CSE 544, Spring 2022, Probability and Statistics for Data Science

Assignment 2: Random Variables Due: 2/21, 8:15pm, via Blackboard

(7 questions, 70 points total)

I/We understand and agree to the following:

(a) Academic dishonesty will result in an 'F' grade and referral to the Academic Judiciary.

(b) Late submission, beyond the 'due' date/time, will result in a score of 0 on this assignment.

(write down the name of all collaborating students on the line below)

---

1. Introduction to Covariance (Total 5 points)

The covariance of two RVs X and Y is defined as: Cov(X,Y) = E[(X - E[X]) (Y - E[Y])] = E[XY] - E[X] E[Y].

Covariance of independent RVs is always zero.

A random variable X has mean 0 and variance 1. A random variable Y has mean 1 and variance 2.

Suppose random variables U and V are given by $U = X + \alpha Y$, $V = X - Y$.

(a) What is the covariance, Cov(U, V)? (3 points)

(b) For what value of α, if any, could U and V be independent? (2 points)

2. Inequalities (Total 10 points)

Let X be a non-negative RV with mean μ and variance σ^2, and let t > 0 be some real number.

(a) Prove that $E[X] \geq \int_t^\infty xf(x)dx$ (3 points)

(b) Using part (a), prove that $Pr\ (X > t)\ \leq \frac{E[X]}{t}$ (3 points)

(c) Using part (b), prove that $Pr\ (|X - \mu| \geq t)\ \leq \frac{\sigma^2}{t^2}$ (4 points)

3. Functions of RVs **(Total 16 points)**

(a) Let $X_1, X_2, \dots, X_k$ be k independent exponential random variables with pdfs given by $f_{X_i}(x) = \lambda_i e^{-\lambda_i x},\ x \geq 0,\ \forall\, i \in \{1, 2, \dots, k\}$. Let $Z = min\,(X_1, X_2, \dots, X_k)$.

i. Find the pdf of Z. (3 points)

ii. Find E[Z]. (1 point)

iii. Find Var(Z). (2 points)

(b) Let X and Y be jointly continuous random variables with joint PDF given by:

$$f_{x,y}(x,y) = \begin{cases} cx + 1, & x, y \geq 0, x + y < 1 \\ 0, & otherwise \end{cases}$$

i. Find the constant c (2 points)

ii. Find $\Pr(Y<2X^2)$ (2 points)

iii. Find $\Pr(Y<2X^2)$ numerically by randomly sampling 100,000 points in the non-zero region of the pdf. Does this value match with the theoretical value in (ii)? Why/why not? (2 points)
(Hint: Sample points uniformly by first randomly sampling an x in [0,1] then randomly sampling a y for that x in [0,1- x]. Calculate $P(Y<2X^2)$ by looking at the fraction of sampled points that meet the requirements of the event, i.e., $Y<2X^2$.)

iv. Find $\Pr(Y<2X^2 \mid X=k)$ (1 point)

v. For each of k=0.25, 0.5, 0.75, randomly sample 100,000 points in the non-zero region of the PDF and calculate $\Pr(Y<2X^2 \mid X=k)$ numerically using the same method as in part iii. Do the numerical values match the answer given by part (iv) for that X=k? Why/why not? (3 points)
(Hint: Sample points for a given X=k by randomly sampling a y in [0,1-k])

For parts (iii) and (v) above, submit your code along with your solution as part of the zip/tar file on BB. Name your python file a2_3b.py.

4. Daenerys returns to King's Landing, almost. (Total 9 points)

In an alternate universe of Game of Thrones (or A Song of Ice and Fire, for fans of the books), Daenerys Targaryen is finally ready to leave Meereen and return to King's Landing. However, she does not know the way. From Meereen, if she goes East, she will wander around for 40 days in the Shadow Lands and return back to Meereen. If she goes West from Meereen, she will immediately arrive at the city of Mantarys. From Mantarys , she can go West by road or South via ship. If she goes South, her ship will get lost in the Smoking Sea and will be swept back to Meereen after 20 days. However, if she goes West from Mantarys, she will eventually reach King's Landing in 10 days. Let X denote the time spent by Daenerys before she reaches King's Landing. Assume that she is equally likely to take either of two paths whenever presented with a choice and has no memory of prior choices.

(a) What is E[X]? (2 points)

(b) What is Var[X]? (7 points)

(Hint: Be careful with Var[X]. You want to use conditioning.)

5. Dependence on past 2 states (Total 13 points)

Consider the Clear-Snowy problem from class. However, this time, assume that the weather tomorrow depends on the weather today AND the weather yesterday. While this does not seem to follow the Markovian property, you can modify the state space to work around this issue. Use the following notation and transition probability values:

Pr[Weather tomorrow is X_{i+1}, given that weather today is X_i and weather yesterday was X_{i-1}]
= Pr[X_{i+1} | X_i X_{i-1}] (note that each X is either *c* or *s*).
Pr[c | c c] = 0.8; Pr[c | c s] = 0.7; Pr[c | s c] = 0.4; Pr[c | s s] = 0.1.

(a) Find the eventual (steady-state) Pr[c c], Pr[c s], Pr[s c], and Pr[s s]. Show your Markov chain and the transition probabilities. (5 points)

(b) In steady-state, what is the probability that it will be snowy 3 days from today. (3 points)

(c) Solve the problem of finding the steady state probability via simulation (in python). You need to find the steady state by raising the transition matrix to a high power ($\pi = P^k; k \gg 1$) and then take any row of the exponentiated matrix ($\pi[i,:]$) as the steady state. For taking power of matrix in python, you can use **np.linalg.matrix_power(matrix, power**). After you obtain the steady state distribution, solve part (b) numerically. (5 points)

Submit your code along with your solution as part of the zip/tar file on BB. Name your python file a2_5.py. The script should have a function **a □ `steady_state_power` (transition matrix)**, where **`steady_state_power` ()** should have the implementation of Power method and the return value **a** is the final steady state. Also, in the hardcopy submission, you should mention the final steady state you obtained in the following format:

`Steady_State:` `Power iteration >> [xx, xx, xx, xx]`

6. Heinz Ketchup **(Total 8 points)**

At the Heinz ketchup factory, the amounts which go into bottles of ketchup are supposed to be normally distributed with mean 36 oz. and standard deviation 0.1 oz. Once every 30 minutes a bottle is selected from the production line, and its contents are noted precisely. If the amount of the bottle goes below 35.8 oz. or above 36.2 oz., then the bottle will be declared out of control.

(a) If the process is in control, meaning $\mu = 36$ oz. and $\sigma = 0.1$ oz., find the probability that a bottle will be declared out of control. (2 points)

(b) In the situation of (a), find the probability that the number of bottles found out of control in an eight-hour day (16 inspections) will be zero. (2 points)

(c) In the situation of (a), find the probability that the number of bottles found out of control in an eight-hour day (16 inspections) will be exactly one. (1 point)

(d) If the process shifts so that $\mu = 37$ oz and $\sigma = 0.4$ oz, find the probability that a bottle will be declared out of control. (3 points)

7. Pokémon Go fanatic (Total 9 points)

Let us assume there are only *n* distinct types of Pokémons to capture in the entire Pokémon world, though there is an infinite supply of each type. Every day, you capture exactly one Pokémon. The Pokémon that you capture could be any one of the *n* types of Pokémons with equal probability. Your goal is to capture at least one Pokémon of all *n* distinct types. Let X denote the number of days needed to complete your goal.

(a) What is E[X]? (4 points)

(b) What is Var[X]? (5 points)

We do not need closed-forms here for parts (a) and (b).