# Probability and Statistics
# for Final Year Engineering Students

By Yoni Nazarathy, Last Updated: May 24, 2011.

## Exercises and Tutorial 4 (covers part A):
## Several Random Variables: Joint Distributions, Correlation

Note: Exercises with a (*) are NOT required for Mechanical Engineering students.

**Joint Distribution**

The joint distribution of two random variables X and Y is $p(x,y)$ (in the discrete case) or $f(x,y)$ in the continuous case allows us to calculate probabilities related to both random variables. We demonstrate by example:

Assume a packet switch with two inputs A and B and two outputs X and Y. In each time unit the following can occur in each of the inputs:

- No arrival of packets, w.p. $\frac{1}{2}$
- An arrival destined to output X, w.p. 1/6.
- An arrival destined to output Y, w.p. 2/6.

Let the R.V. X measure the number of packets destined to output X and let the R.V. Y measure the number of packets destined to output Y.

1. What is the possible range of values that X may take?
2. How about Y?
3. What is the possible values that the pair X,Y may take (e.g. can we have X=2 and Y=2)?
4. Construct the joint probability distribution p(x,y) = P(X=x,Y=y).
5. Assume that a packet is lost if X=2 or Y=2? What is the probability of packet loss?
6. Consider the following joint distribution function $f(x,y) = Cx\ y$ for $0 \le x \le 1$ and $0 \le y \le 1$. It is known that the area between the surface having heights equal to $x * y$ and the plane is ¼ . What is C?

**Independence Revisited:**

X and Y are independent if p(x,y)=p(x)p(y) (discrete case) or f(x,y)=f(x)f(y) (continuous case).

7. Let X have a binomial distribution with n trials and success probability p. Let Y=n-X. Y is thus the number of failures. What is p(x,y)? Are X and Y independent?
8. Let (X,Y) be the coordinates of a uniformly random point in the region [a,b],[c,d]. We have,

$$f(x,y) = \begin{cases} K & a \le x \le b,\ \ c \le y \le d \\ 0 & otherwise \end{cases}.$$

   What is K? Are X and Y independent?

**Covariance**

The covariance between two R.V's X and Y is a measure of their dependence (a number).

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])]$$

If the R.V's are independent Cov(X,Y)=0. If there is a **positive relation** between the RV's the covariance is positive, if there is a **negative relation** between the RV's the covariance is negative.

Covariance is also useful for calculating the variance of a sum of R.V's:

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

9. *)Show that Cov(X,Y) = E[X Y] – E[X] E[Y].
10. *) You know that $Var(aX) = a^2 Var(X)$. Use the formula above for variance of a sum to show that $Var(2X) = 4Var(X)$.
11. What is the covariance of the random variable X and a constant Y=a w.p. 1?
12. *) The sample covariance is:

$$\widehat{cov} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

    For the students in the class, evaluate the sample covariance of the measurements:
    x = Average cumulative time spent driving per week.
    y = Average time spent doing some sort of aerobic activity during the week.

**Correlation**

The correlation of two random variables is a **"normalized form**" of the covariance.

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}.$$

It can be shown that (we don't show it in this subject): $-1 \le \rho(X,Y) \le 1$.

Having $\rho(X,Y) = 1$ indicates that there is a perfect linear relationship between X and Y (without variability): E.g. Y=aX+b with a > 0. (try to show this).

Having $\rho(X,Y) = -1$ is similar yet indicates that a<0.

The **sample correlation** is calculated as follows:

$$\hat{\rho} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}}.$$

13. Calculate the sample correlation of the data collected in exercise 11 above.
14. *) Assume E[X]=0 and E[Y]=0. Write a simplified formula (using only expectations) for $\rho(X,Y)$.
15. *) Following (10), show that if Y=aX then $\rho(X,Y) = sign(a)$.
16. *) Show that for arbitrary RV's (not necessarily zero mean), if Y=aX+b then $\rho(X,Y) = sign(a)$.
17. For each of the following statements state if they are true or false:
    a. The correlation coefficient is zero if and only if the covariance is zero.
    b. The correlation coefficient is not defined if X or Y are non-random.
    c. The sample correlation coefficient is an indication of the relationship of two random quantities. Having a positive value near 1 indicates that as one of the values decreases the other increases.

# Selected Solutions

1) $X \in \{0,1,2\}$
2) Same as X.
3) (0,0), (0,1), (0,2), (1,0), (1,1), (2,0)
4)

$$p(0,0) = \frac{1}{2}\frac{1}{2} = \frac{1}{4} = \frac{9}{36}$$
$$p(0,1) = \frac{1}{2}\frac{1}{3} + \frac{1}{3}\frac{1}{2} = \frac{12}{36}$$
$$p(0,2) = \frac{1}{3}\frac{1}{3} = \frac{4}{36}$$
$$p(1,0) = \frac{1}{2}\frac{1}{6} + \frac{1}{6}\frac{1}{2} = \frac{6}{36}$$
$$p(1,1) = \frac{1}{6}\frac{2}{6} + \frac{2}{6}\frac{1}{6} = \frac{4}{36}$$
$$p(2,0) = \frac{1}{6}\frac{1}{6} = \frac{1}{36}$$

Observe the total is 1.

5) $p(2,0) + p(0,2) = \frac{5}{36}$
6) C=4.

8) $K = \frac{1}{(b-a)(d-c)}$. Yes – the RV's are independent.

11) 0.

17) a) True. b) True (because the variance is 0). c) False, this is what happens at if the correlation is -1.