4. Which transformation will change a r.v X with pdf is as below to a uniform R.V Y whose range is $0 \le x \le 1$ a) $f(x) = \begin{cases} 2e^{-2x}, & x \ge 0 \\ 0 & otherwise \end{cases}$ b) $f(x) = \begin{cases} \frac{1}{2}(x-3), & 3 \le x \le 5 \\ 0 & otherwise \end{cases}$

1.7 Expectation and Variance of a Random Variable

1.7.1 Expected Values

One of the most important things we'd like to know about a random variable is: what value does it take on average? What is the average price of a computer? What is the average value of a number that rolls on a die? The value is found as the average of all possible values, weighted by how often they occur (i.e. probability)

Definition: Let X be a random variable with probability distribution $p(X = x)$. Then the **expected value** of X, denoted $E(X)$ or μ, is given by;

$$E(x) = \mu = \begin{cases} \sum_{x=-\infty}^{\infty} xp(X = x) & \text{if X is discrete} \\ \int_{-\infty}^{\infty} xp(X = x)dx & \text{if X is continuous} \end{cases}.$$

Theorem: Let X be a r.v. with probability distribution $p(X=x)$ and let $g(x)$ be a real-valued function of X. ie $g: \mathbb{R} \to \mathbb{R}$, then the expected value of $g(x)$ is given by

$$E[g(x)] = \begin{cases} \sum_{x=-\infty}^{\infty} g(x)p(X = x) & \text{if X is discrete} \\ \int_{-\infty}^{\infty} g(x)p(X = x)dx & \text{if X is continuous} \end{cases}.$$

Theorem: Let X be a r.v. with probability distribution $p(X = x)$. Then

(i) $E(c) = c$, where c is an arbitrary constant.

(ii) $E[ax + b] == a\mu + b$ where a and b are arbitrary constants

(iii) $E[kg(x)] = kE[g(x)]$ *where* $g(x)$ is a function of X

(iv) $E[ag_1(x) \pm bg_2(x)] = aE[g_1(x)] \pm bE[g_2(x)]$ and in general $E\left[\sum_{i=1}^{n} c_i g_i(x)\right] = \sum_{i=1}^{n} c_i E[g_i(x)]$ are functions of X. This property of expectation is called *linearity property*

Proof

We will sketch the proof using a continuous random variable since the proof using the discrete random variable is similar and was also discussed in probability and statistics I.

(i) $E[c] = \int_{All\ x} cP(X = x)dx = c\int_{All\ x} P(X = x)dx = c(1) = c$

(ii) $E[ax + b] = \int_{All\ x} (ax + b)P(X = x)dx = \int_{All\ x} axP(X = x)dx + \int_{All\ x} bP(X = x)dx$

$$= a\int_{All\ x} xP(X = x)dx + b\int_{All\ x} P(X = x)dx = a\mu + b$$

(iii) $E[kg(x)] = \int_{All\ x} kg(x)P(X = x)dx = k\int_{All\ x} g(x)P(X = x)dx = kE[g(x)]$

(iv) $E[ag_1(x) \pm bg_2(x)] = \int\limits_{All\ x} [ag_1(x) \pm bg_2(x)] P(X=x)dx = \int\limits_{All\ x} ag_1(x) P(X=x)dx \pm \int\limits_{All\ x} bg_2(x) P(X=x)dx$

$= E[ag_1(x)] \pm E[bg_2(x)] = aE[g_1(x)] \pm bE[g_2(x)]$ ie from part iii

1.7.2 Variance and Standard Deviation

Definition: Let X be a r.v with mean $E(X)=\mu$, the **variance** of X, denoted σ^2 or Var(X), is given by $Var(X)=\sigma^2=E(X-\mu)^2$. The units for variance are square units. The quantity that has the correct units is **standard deviation**, denoted σ. It's actually the positive square root of Var(X).

$$\sigma = \sqrt{Var(X)} = \sqrt{E(X-\mu)^2}.$$

Theorem: $Var(X) = E(X-\mu)^2 = E(X)^2 - \mu^2$

Proof:

$Var(X) = E(X-\mu)^2 = E(X^2 - 2X\mu + \mu^2) = E(X)^2 - 2\mu E(X) + \mu^2 = E(X)^2 - \mu^2$ Since $E(X)=\mu$

Theorem: $Var(aX+b) = a^2 \operatorname{var}(X)$

Proof:

Recall that $E[aX+b] = a\mu + b$ therefore

$Var(aX+b) = E[(aX+b)-(a\mu+b)]^2 = E[a(X-\mu)]^2 == E[a^2(X-\mu)^2] = a^2E[(X-\mu)^2] = a^2 \operatorname{var}(X)$

Remark

(i) The expected value of X always lies between the smallest and largest values of X.

(ii) In computations, bear in mind that variance cannot be negative!

Example 1

Given a probability distribution of X as below, find the mean and standard deviation of X.

x	0	1	2	3
P(X=x)	1/8	1/4	3/8	1/4

Solution

x	0	1	2	3	total
$p(X=x)$	$1/8$	$1/4$	$3/8$	$1/4$	1
$xp(X=x)$	0	$1/4$	$3/4$	$3/4$	$7/4$
$x^2p(X=x)$	0	$1/4$	$3/2$	$9/4$	4

$E(X) = \mu = \sum_{x=0}^{3} xp(X=x) = 1.75$ and

standard deviation

$\sigma = \sqrt{E(X^2) - \mu^2} = \sqrt{4 - 1.75^2} \approx 0.968246$

Example 2

The probability distribution of a r.v X is as shown below, find the mean and standard deviation of; a) X b) $Y = 12X + 6$.

x	0	1	2
P(X=x)	1/6	1/2	1/3

Solution

x	0	1	2	total
$p(X=x)$	$1/6$	$1/2$	$1/3$	1
$xp(X=x)$	0	$1/2$	$2/3$	$7/6$
$x^2p(X=x)$	0	$1/2$	$4/3$	$11/6$

$E(X) = \mu = \sum_{x=0}^{2} xp(X=x) = 7/6$ and

$E(X^2) = \sum_{x=0}^{2} x^2 p(X=x) = 11/6$

Standard deviation $\sigma = \sqrt{E(X^2) - \mu^2} = \sqrt{{}^{11}\!/_6 - ({}^{7}\!/_6)^2} = \sqrt{{}^{17}\!/_6} \approx 1.6833$

Now $E(Y) = 12E(X) + 6 = 12({}^{7}\!/_6) + 6 = 20$

$Var(Y) = Var(12X + 6) = 12^2 \times Var(X) = 144 \times \sqrt{{}^{17}\!/_6} \approx 242.38812$

Example 3

A continuous random variable X has a pdf given by $\text{f(x)} = \begin{cases} \frac{1}{2}x, & 0 \le x \le 2 \\ 0, & elsewhere \end{cases}$, find the mean and standard of X

Solution

$$\text{E(x)} = \int_{-\infty}^{\infty} x\text{f}(x)dx = \int_0^2 \tfrac{1}{2}x^2\text{dx} = \left[\frac{x^3}{6}\right]_0^2 = \frac{4}{3} \text{ and } \text{E(x}^2) = \int_{-\infty}^{\infty} x^2\text{f}(x)dx = \int_0^2 \tfrac{1}{2}x^3\text{dx} = \left[\frac{x^4}{8}\right]_0^2 = 2$$

Standard deviation $\sigma = \sqrt{E(X^2) - \mu^2} = \sqrt{2 - ({}^{4}\!/_3)^2} = \frac{\sqrt{2}}{3}$

Exercise

1. Suppose X has a probability mass function given by the table below

x	2	3	4	5	6
P(X=x)	0.01	0.25	0.4	0.3	0.04

   Find the mean and variance of; X

2. Suppose X has a probability mass function given by the table below

x	11	12	13	14	15
P(X=x)	0.4	0.2	0.2	0.1	0.1

   Find the mean and variance of; X

3. Let X be a random variable with $P(X = 1) = 0.2$, $P(X = 2) = 0.3$, and $P(X = 3) = 0.5$. What is the expected value and standard deviation of; a)X *b)* $Y = 5X - 10$?
4. A random variable W has the probability distribution shown below,

w	0	1	2	3
P(W=w)	2d	0.3	d	0.1

   Find the values of the constant d hence determine the mean and variance of W. Also find the mean and variance of $Y = 10X + 25$

5. A random variable X has the probability distribution shown below,

x	1	2	3	4	5
P(X=x)	7c	5c	4c	3c	c

   Find the values of the constant c hence determine the mean and variance of X.

6. The random variable Z has the probability distribution shown below,

z	2	3	5	7	11
P(Z=z)	$^1\!/_6$	$^1\!/_3$	$^1\!/_4$	x	y

   If $E(Z) = 4\frac{2}{3}$, find the values of x and y hence determine the variance of Z

7. A discrete random variable M has the probability distribution $\text{f(m)} = \begin{cases} \frac{m}{36}, & m = 1,2,3,...,8 \\ 0, & elsewhere \end{cases}$, find the mean and variance of M
8. For a discrete random variable Y the probability distribution is $\text{f(y)} = \begin{cases} \frac{5-y}{10}, & y = 1,2,3,4 \\ 0, & elsewhere \end{cases}$, calculate $E(Y)$ and var(Y)
9. Suppose X has a pmf given by $\text{f(x)} = \begin{cases} kx & \text{for } x = 1,2,3,4 \\ 0, & elsewhere \end{cases}$, find the value of the constant k hence obtain the mean and variance of X

10. A team of 3 is to be chosen from 4 girl and 6 boys. If X is the number of girls in the team, find the probability distribution of X hence determine the mean and variance of X
11. A fair six sided die has; '1' on one face, '2' on two of it's faces and '3' on the remaining three faces. The die is rolled twice. If T is the total score write down the probability distribution of T hence determine; a) the probability that T is more than 4. b) the mean and variance of T
12. The pdf of a continuous r.v R is given by $\mathrm{f(r)} = \begin{cases} kr \text{ for } 0 \le r \le 4 \\ 0,\ elsewhere \end{cases}$, (a) Determine *c*. hence Compute $\mathrm{P}(10 \le r \le 2)$, $\mathrm{E(X)}$ and $\mathrm{Var(X)}$.
13. A continuous r.v M has the pdf given by $\mathrm{f(m)} = \begin{cases} k(1-\frac{m}{10}) \text{ for } M \le 10 \\ 0,\ elsewhere \end{cases}$, find the value of the constant k, the mean and the variance of X
14. A continuous r.v X has the pdf given by $\mathrm{f(x)} = \begin{cases} k(1-x) \text{ for } 0 \le x \le 1 \\ 0,\ elsewhere \end{cases}$, findt the value of the constant k. Also find the mean and the variance of X
15. The lifetime of new bus engines, T years, has continuous pdf $\mathrm{f(t)} = \begin{cases} d/_{t^2} \text{ if } x \ge 1 \\ 0, \text{ if } x < 1 \end{cases}$ find the value of the constant d hence determine the mean and standard deviation of T
16. An archer shoots an arrow at a target. The distance of the arrow from the centre of the target is a random variable X whose p.d.f. is given by $\mathrm{f(x)} = \begin{cases} k(3+2\mathrm{x}-\mathrm{x}^2) \text{ if } x \le 3 \\ 0,\ \text{if } x > 3 \end{cases}$ find the value of the constant k. Also find the mean and standard deviation of X
17. A continuous r.v X has the pdf given by $\mathrm{f(x)} = \begin{cases} k(1+x),\ -1 \le x < 0 \\ 2k(1-x),\ 0 \le x \le 1 \\ 0,\ elsewhere \end{cases}$, find the value of the constant k. Also find the mean and the variance of X
18. A continuous r.v X has the pdf given by $\mathrm{f(x)} = \begin{cases} e^{-x} \text{ for } x > 0 \\ 0,\ elsewhere \end{cases}$, find the mean and standard deviation of; a) X b) $\mathrm{Y} = e^{\frac{3}{4}x}$

1.8 Mode, Median, Quartiles and Percentiles

Another measure commonly used to summarize random variables are the mode and median; Mode is the value of x that maximizes the pdf. That is the value of x for which $\mathrm{f}^{\prime}(\mathrm{x}) = 0$. Median is the value m such that "half of the distribution lies to the left of m and half to the right". More formally, m should satisfy $\mathrm{F_X(m)} = 0.5$.

Note: If there is a value m such that the graph of y= f(x) is symmetric about x=m, then both the expected value and the median of X are equal to m.

The lower quartile $\mathrm{Q_1}$ and the upper quartile $\mathrm{Q_3}$ are similarly defined by

$$\mathrm{F_X(Q_1)} = 0.25 \text{ and } \mathrm{F_X(Q_3)} = 0.75$$

Thus, the probability that X lies between $\mathrm{Q_1}$ and $\mathrm{Q_3}$ is $0.75 - 0.25 = 0.5$, so the quartiles give an estimate of how spread-out the distribution is. More generally, we define the $\mathrm{n^{th}}$ percentile