CSE 544, Spring 2021, Probability and Statistics for Data Science

Assignment 2: Random Variables Due: 3/09, 1:15pm, 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) In an experiment, an unbiased/fair coin is flipped 3 times. Let X be the number of heads in the first two flips and Y be the number of heads in the last two flips. Calculate Cov(X,Y). (2 points)

(b) Let X be a fair 5-sided dice with face values {-5, -2, 0, 2, 5}. Let $Y = X^2$. Calculate Cov(X,Y). (2 points)

(c) Does a zero covariance imply that the RVs are independent? Justify your answer. (1 point)

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 10 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 two random variables with joint density function:

$f_{XY}(x, y) = \begin{cases} 2, & 0 \leq x \leq y \leq 1 \\ 0, & otherwise \end{cases}$. Find the pdf of $Z = XY$. (4 points)

4. Daenerys returns to King's Landing, almost. (Total 10 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 20 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 10 days. However, if she goes West from Mantarys, she will eventually reach King's Landing in 5 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]? (3 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 15 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] = 0.9; Pr[c | c s] = 0.8; Pr[c | s c] = 0.5; 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. (7 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 ($\boldsymbol{\pi} = P^k; k \gg 1$) and then take any row of the exponentiated matrix ($\boldsymbol{\pi}[i, :]$) 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. Multivariate Normal **(Total 10 points)**

A random vector $\mathbf{X} = (X_1, \ldots, X_k)$ is said to have a Multivariate Normal distribution if every linear combination of Xj has a Normal distribution. That is, we require $t_1X_1 + t_2X_2 + \ldots + t_kX_K$ to have a Normal distribution for any real values of $t_1, \ldots, t_k$. As a special case, we consider $t_1X_1 + t_2X_2 + \ldots + t_kX_K$ to be a degenerate normal distribution with variance 0 if $t_1X_1 + t_2X_2 + \ldots + t_kX_K$ is a constant (such as when all t_j's are 0).

(a) If $\mathbf{X} = (X_1, \ldots, X_k)$ is a Multivariate Normal, show that the distribution of any X_j is Normal. (1 point)

(b) It is possible to have normally distributed random variables $X_1, \ldots, X_k$ such that $(X_1, \ldots, X_k)$ is not Multivariate Normal: Let X = Normal(0, 1) and S = 1 with probability ½ and -1 with probability ½. Then Y = SX is normal due to the symmetry of the Standard Normal. In this case show that (X, Y) is not a Multivariate Normal. (2 points)

(c) Let Z, W be i.i.d Normal(0, 1) random variables. Show that (Z, W) and (Z + 2W, 3Z + 5W) are Multivariate Normals. (2 points)

(d) If $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_m)$ are Multivariate Normal vectors with X independent of Y, then show that the concatenated vector $W = (X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ is also a Multivariate Normal. (2 points)

(e) Fact 1 (Uncorrelated implies independence): If X is a Multivariate Normal that can be written as $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2})$, where $\mathbf{X}_1$ and $\mathbf{X}_2$ are subvectors, and every component of $\mathbf{X}_1$ is uncorrelated with every component of $\mathbf{X}_2$, then $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent.

Fact 2 (Property of Covariance): For any random variables X, Y, W, and V, we have:
Cov(aX + bY, cW + dV) = ac Cov(X, W) + ad Cov(X, V) + bc Cov(Y, W) + bd Cov(Y, V).

Let X, Y be i.i.d. standard Normals. Use Fact 1 and Fact 2 to show that (X+Y, X-Y) is a Multivariate Normal. (3 points)

7. Pokémon Go fanatic (Total 10 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]? (5 points)

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

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