Probability

G V V Sharma*

Contents

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

***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 $\Pr(X < 2Y)$

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

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

$$\implies \Pr(X < 2Y) = \int_0^2 \frac{x}{8}dx + \int_2^4 \frac{1}{4}dx \tag{1.1.4}$$

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

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,

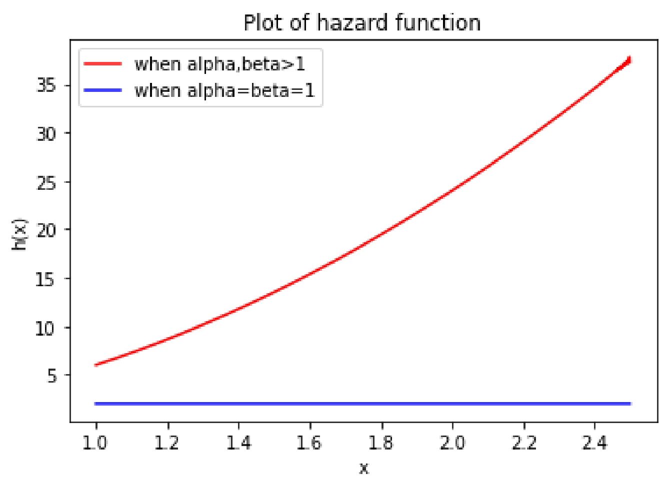
$$h(x) = \frac{f(x)}{1 - F(x)} \tag{1.2.6}$$

$$= \alpha x^{\alpha-1} + \beta x^{\beta-1} \tag{1.2.7}$$

$$h'(x) = \alpha(\alpha-1)x^{\alpha-2} + \beta(\beta-1)x^{\beta-2} \tag{1.2.8}$$

$$h'(x) = \begin{cases} 0 & \alpha = \beta = 1 \\ > 0 & \text{otherwise} \end{cases} \tag{1.2.9}$$

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.

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 respec-

tively.
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	$\frac{1}{2}$
$\Pr(R = 1 \mid B = 1)$	Probability of drawing red ball from Box-1	$\frac{1}{3}$
$\Pr(G = 1 \mid B = 1)$	Probability of drawing green ball from Box-1	$\frac{2}{3}$
$\Pr(R = 1 \mid B = 2)$	Probability of drawing red ball from Box-2	$\frac{2}{3}$
$\Pr(G = 1 \mid B = 2)$	Probability of drawing green ball from Box-2	$\frac{1}{3}$

TABLE 3.2.1: Table 1

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

Substiting values from table (3.2.1) in (3.2.1)

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

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

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

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

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

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

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

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 $n \in N$

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

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 (4.2.6)

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

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

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

Substituting in (4.2.14)

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

Substituting in (4.2.2)

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

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,

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

Using integration by parts,

$$\int_{-\infty}^{\infty} \frac{x^n e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx = \left(x^{n-1} \int \frac{x e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx \right)\Bigg|_{-\infty}^{\infty} - (n-1) \int_{-\infty}^{\infty} x^{n-2} \left(\int \frac{x e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx \right) dx \tag{4.2.24}$$

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

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

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

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

$$= (n-1) \times ... \times 3 \times 1 \tag{4.2.29}$$

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:

a) Given

$$S_n = \sum_{i=1}^{n} X_i, n \geq 1 \tag{4.3.1}$$

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

Dividing by n on both sides

$$\frac{S_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i = \mu_n \tag{4.3.2}$$

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

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

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

$$\therefore \lim_{n\to\infty} \frac{S_n}{n \log n} = 0 \tag{4.3.5}$$

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

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

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

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

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

5 December 2012

5.1. 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 :

$$\Pr(X = k) = {}^{n}C_k(q)^{n-k}(p)^k \tag{5.1.1}$$

Where,

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

$$\Pr(X = k) = {}^{11}C_k \left(\frac{2}{3}\right)^{11-k} \left(\frac{1}{3}\right)^k \tag{5.1.2}$$

For Pr(X = k) to be maximized

$$\Pr(X = k) \geq \Pr(X = k + 1) \tag{5.1.3}$$

$$\frac{\Pr(X = k)}{\Pr(X = k + 1)} = \frac{{}^{11}C_k \left(\frac{2}{3}\right)^{11-k} \left(\frac{1}{3}\right)^k}{{}^{11}C_{k+1} \left(\frac{2}{3}\right)^{10-k} \left(\frac{1}{3}\right)^{k+1}} \geq 1 \tag{5.1.4}$$

$$\frac{2(k + 1)}{11 - k} \geq 1 \tag{5.1.5}$$

$$\implies k \geq 3 \tag{5.1.6}$$

$$\Pr(X = k) \geq \Pr(X = k - 1) \tag{5.1.7}$$

$$\frac{\Pr(X = k)}{\Pr(X = k - 1)} = \frac{{}^{11}C_k \left(\frac{2}{3}\right)^{11-k} \left(\frac{1}{3}\right)^k}{{}^{11}C_{k-1} \left(\frac{2}{3}\right)^{12-k} \left(\frac{1}{3}\right)^{k-1}} \geq 1 \tag{5.1.8}$$

$$\frac{12 - k}{2k} \geq 1 \tag{5.1.9}$$

$$\implies k \leq 4 \tag{5.1.10}$$

From (5.1.6) , (5.1.10) and since k is an integer

$\Pr(X = k)$ is maximized for k=3, k=4

Thus options 2) and 3) are correct