CS200 HW4

Tanjin He

March 2019

1 Question 1

(a) X $\sim B(n, p)$
$E[X] = 5p$

(b) Var(X) = 5p(1-p)

2 Question 2

$$\begin{aligned}
Var(X+Y) &= E((X+Y-\overline{X+Y})^2) \\
&= E((X+Y)^2) - (E(X+Y))^2 \\
&= E(X^2+2XY+Y^2) - (E(X)+E(Y))^2 \\
&= E(X^2) + 2E(XY) + E(Y^2) \\
&\quad - (E(X)^2 + 2E(X)E(Y) + E(Y)^2) \\
&= E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 \\
&\quad + 2E(XY) - 2E(X)E(Y)
\end{aligned}$$

Since X and Y are independent, E(XY)= E(X)E(Y)

$$\begin{aligned}
Var(X+Y) &= E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 \\
&= E((X-\bar{X})^2) + E((Y-\bar{Y})^2) \\
&= Var(X) + Var(Y)
\end{aligned}$$

3 Question 3

(a)

$$\begin{aligned} E(Y) &= E(X_1X_2) \\ &= E(X_1)E(X_2) \end{aligned}$$

Since X_1 is the number of spots form the fair die,

$$\begin{aligned} E(X_1) &= \sum_{i=1}^{6}(p_i i) \\ &= \frac{1}{6}(1+2+3+4+5+6) \\ &= 3.5 \end{aligned}$$

For X_2,

$$\begin{aligned} E(X_2) &= \sum_{i=1}^{6}(p_i i) \\ &= \frac{1}{16}*1+\frac{1}{16}*2+\frac{3}{16}*3+\frac{3}{16}*4+\frac{4}{16}*5+\frac{4}{16}*6 \\ &= 4.25 \end{aligned}$$

Then,

$$\begin{aligned} E(Y) &= 3.5*4.25 \\ &= 14.875 \end{aligned}$$

(b)

$$\begin{aligned} Var(Y) &= Var(X_1X_2) \\ &= E((X_1X_2 - E(Y))^2) \\ &= E(X_1^2X_2^2) - (E(Y))^2 \end{aligned}$$

Since X_1 and X_2 are independent,

$$\begin{aligned} Var(Y) &= E(X_1^2X_2^2) - (E(Y))^2 \\ &= E(X_1^2)E(X_2^2) - (E(Y))^2 \end{aligned}$$

For X_1,

$$\begin{aligned} E(X_1^2) &= \sum_{i=1}^{6}(p_i i^2) \\ &= \frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2) \\ &= \frac{91}{6} \end{aligned}$$

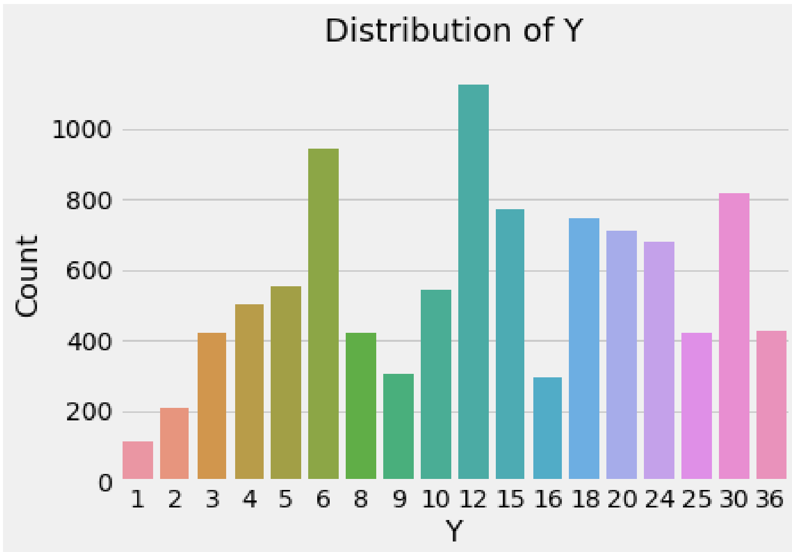
For X_2,

$$\begin{aligned} E(X_2^2) &= \sum_{i=1}^{6}(p_i i^2) \\ &= \frac{1}{16}*1^2+\frac{1}{16}*2^2+\frac{3}{16}*3^2+\frac{3}{16}*4^2+\frac{4}{16}*5^2+\frac{4}{16}*6^2 \\ &= 20.25 \end{aligned}$$

Then,

$$\begin{aligned} Var(Y) &= \frac{91}{6}*20.25-14.875^2 \\ &= 85.86 \end{aligned}$$

(c)
Distribution:



The mean and variance of the sampled Y's are 14.78 and 85.90, respectively, which are similar to the ideal values calculated in (a) and (b).

Code:

```
import numpy as np
np.set_printoptions(threshold=np.nan)

import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

X1 = np.random.choice(np.arange(1, 7), 10000)
X2 = np.random.choice(np.arange(1, 7), 10000,
    p=[1/16, 1/16, 3/16, 3/16, 4/16, 4/16])

Y = X1*X2
e_Y = np.mean(Y)
var_Y = np.var(Y)

fig = plt.figure()
ax = fig.add_subplot(111)
p1 = sns.countplot(Y)
plt.title('Distribution of Y', size=18, y=1.03)
plt.xlabel('Y', size=16)
plt.ylabel('Count', size=16);

print('e_Y: ', e_Y)
```

```python
print('var_Y:', var_Y)
```