Probability & Statistics for EECS: Homework # Solution

Problem 1

*BH CH8 53

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) 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 $j \in \{1, 2, \dots, 110\}$. 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) 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 2

*BH CH9 34

Let $X_1, \ldots, X_n$ be i.i.d. r.v.s with mean μ and variance σ^2, and $n \geq 2$. A bootstrap sample of $X_1, \ldots, X_n$ is a sample of n r.v.s $X_1^*, \ldots, X_n^*$ formed from the $X_j, \forall j \in \{1, \ldots, n\}$ by sampling with replacement with equal probabilities. Let $\bar{X}^*$ denote the sample mean of the bootstrap sample:

$$\bar{X}^* = \frac{1}{n}(X_1^* + \cdots + X_n^*).$$

(a) Calculate $E(X_j^*)$ and $\text{Var}(X_j^*)$ for each $j \in \{1, \ldots, n\}$.

(b) Calculate $E(\bar{X}^* | X_1, \ldots, X_n)$ and $\text{Var}(\bar{X}^* | X_1, \ldots, X_n)$.

Hint: Conditional on $X_1, \ldots, X_n$, the $X_j^*, \forall j \in \{1, \ldots, n\}$ are independent, with a PMF that puts probability $1/n$ at each of the points $X_1, \ldots, X_n$. As a check, your answers should be random variables that are functions of $X_1, \ldots, X_n$.

(c) Calculate $E(\bar{X}^*)$ and $\text{Var}(\bar{X}^*)$.

(d) Explain intuitively why $\text{Var}(\bar{X}) < \text{Var}(\bar{X}^*)$.

Solution

(a) Define random variable I_j that marks what index $i \in \{1, \ldots, n\}$ is the actual value of j, i.e.,

$$X_j^* \mid (I_j = i) = X_i.$$

Conditioning on I_j, we get that

$$E(X_j^*) = E(E(X_j^* \mid I_j)) = E(E(X_{I_j})) = E(\mu) = \mu$$

and

$$\begin{aligned}\text{Var}\left(X_j^*\right) =& \text{Var}\left(E\left(X_j^* \mid I_j\right)\right) + E\left(\text{Var}\left(X_j^* \mid I_j\right)\right) \\ =& \text{Var}\left(E\left(X_{I_j}\right)\right) + E\left(\text{Var}\left(X_{I_j}\right)\right) \\ =& \text{Var}(\mu) + E\left(\sigma^2\right) = \sigma^2.\end{aligned}$$

(b) Using the given hint, we have that

$$\begin{aligned}E\left(\bar{X}^* \mid X_1, \ldots, X_n\right) =& E\left(\frac{1}{n}\left(X_1^* + \cdots + X_n^*\right) \mid X_1, \ldots, X_n\right) \\ =& \frac{1}{n}\sum_{j=1}^{n} E\left(X_j^* \mid X_1, \ldots, X_n\right) \\ =& \frac{1}{n}\sum_{j=1}^{n}\frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n,\end{aligned}$$

where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Problem 2 continued on next page...

Similarly, we can obtain the required variance

$$\begin{aligned}\operatorname{Var}\left(\bar{X}^* \mid X_1, \ldots, X_n\right) =&\operatorname{Var}\left(\frac{1}{n}\left(X_1^*+\cdots+X_n^*\right) \mid X_1, \ldots, X_n\right) \\ =&\frac{1}{n^2} \sum_{j=1}^n \operatorname{Var}\left(X_j^* \mid X_1, \ldots, X_n\right) \\ =&\frac{1}{n^2} \sum_{j=1}^n\left(\frac{1}{n} \sum_{i=1}^n X_i^2-\bar{X}_n^2\right) \\ =&\frac{1}{n}\left(\frac{1}{n} \sum_{i=1}^n X_i^2-\bar{X}_n^2\right) \\ =&\frac{1}{n^2} \sum_{i=1}^n X_i^2-\frac{1}{n} \bar{X}_n^2,\end{aligned}$$

where we have used that

$$\begin{aligned}\operatorname{Var}\left(X_j^* \mid X_1, \ldots, X_n\right) =&E\left(\left(X_j^*\right)^2 \mid X_1, \ldots, X_n\right)-E\left(X_j^* \mid X_1, \ldots, X_n\right)^2 \\ =&\frac{1}{n} \sum_{i=1}^n X_i^2-\bar{X}_n^2.\end{aligned}$$

(c) Use Adam's law (conditioning on $X_1, \ldots, X_n$) to obtain the required mean

$$E\left(\bar{X}^*\right)=E\left(E\left(\bar{X}^* \mid X_1, \ldots, X_n\right)\right)=E\left(\bar{X}_n\right)=\mu,$$

and Eve's law to obtain the variance

$$\begin{aligned}\operatorname{Var}\left(\bar{X}^*\right) =&\operatorname{Var}\left(E\left(\bar{X}^* \mid X_1, \ldots, X_n\right)\right)+E\left(\operatorname{Var}\left(\bar{X}^* \mid X_1, \ldots, X_n\right)\right) \\ =&\operatorname{Var}\left(\bar{X}_n\right)+E\left(\frac{1}{n^2} \sum_{i=1}^n X_i^2-\frac{1}{n} \bar{X}_n^2\right) \\ =&\frac{\sigma^2}{n}+\frac{1}{n^2} \sum_{i=1}^n E\left(X_i^2\right)-\frac{1}{n} E\left(\bar{X}_n^2\right) \\ =&\frac{\sigma^2}{n}+\frac{1}{n^2} \sum_{i=1}^n E\left[\operatorname{Var}\left(X_i\right)+\left[E\left(X_i\right)\right]^2\right]-\frac{1}{n}\left[\operatorname{Var}\left(\bar{X}_n\right)+\left[E\left(\bar{X}_n\right)\right]^2\right] \\ =&\frac{\sigma^2}{n}+\frac{1}{n^2} \sum_{i=1}^n\left(\sigma^2+\mu^2\right)-\frac{1}{n}\left(\frac{\sigma^2}{n}+\mu^2\right) \\ =&\frac{(2 n-1) \sigma^2}{n^2}.\end{aligned}$$

(d) Observe that we have that

$$\operatorname{Var}(\bar{X})=\frac{(2 n-1) \sigma^2}{n^2}<\frac{2 \sigma^2}{n}=\operatorname{Var}\left(\bar{X}^*\right).$$

This is intuitively because of the fact that bootstrap sampling gives another touch of uncertainty to the starting basic uncertainty of variables $X_1, \ldots, X_n$.

Problem 3

*BH CH9 10
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 4

*BH CH9 12
A fair 6-sided die is rolled repeatedly.

(a) Find the expected number of rolls needed to get a 1 followed right away by a 2.

Hint: Start by conditioning on whether or not the first roll is a 1.

(b) Find the expected number of rolls needed to get two consecutive 1's.

(c) Let a_n be the expected number of rolls needed to get the same value n times in a row (*i.e.*, to obtain a streak of n consecutive j's for some not-specified-in-advance value of j). Find a recursive formula for a_{n+1} in terms of a_n.

Hint: Divide the time until there are $n+1$ consecutive appearances of the same value into two pieces: the time until there are n consecutive appearances, and the rest.

(d) Find a simple, explicit formula for an for all $n \geq 1$. What is a_7 (numerically)?

Solution

(a) Let N be the number of rolls needed to get a 1 followed right away by a 2. Denote the outcome of the nth roll as X_n. By the law of total expectation,

$$E(N) = E(N|X_1 = 1)P(X_1 = 1) + E(N|X_1 \neq 1)P(X_1 \neq 1) = \frac{1}{6}E(N|X_1 = 1) + \frac{5}{6}E(N|X_1 \neq 1).$$

in which $E(N|X_1 \neq 1) = E(N) + 1$ and

$$\begin{aligned} E(N|X_1 = 1) =& E(N|X_1 = 1, X_2 = 2)P(X_2 = 2|X_1 = 1) \\ &+ E(N|X_1 = 1, X_2 = 1)P(X_2 = 1|X_1 = 1) \\ &+ E(N|X_1 = 1, X_2 \neq 2, X_2 \neq 1)P(X_2 \neq 2, X_2 \neq 1|X_1 = 1). \end{aligned}$$

Since X_1 and X_2 are independent, we have

$$\begin{aligned} E(N|X_1 = 1) =& E(N|X_1 = 1, X_2 = 1)P(X_2 = 1) + E(N|X_1 = 1, X_2 = 1)P(X_2 = 1) \\ &+ E(N|X_1 = 1, X_2 \neq 2, X_2 \neq 1)P(X_2 \neq 2, X_2 \neq 1) \\ =& \frac{1}{6}E(N|X_1 = 1, X_2 = 2) + \frac{1}{6}E(N|X_1 = 1, X_2 = 1) + \frac{2}{3}E(N|X_1 = 1, X_2 \neq 2, X_2 \neq 1) \end{aligned}$$

in which $E(N|X_1 = 1, X_2 = 2) = 2$, $E(N|X_1 = 1, X_2 = 1) = E(N|X_1 = 1) + 1$, $E(N|X_1 = 1, X_2 \neq 1, X_2 \neq 2) = E(N) + 2$. It follows that

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

from which we get $E(N|X_1 = 1) = \frac{4}{5}E(N) + \frac{11}{5}$. Finally, we have

$$E(N) = \frac{1}{6}(\frac{4}{5}E(N) + \frac{11}{5}) + \frac{5}{6}(E(N) + 1)$$

which yields $E(N) = 36$.

(b) Let M be the number of rolls needed to get two consecutive 1's. By the law of total expectation,

$$E(M) = E(M|X_1 = 1)P(X_1 = 1) + E(M|X_1 \neq 1)P(X_1 \neq 1) = \frac{1}{6}E(M|X_1 = 1) + \frac{5}{6}E(M|X_1 \neq 1).$$

in which $E(M|X_1 \neq 1) = E(M) + 1$ and by low of total expectation and the independence between X_1 and X_2

$$\begin{aligned}E(M|X_1 = 1) =& E(M|X_1 = 1, X_2 = 1)P(X_2 = 1|X_1 = 1) + E(M|X_1 = 1, X_2 \neq 1)P(X_2 \neq 1|X_1 = 1) \\ =& E(M|X_1 = 1, X_2 = 1)P(X_2 = 1) + E(M|X_1 = 1, X_2 \neq 1)P(X_2 \neq 1) \\ =& \frac{1}{6}E(M|X_1 = 1, X_2 = 1) + \frac{5}{6}E(M|X_1 = 1, X_2 \neq 1)\end{aligned}$$

in which $E(M|X_1 = 1, X_2 = 1) = 2$, $E(M|X_1 = 1, X_2 \neq 1) = E(M) + 2$. It follows that

$$E(M|X_1 = 1) = \frac{1}{6} \times 2 + \frac{5}{6}(E(M) + 2) = \frac{5}{6}E(M) + 2.$$

Finally, we have

$$E(M) = \frac{1}{6}(\frac{5}{6}E(M) + 2) + \frac{5}{6}(E(M) + 1)$$

which yields $E(M) = 42$.

(c) Let N_n be the number of roller needed to get the same value n times in a row. By the law of total expectation,

$$\begin{aligned}E(N_{n+1}|N_n) =& E(N_{n+1}|N_n, X_{N_n+1} = X_{N_n})P(X_{N_n+1} = X_{N_n}|N_n) \\ &+ E(N_{n+1}|N_n, X_{N_n+1} \neq X_{N_n})P(X_{N_n+1} \neq X_{N_n}|N_n) \\ =& \frac{1}{6}E(N_{n+1}|N_n, X_{N_n+1} = X_{N_n}) + \frac{5}{6}E(N_{n+1}|N_n, X_{N_n+1} \neq X_{N_n})\end{aligned}$$

Since $E(N_{n+1}|N_n, X_{N_n+1} = X_{N_n}) = N_n + 1$, $E(N_{n+1}|N_n, X_{N_n+1} \neq X_{N_n}) = E(N_{n+1}|N_n) + N_n$, we have

$$E(N_{n+1}|N_n) = \frac{1}{6}(N_n + 1) + \frac{5}{6}(E(N_{n+1}|N_n) + N_n).$$

Taking expectation of both sides, and using Adam's law, we get

$$E(N_{n+1}) = \frac{1}{6}(E(N_n) + 1) + \frac{5}{6}(E(N_{n+1}) + E(N_n))$$

$$\Rightarrow a_{n+1} = \frac{1}{6}(a_n + 1) + \frac{5}{6}(a_{n+1} + a_n)$$

which yields $a_{n+1} = 6a_n + 1$.

(d) Let $b_n = a_n + \frac{1}{5}$, then we have $b_{n+1} = 6b_n$ and $b_1 = a_1 + \frac{1}{5} = 1 + \frac{1}{5} = \frac{6}{5}$. Thus for $n \geq 1$,

$$b_n = \frac{6}{5} \times 6^{n-1} = \frac{1}{5} \times 6^n \Rightarrow a_n = b_n - \frac{1}{5} = \frac{1}{5}(6^n - 1).$$

When $n = 7$, $a_7 = \frac{1}{5}(6^7 - 1)$.

Problem 5

*BH CH9 18

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

(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 $\operatorname{Cov}(X, Y - cX) = 0$.

(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 $E(Y|X)$ and $E(X|Y)$.

(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

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

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

(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.