APSTA Week 03 Exercises

APSTA Week 03 Exercises

1. We are at a train station, waiting for a train. Suppose that the probability of snow is 0.1. If it is snowing, then the probability that the train is delayed is 0.6, otherwise, it is 0.2. Given that the train is delayed, what is the probability that it is snowing? Define appropriate events, and compute the conditional probability.

   Solution:

   First we set up our known probabilities. $S =$ snowing, $D =$ delayed:

   $$P(S) = 0.1$$
   $$P(S^{\text{C}}) = 1 - P(S) = 0.9$$
   $$P(D|S) = 0.6$$
   $$P(D|S^{\text{C}}) = 0.2$$

   Now we want to figure out the probability $P(S|D)$. We can start by using the definition of conditional probability:

   $$P(S|D) = \frac{P(S \cap D)}{P(D)}$$

   Then we can replace $P(S \cap D)$ with $P(D|S) \cdot P(S)$ using the multiplication rule, and $P(D)$ with $P(D|S) \cdot P(S) + P(D|S^{\text{C}}) \cdot P(S^{\text{C}})$ using the law of total probability:

   $$P(S|D) = \frac{P(S \cap D)}{P(D)}$$
   $$= \frac{P(D|S) \cdot P(S)}{P(D|S) \cdot P(S) + P(D|S^{\text{C}}) \cdot P(S^{\text{C}})}$$

   Now we're only left with known quantities and can write out the calculation:

   $$= \frac{0.6 \cdot 0.1}{0.6 \cdot 0.1 + 0.2 \cdot 0.9}$$
   $$= \frac{0.06}{0.06 + 0.18}$$
   $$= \frac{0.06}{0.24}$$
   $$= 0.25$$

   In conclusion, the probability $P(S|D)$ that it is snowing given the train is delayed is 25%.

2. Consider the following experiment. An urn contains 7 black and 13 red balls. You draw 10 balls from the urn with replacement. The outcome X of the experiment the number of red balls you got.

   a) Write down the probability mass function of X.

b) Evaluate the probability $P(X = 7)$.
   Solution:

c) Since we're drawing with replacement, every draw is in practice a disjoint event, and thus, the probability to draw a red ball is 13/20, and black ball 7/20. We can set up our Bernoulli distribution based on these probabilities. For every draw R_i, where $i = 1, 2, ..., 10$:

$$R_i = \begin{cases} 1 & \text{if the ith ball is red} \\ 0 & \text{if the ith ball is black} \end{cases}$$

So, the probability for a red ball is again $P(R_i = 1) = 13/20$. The number of correct answers is the discrete variable $X = R_1 + R_2 + ... + R_{10}$. It is now helpful to calculate $P(X = 0)$. The probability that a ball is *not* red is $P(R_i = 0) = 7/20$. So, the probability that all 10 pulls are black is $P(R_1 = 0)P(R_2 = 0)...P(R_{10} = 0) = (7/20)^{10}$. The probability that one ball was red is

$$P(X = 1) = \frac{13}{20} \cdot \left(\frac{7}{20}\right)^9 \cdot 10$$

(probability of red times probability that all others are black times number of ways this can happen.)
Or generalised:

$$P(X = k) = \left(\frac{13}{20}\right)^k \cdot \left(\frac{7}{20}\right)^{10-k} \cdot C_{10,k}$$

Here $C_{n,k}$ is defined as the following (see Bernoulli distribution notes)

$$C_{n,k} = \frac{n!}{k!(n-k)!}$$

$$C_{10,k} = \frac{10!}{k!(10-k)!}$$

Now we finally have our Bernoulli distribution:

$$P(X = k) = \frac{10!}{k!(10-k)!} \cdot \left(\frac{13}{20}\right)^k \cdot \left(\frac{7}{20}\right)^{10-k}$$

d) Now we can simply evaluate it at $k = 7$:

$$P(X = 7) = \frac{10!}{7!(10-7)!} \cdot \left(\frac{13}{20}\right)^7 \cdot \left(\frac{7}{20}\right)^{10-7}$$

$$= 120 \cdot \left(\frac{13^7}{20^7}\right) \cdot \left(\frac{7^3}{20^3}\right)$$

$$= \frac{64568223993}{256000000000}$$

$$\approx 25.22\%$$

3. A pair of fair dice is thrown until the sum of the two equals 2. Let the random variable X be the number of throws needed for this, where throwing the pair of dice once is counted as a single throw. Find the probability mass function of X.
   Solution:
   Now, this makes no sense, as a random variable is always the probability as a function of some event, and cannot by definition be "the number of throws needed", so I'm going to go with *the probability of getting a sum of 2 as a function of throws.*
   The only scenario in which a sum of two is reached is if both dice are a 1, which has a 1/6 chance for both of them. Together, that's 1/36. Then we just need to multiply that by the amount of throws

we've had, and we have a cumulative function.
Probability to get a sum of 2 on the kth throw.

$$P(X = k) = \frac{1}{36} \cdot k$$

Above is wrong, this is correct

$$\left(\frac{35}{36}\right)^{k-1} \cdot \frac{1}{36}$$

, Geometric distribution. X is the first trial where we get the sum of 2. It is the function of trials until success.

4. Consider you have two fair coins that you toss simultaneously (fair coins have a 0.5 probability of heads). You repeat the trial 15 times. Let X be the random variable indicating the number of cases, where both coins come with heads up. In the following exercises you can use the `dbinom` and `pbinom` functions.

a) Plot the probability mass function of X. What is the probability $P(X = 5)$?

b) What is the probability $P(X \leq 5)$? What is the visual interpretation of this probability in the graphic you plotted in (a)?
   *Solution*:
   First we need to create the Bernoulli (not the PMF). $X = 0$ is the case where none of the trials had two heads. The probability for one trial to not have any heads is $1 - 0.5 \cdot 0.5$, or 0.75. With 15 trials this is 0.75^{15}. The Bernoulli is

$$P(X = k) = 0.25^k \cdot 0.75^{15-k} \cdot \frac{15!}{k!(15-k)!}$$

c) Now we plot it

```
p.X <- function(k) {
  return((0.25^k)*(0.75^(15-k))*(factorial(15)/(factorial(k)*factorial(15-k))))
}
points <- lapply(0:15, FUN=p.X)
plot(
  x=0:15, y=points,
  xlab="Number of trials with two heads", ylab="Probability",
  main="Plot of Bernoulli distribution of X"
)
```

Plot of Bernoulli distribution of X

Number of trials with two heads
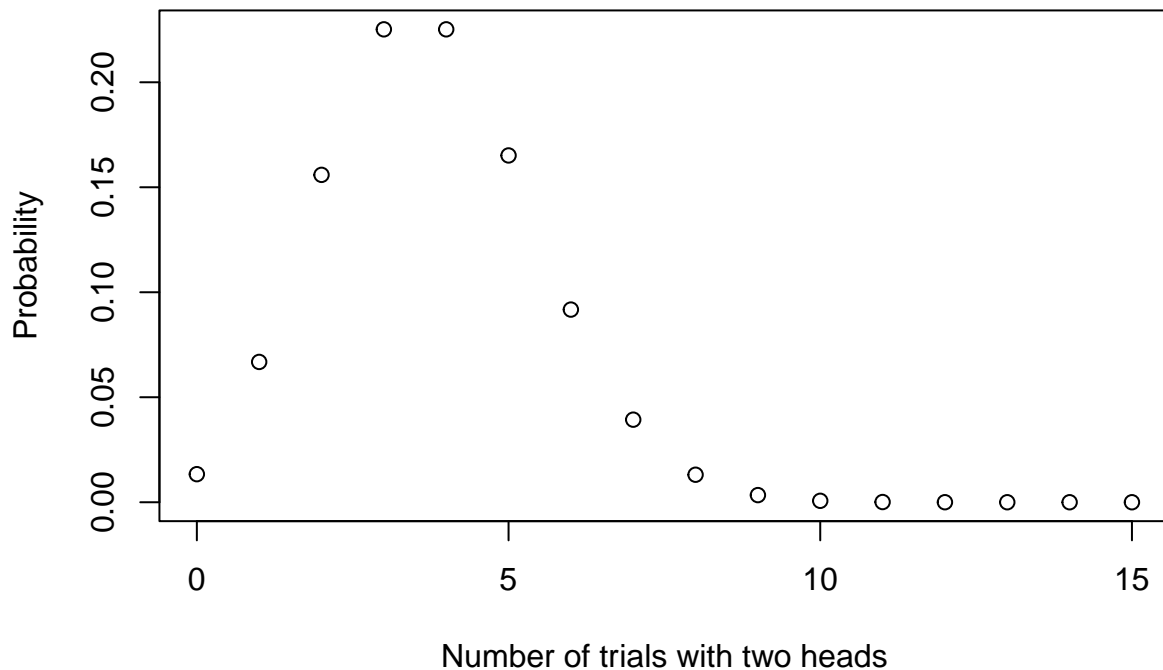
Note how the function is not continuous. And then probability of $P(X = 5)$:

```
k <- 5
sprintf("P(X=%s)~%s%%", k, round(p.X(k)*100, 1))
```

```
## [1] "P(X=5)~16.5%"
```

b) $P(X \leq 5)$ is the sum of the probabilities of $X = 1$ to $X = 5$:

```
sprintf("P(X<=5)~%s%%", round(sum(sapply(0:5, FUN=p.X))*100, 1))
```

```
## [1] "P(X<=5)~85.2%"
```

5. You are a collector of soccer players cards. There is just one card missing from your collection. Every day you buy one, and with the probability $1/100$ it is the one you are missing. Each purchase is independent from the others. Model the number of days it takes to find the missing card by the random variable $X \sim Geo(1/100)$.

   a) Plot the distribution function of X.

   b) How many cards do you have to buy so that the chance of finding the missing card is at least 0.5? How about at least 0.95? (play with different ranges for k).

4

c) Assume you have tried for 20 days, but you have not won yet. For how many days do you need to try further so that you have at least a 0.5$ chance of winning?

Solution:

Definition of the Geo distribution is
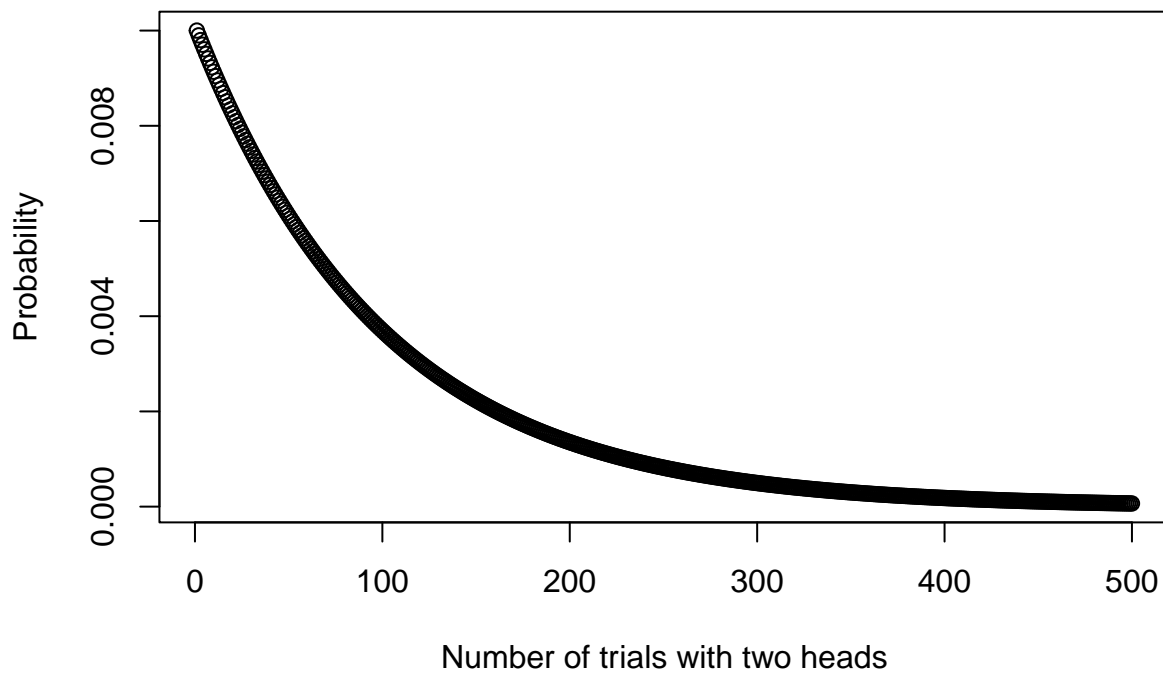
$$p_X(k) = P(X = k) = (1 - p)^{k-1}p$$

The probability to get a card is $1/100$, so $p = 1/100$:

$$p_X(k) = P(X = k) = \left(1 - \frac{1}{100}\right)^{k-1} \cdot \frac{1}{100}$$

d) Plotting the function:

```r
p.X <- function(k) {
  return((1-(1/100))^(k-1)*(1/100))
}
points <- lapply(1:500, FUN=p.X)
plot(
  x=1:500, y=points,
  xlab="Number of trials with two heads", ylab="Probability",
  main="Plot of Bernoulli distribution of X"
)
```

Plot of Bernoulli distribution of X

b) We first need a cumulative function

```r
F.X <- function(k) {
  sum <- 0
  for (i in 1:k) {
    sum <- sum + p.X(i)
  }
}
F.X(100)
```

Now we can create a function that does binary search for a specific probability:

```r
F.X.inv <- function(p) {
  lo <- 1
  hi <- 100

}
```