Student Information

Full Name : Feyza Özlü
Id Number : e2580868

Answer 1

a)

I will use $P(x) = P\{X = x\} = N/x$ formula for each of probabilities.

$P(1) = P\{X = 1\} = \frac{N}{1}$

$P(2) = P\{X = 2\} = \frac{N}{2}$

$P(3) = P\{X = 3\} = \frac{N}{3}$

$P(4) = P\{X = 4\} = \frac{N}{4}$

$P(5) = P\{X = 5\} = \frac{N}{5}$

Since all possibilities are {1,2,3,4,5} sum of their individual probabilities should be equal to 1.

$$\begin{aligned}
P(1) + P(2) + P(3) + P(4) + P(5) &= 1 \\
\frac{N}{1} + \frac{N}{2} + \frac{N}{3} + \frac{N}{4} + \frac{N}{5} &= 1 \\
N \cdot \frac{137}{60} &= 1 \\
N &= \frac{60}{137} \\
N &\approx 0.438
\end{aligned}$$

b)

First let me write all individual probabilities.

$P(1) = \frac{60}{137}$, $P(2) = \frac{30}{137}$, $P(3) = \frac{20}{137}$, $P(4) = \frac{15}{137}$, $P(5) = \frac{12}{137}$

By the expected value formula for discrete values in the textbook page 48 (33.3)

$$E(x) = \sum_x x.P(x) = 1.\frac{60}{137} + 2.\frac{30}{137} + 3.\frac{20}{137} + 4.\frac{15}{137} + 5.\frac{12}{137} = \frac{300}{137} \approx 2.190$$

c)

By the variance formula for discrete random variables in the textbook page 50 definition 3.6

$Var(x) = E(x^2) - (E(x))^2$

Let me first calculate

$$E(x^2) = \sum_x x^2.P(x) = 1^2.\frac{60}{137} + 2^2.\frac{30}{137} + 3^2.\frac{20}{137} + 4^2.\frac{15}{137} + 5^2.\frac{12}{137} = \frac{900}{137}$$

We already know $E(x) = \frac{300}{137}$ from the previous question. Then:

$$\begin{aligned} Var(x) &= E(x^2) - (E(x))^2 \\ &= \frac{900}{137} - \frac{300^2}{137^2} \\ &= \frac{123300 - 90000}{18769} \\ &= \frac{33300}{18769} \\ &\approx 1.774 \end{aligned}$$

d)

First let me write all individual probabilities for all y's.

$P(1) = \frac{1}{15}$, $P(2) = \frac{2}{15}$, $P(3) = \frac{3}{15}$, $P(4) = \frac{4}{15}$, $P(5) = \frac{5}{15}$

And now let me calculate

$$E(Y) = \sum_y y.P(Y) = 1.\frac{1}{15} + 2.\frac{2}{15} + 3.\frac{3}{15} + 4.\frac{4}{15} + 5.\frac{5}{15} = \frac{55}{15} = \frac{11}{3}$$

By part b we know that $E(X) = \frac{300}{137} \approx 2.190$

Since the joint distribution is given by $P(x, y) = P(x)P(y)$

$$
\begin{aligned}
E(XY) &= \sum_{x=1}^{5}\sum_{y=1}^{5} x \cdot y \cdot P(x,y) \\
&= \sum_{x=1}^{5}\sum_{y=1}^{5} x \cdot y \cdot P(x) \cdot P(y) \\
&= \sum_{x=1}^{5}\sum_{y=1}^{5} x \cdot y \cdot \frac{N}{x} \cdot \frac{y}{15} \\
&= \sum_{x=1}^{5}\sum_{y=1}^{5} \frac{N \cdot y^2}{15} \\
&= \frac{N}{15}\sum_{x=1}^{5}\sum_{y=1}^{5} y^2 \qquad \text{since } \frac{N}{15} \text{ is constant we can take it out.} \\
&= \frac{N}{15}\sum_{x=1}^{5}(1+4+9+16+25) \\
&= \frac{N}{15} \cdot 5 \cdot (1+4+9+16+25) \qquad \text{we know N=} \frac{60}{137} \text{ from part a} \\
&= \frac{\frac{60}{137}}{15} \cdot 5 \cdot (1+4+9+16+25) \\
&= \frac{4}{137} \cdot 5 \cdot (55) \\
&= \frac{1100}{137}
\end{aligned}
$$

$$Cov(X,Y) = E(XY) - E(X) \cdot E(Y) = \tfrac{1100}{137} - \tfrac{300}{137} \cdot \tfrac{11}{3} = 0$$

We can say that since $Cov(X,Y) = 0$, X and Y are independent i.e. there is no correlation between them.

Answer 2

a)

Let's say the probability of success for an individual attempt is p. We are looking for the p value such that the probability that at least one attempt is successful in 1000 trials is 95%.

$P(x \geq 1) = 1 - F(0) = \frac{95}{100} \quad (1)$

$F(0)$ means no attempts are successful. We can calculate $F(0) = (1-p)^{1000}$

Back to (1):

$$1 - F(0) = 1 - (1-p)^{1000} = \frac{95}{100}$$

$$(1-p)^{1000} = 1 - \frac{95}{100}$$

$$(1-p)^{1000} = \frac{5}{100}$$

$$(1-p) = \sqrt[1000]{0.05}$$

$$(1-p) \approx 0.997$$

$$p \approx 0.003$$

b)

i)

For this question let X be the number of games played until winning twice. It is a number of trials needed to see 2 successes, hence X has Negative Binomial distribution with k = 2 and p = 0.003.

We need to find:

$$P(X > 500) = 1 - P(X \leq 500)$$

However, there is no table of Negative Binomial distribution in the Appendix, and applying the formula for $P(X \leq 500) = P(1) + P(2) + P(3) + P(4) + ... + P(500)$ is very difficult to calculate.

Although X is not Binomial at all, the probability $P\{X > 500\}$ can be related to some Binomial variable. In our example:

$$\begin{aligned} P\{X > 500\} &= P\{\text{more than 500 trials needed to get 2 successes}\} \\ &= P\{\text{there are fewer than 2 successes in 500 trials}\} \\ &= P\{Y < 2\} \end{aligned}$$

where Y is the number of successes in 500 trials, which is a Binomial variable with parameters $n = 500$ and $p = 0.003$.

If we calculate $binocdf(1, 500, 0.003)$ by using octave we will achieve:

$$P\{X > 500\} = P\{Y < 2\} = P\{Y \leq 1\} = F(1) = 0.558$$

So the possibility of having to play more than 500 games in order to win twice against an IM for an average player is 0.558.

ii)

Similar to the part b we will again do the same thing. But this time our $p = 10^{-4}$ and number of trials is 10000. Let X be the number of games played until winning twice. It is a number of trials needed to see 2 successes, hence X has Negative Binomial distribution with k = 2 and p = 0.0001.

We need to find:

$$P(X > 10000) = 1 - P(X \leq 10000)$$

However, there is no table of Negative Binomial distribution in the Appendix, and applying the formula for $P(X \leq 10000) = P(1)+P(2)+P(3)+P(4)+...+P(10000)$ is very difficult to calculate.

Although X is not Binomial at all, the probability $P\{X > 10000\}$ can be related to some Binomial variable. In our example:

$$\begin{aligned} P\{X > 10000\} &= P\{\text{more than 10000 trials needed to get 2 successes}\} \\ &= P\{\text{there are fewer than 2 successes in 10000 trials}\} \\ &= P\{Y < 2\} \end{aligned}$$

where Y is the number of successes in 10000 trials, which is a Binomial variable with parameters $n = 10000$ and $p = 0.0001$.

If we calculate $binocdf(1, 10000, 0.0001)$ by using octave we will achieve:

$$P\{X > 10000\} = P\{Y < 2\} = P\{Y \leq 1\} = F(1) = 0.736$$

So the possibility of having to play more than 1000 games in order to win twice against an GM for an average player is 0.736.

c)

In this question lets assume feeling sick as a success. Then the probability of success will be

$$p = 1 - 0.98 = 0.02$$

And our number of trials will be the number of days in a year which is n=366. Since n is large and p is small we can use poisson distribution to approximate binomial distribution.
First let's calculate the frequency $\lambda = n \cdot p = 336 \cdot 0.02 = 7.32$
Then by looking at the Table A3 Poisson distribution in the textbook page 415 we need to find most fitting values. We have most close value as 7.5 for λ. And since throughout the year, which has 366 days, a person can feel sick at most 6 days we need to look at 6 for x. Since the corresponding value is 0.378 we can conclude that the probability of a person not feeling sick for at least 360 days is approximately 0.378.

Answer 3

a)

Here probability of success (feeling sick) will be 0.02 and number of trials (days in a year) will be 366. Since we want to feel not sick at least 360 days which means we want to feel sick at most 6 days. So we need to find $P(X \leq 6)$. We can use binocdf in order to find this value. For more accurate value via octave online we can write $binocdf(6, 366, 0.02)$ which is equal to 0.4013.

This value is higher then the value we found in question2 part-c. This is because the frequency we used for approximation in part 2-c was (7.5) greater then the real frequency (7.32). This is the reason for the difference between our assumption and calculation.

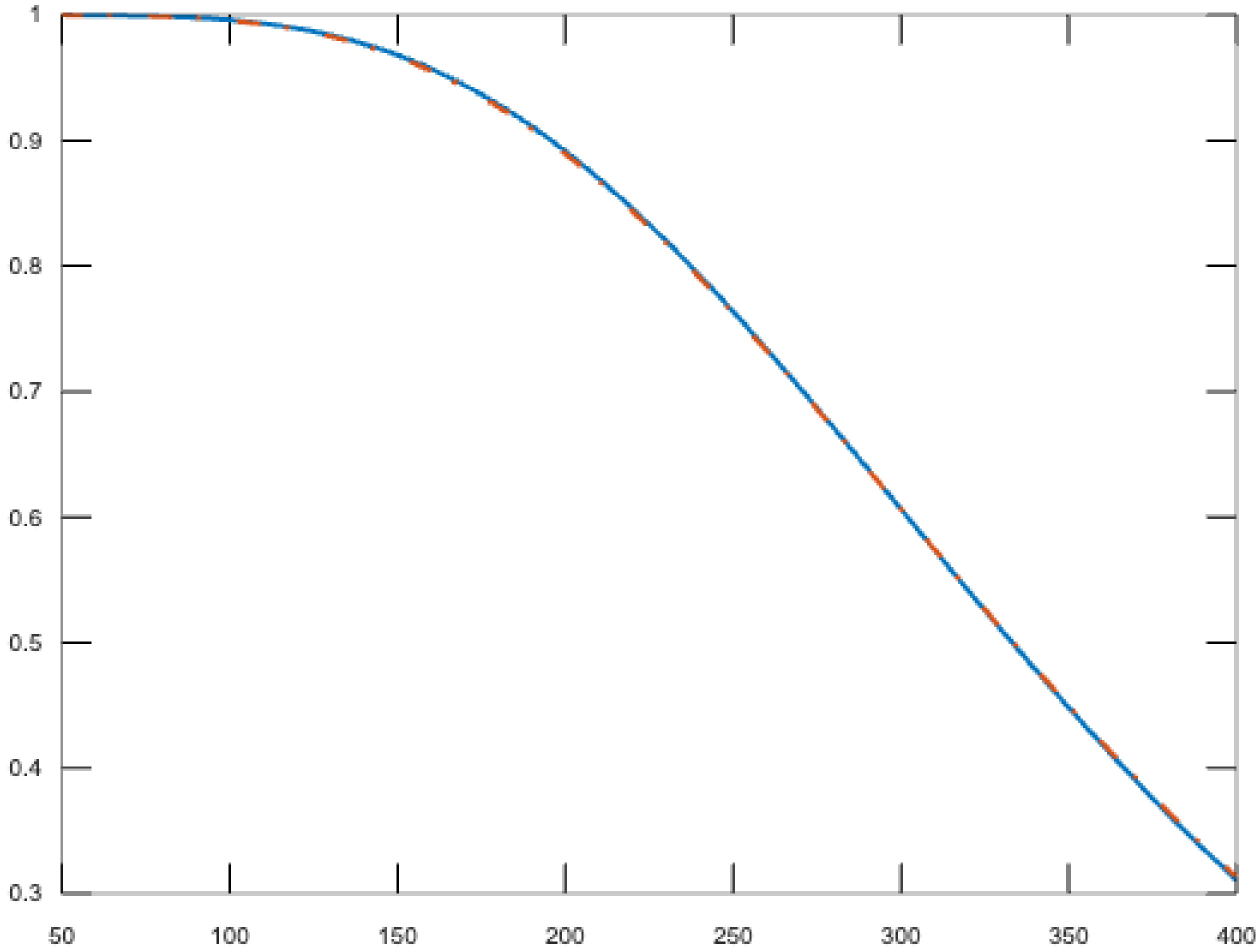
b)

Here is the octave code for plotting this graph:

```
>> p = 0.98;
>> ns = 50:400;
>> binomial_probabilities = binocdf(6, ns, (1-p));
>> poisson_probabilities = poisscdf(6, ns * (1-p));
>> close all;
>> plot(ns, binomial_probabilities, 'linewidth', 2);
>> hold on;
>> plot(ns, poisson_probabilities, '-.', 'linewidth', 2);
```

```
>> saveas(1, "p=0.98.png");
```

And the graph:



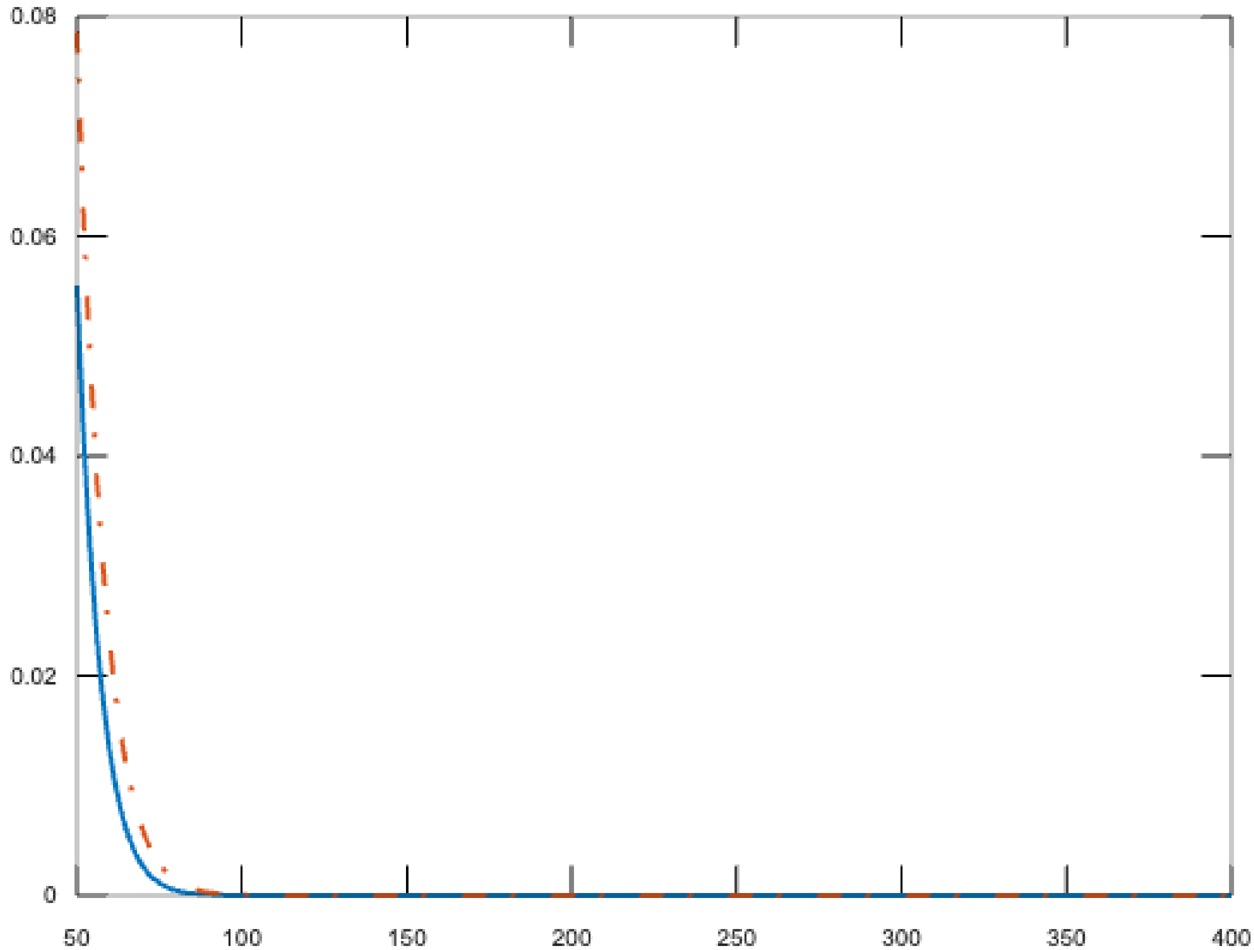
Here blue line represents binomial cdf and red line represents the poisson cdf.

c)

Here is the octave code for plotting this graph:

```
>> p = 0.78;
>> ns = 50:400;
>> binomial_probabilities = binocdf(6, ns, (1-p));
>> poisson_probabilities = poisscdf(6, ns * (1-p));
>> close all;
>> plot(ns, binomial_probabilities, 'linewidth', 2);
>> hold on;
>> plot(ns, poisson_probabilities, '-.', 'linewidth', 2);
>> saveas(1, "p=0.78.png");
```

And the graph:



Here blue line represents binomial cdf and red line represents the poisson cdf.

We can clearly see that as the success rate gets larger, the difference between binomial cdf and poisson cdf became larger too. The reason for this is poisson approximation will be more close to the real values when the success rate is smaller. In our examples first success rate was $1 - 0.98 = 0.02$ and the second success rate was $1 - 0.78 = 0.22$. Since success rate gets larger in the second graph the difference between binomial cdf and poisson cdf became larger too.