# Problem Set 6 - Solutions

Statistics - NYU, Summer 2016
Ercan Karadas

## Section 1

**[1]** a) X has a Poisson distribution with $\lambda = 5$ and number of successes we would like to find is 7, i.e. $x = 7$.
$P(X = 7) = \frac{e^{-5}5^7}{7!} = 0.1044$

b) $P(X \le 7) = 0.867$.

c) $P(X > 7) = 1 - P(X \le 7) = 1 - 0.867 = .133$

d) $P(4 \le X \le 9) = P(X \le 9) - P(X \le 3) = .968 - .265 = .703$.

**[2]** a) We should use the Poisson probability distribution because the random variable $X$ is the number of occurrences of a certain event (delivery failures) in a given continuous interval (one day).

b) First, use the given data to determine an estimate for $\lambda$, the expected number of failures per day:

$$\lambda = \frac{15 + 16 + \ldots + 11}{20} = 10.75.$$

Then

$$\begin{aligned} P(X \le 10) &= 1 - P(X < 10) \\ &= 1 - P(X = 0) - P(X = 1) - \ldots - P(X = 9) \\ &= 0.6318 \end{aligned}$$

c)

$$\begin{aligned} P(X < 6) &= P(X = 0) + P(X = 1) + \ldots + P(X = 5) \\ &= 0.0435 \end{aligned}$$

**[3]** a) $Y_1 = 3 - 7X$

| $Y_1$ | 17 | -4 | -11 |
|---|---|---|---|
| $P(Y_1)$ | 0.5 | 0.3 | 0.2 |

b) $Y_2 = 0.5X$

| $Y_2$ | -1 | 0.5 | 1 |
|---|---|---|---|
| $P(Y_2)$ | 0.5 | 0.3 | 0.2 |

c) $Y_3 = X^2 + X - 2$

| $Y_3$ | 0 | 4 |
|---|---|---|
| $P(Y_3)$ | 0.8 | 0.2 |

Note: $Y_3 = 0$ when $X = -2$ or $X = 1$. Therefore $P(Y_3 = 0) = P(X = -2 \text{ or } X = 1) = 0.5 + 0.3 = 0.8$.

d) $Y_4 = X^2$

| $Y_4$ | 4 | 1 |
|---|---|---|
| $P(Y_4)$ | 0.7 | 0.3 |

Note: $Y_4 = 4$ when $X = -2$ or $X = 2$. Therefore $P(Y_4 = 4) = P(X = -2 \text{ or } X = 2) = 0.5 + 0.2 = 0.7$.

**[4]** a) $\mu_{Y_1} = \sum_{y_1 \in \{17,-4,-11\}} y_1 P(Y_1 = y_1) = 17(0.5) - 4(0.3) - 11(0.2) = 5.1$

$$\begin{aligned}\sigma^2_{Y_1} &= \sum_{y_1 \in \{17,-4,-11\}} (y_1 - \mu_{Y_1})^2 P(Y_1 = y_1) \\ &= (17 - 5.1)^2(0.5) + (-4 - 5.1)^2(0.3) + (-11 - 5.1)^2(0.2) \\ &= 147.49\end{aligned}$$

b) $\mu_{Y_2} = \sum_{y_2 \in \{-1,0.5,1\}} y_2 P(Y_2 = y_2) = -1(0.5) + 0.5(0.3) + 1(0.2) = -0.15$

$$\begin{aligned}\sigma^2_{Y_2} &= \sum_{y_2 \in \{-1,0.5,1\}} (y_2 - \mu_{Y_2})^2 P(Y_2 = y_2) \\ &= (-1 - (-0.15))^2(0.5) + (0.5 - (-0.15))^2(0.3) + (1 - (-0.15))^2(0.2) \\ &= 0.7525\end{aligned}$$

c) $\mu_{Y_3} = \sum_{y_3 \in \{0,4\}} y_3 P(Y_3 = y_3) = 0(0.8) + 4(0.2) = 0.8$

$$\begin{aligned}\sigma^2_{Y_3} &= \sum_{y_3 \in \{0,4\}} (y_3 - \mu_{Y_3})^2 P(Y_3 = y_3) \\ &= (0 - 0.8)^2(0.8) + (4 - 0.8)^2(0.2) \\ &= 2.56\end{aligned}$$

d) $\mu_{Y_4} = \sum_{y_4 \in \{4,1\}} y_4 P(Y_4 = y_4) = 4(0.7) + 1(0.3) = 3.1$

$$\begin{aligned}\sigma^2_{Y_4} &= \sum_{y_4 \in \{4,1\}} (y_4 - \mu_{Y_4})^2 P(Y_4 = y_4) \\ &= (4 - 3.1)^2(0.7) + (1 - 3.1)^2(0.3) \\ &= 1.89\end{aligned}$$

**[5]** a) $E[Y_1] = E[3 - 7X] = 3 - 7\mu_X = 3 - 7(3) = -18$
$Var(Y_2) = Var(3 - 7X) = 7^2\sigma^2_X = 49(1) = 49$

b) $E[Y_2] = E[0.5X] = 0.5\mu_X = 0.5(3) = 1.5$
$Var(Y_2) = Var(0.5X) = 0.5^2\sigma^2_X = 0.25(1) = 0.25$

c) $E[Y_3] = E[X^2 + X - 2] = E[X^2] + \mu_X - 2$

We can use the following useful expression to help us find the value of $E[X^2]$:

$$\begin{aligned}Var(X) &= E[X^2] - (E[X])^2 \\ 1 &= E[X^2] - 3^2 \\ E[X^2] &= 10\end{aligned}$$

Thus $E[Y_3] = 10 + 3 - 2 = 11$.

$Var(Y_3)$ cannot be calculated with the given information.

d) $E[Y_4] = E[X^2] = 10$ (calculated above in c).
$Var(Y_4)$ cannot be calculated with the given information.

We were lucky we were able to calculate the value of $E[X^2]$, but calculating the expected value of a non-linear transformation is not possible in general. For those of you interested in the derivation of the expression used, note that:

$$\begin{aligned}\sigma^2_X &= \sum_x (x - \mu_X)^2 P(X = x) \\ &= \sum_x (x^2 - 2x\mu_X + \mu^2_X) P(X = x) \\ &= \sum_x x^2 P(X = x) - 2\mu_X \sum_x x P(X = x) + \mu^2_X \sum_x P(X = x) \\ &= E[X^2] - 2\mu^2_X + \mu^2_X \\ &= E[X^2] - \mu^2_X\end{aligned}$$

**Note:** The whole point of this exercise is to show that once we have the p.d.f. of a discrete r.v. we can calculate p.d.f of any generated new r.v., no matter it is linear or not. And since we have the p.d.f. of the generated r.v. we can compute its expected value and variance as usual (that is what we did in problem 2 and 3). But when we only know $\mu_x$ and $\sigma_x^2$, but not the p.d.f. of $X$, we have the following observations regarding the generated r.v., say $Y$:

- $P(y)$ : we cannot compute p.d.f. of the generated r.v., no matter whether it is linear nor nonlinear

- $E(Y)$: if the generated r.v. is linear, i.e $Y = a + bX$, we can compute its expected value from the formula $E(Y) = a + bE(X)$ but when it is nonlinear we cannot compute the expected value in general, except in some special cases like in 4.c.

- $Var(Y)$: if the generated r.v. is linear, i.e. like $Y = a + bX$, we can compute its variance from the formula $Var(Y) = b^2Var(X)$ but when it is nonlinear we cannot compute its variance in general.

These observations apply also to the random variables that are generated from jointly distributed two random variables (see problems 6 and 7).

**[6]** a)

| $X$ | 1 | 3 |
|---|---|---|
| $P(X)$ | 0.5 | 0.5 |

| $Y$ | 0 | 2 |
|---|---|---|
| $P(Y)$ | 0.55 | 0.45 |

b)

$$\mu_X = \sum_{x\in\{1,3\}} xP(X=x) = 1(0.5) + 3(0.5) = 2$$
$$\mu_Y = \sum_{y\in\{0,2\}} yP(Y=y) = 0(0.55) + 2(0.45) = 0.9$$

$$\sigma_X^2 = \sum_{x\in\{1,3\}} (x-\mu_X)^2P(X=x) = (1-2)^2(0.5) + (3-2)^2(0.5) = 1$$
$$\sigma_Y^2 = \sum_{y\in\{0,2\}} (y-\mu_Y)^2P(Y=y) = (0-0.9)^2(0.55) + (2-0.9)^2(0.45) = 0.99$$

$$\begin{aligned} cov(X,Y) &= \sum_x\sum_y(x-\mu_X)(y-\mu_Y)P(X=x,Y=y) \\ &= (1-2)(0-0.9)(0.3) + (1-2)(2-0.9)(0.2) \\ &= +(3-2)(0-0.9)(0.25) + (3-2)(2-0.9)(0.25) \\ &= 0.1 \end{aligned}$$

$$corr(X,Y) = \frac{cov(X,Y)}{\sigma_X\sigma_Y} = \frac{0.1}{\sqrt{1}\sqrt{0.99}} = 0.101$$

c) No, they are not independent because

$$P(X = 1, Y = 0) \neq P(X = 1) \cdot P(Y = 0)$$
$$0.3 \neq (0.5) \cdot (0.55)$$

Alternatively, since we have already computed the covariance, we could simply say they are not independent because the covariance is not 0.

d) $E[W] = E[2X + Y] = 2\mu_X + \mu_Y = 2(2) + 0.9 = 4.9$

$$\begin{aligned} Var(W) &= Var(2X + Y) \\ &= 2^2\sigma_X^2 + \sigma_Y^2 + 2(2)(1)cov(X, Y) \\ &= 4(1) + 0.99 + 4(0.1) \\ &= 5.39 \end{aligned}$$

e) The probability distribution function for $W_3$ is as follows:

| $W_3$ | 1 | 3 | 9 | 11 |
|---|---|---|---|---|
| $P(W_3)$ | 0.3 | 0.2 | 0.25 | 0.25 |

Then the mean and variance are as follows:

$$\mu_{W_3} = \sum_{w_3 \in \{1,3,9,11\}} w_3 P(W_3 = w_3) = 1(0.3) + 3(0.2) + 9(0.25) + 11(0.25) = 5.9$$

Alternatively, we could have done the following:

$E[W_3] = E[X^2 + Y] = E[X^2] + \mu_Y$

$$\begin{aligned} Var(X) &= E[X^2] - (E[X])^2 \\ 1 &= E[X^2] - 2^2 \\ E[X^2] &= 5 \end{aligned}$$

Thus $E[W_3] = 5 + 0.9 = 5.9$.

$$\begin{aligned} \sigma_{W_3}^2 &= \sum_{w_3 \in \{1,3,9,11\}} (w_3 - \mu_{W_3})^2 P(W_3 = w_3) \\ &= (1 - 5.9)^2(0.3) + \cdots + (11 - 5.9)^2(0.25) \\ &= 17.79 \end{aligned}$$

**[7]** a)

| $A(row)/B$ | \$35 | \$55 | \$80 | Marg. of A |
|---|---|---|---|---|
| \$50 | .10 | .15 | .05 | .30 |
| \$60 | .10 | .20 | .10 | .40 |
| \$70 | .05 | .15 | .10 | .30 |
| Marg. of B | .25 | .50 | .25 | 1.00 |

b) Since the portfolio can be expressed as $W = 5A + 10B$, we have $\mu_w = 5\mu_A + 10\mu_B$. Therefore, first we need to find $\mu_A$ and $\mu_B$:
$\mu_A = 50(.3) + 60(.4) + 70(.3) = 60$, and
$\mu_B = 35(.25) + 55(.5) + 80(.25) = 56.3$.
Now we can compute the expected value of the portfolio as
$\mu_w = 5(60) + 10(56.3) = 862.5$

c) $\sigma_w^2 = 5^2\sigma_B + 10^2\sigma_B + 2(5)(10)Cov(A, B)$.
Let us first find Var(A), Var(B), and Cov(A,B):
$\sigma_A^2 = (50 - 60)^2 0.3 + (60 - 60)^2 0.4 + (70 - 60)^2 0.3 = 60$, $\sigma_B^2 = 254.7$
$\sigma_B^2 = (35 - 56.3)^2(0.25) + (55 - 56.3)^2(0.5) + (80 - 56.3)^2(0.25) = 254.7$

$$\begin{aligned} Cov(A, B) =& (50 - 60)(35 - 56.3)(0.10) + (60 - 60)(35 - 56.3)(0.10) \\ &+ (70 - 60)(35 - 56.3)(0.05) + \ldots + (70 - 60)(80 - 56.3)(0.10) \\ &= 22.5 \end{aligned}$$

d) Finally,

$$\begin{aligned} Var(W) =& 5^2 Var(A) + 10^2 Var(B) + 2(5)(10)Cov(A, B) \\ =& 5^2(60) + 10^2(254.69) + 2(5)(10)(22.5) \\ =& 29,219 \end{aligned}$$

$\sigma_w = \sqrt{Var(W)} = \sqrt{29,219} = 170.9$

e) Let us first construct the probability distribution table for the random variable $A|B = 55\$$

| $A\|B = 55$ | 50 | 60 | 70 |
|---|---|---|---|
| $P(A\|B = 55)$ | $.15/.50$ | $.20/.50$ | $.15/.50$ |

Then, we can compute the expected value of this r.v. as usual:

$$E(A|B = 55) = 50(\frac{0.15}{0.50}) + 60(\frac{0.20}{0.50}) + 70(\frac{0.15}{0.50}) = 60$$

# Section 2

**[8]**

a) $\lambda = 7$ per-minute

$$P(X = 3) = \frac{e^{-7}7^3}{3!} = 0.0521$$

b)

$$P(X = 0) + P(X = 1) + P(X = 2) = \frac{e^{-7}7^0}{0!} + \frac{e^{-7}7^1}{1!} + \frac{e^{-7}7^2}{2!} = 0.0287$$

c) $\lambda = 35$ per-five minute, so

$$P(X = 10) = \frac{e^{-35}35^{10}}{10!}$$

d) To get 3 calls, the web should be visited by 30 visitors in the previous ten minutes. $\lambda = 7 * 10 = 70$ in this case, so answer:

$$P(X = 30) = \frac{e^{-70}70^{30}}{30!}$$

**[9]** Consider a r.v. $X$

| $X$ | -1 | 1 |
|---|---|---|
| P(X) | .4 | .6 |

And define a new r.v. as $Y = X^2$. Then $Y$ is not random anymore because $Y = 1$ with probability 1, no matter whether $X = -1$ or $X = 1$.

**[10]** a) $E[W_1] = E[3X + Y] = 3\mu_X + \mu_Y = 3(0.7) + 1.1 = 3.2$

$$\begin{aligned} Var(W_1) &= Var(3X + Y) \\ &= 3^2\sigma_X^2 + \sigma_Y^2 + 2(3)(1)cov(X, Y) \\ &= 9(0.04) + 0.09 + 6(-0.03) \\ &= 0.27 \end{aligned}$$

b) $E[W_2] = E[Y - 0.5X] = \mu_Y - 0.5\mu_X = 1.1 - 0.5(0.7) = 0.75$

$$\begin{aligned} Var(W_2) &= Var(Y - 0.5X) \\ &= \sigma_Y^2 + 0.5^2\sigma_X^2 - 2(1)(0.5)cov(X, Y) \\ &= 0.09 + 0.25(0.04) - (-0.03) \\ &= 0.13 \end{aligned}$$

c) $E[W_3] = E[X^2 + Y] = E[X^2] + \mu_Y$
As we did before in Question 4, we can use the expression:

$$\begin{aligned} Var(X) &= E[X^2] - (E[X])^2 \\ 0.04 &= E[X^2] - 0.7^2 \\ E[X^2] &= 0.53 \end{aligned}$$

Thus $E[W_3] = 0.53 + 1.1 = 1.63$.
$Var(W_3)$ cannot be calculated with the given information.

d) Knowing the probability distribution function (p.d.f) of $W_3$ allowed us to compute its mean and variance in 6.e. But here, we do not have that information, so in general we know that we cannot compute its expected value or variance (see the note at the end of problem 5). But because of its special form, we were still able to compute its expected value thanks to the formula $Var(X) = E[X^2] - (E[X])^2$.

**[11]**

| $Y(row)/X$ | 1 | 2 | 3 |
|---|---|---|---|
| 0 | .10 | .12 | .06 |
| 1 | .05 | .10 | .11 |
| 2 | .02 | .16 | .28 |

a)

| $X$ | 1 | 2 | 3 |
|---|---|---|---|
| $P(x)$ | .17 | .38 | .45 |

b) $\mu_X = \sum_{x \in \{1,2,3\}} xP(X = x) = 1(0.17) + 2(0.38) + 3(0.45) = 2.28$

c)

$$\begin{aligned} \sigma_X^2 &= \sum_{x \in \{1,2,3\}} (x - \mu_X)^2 P(X = x) \\ &= (1 - 2.28)^2(0.17) + (2 - 2.28)^2(0.38) + (3 - 2.28)^2(0.45) \\ &= 0.54 \end{aligned}$$

Therefore, $\sigma_X = \sqrt{0.54} = 0.73$

d) First calculate $\mu_Y = 1.18$ as in part (a). Then

$$
\begin{aligned}
Cov(X,Y) &= \sum_{x\in\{1,2,3\}} \sum_{y\in\{0,1,2\}} (x-\mu_X)(y-\mu_Y)P(x,y) \\
&= (1-2.28)(0-1.18)(0.1) + (2-2.28)(0-1.18(0.12) \\
&+ \ldots + (3-2.28)(2-1.18)(0.28) \\
&= 0.2496
\end{aligned}
$$

e) First construct the prob. dist. table for the r.v. $X|Y=2$

| $X\|Y=2$ | 1 | 2 | 3 |
|---|---|---|---|
| $P(x\|Y=2)$ | 0.2/0.46 | 0.16/0.46 | 0.28/0.46 |

Then

$$E(X|Y=2) = 1(0.2/0.46) + 2(0.16/0.46) + 3(0.28/0.46) = 2.565$$

Alternatively, you can directly apply the formula

$$
\begin{aligned}
E(X|Y=2) &= \sum_{x\in\{1,2,3\}} x \cdot P(X=x|Y=2) \\
&= 1(0.2/0.46) + 2(0.16/0.46) + 3(0.28/0.46) \\
&= 2.565
\end{aligned}
$$

f)

$$\mu_W = 2\mu_X + \mu_Y = 2(2.28) + 1(1.18) = 5.74$$

g) There are three combinations of $(X,Y)$ that gives $X+Y=3$:

$$(X=1, Y=2), (X=2, Y=1), \text{ and } (X=3, Y=0)$$

To calculate $E(W|X+Y=3)$: First compute $W = 2X+Y$ for each of these three pairs, and then weight with the conditional probabilities and add them up:

$$
\begin{aligned}
E(W|X+Y=3) &= [2(1)+2]\frac{0.02}{0.02+0.1+0.06} + [2(2)+1]\frac{0.1}{0.02+0.1+0.06} + \ldots \\
&+ [2(3)+0]\frac{0.06}{0.02+0.1+0.06} \\
&= 5.22
\end{aligned}
$$

**[12]**

| $C(row)/R$ | \$10 | \$15 | \$20 |
|---|---|---|---|
| \$12 | .15 | .10 | .05 |
| \$14 | .20 | .15 | .10 |
| \$16 | .05 | .05 | .15 |

a)

| $C$ | \$12 | \$14 | \$16 |
|---|---|---|---|
| $P(C)$ | .30 | .45 | .25 |

| $R$ | \$10 | \$15 | \$20 |
|---|---|---|---|
| $P(R)$ | .40 | .30 | .30 |

b)

$$E(C) = 12 \times .30 + 14 \times .45 + 16 \times .25 = 13.9$$

c) Since $E(\pi) = E(R) - E(C)$, first compute

$$E(R) = 10 \times .40 + 15 \times .30 + 20 \times .30 = 14.5$$

then conclude

$$E(\pi) = 14.5 - 13.9 = 0.6$$

d)

$$P(loss = 6|\pi < 0) = \frac{.05}{.15 + .20 + .05 + .05} = \frac{.05}{.45} = \frac{1}{9} = 0.11$$

e) Firm makes loss with .45 probability so if we define a r.v. as

$$X = \# \text{ of days with loss}$$

then the answer is

$$P(X = 2) = \binom{3}{2}(.45)^2(.55)^1 = .334$$

f) For a randomly chosen day, find the expected profit of the company given that its cost is \$14 in that day, $E(\pi|C = 14) =?$ (5p)

$$E(\pi|C = 14) = \frac{(10 - 14) \times (.2) + (15 - 14) \times (.15) + (20 - 14) \times (.10)}{.20 + .15 + .10} = -0.11$$