5}.$$
As for the a_7, we have that
$$a_7 = \frac{6^7 - 1}{5} = 55987.$$

Problem 5

(a) As $y = ax + b$ is the equation of the best line for predicting Y from X, then as now we want to use Y to predict X, as we have that $x = \dfrac{y-b}{a}$, we then have that an intuitive guess of the slope is $\dfrac{1}{a}$, which is the revert.

(b) As for a constant c and an r.v. V such that $Y = cX + V$, we firstly set that

$$Cov(X, Y - cX) = Cov(X, Y) - cCov(X, X) = \rho - c = 0,$$

we then have that $c = \rho$, then we define $V = Y - cX$ as we have $Y = cX + V$. Then we have that (X, V) is Bivariate Normal, so we get that X and V are independent. Then we get that the constant is that

$$c = \rho.$$

and the r.v. V is that

$$V = Y - \rho X.$$

(c) As for a constant d and an r.v. W such that $X = dY + W$, we firstly set that

$$Cov(X - dY, Y) = Cov(X, Y) - dCov(Y, Y) = \rho - d = 0,$$

we then have that $d = \rho$, then we define $W = X - dY$ as we have $X = dY + W$. Then we have that (Y, W) is Bivariate Normal, so we get that Y and W are independent. Then we get that the constant is that

$$d = \rho.$$

and the r.v. W is that

$$W = X - \rho Y.$$

(d) As for the $E(Y|X)$ and $E(X|Y)$, we have that as $V = Y - \rho X$ and $W = X - \rho Y$, we have that

$$E(Y|X) = E(\rho X + V|X) = \rho E(X|X) + E(V|X) = \rho X + E(V) = \rho X.$$

$$E(X|Y) = E(\rho Y + W|Y) = \rho E(Y|Y) + E(W|Y) = \rho Y + E(W) = \rho Y.$$

(e) Due to the symmetry of the correlation coefficient, the slope of the best linear equation for predicting X from Y is the same as the slope for predicting Y from X, not the reciprocal. This is due to the phenomenon of regression toward the mean, which involves a trade-off between the father and son. If the correlation coefficient is close to 1, we give more weight to the father's height; if the correlation coefficient is close to 0, we give more weight to the population mean. This is because height is not entirely hereditary, and by considering regression toward the mean, we obtain a more reasonable prediction. Therefore, the best guess for X given Y is $E(X) = 0$.