Dept of AIML, RVCE Bangalore

Statistics for Data Science

Module 2: Joint Distribution, Correlation, Covariance and Independent Random Variables

1. Suppose that n people throw their hats in a box and then each picks one hat at random. Each hat can be picked by only one person and each assignment of hats to persons is equally likely. What is the expected value of X, the number of people that get back their own hat? *This does not fall under the topics I have listed above. I just thought I will include this anyways!*

2. Suppose that some customers at a restaurant take a simple survey. They are asked to rate the quality of their experience on a scale of 1 to 5, and they are asked whether or not there waiter/waitress was attentive. Let X be a random variable describing the quality of the experience, and Y a random variable which is 1 if the waiter/waitress was attentive and 0 otherwise

	x = 1	x = 2	x = 3	x = 4	x = 5
y = 0	0.05	0.2	0.1	0.04	0.01
y = 1	0.01	0.09	0.15	0.2	0.15

   (a) Is the table a valid PMF?

   (b) Find the marginal PMF of X and Y.

   (c) Are X and Y independent?

3. Suppose X and Y are two random variables with means $\mu_X = 1$ and $\mu_Y = -3$ and variances $\sigma_X^2 = 2$ and $\sigma_Y^2 = 4$ respectively. Suppose the covariance between X and Y is known to be $Cov(X, Y) = 1/2$, find the mean and variance of $Z = 3X + Y$.

4. Suppose X and Y are two random variables with $E(X) = E(Y) = 0$, and $VAR(X) = 1$, and suppose we know X is independent from $X + Y$. What is the covariance between X and Y ?

5. Identify the following statement as YES or NO with proper justification.
   Statement: If X and Y are independent random variables, then $VAR(3X + 2Y + 1) = VAR(3X - 2Y + 3)$.

6. Let X, Y, and Z be random variables, where X and Y are uncorrelated. The means of the RVs are $E(X) = 1, E(Y) = 2$, and $E(Z) = -1$, and $E(XZ) = 5$. What is $COV(X, Y + 2Z)$?

7. List out the conditions that a matrix has to satisfy for it to be deemed as a valid covariance matrix.

8. Which of the following matrices can be a valid covariance matrix and why? Indicate clearly which property of covariance matrix do the other matrices violate?

   A. $M_1 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ B. $M_2 = \begin{pmatrix} 1 & 1 \\ 3 & 2 \end{pmatrix}$ C. $M_3 = \begin{pmatrix} 5 & 0 \\ 0 & -2 \end{pmatrix}$ D. Matrix M_4 whose determinant is 15 and the eigenvalues are -3 and -5.

9. You toss a fair coin 100 times. (*No no... not ordering you! Assume you do this when u really feel totally jobless!!!*. Let X denote the number of heads in the 100 tosses and Y be the number of tails in the 100 tosses. Compute the correlation coefficient $\rho_{X,Y}$. With the given information can you calculate $Var(X+Y)$? If yes, find the value. If not, what more information you need to compute $Var(X+Y)$?.

10. Let X, Y and Z be random variables. The covariance matrix of (X, Y, Z) is given below:

$$C_{X,Y,Z} = \begin{pmatrix} Var(X) & Cov(X,Y) & Cov(X,Z) \\ Cov(X,Y) & Var(Y) & Cov(Y,Z) \\ Cov(X,Z) & Cov(Y,Z) & Var(Z) \end{pmatrix} = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 4 & -1 \\ 1 & -1 & 4 \end{pmatrix}$$

Given the above information alone, is it possible to compute $Var(X+Y+Z)$? If yes, obtain the value. If your answer is a NO, then list out the information that you need to compute $Var(X+Y+Z)$.

11. Let X, Y and Z be random variables. The covariance matrix of (X, Y, Z) is given below:

$$C_{X,Y,Z} = \begin{pmatrix} Var(X) & Cov(X,Y) & Cov(X,Z) \\ Cov(X,Y) & Var(Y) & Cov(Y,Z) \\ Cov(X,Z) & Cov(Y,Z) & Var(Z) \end{pmatrix} = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 4 & -1 \\ 1 & -1 & 4 \end{pmatrix}$$

   (a) Find $Cov(X, Y+Z)$.
   (b) Find the covariance matrix corresponding to $(X, X+Z, Y+Z)$

12. Two fair six sided dice, a black one and a red one, are rolled, with outcomes X and Y respectively for the black and red die respectively. Are $X+Y$ and $X-Y$ independent? Justify your choice.

13. A chicken lays a Poisson(λ) number N of eggs. Each egg, independently, hatches a chick with probability p. Let X be the number which hatch, so $X|N \sim Bin(N,p)$. Find the correlation between N (the number of eggs) and X (the number of eggs which hatch). Your final answer should work out to a simple function of p.

14. A device runs until either of two components fails, at which point the device stops running. The joint density function of the lifetimes of the two components, both measured in hours, is

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{x+y}{8} & 0 < x < 2, 0 < y < 2 \\ 0 & \text{otherwise.} \end{cases}$$

What is the probability that the device fails during the first hour of the operation?

15. Let the joint PDF of (X, Y) be

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{x+y}{3} & \text{if} 0 < x < 2, 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Calculate the covariance $Cov(X,Y)$.