Probability

G V V Sharma*

Contents

***Abstract*—This book provides solved examples on Probability**

1 December 2018

1.1. Let X and Y be i.i.d random variables uniformly distributed on (0,4).Then $\Pr(X > Y|X < 2Y)$ is

a) 1/3
b) 5/6
c) 1/4
d) 2/3

Solution:
The PDF is given by

$$f_X(x) = f_Y(x) = \begin{cases} \frac{1}{4}, & \text{if } 0 < x < 4 \\ 0, & \text{otherwise} \end{cases}$$

The CDF is given by

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

$$F_X(x) = F_Y(x) = \begin{cases} 0, & x \leq 0 \\ \frac{x}{4}, & \text{if } 0 < x < 4 \\ 1, & x \geq 4 \end{cases}$$

Using definition of conditional probability

$$\Pr(X > Y|X < 2Y) = \frac{\Pr(Y < X < 2Y)}{\Pr(X < 2Y)} \tag{1.1.1}$$

Now finding $\Pr(X < 2Y)$

$$\Pr(X < 2y) = F_X(2y) \tag{1.1.2}$$

$$\implies \Pr(X < 2Y) = \int_{-\infty}^{\infty} f_Y(x) \times F_X(2x)dx \tag{1.1.3}$$

$$\implies \Pr(X < 2Y) = \int_{0}^{2} \frac{x}{8}dx + \int_{2}^{4} \frac{1}{4}dx \tag{1.1.4}$$

$$\implies \Pr(X < 2Y) = \frac{3}{4} = 0.75 \tag{1.1.5}$$

*The author is with the Department of Electrical Engineering, Indian Institute of Technology, Hyderabad 502285 India e-mail: gadepall@iith.ac.in.

Now to find $\Pr(Y < X < 2Y)$

$$\Pr(y < X < 2y) = F_X(2y) - F_X(y) \tag{1.1.6}$$

$$\implies \Pr(Y < X < 2Y) \tag{1.1.7}$$
$$= \int_{-\infty}^{\infty} f_Y(x)(F_X(2x) - F_X(x))dx$$
$$\implies \int_0^2 \frac{1}{4}\left(\frac{x}{2} - \frac{x}{4}\right)dx + \int_2^4 \frac{1}{4}\left(1 - \frac{x}{4}\right)dx \tag{1.1.8}$$

$$\implies \Pr(Y < X < 2Y) = \frac{1}{4} = 0.25 \tag{1.1.9}$$

Now using (1.1.1),(1.1.5) and (1.1.9)

$$\Pr(X > Y|X < 2Y) = \frac{1/4}{3/4} = \frac{1}{3} \tag{1.1.10}$$

Hence final solution is option 1) or 1/3

1.2. Suppose X is a positive random variable with the following probability density function,

$$f(x) = (\alpha x^{\alpha-1} + \beta x^{\beta-1})e^{-x^{\alpha}-x^{\beta}}; x > 0$$

for $\alpha > 0, \beta > 0$. Then the hazard function of X for some choices of α and β can be

a) an increasing function.
b) a decreasing function.
c) a constant function.
d) a non monotonic function

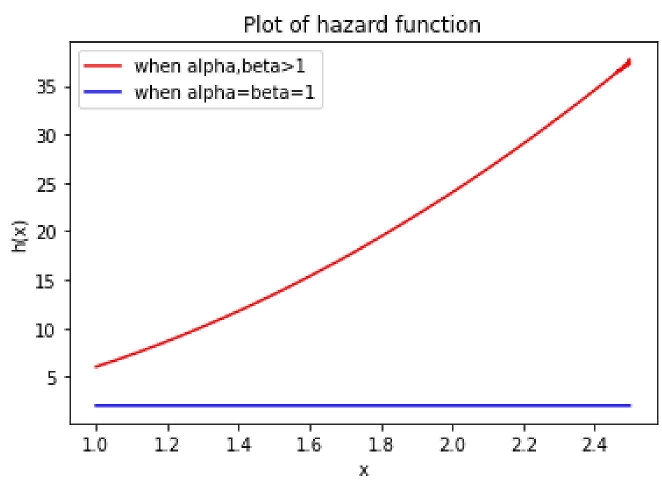
Solution:
CDF of X,

$$F(x) = \int_{-\infty}^{x} f(t)dt \tag{1.2.1}$$
$$= \int_0^x f(t)dt \qquad \text{as } x > 0 \tag{1.2.2}$$
$$= \int_{-\infty}^{t} \left((\alpha t^{\alpha-1} + \beta t^{\beta-1}) \times e^{-t^{\alpha}-t^{\beta}}\right) dt \tag{1.2.3}$$
$$= -e^{-t^{\alpha}-t^{\beta}}\Big|_0^x \tag{1.2.4}$$
$$= 1 - e^{-x^{\alpha}-x^{\beta}} \tag{1.2.5}$$

Hazard function,

$$h(x) = \frac{f(x)}{1 - F(x)} \tag{1.2.6}$$
$$= \alpha x^{\alpha-1} + \beta x^{\beta-1} \tag{1.2.7}$$
$$h'(x) = \alpha(\alpha-1)x^{\alpha-2} + \beta(\beta-1)x^{\beta-2} \tag{1.2.8}$$
$$h'(x) = \begin{cases} 0 & \alpha = \beta = 1 \\ > 0 & \text{otherwise} \end{cases} \tag{1.2.9}$$

Thus $h(x)$ can be either constant function or an increasing function.



From the above figure, it is verified that $h(x)$ can be either constant function or an increasing function.
Correct options are 1,3.

1.3. Suppose n units are drawn from a population of N units sequentially as follows. A random sample

$$U_1, U_2, ...U_N \text{ of size N, drawn from } U(0,1) \tag{1.3.1}$$

The k-th population unit is selected if

$$U_k < \frac{n - n_k}{N - k + 1}, k = 1, 2, ..N.\text{where, } n_1 = 0, n_k = \tag{1.3.2}$$

number of units selected out of first k-1 units for each k = 2, 3, ..N. Then,

a) The probability of inclusion of the second unit in the sample

$$\text{is } \frac{n}{N} \tag{1.3.3}$$

b) The probability of inclusion of the first and the second unit in the sample

$$\text{is } \frac{n(n-1)}{N(N-1)} \tag{1.3.4}$$

c) The probability of not including the first and including the second unit in the sample

$$\text{is } \frac{n(N-n)}{N(N-1)} \tag{1.3.5}$$

d) The probability of including the first and not including the second unit in the sample

$$\text{is } \frac{n(n-1)}{N(N-1)} \tag{1.3.6}$$

Solution:

2 June 2018

2.1. Two students are solving the same problem independently,if the probability of first one solves the problem is $\frac{3}{5}$ and the probability that the second one solves the problem is $\frac{4}{5}$,what is the probability that atleast one of them solves the problem?

a) $\frac{17}{25}$

b) $\frac{19}{25}$

c) $\frac{21}{25}$

d) $\frac{23}{25}$

Solution: Let X,Y be two events representing solving the problem by students A,B respectively.
Given

$$\Pr(X) = \frac{3}{5} \tag{2.1.1}$$

$$\Pr(Y) = \frac{4}{5} \tag{2.1.2}$$

Since students solve the problem independently,So events X and Y are independent, For independent events

$$\Pr(XY) = \Pr(X) \times \Pr(Y) \tag{2.1.3}$$

from (2.1.1) and (2.1.2)

$$\Pr(XY) = \frac{3}{5} \times \frac{4}{5} \tag{2.1.4}$$

$$\Pr(XY) = \frac{12}{25} \tag{2.1.5}$$

Now we have to find probability of solving the problem by atleast one of them i.e $\Pr(X+Y)$. As,

$$\Pr(X+Y) = \Pr(X) + \Pr(Y) - \Pr(XY) \tag{2.1.6}$$

from (2.1.1), (2.1.2), (2.1.5)

$$\Pr(X+Y) = \frac{3}{5} + \frac{4}{5} - \frac{12}{25} \tag{2.1.7}$$

$$\Pr(X+Y) = \frac{23}{25} \tag{2.1.8}$$

Hence the required probability is $\frac{23}{25}$

3 December 2016

3.1. $X_1, X_2, \ldots, X_n$ are independent and identically distributed as $N(\mu, \sigma^2)$, $-\infty < \mu < \infty$, $\sigma^2 > 0$. Then

a) $\sum_1^n \frac{(X_i-\bar{X})^2}{n-1}$ is the Minimum Variance Unbiased Estimate of σ^2

b) $\sqrt{\sum_1^n \frac{(X_i-\bar{X})^2}{n-1}}$ is the Minimum Variance Unbiased Estimate of σ

c) $\sum_1^n \frac{(X_i-\bar{X})^2}{n}$ is the Maximum Likelihood Estimate of σ^2

d) $\sqrt{\sum_1^n \frac{(X_i-\bar{X})^2}{n}}$ is the Maximum Likelihood Estimate of σ

Solution: The pdf for each random variable is same as they are all identical and independent Normal Distributions with same μ and σ^2.

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(x-\mu)^2}{2\sigma^2} \tag{3.1.1}$$

Let us take our maximum likelihood function for given random variable X_i

$$L(\mu;\sigma|X_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(X_i-\mu)^2}{2\sigma^2} \tag{3.1.2}$$

Since all the random variables are i.i.d

$$L(\mu;\sigma|X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} L(\mu;\sigma|X_i) \tag{3.1.3}$$

Let us denote:

$$L_m : L(\mu;\sigma|X_1, X_2, \ldots, X_n) \tag{3.1.4}$$

Substituting (3.1.2) for each Random Variable in (3.1.3)

$$L_m = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\frac{(X_i-\mu)^2}{2\sigma^2} \tag{3.1.5}$$

Taking natural log on both sides and simplifying

$$\ln L_m = \frac{-n}{2}\ln 2\pi - n\ln\sigma - \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{2\sigma^2} \tag{3.1.6}$$

In order to find Maximum Likelihood we need to maximise μ and σ w.r.t. all Random variables. Taking partial derivative w.r.t μ and taking σ as constant

$$\frac{\partial \ln L_m}{\partial\mu} = \sum_{i=1}^{n}\frac{(X_i-\mu)}{\sigma^2} \tag{3.1.7}$$

The value for μ at which L_m achieves maximum value is same in $\ln L_m$

$$\because \frac{\partial \ln L_m}{\partial\mu} = 0 \tag{3.1.8}$$

$$\therefore \sum_{i=1}^{n}\frac{(X_i-\mu)}{\sigma^2} = 0 \tag{3.1.9}$$

On simplifying the expression we get:

$$n\mu = \sum_{i=1}^{n} X_i \tag{3.1.10}$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{3.1.11}$$

Let us denote the value achieved in (3.1.11) as $\bar{X}$. Taking partial derivative w.r.t σ and taking μ as constant

$$\frac{\partial \ln L_m}{\partial\sigma} = \frac{-n}{\sigma} + \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^3} \tag{3.1.12}$$

The value for σ at which L_m achieves maximum value is same in $\ln L_m$

$$\frac{\partial \ln L_m}{\partial\sigma} = 0 \tag{3.1.13}$$

$$\frac{-n}{\sigma} + \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^3} = 0 \tag{3.1.14}$$

Upon simplifying the expression

$$\frac{n}{\sigma} = \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^3} \tag{3.1.15}$$

$$\sigma^2 = \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{n} \tag{3.1.16}$$

Substituting (3.1.11) in (3.1.16)

$$\sigma^2 = \sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{n} \tag{3.1.17}$$

$$\sigma = \sqrt{\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{n}} \tag{3.1.18}$$

Hence **Option 3** and **Option 4** are correct

3.2. There are two boxes. Box-1 contains 2 red balls and 4 green balls. Box-2 contains 4 red balls and 2 green balls. A box is selected at random and a ball is chosen randomly from the selected box. If the ball turns out to be red, what is the probability that Box-1 had been selected?
Solution: Box-1 has 2 red balls and 4 green balls.
Box-2 has 4 red balls and 2 green balls.
Let $B \in \{1,2\}$ represent a random variable where 1 represents selecting box-1 and 2 represents selecting box-2. From Baye's theorem

Event	definition	value
$\Pr(B=1)$	Probability of selecting Box-1	$\frac{1}{2}$
$\Pr(B=2)$	Probability of selecting Box-2	$\frac{1}{2}$
$\Pr(R=1\|B=1)$	Probability of drawing red ball from Box-1	$\frac{1}{3}$
$\Pr(G=1\|B=1)$	Probability of drawing green ball from Box-1	$\frac{2}{3}$
$\Pr(R=1\|B=2)$	Probability of drawing red ball from Box-2	$\frac{2}{3}$
$\Pr(G=1\|B=2)$	Probability of drawing green ball from Box-2	$\frac{1}{3}$

TABLE 3.2.1: Table 1

$$\Pr(R=1) = \Pr(R=1|B=1)\times\Pr(B=1) + \Pr(R=1|B=2)\times\Pr(B=2) \tag{3.2.1}$$

Substiting values from table (3.2.1) in (3.2.1)

$$\Pr(R = 1) = \frac{1}{2} \tag{3.2.2}$$

$$\Pr((R = 1)(B = 1)) = \Pr(R = 1|B = 1) \times \Pr(B = 1) \tag{3.2.3}$$

$$= \frac{1}{6} \tag{3.2.4}$$

We need to find $\Pr(B = 1|R = 1)$

$$\Pr(B = 1|R = 1) = \frac{\Pr((R = 1)(B = 1))}{\Pr(R = 1)} \tag{3.2.5}$$

$$= \frac{1}{3} \tag{3.2.6}$$

$\therefore$ The desired probability that box-1 is selected $= \frac{1}{3}$

3.3. Suppose customers arrive in a shop according to a Poisson process with rate 4 per hour. The shop opens at 10 : 00 am. If it is given that the second customer arrives at 10 : 40 am, what is the probability that no customer arrived before 10 : 30 am?

a) $\frac{1}{4}$
b) e^{-2}
c) $\frac{1}{2}$
d) $e^{\frac{1}{2}}$

Solution: We need to find

Random Variable	Time at which people arrive
X_p	$p = 10:00 - 10:30$
X_q	$q = 10:30 - 10:40$
X_r	$r = 10:00 - 10:40$
Y	$10:40$

TABLE 3.3.1: Random Variables

$$\Pr\left(X_p = 0|Y = 2\right) \tag{3.3.1}$$

In the world where the 2^{nd} person arrives at 10 : 40 am the (3.3.1) becomes:

$$= \frac{\Pr\left(X_p = 0, X_q = 1\right)}{\Pr(X_r = 1)} \tag{3.3.2}$$

$$= \frac{\Pr\left(X_p = 0\right) \times \Pr\left(X_q = 1\right)}{\Pr(X_r = 1)} \tag{3.3.3}$$

The Poisson function distribution for time interval t and rate λ for a random variable X:

$$f_X(x;t) = \frac{(\lambda t)^x \exp(-\lambda t)}{x!}$$

For the time interval p:

$$\lambda = 4, t = 0.5, x = 0 \tag{3.3.4}$$

$$\Pr\left(X_p = 0\right) = f_X\left(0; \frac{1}{2}\right) \tag{3.3.5}$$

$$= e^{-2} \tag{3.3.6}$$

$$\tag{3.3.7}$$

For the time interval q:

$$\lambda = 4, t = \frac{1}{6}, x = 1 \tag{3.3.8}$$

$$\Pr\left(X_q = 1\right) = f_X\left(1; \frac{1}{6}\right) \tag{3.3.9}$$

$$= \frac{2}{3}e^{\frac{-2}{3}} \tag{3.3.10}$$

For the time interval r:

$$\lambda = 4, t = \frac{2}{3}, x = 1 \tag{3.3.11}$$

$$\Pr(X_r = 1) = f_X\left(1; \frac{2}{3}\right) \tag{3.3.12}$$

$$= \frac{8}{3}e^{\frac{-8}{3}} \tag{3.3.13}$$

Substituting (3.3.6) (3.3.10) (3.3.13) in (3.3.3):

$$\Pr\left(X_p = 0|Y = 2\right) = \frac{1}{4} \tag{3.3.14}$$

3.4. A fair die is thrown two times independently. Let X, Y be the outcomes of these two throws and $Z = X+Y$. Let U be the remainder obtained when Z is divided by 6. Then which of the following statement(s) is/are true?

a) X and Z are independent
b) X and U are independent
c) Z and U are independent
d) Y and Z are not independent

Solution: Let $X \in \{1, 2, 3, 4, 5, 6\}$ represent the random variable which represents the outcome of the first throw of a dice. Similarly, $Y \in \{1, 2, 3, 4, 5, 6\}$ represents the random variable which represents the outcome of the second throw of a dice.

$$n(X = i) = 1, \quad i \in \{1, 2, 3, 4, 5, 6\} \tag{3.4.1}$$

$$\Pr(X = i) = \begin{cases} \frac{1}{6} & i \in \{1,2,3,4,5,6\} \\ 0 & \text{otherwise} \end{cases} \tag{3.4.2}$$

Similarly,

$$\Pr(Y = i) = \begin{cases} \frac{1}{6} & i \in \{1,2,3,4,5,6\} \\ 0 & \text{otherwise} \end{cases} \tag{3.4.3}$$

$$Z = X + Y \tag{3.4.4}$$

$$\text{Let } z \in \{1, 2, \ldots, 11, 12\} \tag{3.4.5}$$

$$\Pr(Z = z) = \Pr(X + Y = z) \tag{3.4.6}$$

$$= \sum_{x=0}^{z} \Pr(X = x) \Pr(Y = z - x) \tag{3.4.7}$$

$$= (6 - |z - 7|) \times \frac{1}{6} \times \frac{1}{6} \tag{3.4.8}$$

$$= \frac{6 - |z-7|}{36} \tag{3.4.9}$$

$$\Pr(Z = z) = \begin{cases} \frac{6-|z-7|}{36} & z \in \{1, 2, \ldots, 11, 12\} \\ 0 & \text{otherwise} \end{cases} \tag{3.4.10}$$

U is the remainder obtained when Z is divided by 6.

$$\text{Let } u \in \{0, 1, 2, 3, 4, 5\} \tag{3.4.11}$$

$$\Pr(U = u) = \sum_{k=0}^{2} \Pr(Z = 6k + u) \tag{3.4.12}$$

$$\Pr(U = 0) = \Pr(Z = 0) + \Pr(Z = 6) + \Pr(Z = 12) \tag{3.4.13}$$

$$= 0 + \frac{5}{36} + \frac{1}{36} = \frac{1}{6} \tag{3.4.14}$$

$$\text{for } u \in \{1, 2, 3, 4, 5\} \tag{3.4.15}$$

$$\Pr(U = u) = \Pr(Z = 0 + u) + \Pr(Z = 6 + u) \tag{3.4.16}$$

$$= \frac{6 - |u - 7|}{36} + \frac{6 - |6 + u - 7|}{36} \tag{3.4.17}$$

$$= \frac{6 - (7 - u)}{36} + \frac{6 - (u - 1)}{36} \tag{3.4.18}$$

$$= \frac{u - 1 + 7 - u}{36} = \frac{6}{36} \tag{3.4.19}$$

$$= \frac{1}{6} \tag{3.4.20}$$

$$\Pr(U = u) = \begin{cases} \frac{1}{6} & u \in \{0, 1, 2, 3, 4, 5\} \\ 0 & \text{otherwise} \end{cases} \tag{3.4.21}$$

Now, for checking each option,

a) Checking if X and Z are independent

$$p_1 = \Pr(Z = z, X = x) \tag{3.4.22}$$

$$= \Pr(Y = z - x, X = x) \tag{3.4.23}$$

$$= \Pr(Y = z - x) \times \Pr(X = x) \tag{3.4.24}$$

$$= \begin{cases} \frac{1}{36} & z - x \in \{1, 2, 3, 4, 5, 6\} \\ 0 & \text{otherwise} \end{cases} \tag{3.4.25}$$

$$\Pr(Z = z) \times \Pr(X = x) = \frac{6 - |z - 7|}{36} \times \frac{1}{6} \tag{3.4.26}$$

$$= \frac{6 - |z - 7|}{216} \tag{3.4.27}$$

$$\Pr(Z = z) \Pr(X = x) \neq \Pr(Z = z, X = x) \tag{3.4.28}$$

X and Z are not independent from (3.4.28) and hence option (**??**) is false.

b) Checking if X and U are independent

$$p_2 = \Pr(U = u, X = x) \tag{3.4.29}$$

$$p_2 = \Pr((Z = u) + (Z = 6 + u) + (Z = 12 + u), X = x) \tag{3.4.30}$$

$$p_2 = \Pr((Y = u - x) + (Y = 6 + u - x) + (Y = 12 + u - x), X = x) \tag{3.4.31}$$

$$p_2 = \frac{1}{6} \times \frac{1}{6} \tag{3.4.32}$$

$$= \frac{1}{36} \tag{3.4.33}$$

$$\Pr(U = u) \times \Pr(X = x) = \frac{1}{6} \times \frac{1}{6} \tag{3.4.34}$$

$$= \frac{1}{36} \tag{3.4.35}$$

$$\Pr(U = u)\Pr(X = x) = \Pr(U = u, X = x) \tag{3.4.36}$$

X and U are independent from (3.4.36) and hence option (**??**) is true.

c) Checking if Z and U are independent

$$p_3 = \Pr(Z = z|U = u) \tag{3.4.37}$$

$$p_3 = \begin{cases} 1 & u = 1 \text{ and } z = 7 \\ \frac{1}{2} & u = 0 \text{ and } z \in \{6, 12\} \\ \frac{1}{2} & u \in \{2, 3, 4, 5\} \text{ and } \\ & z = u \text{ or } z = 6 + u \\ 0 & \text{otherwise} \end{cases} \tag{3.4.38}$$

$$\Pr(Z = z) = \frac{6 - |z - 7|}{36} \tag{3.4.39}$$

If Z and U are independent, then

$$\Pr(Z = z|U = u) = \frac{\Pr(Z = z, U = u)}{\Pr(U = u)} \tag{3.4.40}$$

$$= \frac{\Pr(Z = z)\Pr(U = u)}{\Pr(U = u)} \tag{3.4.41}$$

$$= \Pr(Z = z) \tag{3.4.42}$$

But,

$$\Pr(Z = z|U = u) \neq \Pr(Z = z) \tag{3.4.43}$$

X and U are not independent from (3.4.43) and hence option (**??**) is false.

d) Checking if Y and Z are independent

$$p_1 = \Pr(Z = z, Y = y) \tag{3.4.44}$$

$$= \Pr(X = z - y, Y = y) \tag{3.4.45}$$

$$= \Pr(X = z - y) \times \Pr(Y = y) \tag{3.4.46}$$

$$= \begin{cases} \frac{1}{36} & z - y \in \{1, 2, 3, 4, 5, 6\} \\ 0 & \text{otherwise} \end{cases} \tag{3.4.47}$$

$$\Pr(Z = z) \times \Pr(Y = y) = \frac{6 - |z - 7|}{36} \times \frac{1}{6} \tag{3.4.48}$$

$$= \frac{6 - |z - 7|}{216} \tag{3.4.49}$$

$$\Pr(Z = z)\Pr(Y = y) \neq \Pr(Z = z, Y = y) \tag{3.4.50}$$

X and Z are not independent from (3.4.50) and hence option (**??**) is true.

Thus, options (**??**) and (**??**) are true.

4 December 2015

4.1. The probability that a ticketless traveler is caught during a trip is 0.1. If the traveler makes 4 trips , the probability that he/she will be caught during at least one of the trips is:

a) $1 - (0.9)^4$
b) $(1 - 0.9)^4$
c) $1 - (1 - 0.9)^4$
d) $(0.9)^4$

Solution: Let $X_i \in \{0, 1\}$ represent the ith trip where 1 denotes a ticketless traveller is caught. Given,

$$\Pr(X_i = 1) = p = 0.1 \tag{4.1.1}$$

Let,

$$X = \sum_{i=1}^{n} X_i \tag{4.1.2}$$

where n is the number of trips and X has a binomial distribution.

$$p_X(k) = \begin{cases} {}^nC_k\, p^K (1-p)^{n-k}, & 0 \le k \le n \\ 0, & otherwise \end{cases} \tag{4.1.3}$$

As he/she makes 4 trips in total, Using (4.1.1) and (4.1.3),

$$\Pr(X=0) = p_X(0) \tag{4.1.4}$$
$$= {}^4C_0\, p^0 (1-p)^4 \tag{4.1.5}$$
$$\Pr(X=0) = (0.9)^4 \tag{4.1.6}$$

Then probability of being caught in atleast one trip is,(Using (4.1.6))

$$\Pr(X \ge 1) = 1 - \Pr(X < 1) \tag{4.1.7}$$
$$= 1 - \Pr(X = 0) \tag{4.1.8}$$
$$= 1 - (0.9)^4 \tag{4.1.9}$$

4.2. Suppose that (X, Y) has a joint probability distribution with the marginal distribution of X being N(0,1) and $E(Y|X = x) = x^3$ for all $x \in R$. Then, which of the following statements are true?

a) $\text{Corr}(X, Y) = 0$
b) $\text{Corr}(X, Y) > 0$
c) $\text{Corr}(X, Y) < 0$
d) X and Y are independent

Solution: The following result shall be useful later. For $n \in N$

$$\int_{-\infty}^{\infty} \frac{x^n e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx = \begin{cases} 0 & n \text{ is odd} \\ (n-1) \times ... \times 3 \times 1 & n \text{ is even} \end{cases} \tag{4.2.1}$$

The proof for the above can be found at the end of the solution.

$$Corr(X, Y) = \frac{\sigma_{XY}^2}{\sigma_X \sigma_Y} \tag{4.2.2}$$

We know $X \sim N(0, 1)$. Thus,

$$f_X(x) = \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} \tag{4.2.3}$$
$$E(X) = 0 \tag{4.2.4}$$
$$\sigma_X^2 = 1 \tag{4.2.5}$$

$$\sigma_Y^2 = E(Y^2) - E(Y)^2 \tag{4.2.6}$$

$$E(Y) = \int_{-\infty}^{\infty} E(Y|X = x) f_X(x) dx \tag{4.2.7}$$
$$= \int_{-\infty}^{\infty} \frac{x^3 e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx \tag{4.2.8}$$
$$= 0 \tag{4.2.9}$$

$$E(Y^2) = \int_{-\infty}^{\infty} E(Y^2|X = x) f_X(x) dx \tag{4.2.10}$$
$$= \int_{-\infty}^{\infty} \frac{x^6 e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx \tag{4.2.11}$$
$$= 15 \tag{4.2.12}$$

Substituting in (4.4.6)

$$\sigma_Y^2 = 15 \tag{4.2.13}$$

$$\sigma_{XY}^2 = E(XY) - E(X)E(Y) \tag{4.2.14}$$

$$E(XY) = \int_{-\infty}^{\infty} E(XY|X = x) f_X(x) dx \tag{4.2.15}$$
$$= \int_{-\infty}^{\infty} E(xY|X = x) f_X(x) dx \tag{4.2.16}$$
$$= \int_{-\infty}^{\infty} xE(Y|X = x) f_X(x) dx \tag{4.2.17}$$
$$= \int_{-\infty}^{\infty} \frac{x^4 e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx \tag{4.2.18}$$
$$= 3 \tag{4.2.19}$$

Substituting in (4.4.14)

$$\sigma_{XY}^2 = 3 \tag{4.2.20}$$

Substituting in (4.4.2)

$$Corr(X, Y) = \frac{3}{\sqrt{15}} > 0 \tag{4.2.21}$$

Since $Corr(X, Y) \neq 0$, X and Y are dependent. Thus option 2 is the only correct option. **Proof for the integral:** If n is odd, $\frac{x^n e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}$ is an odd function, thus

$$\int_{-\infty}^{\infty} \frac{x^n e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx = 0 \tag{4.2.22}$$

If n is even,

$$\int_{-\infty}^{\infty} \frac{x^n e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx = \int_{-\infty}^{\infty} (x^{n-1})(\frac{xe^{\frac{-x^2}{2}}}{\sqrt{2\pi}}) dx \quad (4.2.23)$$

Using integration by parts,

$$\int_{-\infty}^{\infty} \frac{x^n e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx = \left(x^{n-1} \int \frac{xe^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx \right)\Bigg|_{-\infty}^{\infty}$$

$$-(n-1) \int_{-\infty}^{\infty} x^{n-2} \left(\int \frac{xe^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx \right) dx \quad (4.2.24)$$

$$= \left(x^{n-1}(-\frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}) \right)\Bigg|_{-\infty}^{\infty} - (n-1) \int_{-\infty}^{\infty} x^{n-2}(-\frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}) dx \quad (4.2.25)$$

$$= (n-1) \int_{-\infty}^{\infty} \frac{x^{n-2} e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx \quad (4.2.26)$$

$$= (n-1)(n-3) \int_{-\infty}^{\infty} \frac{x^{n-4} e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx \quad (4.2.27)$$

$$= (n-1) \times ... \times 3 \times 1 \int_{-\infty}^{\infty} \frac{x^0 e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx \quad (4.2.28)$$

$$= (n-1) \times ... \times 3 \times 1 \quad (4.2.29)$$

Alternative proof for the integral:

If n is odd, $\frac{x^n e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}$ is an odd function, thus

$$\int_{-\infty}^{\infty} \frac{x^n e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx = 0 \quad (4.2.30)$$

If n is even, let $n = 2k$. We differentiate the following identity k times w.r.t. α.

$$\int_{-\infty}^{\infty} e^{-\alpha x^2} dx = \sqrt{\left(\frac{\pi}{\alpha}\right)} \quad (4.2.31)$$

On differentiating k times, we get

$$\int_{-\infty}^{\infty} x^{2k} e^{-\alpha x^2} = \frac{1 \times 3 \times ... \times (2k-1)}{2^k} \sqrt{\left(\frac{\pi}{\alpha^{2k+1}}\right)} \quad (4.2.32)$$

On substituting $\alpha = \frac{1}{2}$, we get

$$\int_{-\infty}^{\infty} x^n e^{-\frac{x^2}{2}} = 1 \times 3 \times ... \times (n-1) \sqrt{2\pi} \quad (4.2.33)$$

Thus

$$\int_{-\infty}^{\infty} \frac{x^n e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx = (n-1) \times ... \times 3 \times 1 \quad (4.2.34)$$

4.3. Let $X_1, X_2, ..., X_n$ be independent and identically distributed, each having a uniform distribution on $(0, 1)$. Let $S_n = \sum_{i=1}^{n} X_i$ for $n \geq 1$. Then, which of the following statements are true?

A) $\frac{S_n}{n \log n} \to 0$ as $n \to \infty$ with probability 1.

B) $\Pr\left(\left(S_n > \frac{2n}{3}\right) \text{occurs for infinitely many n}\right) = 1$

C) $\frac{S_n}{\log n} \to 0$ as $n \to \infty$ with probability 1.

D) $\Pr\left(\left(S_n > \frac{n}{3}\right) \text{occurs for infinitely many n}\right) = 1$

Solution:

Symbol	expression/definition
S_n	$\sum_{i=1}^{n} X_i$
μ_n	$\frac{1}{n}\sum_{i=1}^{n} X_i$
X	Independent continuous random variable identical to $X_1, X_2, ..., X_n$

TABLE 4.3.1: Variables and their definitions

a) Given

$$S_n = \sum_{i=1}^{n} X_i, n \geq 1 \quad (4.3.1)$$

Dividing by n on both sides

$$\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i = \mu_n \quad (4.3.2)$$

It can be said that $X_1, X_2, ..., X_n$ are the trials of X. By definition

$$E[X] = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i}{n} = \lim_{n\to\infty} \frac{S_n}{n} \quad (4.3.3)$$

$$\lim_{n\to\infty} \frac{S_n}{n} = E[X] = \frac{1}{2} \quad (4.3.4)$$

$$\therefore \lim_{n\to\infty} \frac{S_n}{n \log n} = 0 \quad (4.3.5)$$

b) Using weak law, (4.3.4), and table (4.3.1)

$$\lim_{n\to\infty} \Pr\left(|\mu_n - E[X]| > \epsilon\right) = 0, \forall \epsilon > 0 \tag{4.3.6}$$

$$\lim_{n\to\infty} \Pr\left(S_n = \frac{n}{2}\right) = 1 \tag{4.3.7}$$

It can be easily implied from (4.3.7) that option B is false.

c) It is easy to observe from (4.3.4) that option C is false.

d) Using (4.3.7), we get

$$\Pr\left(\left(S_n > \frac{n}{3}\right) \text{occurs for infinitely many n}\right) = 1 \tag{4.3.8}$$

4.4. Suppose that (X, Y) has a joint probability distribution with the marginal distribution of X being N(0,1) and $E(Y|X = x) = x^3$ for all $x \in R$. Then, which of the following statements are true?

a) $\mathrm{Corr}(X, Y) = 0$
b) $\mathrm{Corr}(X, Y) > 0$
c) $\mathrm{Corr}(X, Y) < 0$
d) X and Y are independent

Solution: The following result shall be useful later. For $n \in N$

$$\int_{-\infty}^{\infty} \frac{x^n e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx = \begin{cases} 0 & n \text{ is odd} \\ (n-1) \times \dots \times 3 \times 1 & n \text{ is even} \end{cases} \tag{4.4.1}$$

The proof for the above can be found at the end of the solution.

$$Corr(X, Y) = \frac{\sigma_{XY}^2}{\sigma_X \sigma_Y} \tag{4.4.2}$$

We know $X \sim N(0, 1)$. Thus,

$$f_X(x) = \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} \tag{4.4.3}$$

$$E(X) = 0 \tag{4.4.4}$$

$$\sigma_X^2 = 1 \tag{4.4.5}$$

$$\sigma_Y^2 = E(Y^2) - E(Y)^2 \tag{4.4.6}$$

$$E(Y) = \int_{-\infty}^{\infty} E(Y|X = x) f_X(x) dx \tag{4.4.7}$$

$$= \int_{-\infty}^{\infty} \frac{x^3 e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx \tag{4.4.8}$$

$$= 0 \tag{4.4.9}$$

$$E(Y^2) = \int_{-\infty}^{\infty} E(Y^2|X = x) f_X(x) dx \tag{4.4.10}$$

$$= \int_{-\infty}^{\infty} \frac{x^6 e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx \tag{4.4.11}$$

$$= 15 \tag{4.4.12}$$

Substituting in (4.4.6)

$$\sigma_Y^2 = 15 \tag{4.4.13}$$

$$\sigma_{XY}^2 = E(XY) - E(X)E(Y) \tag{4.4.14}$$

$$E(XY) = \int_{-\infty}^{\infty} E(XY|X = x) f_X(x) dx \tag{4.4.15}$$

$$= \int_{-\infty}^{\infty} E(xY|X = x) f_X(x) dx \tag{4.4.16}$$

$$= \int_{-\infty}^{\infty} x E(Y|X = x) f_X(x) dx \tag{4.4.17}$$

$$= \int_{-\infty}^{\infty} \frac{x^4 e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx \tag{4.4.18}$$

$$= 3 \tag{4.4.19}$$

Substituting in (4.4.14)

$$\sigma_{XY}^2 = 3 \tag{4.4.20}$$

Substituting in (4.4.2)

$$Corr(X, Y) = \frac{3}{\sqrt{15}} > 0 \tag{4.4.21}$$

Since $Corr(X, Y) \neq 0$, X and Y are dependent. Thus option 2 is the only correct option. **Proof for the integral:** If n is odd, $\frac{x^n e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}$ is an odd function, thus

$$\int_{-\infty}^{\infty} \frac{x^n e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx = 0 \tag{4.4.22}$$

If n is even,

$$\int_{-\infty}^{\infty} \frac{x^n e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx = \int_{-\infty}^{\infty} (x^{n-1})(\frac{x e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}) dx \quad (4.4.23)$$

Using integration by parts,

$$\int_{-\infty}^{\infty} \frac{x^n e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx = \left(x^{n-1} \int \frac{x e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx \right) \Bigg|_{-\infty}^{\infty} - (n-1) \int_{-\infty}^{\infty} x^{n-2} \left(\int \frac{x e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx \right) dx \quad (4.4.24)$$

$$= \left(x^{n-1} (-\frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}) \right) \Bigg|_{-\infty}^{\infty} - (n-1) \int_{-\infty}^{\infty} x^{n-2} (-\frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}) dx \quad (4.4.25)$$

$$= (n-1) \int_{-\infty}^{\infty} \frac{x^{n-2} e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx \quad (4.4.26)$$

$$= (n-1)(n-3) \int_{-\infty}^{\infty} \frac{x^{n-4} e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx \quad (4.4.27)$$

$$= (n-1) \times ... \times 3 \times 1 \int_{-\infty}^{\infty} \frac{x^0 e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx \quad (4.4.28)$$

$$= (n-1) \times ... \times 3 \times 1 \quad (4.4.29)$$

Alternative proof for the integral:

If n is odd, $\frac{x^n e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}$ is an odd function, thus

$$\int_{-\infty}^{\infty} \frac{x^n e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx = 0 \quad (4.4.30)$$

If n is even, let $n = 2k$. We differentiate the following identity k times w.r.t. α.

$$\int_{-\infty}^{\infty} e^{-\alpha x^2} dx = \sqrt{\left(\frac{\pi}{\alpha}\right)} \quad (4.4.31)$$

On differentiating k times, we get

$$\int_{-\infty}^{\infty} x^{2k} e^{-\alpha x^2} = \frac{1 \times 3 \times ... \times (2k-1)}{2^k} \sqrt{\left(\frac{\pi}{\alpha^{2k+1}}\right)} \quad (4.4.32)$$

On substituting $\alpha = \frac{1}{2}$, we get

$$\int_{-\infty}^{\infty} x^n e^{-\frac{x^2}{2}} = 1 \times 3 \times ... \times (n-1) \sqrt{2\pi} \quad (4.4.33)$$

Thus

$$\int_{-\infty}^{\infty} \frac{x^n e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx = (n-1) \times ... \times 3 \times 1 \quad (4.4.34)$$

5 June 2013

5.1. Let X be a non-negative integer valued random variable with probability mass function $f(x)$ satisfying $(x+1)f(x+1) = (\alpha + \beta x)f(x)$, $x = 0, 1, 2, ...$; $\beta \neq 1$. You may assume that $E(X)$ and $Var(X)$ exist. Then which of the following statements are true?

a) $E(X) = \frac{\alpha}{1-\beta}$

b) $E(X) = \frac{\alpha^2}{(1-\beta)(1+\alpha)}$

c) $Var(X) = \frac{\alpha^2}{(1-\beta)^2}$

d) $Var(X) = \frac{\alpha}{(1-\beta)^2}$

Solution: For a discrete random variable X with P.D.F. $f(x)$ and which can take values from a set $\mathbb{S}$,

$$E(X) = \sum_{x \in \mathbb{S}} x f(x) \quad (5.1.1)$$

And,

$$E(X^2) = \sum_{x \in \mathbb{S}} x^2 f(x) \quad (5.1.2)$$

Also, as $f(x)$ is the P.D.F.,

$$\sum_{x \in \mathbb{S}} f(x) = 1 \quad (5.1.3)$$

Given, for $x \in \mathbb{S} = \{0, 1, 2, ...n\}$,

$$(x+1)f(x+1) = (\alpha + \beta x)f(x) \quad (5.1.4)$$

Summing both sides for $x \in \mathbb{S}$ we get,

$$\sum_{x=0}^{n} (x+1)f(x+1) = \sum_{x=0}^{n} (\alpha + \beta x)f(x) \quad (5.1.5)$$

Replacing $x+1$ with x in L.H.S. we get,

$$\sum_{x=1}^{n+1} x f(x) = \sum_{x=0}^{n} (\alpha + \beta x)f(x) \quad (5.1.6)$$

Rewriting LHS, we get,

$$\sum_{x=0}^{n} xf(x) + (n+1)f(n+1) = \sum_{x=0}^{n}(\alpha + \beta x)f(x) \tag{5.1.7}$$

But as $x \in \{0, 1, 2...n\}$, $f(n+1) = 0$. So the equation becomes

$$\sum_{x=0}^{n} xf(x) = \alpha \sum_{x=0}^{n} f(x) + \beta \sum_{x=0}^{n} xf(x) \tag{5.1.8}$$

Using (**??**) and (**??**), we get,

$$E(X) = \alpha(1) + \beta E(X) \tag{5.1.9}$$

So,

$$E(X) = \frac{\alpha}{1-\beta} \tag{5.1.10}$$

Now in (**??**), multiplying both sides by $(x+1)$, we get,

$$(x+1)^2 f(x+1) = (\alpha + \beta x)(x+1)f(x) \tag{5.1.11}$$

Summing both sides for $x \in \mathbb{S}$ we get,

$$\sum_{x=0}^{n}(x+1)^2 f(x+1) = \sum_{x=0}^{n}(\alpha + \beta x)(x+1)f(x) \tag{5.1.12}$$

Replacing $x+1$ with x in L.H.S. we get,

$$\sum_{x=1}^{n+1} x^2 f(x) = \sum_{x=0}^{n}(\beta x^2 f(x) + (\alpha+\beta)xf(x) + \alpha f(x)) \tag{5.1.13}$$

Rewriting LHS similarly as before, we get,

$$\sum_{x=0}^{n} x^2 f(x) = \beta \sum_{x=0}^{n} x^2 f(x) + (\alpha+\beta)\sum_{x=0}^{n} xf(x) + \alpha \sum_{x=0}^{n} f(x) \tag{5.1.14}$$

Using (**??**), (**??**) and (**??**), we get,

$$E(X^2) = \beta E(X^2) + (\alpha+\beta)E(X) + \alpha(1) \tag{5.1.15}$$

Using (**??**)

$$E(X^2)(1-\beta) = \frac{\alpha(\alpha+\beta)}{1-\beta} + \alpha \tag{5.1.16}$$

So,

$$E(X^2) = \frac{\alpha^2 + \alpha}{(1-\beta)^2} \tag{5.1.17}$$

Now,

$$Var(X) = E(X^2) - (E(X))^2 \tag{5.1.18}$$

Using (**??**) and (**??**),

$$Var(X) = \frac{\alpha^2 + \alpha}{(1-\beta)^2} - \frac{\alpha^2}{(1-\beta)^2} \tag{5.1.19}$$

So,

$$Var(X) = \frac{\alpha}{(1-\beta)^2} \tag{5.1.20}$$

So, options 1 and 4 are correct.

6 December 2012

6.1. Let X be a binomial random variable with parameters $\left(11, \frac{1}{3}\right)$. At which value(s) of k is $\Pr(X = k)$ maximized?

a) k=2
b) k=3
c) k=4
d) k=5

Solution: X has a binomial distribution :

$$\Pr(X = k) = {}^{n}C_k (q)^{n-k} (p)^k \tag{6.1.1}$$

Where,

- n=11
- $p = \frac{1}{3}$
- $q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}$

$$\Pr(X = k) = {}^{11}C_k \left(\frac{2}{3}\right)^{11-k} \left(\frac{1}{3}\right)^k \tag{6.1.2}$$

For Pr(X = k) to be maximized

$$\Pr(X = k) \geq \Pr(X = k+1) \tag{6.1.3}$$

$$\frac{\Pr(X = k)}{\Pr(X = k+1)} = \frac{{}^{11}C_k \left(\frac{2}{3}\right)^{11-k} \left(\frac{1}{3}\right)^k}{{}^{11}C_{k+1} \left(\frac{2}{3}\right)^{10-k} \left(\frac{1}{3}\right)^{k+1}} \geq 1 \tag{6.1.4}$$

$$\frac{2(k+1)}{11-k} \geq 1 \tag{6.1.5}$$

$$\implies k \geq 3 \tag{6.1.6}$$

$$\Pr(X = k) \geq \Pr(X = k-1) \tag{6.1.7}$$

$$\frac{\Pr(X = k)}{\Pr(X = k-1)} = \frac{{}^{11}C_k \left(\frac{2}{3}\right)^{11-k} \left(\frac{1}{3}\right)^{k}}{{}^{11}C_{k-1} \left(\frac{2}{3}\right)^{12-k} \left(\frac{1}{3}\right)^{k-1}} \geq 1 \tag{6.1.8}$$

$$\frac{12-k}{2k} \geq 1 \tag{6.1.9}$$

$$\implies k \leq 4 \tag{6.1.10}$$

From (**??**) , (**??**) and since k is an integer
$\Pr(X = k)$ is maximized for k=3, k=4
Thus options 2) and 3) are correct