Sekhar_Mekala_HW2

Sekhar Mekala

Monday, September 12, 2016

1. Suppose that X is a discrete random variable having probability function Pr(X = k) = ck^2 for k = 1,2,3. Find c, Pr(X ≤ 2), E[X], and Var(X).

Answer:

Given the following probability mass function (pmf):

$$Pr(X = k) = ck^2, \text{ where } k = 1, 2, 3$$

Therefore the sum of probabilities on all the values of $k = 1, 2, 3$ will be 1.

$$\sum_{k=1}^{3} ck^2 = 1$$

$$c(1) + c(2^2) + c(3^2) = 1$$

$$c = \frac{1}{14}$$

Therefore, the pmf can be written as:

$$Pr(X = k) = \frac{k^2}{14}, \text{ where } k = 1, 2, 3$$

Let us find the $Pr(X \leq 2)$

$$Pr(x \leq 2) = Pr(x = 1) + Pr(x = 2) = \frac{1}{14} + \frac{4}{14} = \frac{5}{14}$$

Let us find the E[X]

$$E[X] = \sum_{k=1}^{3} k.k^2/14 = \sum_{k=1}^{3} k^3/14 = \frac{1^3 + 2^3 + 3^3}{14} = \frac{36}{14} = \frac{18}{7}$$

Deriving V[X]

$$V[X] = E[X^2] - (E[X])^2$$

$$E[X^2] = \sum_{k=1}^{3} k^2.k^2/14 = \sum_{k=1}^{3} k^4/14 = \frac{1^4 + 2^4 + 3^4}{14} = \frac{1 + 16 + 81}{14} = 7$$

We already know that $E[X] = \frac{18}{7}$

Therefore,

$$V[X] = E[X^2] - (E[X])^2 = 7 - (\frac{18}{7})^2 = 0.38775$$

2. Suppose that X is a continuous random variable having p.d.f. *f(x)* = cx^2 for $1 \le x \le 2$. Find c, $\Pr(X \ge 1)$, E[X], and Var(X).

Answer:

Given that $p.d.f$ as

$$f(x) = cx^2 \text{ where } x \in [1, 2]$$

Finding the value of c

$$\int_1^2 cx^2 \, dx = 1$$
$$\frac{c.2^3}{3} - \frac{c.1^3}{3} = 1$$
$$c = \frac{3}{7}$$

Finding $pr(X \ge 1)$

$$pr(x \ge 1) = \int_1^{\infty} (3/7)x^2 \, dx$$

But $x \in [1, 2]$, hence

$$pr(x \ge 1) = \int_1^2 (3/7)x^2 \, dx = \frac{2^3}{7} - \frac{1^3}{7} = 1$$

Finding E[X]

$$E[X] = \int_1^2 (3/7)x^2.x \, dx = \left[(3/7)(x^4/4)\right]_1^2 = \left[3x^4/28\right]_1^2 = \frac{45}{28}$$

Finding V[X]

$$V[X] = E[X^2] - (E[X])^2$$

$$E[X^2] = \int_1^2 (3/7)x^2.x^2 \, dx = \int_1^2 (3/7)x^4 \, dx = \left[(3/35)x^5\right]_1^2 = \frac{93}{35}$$

We know that $E[X] = \frac{45}{28}$

Therefore,

$$V[X] = E[X^2] - (E[X])^2 = \frac{93}{35} - (\frac{45}{28})^2 = 2.657 - 2.583 = 0.074$$

3. Suppose that X and Y are jointly continuous random variables with

$$\begin{cases} y - x & \text{for } 0 < x < 1 \quad \text{and } 1 < y < 2 \\ 0 & \text{otherwise} \end{cases}$$

a. Compute and plot $f_X(x)$ and $f_Y(y)$.
b. Are X and Y independent?
c. Compute $F_X(x)$ and $F_Y(y)$.
d. Compute E[X], Var(X), E[Y], Var(Y), Cov(X,Y), and Corr(X,Y).

Answer:

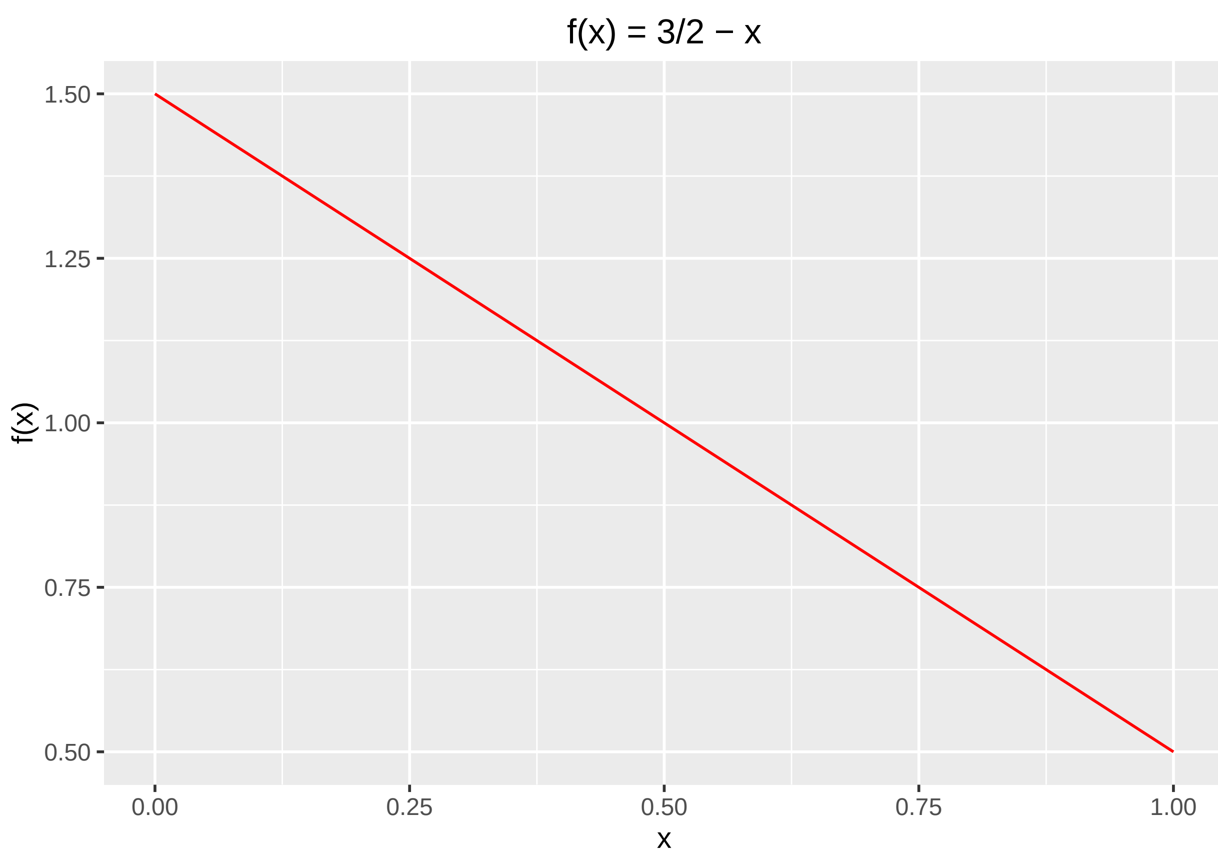
Finding $f_x(x)$

$$\int_1^2 (y-x)dy = \left[\frac{y^2}{2} - xy\right]_1^2 = \frac{3}{2} - x$$

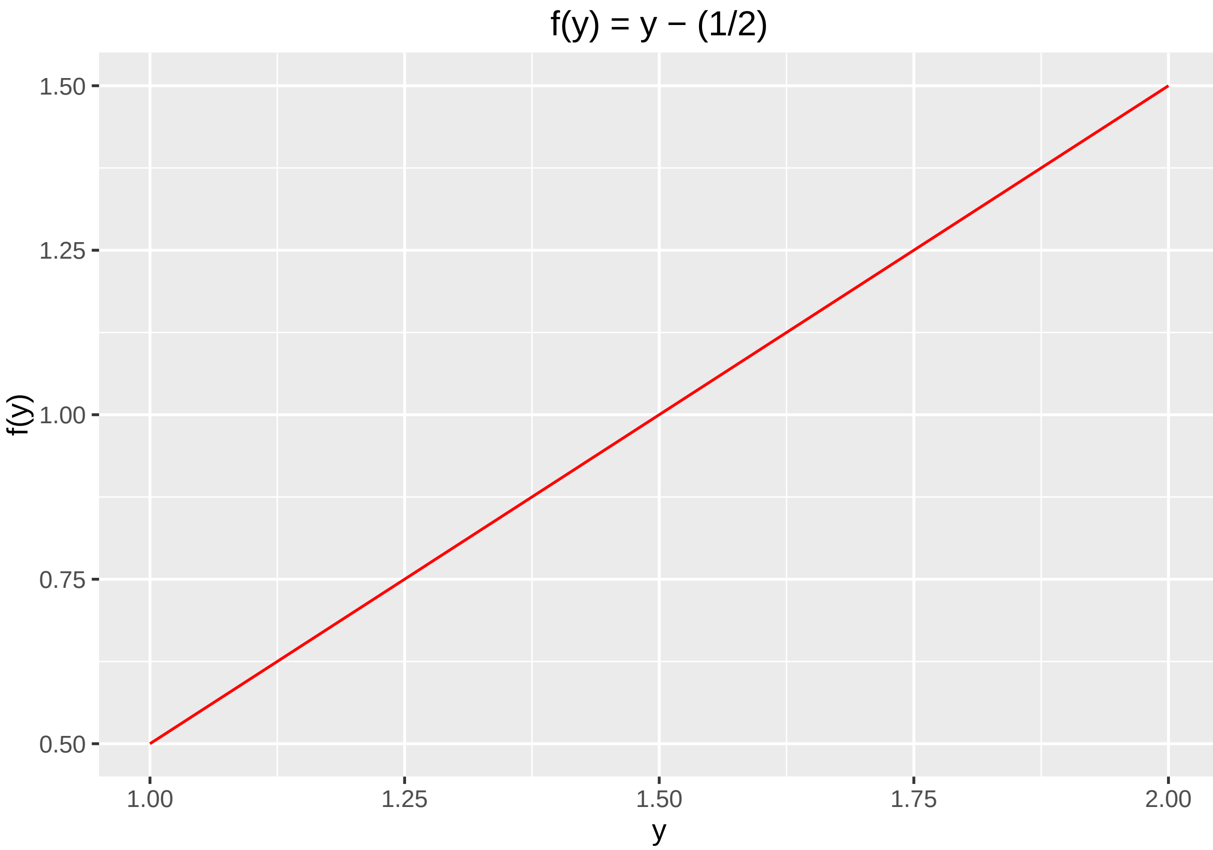
Finding $f_y(y)$

$$\int_0^1 (y-x)dx = \left[yx - \frac{x^2}{2}\right]_0^1 = y - \frac{1}{2}$$

Drawing $f_x(x)$ *in the interval* x*=[0,1]*



Drawing $f_y(y) in the interval y = [1, 2]$



Checking if X and Y are independent

X and Y are independent, if $f(X, Y) = f_x(x)f_y(y)$

Given that $f(x, y) = y - x$ and $f_x(x)f_y(y) = ((3/2) - x)(y - (1/2)) = (3/2)y - xy - (3/4) + (x/2)$

Since $f(X, Y) \neq f_x(x)f_y(y)$, x and y are not independent.

Computing $F_x(x)$

$$F_x(x) = \int_0^x f_x(x)dx$$

$$F_x(x) = \int_0^x ((3/2) - t)dt = \left[1.5t - (t^2/2)\right]_0^x = 1.5x - \frac{x^2}{2}$$

Therefore,

$$F_x(x) = 1.5x - \frac{x^2}{2} \text{ when } x \in [0, 1]$$

$$F_x(x) = 0 \text{ when } x \leq 0$$

$$F_x(x) = 1 \text{ when } x \geq 1$$

Computing $F_y(y)$

$$F_y(y) = \int_1^y f_y(y)dy$$

$$F_y(y) = \int_1^y (t-(1/2))dt = \left[t^2/2-(1/2)t\right]_1^y = 0.5y^2 - 0.5y$$

Therefore,

$$F_y(y) = 0.5y^2 - 0.5y \text{ when } y \in [1,2]$$
$$F_y(y) = 0 \text{ when } y \leq 1$$
$$F_y(y) = 1 \text{ when } y \geq 2$$

Computing $E[X]$

$$E[X] = \int_0^1 x.f_X(x)\ dx = \int_0^1 (3/2)x - x^2\ dx = \left[(3/4)x^2 - (x^3/3)\right]_0^1 = (3/4) - (1/3) = 5/12 = 0.417$$

Computing $V[X]$

$$V[X] = E[X^2] - (E[X])^2$$

$$E[X^2] = \int_0^1 x^2.f_X(x)\ dx = \int_0^1 (3/2)x^2 - x^3\ dx = \left[(1/2)x^3 - (x^4/4)\right]_0^1 = (1/2) - (1/4) = 1/4 = 0.25$$

$$V[X] = E[X^2] - (E[X])^2 = 0.25 - (5/12)^2 = 0.07638889$$

Computing $E[Y]$

$$E[Y] = \int_1^2 y.f_Y(y)\ dy = \int_1^2 y^2 - (y/2)\ dy$$
$$E[Y] = \left[(y^3/3) - (y^2/4)\right]_1^2 = (8/3) - (4/4) - (1/3) + (1/4) = (7/3) - (3/4) = 19/12 = 1.583333$$

Computing $V[Y]$

$$V[Y] = E[Y^2] - (E[Y])^2$$

$$E[Y^2] = \int_1^2 y^2.f_Y(y)\ dy = \int_1^2 y^3 - (1/2)y^2\ dy = \left[y^4/4 - (y^3/6)\right]_1^2 = 31/12$$

$$V[Y] = E[y^2] - (E[y])^2 = 31/12 - (19/12)^2 = 0.07638889$$

Computing $Cov(X,Y)$

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

$$E[XY] = \int_0^1 \int_1^2 (x.y.(y-x)\ dy)\ dx$$
$$E[XY] = \int_0^1 \int_1^2 (x.y^2 - x^2.y)\ dy)\ dx$$
$$E[XY] = \int_0^1 \left[(x.y^3/3) - (x^2.y^2/2)\right]_1^2\ \ dx = \int_0^1 (7x/3) - (3x^2/2)\ dx$$
$$E[XY] = \left[(7x^2/6 - (3x^3/6)\right]_0^1 = 2/3$$

$$Cov(X,Y) = E[XY] - E[X].E[Y] = (2/3) - (5/12)(19/12) = 0.006944444$$

Computing $Corr(X,Y)$

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{V[X]V[Y]}} = \frac{0.006944444}{\sqrt{(0.07638889)(0.07638889)}} = 0.09090908$$

4. Suppose that the following 10 observations come from some distribution (not highly skewed) with unknown mean μ.

   7.3 6.1 3.8 8.4 6.9 7.1 5.3 8.2 4.9 5.8

   Compute $\bar{X}$, S^2, and an approximate 95% confidence interval for μ.

Answer:

Let us read the given data to an R Vector

```
x <- c(7.3,6.1, 3.8, 8.4, 6.9, 7.1, 5.3, 8.2, 4.9, 5.8)
mean(x)
```

```
## [1] 6.38
```

```
var(x)
```

```
## [1] 2.161778
```

Therefore mean, $\overline{x} = 6.38$ and $s^2 = 2.161778$

To compute the 95% confidence interval for the population mean (μ), we need to obtain the z-scores at the p-values 0.025 and 0.975

```
z_1 = qnorm(0.025)
z_2 = qnorm(0.975)

x_1 = z_1*sd(x)+mean(x)
x_2 = z_2*sd(x)+mean(x)

print(x_1)
```

```
## [1] 3.498268
```

```r
print(x_2)
```

```
## [1] 9.261732
```

Therefore, the 95% confidence interval for μ is $[3.498268, 9.261732]$

5. A random variable X has the *memoryless property* if, for all s,t > 0,

$$\Pr(X > t + s|X > t) = \Pr(X > s)$$

Show that the exponential distribution has the memoryless property.

Answer:

We can express $Pr(X > t + s|X > t)$ as

$$Pr(X > t + s|X > t) = \frac{Pr(X > t + s \cap x > t)}{Pr(X > t)}$$

Since t and s are always positive, we can say that $(s+t)$ is always greater than t. Hence $Pr(X > s+t \cap X > t)$ must be equal to $Pr(X > s + t)$

Therefore,

$$Pr(X > t + s|X > t) = \frac{Pr(X > t + s)}{Pr(X > t)}$$

For the exponential distribution, the $c.d.f$ is

$$F(X \leq x) = 1 - e^{\frac{-x}{\beta}}$$

Therefore

$$F(X \geq x) = e^{\frac{-x}{\beta}}$$

Where β is the average of the exponential distribution.

Using exponential distribution $c.d.f$, we can write the above expression as:

$$Pr(X > t + s|X > t) = \frac{e^{\frac{-(t+s)}{\beta}}}{e^{\frac{-(t)}{\beta}}} = e^{\frac{-s}{\beta}} = Pr(X > s)$$

Hence, $Pr(X > t + s|X > t) = Pr(X > s)$, and the distribution is memoryless.

6. Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. Exp(λ=1). Use the Central Limit Theorem to find the approximate value of $\Pr(100 \leq \sum_{i=1}^{100} X_i \leq 110)$.

Answer:

Given that $\lambda = 1$. Therefore the mean $\mu = \frac{1}{\lambda} = 1$ and standard deviation $\sigma = \frac{1}{\lambda} = 1$

As per central limit theorem, if we draw random samples of size "n" from a distribution, then the mean of the samples will be approximately equal to the population mean, and the standard deviation of the sampling distribution will be $\sigma/\sqrt{n}$

In the problem, n = 100. Therefore standard deviation of the sampling distribution = $1/\sqrt{100} = 0.1$

$$Pr(100 \leq \sum_{i=1}^{100} X_i \leq 110) = Pr(1 \leq \sum_{i=1}^{100} \frac{X_i}{100} \leq 1.1)$$

(When dividing $100 \leq \sum_{i=1}^{100} X_i \leq 110$ by 100)

The z-scores of 1 and 1.1 (for a std. normal distribution with mean 1 and std. deviation of 0.1) are 0 and 1 respectively. The associated p-values for z-scores of 0 and 0.1 are 0.5 and 0.8413447 respectively. Therefore the $Pr(1 \leq \sum_{i=1}^{100} \frac{X_i}{100} \leq 1.1) = 0.8413447 - 0.5 = 0.3413447$

Question 7 - 5.13 problem in DES book

(a)

Given the $p.m.f$ as

$$p(x) = 1/(n+1)$$

where x = {0,1,2,3. . . n}

$$E[X] = \sum_{x=0}^{x=n} x.(1/(n+1)) = \sum_{x=0}^{x=n} x/(n+1)$$

$$E[X] = \frac{1}{n+1} + \frac{2}{n+1} + \frac{3}{n+1} + ... + \frac{n}{n+1}$$

$$E[X] = \frac{1}{n+1}(1+2+3+...+n) = \frac{1}{n+1}(\frac{n(n+1)}{2}) = n/2$$

$$E[X] = n/2$$

$$V[X] = E[X^2] - (E[X])^2$$

$$E[X^2] = \sum_{x=0}^{x=n} x^2.(1/(n+1)) = \sum_{x=0}^{x=n} x^2/(n+1)$$

$$E[X^2] = \frac{1^2}{n+1} + \frac{2^2}{n+1} + \frac{3^2}{n+1} + ... + \frac{n^2}{n+1}$$

$$E[X^2] = \frac{1}{n+1}(1^2+2^2+3^3+...+n^2) = \frac{1}{n+1}(\frac{n(n+1)(2n+1)}{6}) = n(2n+1)/6$$

$$V[X] = E[X^2] - (E[X])^2 = \frac{n(2n+1)}{6} - (\frac{n}{2})^2 = (2n^2/6) + (n/6) - (n^2/4) = \frac{n^2+2n}{12}$$

(b)

Given that $R_X = \{a, a+1, a+2, a+3, ..., b\}$

If this is a discrete uniform distribution, then the $p.m.f$ is the reciprocal of the number of elements in the sample space, and hence for the given distribution, it can be written as:

$$P(X = x) = \frac{1}{b - a + 1}$$

Let us find the E[X]:

$$E[X] = \sum_{i=a}^{b} i.\frac{1}{b - a + 1}$$

$$E[X] = a.\frac{1}{b-a+1} + (a+1).\frac{1}{b-a+1} + (a+2).\frac{1}{b-a+1} ... + b.\frac{1}{b-a+1}$$

$$E[X] = \frac{1}{b-a+1}[a + (a+1) + (a+2) + (a+3) + ... + b]$$

In the above expression, the sum of the consecutive numbers $[a + (a+1) + (a+2) + (a+3) + ... + b]$ can be written as the difference between the sum of first b natural numbers and the sum of first $(a-1)$ natural numbers.

$$[a + (a+1) + (a+2) + (a+3) + ... + b] = \frac{b.(b+1)}{2} - \frac{a.(a-1)}{2} = \frac{b^2 + b - a^2 + a}{2} = \frac{(b+a)[b-a+1]}{2}$$

Hence,

$$E[X] = \frac{1}{b-a+1}[a + (a+1) + (a+2) + (a+3) + ... + b] = \frac{1}{b-a+1}.\frac{(b+a)[b-a+1]}{2} = \frac{b+a}{2}$$

$$V[X] = E[X^2] - (E[X])^2$$

$$E[X^2] = \sum_{i=a}^{b} i^2 \frac{1}{b-a+1}$$

$$E[X^2] = a^2 \frac{1}{b-a+1} + (a+1)^2 \frac{1}{b-a+1} + (a+2)^2 \frac{1}{b-a+1} + ... + b^2 \frac{1}{b-a+1}$$

$$E[X^2] = \frac{1}{b-a+1}[a^2 + (a+1)^2 + (a+2)^2 + ... + b^2]$$

$$\frac{1}{b-a+1}[a^2 + (a+1)^2 + (a+2)^2 + ... + b^2] = \frac{1}{b-a+1}[\frac{b.(b+1)(2b+1)}{6} - \frac{(a-1)(a)(2a-1)}{6}]$$

$$V[X] = E[X^2] - E[X] = \frac{1}{b-a+1}[\frac{b.(b+1)(2b+1)}{6} - \frac{(a-1)(a)(2a-1)}{6}] - \frac{b^2 + a^2 + 2ab}{4}$$

Upon simplifying, we will get:

$$V[X] = \frac{(b-a+1)^2 - 1}{12}$$

Alternate proof:

If we have $R_x = \{0, 1, 2, 3...n\}$ as discrete uniform distribution sample space, then the expected value was computed as $n/2$. Now if we increase/decrease each element by a constant in the sample space by a constant a, then the E[X] will also get effected by the same value (increase/decrease by a). If we add a to all the elements of $R_x = \{0, 1, 2, 3...n\}$, we get $R_x = \{a, a+1, a+2...a+n\}$, which can be written as $R_x = \{a, a+1, a+2...b\}$, when you assume $a+n$ as b.

The number of elements will be $b-a+1$ (The number of elements in the distribution $R_x = \{0, 1, 2, 3...n\}$ is $n+1$. Therefore $b-a+1 = n+1$, and $b-a = n$).

The expected value will be (substitute $n = b-a$ in $n/2$:

$$E[X] = ((b-a+1)-1)/2 + a$$

(since the average of n+1 elements $R_x = \{0, 1, 2, 3...n\}$ is $n/2$, and also the resulting mean has to be transformed by a)

$$E[X] = ((b-a)/2 + a = (b+a)/2$$

The V[X] does not change, even though when we increment the distribution by a constant number a. So for the uniform discrete distribution (with $n+1$ elements) $R_x = \{0, 1, 2, 3...n\}$, we know that the $V[X] = \frac{n^2+2n}{12}$. So when we increment each value of $R_x = \{0, 1, 2, 3...n\}$ by a, the number of elements will be $b-a+1$. Hence, the V[X] will be

$$V[X] = \frac{(b-a)^2 + 2(b-a)}{12}$$

$$V[X] = \frac{(b-a)^2 + 2(b-a)}{12}$$

Adding and subtracting 1 to the numerator, will give:

$$V[X] = \frac{(b-a)^2 + 2(b-a) + 1 - 1}{12}$$

$$V[X] = \frac{(b-a+1)^2 - 1}{12}$$

Question 7 - 5.14 problem in DES book

Given the exponential distribution with $\lambda = 0.4$. Therefore mean $= 1/\lambda = 1/0.4 = 2.5 years$

(a)

The probability that the satillite lives more than 5 years is:

$$Pr(X > 5) = \int_5^\infty \frac{e^{\frac{-x}{2.5}}}{2.5} \text{ dx} = e^{-2} = 0.1353353$$

(b)

The probability that the satillite dies between 3 and 6 years is:

$$Pr(3 < X < 6) = \int_3^6 \frac{e^{\frac{-x}{2.5}}}{2.5} \text{ dx}$$

$$Pr(3 < X < 6) = e^{\frac{-3}{2.5}} - e^{\frac{-6}{2.5}} = 0.2104763$$

Question 7 - 5.39 problem in DES book

(a).

The combined linkage will have normal distribution $N(150, 0.25)$. This normal distribution's mean and variance are obtained by adding together the means and variances or the given 3 distributions respectively.

(b)

The $P(Linkage > 150.2)$ can be found, by first identifying the z-score of 150.2 cm (for the mean 150 and variance of 0.25 normal distribution).

$$Z_{150.2} = \frac{150.2 - 150}{0.5} = 0.4$$

The area under the std. normal curve (for Z>0.4) is $1 - p(Z < 0.4) = 1 - 0.6554217 = 0.3445783$

(c)

$$Z_{149.83} = \frac{149.83 - 150}{0.5} = -0.34$$

$$Z_{150.21} = \frac{150.21 - 150}{0.5} = 0.42$$

p-value of -0.34 is $p(Z < -0.34) = 0.3669283$ and p-value of 0.42 is $P(Z < 0.42) = 0.6627573$

Therefore the proportion of links which are between 149.83 cm and 150.2 cm is: $0.6627573 - 0.3669283 = 0.295829$