Book of Prob

Contents

Introduction

This book is the combined lecture notes of my course IE-231 Introduction to Probability at Istanbul Bilgi University, Turkey. You can find the original lecture notes and archives here. As the name suggests, these notes consist of an abbreviated version of the course as the objective is to give the essence of very significant parts of Probability. The course mainly follows (Walpole et al., 2012).

This book is periodically updated. Latest update is on 2018-03-25.

About the Author

The author of this book is Berk Orbay. I am the co-founder of Algopoly, a data science firm specialized on large scale forecasting problems, and a part-time academic (yes, still writing papers). My PhD was about pricing financial options with multiple models. I taught or am currently teaching Introduction to Probability, Computational Finance, Business Analytics and Essentials of Data Analysis courses.

Acknowledgements

Special thanks section.

Chapter 1

Initial Concepts of Probability

- **Probability** is the quantification of event uncertainty. For instance, probability of getting (H)eads in a coin toss is 1/2. Deterministic models will give the same results given the same inputs (e.g. 2 times 2 is 4), but probabilistic models might yield different outcomes.

- An **experiment** is a process that generates data. For instance, tossing a coin is an experiment. **Outcome** is the realization of an experiment. Possible outcomes for a coin toss is Heads and Tails.

- **Sample space** ($\mathbb{S}$) is the collection of all the possible outcomes of an experiment. Sample space of the coin toss is $\mathbb{S} = \{H, T\}$. Sample space of two coin tosses experiment is $\mathbb{S} = \{HH, HT, TH, TT\}$. Sample space can be discrete (i.e. coin tosses) as well as continuous (i.e. All real numbers between 1 and 3. $\mathbb{S} = \{x | 1 \leq x \leq 3, x \in \mathbb{R}\}$) *(Side note: Sample space is not always well defined.)*

- An **event** is a subset of sample space. While outcome represents a realization, event is an information. Probability of an event $P(A)$, say getting two Heads in two coin tosses is $P(A) = 1/4$.

- A **random variable** represents an event is dependent on a probabilistic process. On the other hand, a **deterministic variable** is either a constant or a decision variable. For instance, value of the dollar tomorrow can be considered a random variable but the amount I will invest is a decision variable (subject to no probabilistic process) and spot (current) price of the dollar is a constant.

1.1 Set Operations

- **Complement** of an event (A') with respect to the sample space represents all elements of the sample space that are not included by the event (A). For instance, complement of event $A = \{HH\}$ is $A' = \{HT, TH, TT\}$

- **Union** of two events A and B ($A \cup B$) is a set of events which contains all elements of the respective events. For example, say A is the set that contains events which double Heads occur ($A = \{HH, HT, TH\}$) and B is the set which Tails occur at least once ($B = \{TT, HT, TH\}$). The union is $A \cup B = \{HH, TH, HT, TT\}$.

- **Intersection** of two events A and B ($A \cap B$) contains the common elements of the events. For example, say A is the set that contains events which Heads occur at least once ($A = \{HH, HT, TH\}$) and B is the set which Tails occur at least once ($B = \{TT, HT, TH\}$). The intersection is $A \cap B = \{TH, HT\}$.

- **Mutually exclusive** or disjoint events mean that two events have empty intersection ($A \cap B = \emptyset$) and their union ($A \cup B$) contains the same amount of elements as the sum of their respective number of elements. Also $P(A \cap B) = 0$ and $P(A \cup B) = P(A) + P(B)$. For example getting double Heads (HH) and double Tails (TT) are mutually exclusive events.

1.2   Axioms of Probability

1. Any event A belonging to the sample space $A \in \mathbb{S}$ should have nonnegative probability ($P(A) \geq 0$).
2. Probability of the sample space is one ($P(\mathbb{S}) = 1$).
3. Any disjoint events ($A_i \cap A_j = \emptyset \ \forall_{i,j \in 1\ldots n}$) satisfies $P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$.

1.3   Other Set and Probability Rules

- $(A')' = A$

- $S' = \emptyset$

- $\emptyset' = S$

- $(A \cap B)' = A' \cup B'$

- $(A \cup B)' = A' \cap B'$

- $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

- $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

- $(A \cup B) \cup C = A \cup (C \cup B)$

- $(A \cap B) \cap C = A \cap (C \cap B)$

- $A \cup A' = \mathbb{S}$ and $A \cap A' = \emptyset$ so $P(A) = 1 - P(A')$. This is especially useful for many problems. For example the probability of getting at least one Heads in a three coin tosses in a row is $1 - P(\{TTT\}) = 7/8$, the complement of no Heads in a three coin tosses in a row. Otherwise, you should calculate the following expression.

$$P(\{HTT\}) + P(\{THT\}) + P(\{TTH\}) + P(\{HHT\}) + P(\{HTH\}) + P(\{THH\}) + P(\{HHH\}) = 7/8$$

- If $A \subseteq B$ then $P(A) \leq P(B)$.

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

- $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$

1.4   Counting

Counting rules will help us enumerate the sample space. It will include multiplication rule, permutation and combination.

1.4.1   Multiplication Rule

If I have a series of independent events, say 1 to k, and number of possible outcomes are denoted with n_1 to n_k; total number of outcomes in the sample space would be $n_1 n_2 \ldots n_k$.

Take a series of coin tosses in a row. If I toss a coin its sample space consists of 2 elements such as $\{H, T\}$. If I toss 2 coins the sample space would be 2*2 $\{HH, HT, TH, TT\}$. If I toss 3 coins, the sample space would be 2*2*2 $\{HHH, HTH, THH, TTH, HHT, HTT, THT, TTT\}$.

A poker card consists of a type and a rank. There are four types of playing cards (clubs, diamonds, hearts and spades) and 13 ranks (A - 2 to 10 - J - Q - K). Number of cards in a deck is 4*13 = 52.

1.4.2 Permutation Rule

Permutation is the arrangement of all or a subset of items.

- Given a set of items, say $A = a, b, c$ in how many different ways I can order the elements? Answer is n!. In our case it is, $3! = 3.2.1 = 6$.

$$A = \{a, b, c\}, \{b, a, c\}, \{b, c, a\}, \{c, a, b\}, \{c, b, a\}, \{a, c, b\}$$

- Suppose there are 10 (n) participants in a competition and 3 (r) medals (gold, silver and bronze). How many possible outcomes are there? Answer is $n(n-1)(n-2)\ldots(n-r+1) = \dfrac{n!}{(n-r)!} = \dfrac{10!}{(10-3)!} = 720$.

- If there are more than one same type items in a sample, then the permutation becomes $\dfrac{n!}{n_1!n_2!\ldots n_k!}$ where $\sum n_i = n$.

  For example enumerate the different outcomes of four coin tosses which result in 2 heads and 2 tails. Answer is $\dfrac{4!}{2!2!} = 6$

$$A = \{HHTT, HTTH, HTHT, THTH, THHT, TTHH\}$$

1.4.3 Combination Rule

Suppose we want to select r items from n items and the order does not matter. So the number of different outcomes can be found using $\binom{n}{r} = \dfrac{n!}{(n-r)!r!}$.

Out of 10 students how many different groups of 2 students can we generate? Answer $\dfrac{10!}{8!2!} = 45$

1.5 Chapter Problems

1. Suppose I toss a coin, roll a die and draw a card from the deck. How many different number of outcomes are there for this experiment?

   Solution: Multiplication rule. $n_1 n_2 n_3 = 2.6.52 = 624$.

   ```
   n1 <- 2 #A coin toss has two potential outcomes.
   n2 <- 6 #A die roll has six potential outcomes.
   n3 <- 52 #A card draw has 52 potential outcomes.
   ```

2. In how many ways can I order the Teletubbies? (Tinky-Winky, Dipsy, Laa Laa and Po) For instance, (TW - Dipsy - Po - Laa Laa) is an ordering and (Dipsy - Po - TW - Laa Laa) is another.

   Solution: Permutation rule. $n! = 4! = 24$

   ```
   n_tubbies <- 4 #Number of teletubbies
   factorial(n_tubbies) #By permtuation it is 4!
   ```

   ```
   ## [1] 24
   ```

3. I want to reorder the letters of the phrase "GOODGRADES". In how many ways can I do it?.

   Solution: Remember the permutation rule with identical items. There are two "G"s, two "D"s and two "O"s. Remember the formula $\dfrac{n!}{n_1!n_2!\ldots n_k!}$. So the result should be $\dfrac{10!}{2!2!2!1!1!1!1!} = 453600$.

```
the_phrase <- "GOODGRADES"
freq_table <- table(strsplit(the_phrase, split = "")[[1]])  #Let's create a frequency table first
print(freq_table)  #Let's show it
```

```
##
## A D E G O R S
## 1 2 1 2 2 1 1
```

```
the_dividend <- factorial(nchar(the_phrase))  #Dividend part is 10 characters so 10!
the_divisor <- prod(factorial(freq_table))  #Get multiplication of factorials for the divisor
the_dividend/the_divisor
```

```
## [1] 453600
```

4. I want to make two letter words from "GRADES" such as "GA", "ED" or "DE" (it doesn't have to make sense). Find the number of permutations.

   Solution: Permutation of r items from n items is $\dfrac{n!}{(n-r)!}$. So the result is $\dfrac{6!}{4!} = 30$.

```
the_phrase<-"GRADES"
letter_length <- 2 #We want two letter words
#Since all letters are different no need for special permutation.
factorial(nchar(the_phrase))/factorial(nchar(the_phrase)-letter_length)
```

```
## [1] 30
```

5. Suppose I am drawing a hand of 5 cards from a playing deck of 52 cards. How many different hands there can be? (Each card is different. See the bottom of this document for details.)

   Solution: Since in a hand you do not care for the order, it is the combination $\binom{52}{5} = \dfrac{52!}{(52-5)!5!} = 2598960$.

```
#Combination (a.k.a binomial coefficient) function is choose
choose(52,5)
```

```
## [1] 2598960
```

1.6 Extra Problems

- Question 1 Suppose we draw three cards from a deck and roll two dice. Answer the following questions.

    a) What is the experiment?
       The experiment is "drawing three cards from a deck and rolling two dice".

    b) What is "getting two-sixes and three-kings or five-one (in any order) one queen one king one ace"?
       Pick one (Event / Outcome / Sample Space)
       Event.

c) Give an example of two mutually exclusive events. (6 pts)

Event A: Queen of Hearts / Queen of Spades / Queen of Diamonds / 6 / 5 Event B: Ace of Clubs / King of Clubs / Queen of Clubs / 4 / 4

d) What is the probability of getting four-three (in any order) in dice roll and three queens in card draw?

```
#First roll can either be 3 or 4 and second roll should be the other
# 2/6 * 1/6
#Getting the first Queen has probability of 4/52
#Getting the second Queen has probability of 3/51
#Getting the third Queen has probability of 2/50
2/6*1/6*4/52*3/51*2/50
```

```
## [1] 1.00553e-05
```

e) How many different outcomes can there be? This time assume ordering is important (e.g. 6-1 and 1-6 are different outcomes).

```
#Six outcomes per die
#52 outcomes for the first card draw
#51 outcomes for the second card draw
#50 outcomes for the second card draw
#Multiplication rule
#You can also use permutation rule for cards
6*6*52*51*50
```

```
## [1] 4773600
```

- Question 2 In how many ways can you arrange the letters of "HOUSEPARTY"?

a) Any order.

```
the_phrase<-"HOUSEPARTY"
#No repetitive letters
#Permutation rule
factorial(nchar(the_phrase))
```

```
## [1] 3628800
```

b) Vowels together?

```
#4 vowels, 6 consonants
#Assume all vowels are a single "letter". So 8 characters.
#But vowels permutate within the single "letter".
#Multiplication rule
factorial(6+1)*factorial(4)
```

```
## [1] 120960
```

c) Vowels in alphabetical order?

```
#We start with all the permutations 10!
#For any permutation there can be only one ordering of vowels.
#For instance HOUSEPARTY is not valid but HAESOPURTY is valid
#So remove invalid permutations with division
factorial(10)/factorial(4)
```

```
## [1] 151200
```

d) There should be no consecutive vowels?

```
#There are 6 consonants, 4 vowels.
# Assume Xs are consonants and .s are potent vowel places.
# .X.X.X.X.X.X.
#Consonants can permutate in any order so 6! there
#7 places for vowels but only 4 vowels.
# So it is a permutation of 4 out of 8 places.
factorial(6)*(factorial(7)/factorial(7-4))
```

```
## [1] 604800
```

- Question 3 In how many ways can you arrange the letters of "CAMARADERIE"?

```
# 11 characters.
# 6 vowels, 5 consonants
# 3 As, 2 Es, 2 Rs
```

a) Any order.

```
#By the formula of permutation with repetitive letters
#Assign the value to all_perms object
all_perms<-factorial(11)/(factorial(3)*factorial(2)*factorial(2))
all_perms
```

```
## [1] 1663200
```

b) Vowels together?

```
#Assume all vowels are single "character" again. So 6 characters
(factorial(5+1)/(factorial(2)))*(factorial(6)/(factorial(3)*factorial(2)))
```

```
## [1] 21600
```

c) Vowels in alphabetical order?

```
#Same as the last question. But be careful about identical vowels.
all_perms/(factorial(6)/(factorial(3)*factorial(2)))
```

```
## [1] 27720
```

d) There should be no consecutive vowels?

```
#Same as the last question. But be careful about identical vowels.
(factorial(5)/factorial(2))*(factorial(6)/factorial(6-6))/(factorial(3)*factorial(2))
```

```
## [1] 3600
```

- Question 4

  Suppose you are putting the top 12 basketball teams in 4 groups evenly (each group should consist of 3 teams). In how many different ways can you arrange the teams?

```
#It is either a chain of combinations or just grouping combination
choose(12,3)*choose(9,3)*choose(6,3)
```

```
## [1] 369600
```

- Question 5 (20 pts - all equal)

  There are 18 people; 10 from Izmir, 8 from Mugla.

  a) Suppose you want to form a group of 5 people with at least 1 person from Izmir and Mugla. In how many ways can you form such a group?

  ```
  #Calculate as if no rules. It is the combination of 18 to 5.
  #Then remove the combinations of all Izmir or all Mugla people
  choose(18,5) - choose(10,5) - choose(8,5)
  ```

  ```
  ## [1] 8260
  ```

  b) In how many ways can you form a group of 3 people from Izmir and 4 people from Mugla?

  ```
  #Simply separate combinations with multiplication rule.
  choose(10,3)*choose(8,4)
  ```

  ```
  ## [1] 8400
  ```

1.7 Coins, Dice and Cards

When questions mention about coins, dice and cards they are commonly referred items. Nevertheless, you can refer to .

- Coin tosses: Two possible outcomes. Heads or Tails.
- Dice rolling: Six possible outcomes. 1-2-3-4-5-6.
- Card drawing: 52 possible outcomes. There are 4 types (clubs, diamonds, spades and hearts) and 13 ranks for each type. (A)ce - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10 - (J)ack - (Q)ueen - (K)ing.

Chapter 2

Conditional Probability and Bayes' Rule

Let's recall some probability concepts.

- Probability is the quantification of uncertainty. For instance in a coin toss, $P(H) = 1/2$.
- Set of the union of all possible events is the sample space. For instance single coin toss sample space is $S = \{H, T\}$.
- Probability of an event is actually the relative "size" of the event to the sample space.
- In an experiment, that has equally likely outcomes, number of potential outcomes in an event relative to the total number of outcomes in the sample space corresponds to the probability of that event.
- Some events are mutually exclusive and some events are dependent.
- Sample space is not always easy to measure and is not always finite. Counting is one way of estimating a sample space (especially discrete ones).
- Estimation of probability of events is not always easy. One way to estimate is to repeat the experiment for a number of times. As we increase the number of repetitions, the probability will become more stable. For instance suppose we don't know the probability of getting (H)eads in a coin toss and we repeat coin tossing for a number of times.

```r
set.seed(231)
coin_toss_probability<-function(repetitions){
    experiment<-sample(c("H","T"),size=repetitions,replace=TRUE)
    return(round(sum(experiment=="H")/repetitions,4))
}
#Do the experiment with increasing repetitions
coin_toss_probability(repetitions=10)
```

```
## [1] 0.6
```

```r
coin_toss_probability(repetitions=100)
```

```
## [1] 0.57
```

```r
coin_toss_probability(repetitions=1000)
```

```
## [1] 0.505
```

```
coin_toss_probability(repetitions=50000)
```

```
## [1] 0.4989
```

2.1   Some Examples

1. In a sports bar, there are 10 football fans, 15 basketball fans, 12 tennis fans and 6 curling fans. Suppose there is a single TV and there are matches of football, basketball, tennis and curling at the same time. If the TV remote is handed to one of them randomly, what is the probability that curling will be the show on TV?

   Answer: There are a total of 43 fans in the sports bar and 6 curling fans. P(Curling) = 6/43.

2. Suppose there is a group of 20 people; 10 from Ankara, 10 from Istanbul. Suppose we randomly choose 5 people from that group.

   a. What is the probability that all of this subgroup is from Ankara?

      Answer: (# of combinations including people only from Ankara) / (# of total combinations)

      $$\frac{\binom{10}{5}}{\binom{20}{5}} = 0.01625387$$

      Alternative answer is: $(10/20) * (9/19) \cdots * (6/16) = 0.01625387$.

   b. What is the probability that at least two from Ankara and at least two from Istanbul?

      Answer: P(Ankara = 2, Istanbul = 3 OR Ankara = 3, Istanbul = 2) = (P(Ankara = 2, Istanbul = 3) + P(Ankara = 3, Istanbul = 2)) = ((# of comb. Ank 2, Ist 3) + (# of comb. Ank 3, Ist 2))/(# of total combinations)

      $$\frac{\binom{10}{2} * \binom{10}{3} + \binom{10}{3} * \binom{10}{2}}{\binom{20}{5}} = 0.6965944$$

3. ("Blackjack") Suppose you are drawing two cards from the deck. Assume all number rank cards (2-10) are points, Ace (A) is 11 points and (J)ack, (Q)ueen and (K)ing are 10 points. What is the probability that total points of these two cards exceed 20.

   Answer: Max score for two cards is 22 (two Aces). To exceed 20, one needs Ace-9 pairs or any pair of 10-Ace-J-Q-K cards. We know that there are 4 of each rank. So $P(Score \geq 20) = P(A and 9) + P(any\ two\ of\ 10 - Ace - J - Q - K)$. We know that total number of hands is $\binom{52}{2} = 1326$. So # of combinations with either A-9 or 9-A is $4 * 4 = 16$. # of combinations of getting two cards from 10-Ace-J-Q-K ranks is $\binom{20}{2} = 190$. So $P(Score \geq 20) = \dfrac{190 + 16}{1326} = 0.1553544$

   Alternative answer: $P(A - 9 or 9 - A) = 4/52 * 4/51 + 4/52 * 4/51 = 0.01206637$ and $P(any\ two\ of\ 10 - Ace - J - Q - K) = 20/52 * 19/51 = 0.1432881$. Total probability is $P(Score \geq 20) = 0.01206637 + 0.1432881 = 0.1553545$

4. Suppose you are the gatekeeper of a playing ground for 10 year old boys and try to assess their age by their height. Height of 90% of boys aged 10 falls between 130 and 150 cms. It is equally likely that a boy aged 10 has the height lower than 130 cm or higher than 150 cm. What is the probability of rejecting a child because he's too long?

   Answer: $P(130 \leq X \leq 150) = 0.9$ and $P(X \leq 130) = P(X \geq 150)$. So $P(X \geq 150) = 1 - P(130 \leq X \leq 150) - P(X \leq 130)$. $2.P(X \geq 150) = 0.1$ and $P(X \geq 150) = 0.05$.

2.2 Conditional Probability

Definition: $P(A|B)$ means that probability of event A given that event B already occured or probability of A **conditional on** B.
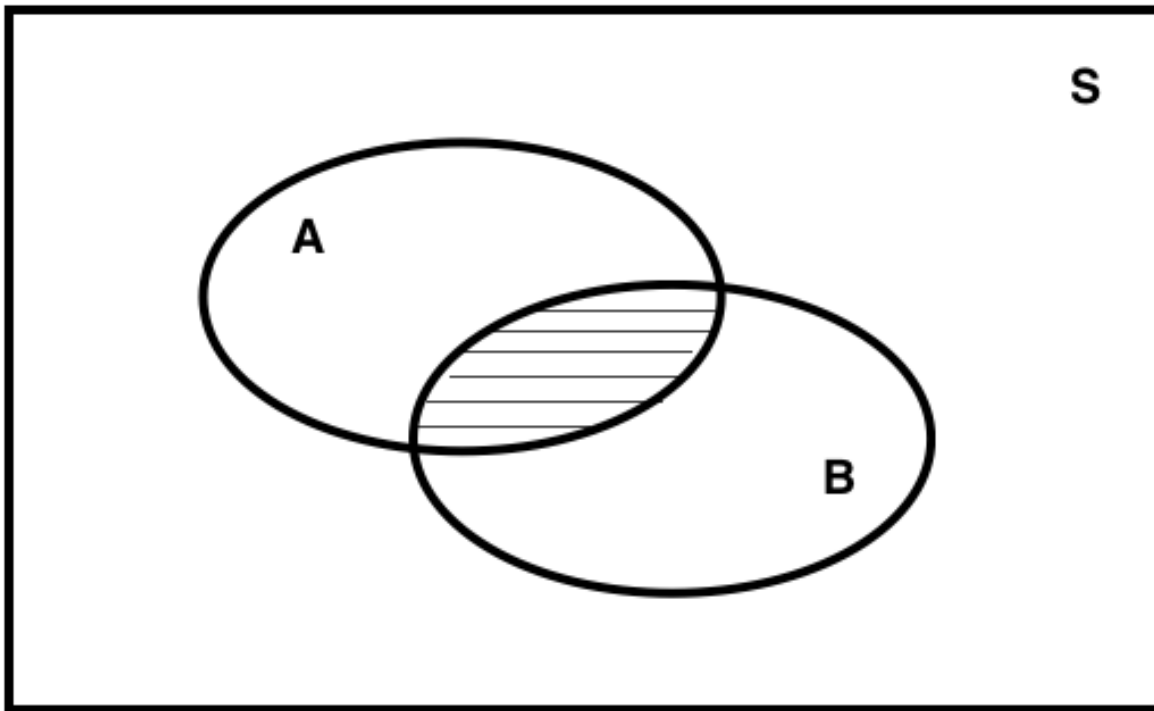
Example: What is the probability of getting two heads (HH) if we know that the first coin is H? (1/2)

Example: Suppose we are rolling two dice.

    a. What is the probability of getting the sum equal to 8?

       $P(Sum = 8) = P(D_1 = 2, D_2 = 6) + P(3, 5) + P(4, 4) * 2 + P(5, 3) + P(6, 2) = 6/36 = 1/6$

    b. What is the probability of getting the sum more than or equal to 8 given the first roll is 3?

       $P(Sum \geq 8|D_1 = 3) = P(3, 5) + P(3, 6) = 2/6 = 1/3$

Conditional probability of event A given B, can be defined as $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ if $P(B) > 0$.

Example: We asked 1000 people, 400 men and 600 women: 'Where do you want to visit this summer? Amsterdam or London?'. The answers are as follows.

```
##         Amsterdam London
## Men          150    250
## Women        450    150
```

    a. What is the probability of a randomly chosen person is a woman?

       Answer: $P(W) = (n(W|L) + n(W|P))/n(S) = (150 + 450)/1000 = 0.6$

b. What is the probability of a randomly chosen person chooses to to visit Amsterdam?

Answer: $P(A) = (n(A|W) + n(A|M))/n(S) = (450 + 150)/1000 = 0.6$

c. What is the probability of a London visitor is a man?

Answer: $P(M|L) = \dfrac{P(M \cap L)}{P(L)} = \dfrac{n(M \cap L)/n(S)}{n(L)/n(S)} = 250/400 = 0.625$

2.3 Independence

Two events can be independent. It means $P(A|B) = P(A)$ or $P(B|A) = P(B)$.

If the events (A_i) are independent then $P(A_1 \cap A_2 \ldots A_k) = P(A_1)P(A_2)\ldots P(A_k)$.

Example: Suppose you rolled a die 5 times and they all ended up 4. What is the probability that, at the 6th time, the die will get 4 again?

Solution: Die rolls are independent events. So outcome of a roll does not affect the outcome of the other rolls. It is 1/6.

2.4 Product rule

If both $P(A) > 0$ and $P(B) > 0$, then $P(A \cap B) = P(A)P(B|A)$. We can also say $P(A \cap B) = P(B \cap A) = P(B)P(A|B)$

Example: You draw two cards from a deck. What is the probability that you get two (A)ces?

Solution: $P(A \cap B) = 4/52 * 3/52 = 0.00443787$

Example: You are at a tea shop. There are 10 (W)hite teas, 12 (B)lack teas and 8 (G)reen teas on the menu. According to the waiter all teas are equally tasty. You ask the waiter bring a tea but it should not be green tea. What is the probability that you will get a white tea?

Solution: $P(W|G') = P(W \cap G')/P(G') = P(W)/P(G') = \dfrac{10/30}{22/30} = 10/22$

2.5 Bayes' Rule

Theorem of total probability: Given the events B_i are collective parts of the sample space and $P(B_i) > 0$, then for any event A, $P(A) > 0$

$$P(A) = \sum_{i}^{k} P(B_i \cap A) = \sum_{i}^{k} P(B_i)P(A|B_i)$$

Bayes'rule: Given the events B_i are collective parts of the sample space and $P(B_i) > 0$, then for any event A, $P(A) > 0$

$$P(B_r|A) = \dfrac{P(B_r|A)}{\sum_{i}^{k} P(B_i \cap A)} = \dfrac{P(B_r)P(A|B_r)}{\sum_{i}^{k} P(B_i)P(A|B_i)}$$

Example 2.41 from the book: In an assembly plant, there are three machines B_1, B_2 and B_3. These machines make the 30%, 45% and 25% of the products. 2%, 3% and 2% of their output is known to be defective. So, what is the probability of getting a defective product at a random time?

Solution:

Say, event of getting a defective item is A and we want to know $P(A)$. $P(B_i)$ is the probability of a product being manufactured by machine i. $P(B_1) = 0.3$, $P(B_2) = 0.45$, $P(B_3) = 0.25$. The defect probabilities given the machine are $P(A|B_1) = 0.02$, $P(A|B_1) = 0.03$, $P(A|B_1) = 0.02$.

So, $P(A) = \sum_{i=1}^{3} P(B_i)P(A|B_i) = 0.3 * 0.02 + 0.45 * 0.03 + 0.25 * 0.02 = 0.0245$.

Example 2.42: If the product is defective, what is the probability that it came from machine 3?

$$P(B_3|A) = \frac{P(B_3 \cap A)}{\sum_{i=1}^{3} P(B_i)P(A|B_i)} = \frac{P(B_3)P(A|B_3)}{\sum_{i=1}^{3} P(B_i)P(A|B_i)} = \frac{0.25 * 0.02}{0.0245} = 10/49$$

Chapter 3

Random Variables and Distributions

3.1 Scales on Measurement

- Nominal scale: These are categorical values that has no relationship of order or rank among them. (e.g. colors, species)
- Ordinal scale: These are categorical values that has relationship of order or rank among them (e.g. military ranks, competiton results). Though the relative order has no defined magnitude (e.g. Champion can get 40 points, runner up 39 and third place 30).
- Interval scale: There is a numerical order but the difference can only be defined in intervals, *since there is no absolute minimum*. We cannot compare in relative values. For instance, we cannot say 10 degree celsius is twice as hot as 5 degree celsius; what about -5 vs +5?
- Ratio scale: Scale with an absolute minimum. (e.g. If I have 50TL and my friend has 100TL, I can say that she has twice the money that I have.) Height, weight, age are similar examples.

See more on https://en.wikipedia.org/wiki/Level_of_measurement. (p.s. Wikipedia wasn't banned when I prepared these notes)

3.2 Infinity

The concept of infinity is very broad. Currently, you just need to keep the distinction of countable and uncountable infinities in mind.

- Countably infinite: 1, 2, 3, 4, … (e.g. natural numbers, integers, **rational numbers**)
- Uncountably infinite: 1, 1.01, 1.001, 1.0001, 1.00001, … (e.g. real numbers)

How many real numbers are there between 0 and 1?

3.3 Descriptive Statistics

Here are brief descriptions of mean (expectation), median, mode, variance, standard deviation, quantile.

- Mean: $\bar{X} = \sum_i^N X_i$

- Median: Let's say X_k are ordered from smallest to largest and there are n values in the sample. Median$(X)= X_{(n+1)/2}$ if n is odd and (usually) Median$(X)= \dfrac{X_{(n/2)} + X_{(n/2+1)}}{2}$.

- Quantile: On an ordered list of values for quantile (α) provides the $(\alpha * n)^{th}$ smallest value of the list. For instance, if $\alpha = 70\% = 0.7$ quantile value is the 7th smallest value in a list of 10 values. $\alpha = 1$ means the maximum. Quantile is an important parameter in especially statistics.

- Mode: X_k with the highest frequency in the sample. In a sample of $(1, 2, 2, 3, 4, 5)$, 2 is the mode.

- Variance: $V(X) = \dfrac{\sum_i^N (X_i - \bar{X})^2}{n - 1}$

- Standard Deviation: $\sigma(X) = \sqrt{\dfrac{\sum_i^N (X_i - \bar{X})^2}{n - 1}}$

```r
set.seed(231)
#Let's pick 10 values from the numbers between 1 and 50.
numbers <- sample(1:50,10,replace=TRUE)
#The sorted version of the numbers
sort(numbers)
```

```
## [1]  1  9 15 16 16 18 26 31 32 35
```

```r
#The mean values of the numbers
sum(numbers)/10
```

```
## [1] 19.9
```

```r
#or in R
mean(numbers)
```

```
## [1] 19.9
```

```r
#Median of the numbers
median(numbers)
```

```
## [1] 17
```

```r
#Quantile 7/9 of the numbers
quantile(numbers,7/9)
```

```
## 77.77778%
##        31
```

```r
#Quantile 0 of the numbers (also the min)
quantile(numbers,0)
```

```
## 0%
##  1
```

```
#Quantile 1 of the numbers (also the max)
quantile(numbers,1)
```

```
## 100%
##   35
```

```
#No simple solution for mode in R
freq_table<-table(numbers)
freq_table
```

```
## numbers
##  1  9 15 16 18 26 31 32 35
##  1  1  1  2  1  1  1  1  1
```

```
names(freq_table[which.max(freq_table)])
```

```
## [1] "16"
```

```
#Sample variance of numbers
sum((numbers - mean(numbers))^2)/(10-1)
```

```
## [1] 118.7667
```

```
#For large values you can take n ~ n-1
#in R
var(numbers)
```

```
## [1] 118.7667
```

```
#Sample standard deviation of values
sqrt(sum((numbers - mean(numbers))^2)/(10-1))
```

```
## [1] 10.89801
```

```
#in R
sd(numbers)
```

```
## [1] 10.89801
```

3.4 Random Variables

A random variable (usually defined with a capital letter or symbol i.e. X) is a quantity determined by the outcome of the experiment. Its realizations are usually symbolized with lowercase letter (x).

Example: Suppose there are 10 balls in an urn, 5 black and 5 red. Two balls are randomly drawn from the urn without replacement. Define the random variable X as the number of black balls. Then, $X = x$ can get the values of 0, 1 and 2. Let's enumerate $P(X = x)$.

$$P(X = 0) = P(RR) = 5/10 * 4/9 = 2/9$$
$$P(X = 1) = P(BR) + P(RB) = 5/10 + 5/9 + 5/10 * 5/9 = 5/9$$
$$P(X = 2) = P(BB) = 5/10 * 4/9 = 2/9$$

3.5   Discrete Random Variables and Distributions

If a sample space has finite number of possibilities or countably infinite number of elements, it is called a discrete sample space. Discrete random variable probabilities are shown as point probabilities $P(X = x)$. The probability distribution of discrete random variables is also called probability mass function (pmf).

$$P(X = x) = f(x)$$

$$\sum_x f(x) = 1$$

$$f(x) \geq 0$$

Example: (Same as above) Enumerate the probability distribution.

Solution: Random variable X can take values (x) 0, 1 and 2. So $f(1) = 2/9$, $f(2) = 5/9$ and $f(3) = 2/9$.

3.5.1   Cumulative Distribution Function (CDF)

Cumulative distribution function is a special defined function yielding the cumulative probability of random variables up to a value. It is usually symbolised as $F(x)$

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

$$F(x) = P(X \leq x) = \sum_{\infty} f(t) = 1$$

Example: (Same as above) Enumerate the cdf.

$$F(0) = P(X \leq 0) = P(X = 0) = 2/9$$

$$F(1) = P(X \leq 1) = P(X = 0) + P(X = 1) = 7/9$$

$$F(2) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 1$$

3.6   Continuous Random Variables and Distributions

If a sample space has uncountably infinite number of possibilities, it is called a continuous sample space. Continuous random variables' probabilities are defined in intervals $P(a < X < b)$. The probability distribution of continuous random variables is also called probability density function (pdf).

$$f(x) \geq 0$$

$$P(a < X < b) = f(x) = \int_a^b f(x)dx$$

$$\int_{-\infty}^{\infty} f(x) = 1$$

Example (from the book): Suppose the probability function of a continuous distribution $f(x) = x^2/3$ defined between $-1 < x < 2$ and 0 everywhere else. Verify that it is a density function (i.e. the integral in the defined interval is 1) and calculate $P(0 < x < 1)$.

a. $\int_{-1}^{2} x^2/3dx = x^3/9|_{-1}^{2} = 8/9 - (-1/9) = 1$. Verified.
b. $\int_{0}^{1} x^2/3dx = x^3/9|_{0}^{1} = 1/9 - (0) = 1/9$. Verified.

Cumulative distribution function (CDF) for continuous random variables is defined with the integral.

$$F(x) = P(X < x) = \int_{-\infty}^{a} f(x)dx$$

Example: (same as above) Calculate the cdf $F(3/2)$

Solution: $F(1.5) = \int_{-\infty}^{3/2} f(x) = x^3/9|_{-1}^{3/2} = 3/8 - (-1/9) = 35/72$

Example: Calculate $P(X > 1)$.

Solution: $P(X > 1) = 1 - P(X < 1) = 1 - F(1) = 1 - \int_{-\infty}^{1} = 1 - ((1/9) - (-1/9)) = 7/9$.

3.7 Joint Distribution

So far we had distributions with only one random variable. What if we had more than one random variable in a distribution? It is not that different from univariate distributions.

$$f(x, y) \geq 0$$

$$\sum_{x} \sum_{y} f(x, y) = 1$$

$$P(X = x, Y = y) = f(x, y)$$

Example (from the book): Two pens are selected at random from a box of 3 blue, 2 red and 3 green pens. Define X as the number of blue pens and Y as the red pens selected. Find

a. Joint probability function $f(x, y)$
b. $P[(X, Y) \ in A]$ where A is the region $\{(x, y)|x + y \leq 1\}$.

Solution:

a. The possible cases for (x, y) are $(0, 0), (0, 1), (0, 2), (1, 0), (2, 0), (1, 1)$. For instance $(0, 1)$ is one green and one red pen selected. There are a total of 8 pens. Then sample space size for two pens selected is $\binom{8}{2} = 28$. There are $\binom{2}{1}\binom{3}{1} = 6$ ways of selecting 2 pens from green and red pens. So the probability is $f(0, 1) = P(X = 0, Y = 1) = 6/28 = 3/14$. It is possible to calculate other possible outcomes in a similar way. A generalized formula would be as follows.

$$\frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}}$$

b. Possible outcomes satisfying $A = (x, y) \leq 1$ are $(0, 0), (0, 1), (1, 0)$. So $P(X + Y \leq 1) = P(0, 0) + P(0, 1) + P(1, 0) = 9/14$.

In the continuous case it is similar. It is now called joint probability density function.

$$f(x,y) \geq 0$$

$$\int_x \int_y f(x,y)dxdy = 1$$

$$P(X = x, Y = y) \in A = \int \int_A f(x,y)dxdy$$

Example: (from the book)

A privately owned business operates both a drive in and a walk in facility. Define X and Y as the proportions of using the drive in and walk in facilities. Suppose the joint density function is $f(x,y) = 2/5 * (2x + 3y)$ where $0 \leq x \leq 1$ and $0 \leq y \leq 1$ (0, otherwise).

   a. Verify it is a distribution function.
   b. Find $P[(X,Y)] \in A$, where $A = \{(x,y)|0 < x < 1/2, 1/4 < y < 1/2\}$.

Solution:

   a. $\int \int f(x,y)dxdy = \int 2/5 * (x^2 + 3xy)dy|_0^1 = \int (2/5 + 6/5 * y)dy = 2y/5 + 3/5 * y^2|_0^1 = 2/5 + 3/5 = 1$.

   b. $\int_{1/4}^{1/2} \int_0^{1/2} f(x,y)dxdy = \int 2/5 * (x^2 + 3xy)dy|_0^{1/2} = \int_{1/4}^{1/2}(1/10 + 3/5 * y)dy, \ y/10 + 3y^2/10|_{1/4}^{1/2} = 13/160$.

3.8   Marginal Distribution

In a joint distribution, marginal distribution is the probability distributions of individual random variables. Define $g(x)$ and $h(y)$ as the marginal distributions of X and Y.

$$g(x) = \sum_y f(x,y), h(y) = \sum_y f(x,y)$$

$$g(x) = \int_y f(x,y)dy, h(y) = \int_x f(x,y)dx$$

Example: Go back to pen and walk in examples and calculate marginal probabilities.

3.9   Conditional Distribution

Remember the conditional probability rule $P(A|B) = P(A \cap B)/P(B)$ given $P(B) > 0$. We can define conditional distribution as $f(y|x) = f(x,y)/g(x)$, provided $g(x) > 0$ whether they are discrete or continuous.

Example: (pen example) Calculate $P(X = 0|Y = 1)$

Solution: We know that $h(1) = 3/7$. $f(x = 0, y = 1) = 3/14$. $f(x = 0|y = 1) = f(x = 0, y = 1)/h(1) = 1/2$

3.10   Statistical Independence

Two random variables distributions are statistically independent if and only if $f(x,y) = g(x)h(y)$.

Proof:

$$f(x,y) = f(x|y)h(y)$$

$$g(x) = \int f(x,y)dy = \int f(x|y)h(y)dy$$

If $f(x|y)$ does not depend on y we can write $f(x|y) \int h(y)dy$. $f(x|y) * 1$. Therefore, $g(x) = f(x|y)$ and $f(x,y) = g(x)h(y)$.

Any number of random variables $(X_1 \ldots X_n)$ are statistically independent if and only if $f(x_1, \ldots, x_n) = f(x_1) \ldots f(x_n)$.

Chapter 4

Expectation and Variance

This chapter will be improved

4.1 Mathematical Expectation

$$E[X] = \sum x f(x)$$
$$E[X] = \int x f(x) dx$$

4.2 Variance

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

Chapter 5

Some Discrete Distributions

5.1 Bernoulli Distribution

It can also be called "single coin toss distribution". For a single event with probability of success p and failure $q = 1 - p$, the distribution is called Bernoulli.

- pmf: $f(x = 0; p) = q$, $f(x = 1) = p$
- $E[X] = 0 * (1 - p) + 1 * p = p$
- $V[X] = pq$

Example: Coin Toss

- $p = 0.5$, $q = 1 - p = 0.5$
- pmf: $f(x = 0) = 0.5$, $f(x = 1) = 0.5$
- $E[X] = 0 * (1 - 0.5) + 1 * 0.5 = 0.5$
- $V(X) = 0.5 * 0.5 = 0.25$

5.2 Binomial Distribution

Think of multiple Bernoulli trials (e.g. several coin tosses).

- pmf: $f(x; p, n) = \binom{n}{x} p^x q^{(n-x)}$
- $E[X] = np$
- $V(X) = npq$
- cdf: $F(X \leq x) = \sum_{i=0}^{n} f(i)$

Example: Multiple Coin Tosses (x5 coins, p = 0.5)

- pmf: $f(x = 3; n = 5) = \binom{5}{3}(0.5)^3(1 - 0.5)^{(5-3)} = 0.3125$

```
#R way
#(d)ensity(binom)ial
dbinom(x=3,size=5,prob=0.5)
```

[1] 0.3125

- $E[X] = 5 * 0.5 = 2.5$

- $V(X) = 5 * 0.5 * 0.5 = 1.25$

- cdf: $F(X \leq 3; n = 5) = \sum_{i=0}^{5} f(i) = 0.8125$

```
#R way
pbinom(q=3,size=5,prob=0.5)
```

[1] 0.8125

5.3   Multinomial Distribution

Now suppose there is not one probability (p) but there are many probabilities $(p_1, p_2, \ldots, p_k)$.

- pmf: $f(x_1, \ldots, x_k; p_1, \ldots, p_k; n) = \binom{n}{x_1, \ldots, x_k} p_1^{x_1} * \cdots * p_k^{x_k}$

  where $\binom{n}{x_1, \ldots, x_k} = \dfrac{n!}{x_1! \ldots x_k!}$, $\sum_i^k x_i = n$ and $\sum_i^k p_i = 1$.

Example: Customers of a coffee shop prefer Turkish coffee with probability 0.4, espresso 0.25 and filter coffee 0.35. What is the probability that out of the first 10 customers, 3 will prefer Turkish coffee, 5 will prefer espresso and 2 will prefer filter coffee?

$f(3, 5, 2; 0.4, 0.25, 0.35; 10) = \binom{10}{3,5,2} * 0.4^3 * 0.25^5 * 0.35^10 = 4.3 * 10^{-6} = 0.0193$

```
#Explicit form
factorial(10)/(factorial(3)*factorial(5)*factorial(2))*0.4^3 * 0.25^5 * 0.35^2
```

[1] 0.01929375

```
#Density multinomial
dmultinom(x=c(3,5,2),prob=c(0.4,0.25,0.35))
```

[1] 0.01929375

Binomial distribution is a special case of multinomial distribution.

5.4 Hypergeometric Distribution

Hypergeometric distribution can be used in case the sample is divided in two such as defective/nondefective, white/black, Ankara/Istanbul. Suppose there are a total of N items, k of them are from group 1 and $N - k$ of them are from group 2. We want to know the probability of getting x items from group 1 and $n - k$ items from group 2.

- pmf: $f(x, n; k, N) = \dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$

- $E[X] = \dfrac{nk}{N}$

- $V[X] = \dfrac{N-n}{N-1} * n * \dfrac{k}{N} * (1 - \dfrac{k}{N})$

Example: Suppose we have a group of 20 people, 12 from Istanbul and 8 from Ankara. If we randomly select 5 people from it what is the probability that 1 of them is from Ankara and 4 of them from Istanbul.

$$f(1, 5; 8, 20) = \frac{\binom{8}{1}\binom{20-8}{5-1}}{\binom{20}{5}} = 0.256$$

```
#Explicit form
x=1
n=5
k=8
N=20
(choose(k,x)*choose(N-k,n-x))/choose(N,n)
```

```
## [1] 0.255418
```

```
#Density hypergeometric, see ?dhyper for explanations
dhyper(x=1,m=8,n=12,k=5)
```

```
## [1] 0.255418
```

5.5 Negative Binomial Distribution

Negative Binomial distribution answers the question "What is the probability that k-th success occurs in n trials?". Differently from the binomial case, we fix the last attempt as success.

- pmf: $f(x; p, n) = \binom{n-1}{x-1} p^x q^{(n-x)}$

Example: Suppose I'm repeatedly tossing coins. What is the probability that 3rd Heads come in the 5th toss?

$$f(3; 0.5, 5) = \binom{5-1}{3-1} 0.5^3 0.5^{(5-3)} = 0.1875$$

```
#Explicit form
choose(5-1,3-1)*0.5^3*0.5^(5-3)
```

```
## [1] 0.1875
```

```
#Binomial way
dbinom(3-1,5-1,0.5)*0.5
```

[1] 0.1875

```
#Negative binomial way
dnbinom(x=5-3,size=3,prob=0.5)
```

[1] 0.1875

5.6   Geometric Distribution

Geometric distribution answers "What is the probability that first success comes in the n-th trial?"

- pmf: $f(x; p, n) = q^{(n-1)}p$

- $E[X] = 1/p$

- $V[X] = \dfrac{1-p}{p^2}$

5.7   Chapter Questions

1. A boutique shop offers three types of breads; olive, rye and white. 15% of its customers buy olive bread, 55% rye bread and the rest white bread. What is the probability that the first 10 customers of the shop buys 2 olive, 5 rye and 3 white bread?

   Solution: Multinomial distribution.

   $$\binom{10}{2,5,3} 0.15^2 * 0.55^5 * 0.3^3 = 0.0770478$$

   ```
   n_total = 10
   # Olive bread
   p_1 = 0.15
   n_1 = 2
   # Rye bread
   p_2 = 0.55
   n_2 = 5
   # White bread
   p_3 = 1 - p_2 - p_1
   n_3 = 3

   factorial(n_total)/(factorial(n_1)*factorial(n_2)*factorial(n_3)) * p_1^n_1 * p_2^n_2 * p_3^n_3
   ```

   ## [1] 0.0770478

2. An urn contains 32 balls; 18 red, 14 white. If I take 8 balls randomly from the urn, what is the probability of getting 4 red balls?

   Solution: Hypergeometric distribution.

   $$\frac{\binom{18}{4}\binom{14}{4}}{\binom{32}{8}} = 0.2912125$$

   ```
   n_total = 32 #N
   k_total = 8
   n_1 = 18
   n_2 = 14
   k_1 = 4
   k_2 = 4

   (choose(n_1,k_1)*choose(n_2,k_2))/choose(n_total,k_total)
   ```

   ```
   ## [1] 0.2912125
   ```

3. A UNICEF activist asks bypassers to contribute to their efforts. Each of her attempts has a 10% chance of success. What is the probability that she got the first contribution at the 10th attempt?

   Solution: Geometric distribution.

   $$(0.9)^9 * (0.1) = 0.03874205$$

   ```
   (0.9)^9 * (0.1)
   ```

   ```
   ## [1] 0.03874205
   ```

4. Suppose there are 10 cards in a deck numbered from 1 to 10. If I draw four cards out of the deck without putting them back, what is the probability that they are in increasing order (e.g. 3-4-5-8)?

   Solution: Increasing order is a special order. There is only one decreasing order in every permutation set. For instance

   $$8 - 5 - 4 - 3$$
   $$8 - 5 - 3 - 4$$
   $$8 - 4 - 5 - 3$$
   $$4 - 5 - 3 - 8$$
   $$\dots$$
   $$**\,3 - 4 - 5 - 8\,**$$

   So, the answer is $1/(4!) = 1/24$.

5. A tennis player has the probability of 0.15 to ace a serve (i.e. he serves and gets the point without the opponent touching the ball) at each shot. His trainer challenged him that he will pay the square of the aces he makes out of 10 serves (e.g. if he serves 8 aces out of 10, the trainer pays him $8^2 = 64TL$). But, if he serves less than or equal to 6 aces, the player should pay the trainer 5 TL for each ace he missed (e.g. 4 aces 6 misses, player pays up 6*5 = 30TL).

   a) What is the maximum amount of money he can get?
   b) What is the expected earnings of the player?

c) What is the variance of his earnings?

Solution:

a) 10 out of 10 means 10^2 = 100.

b) $g(X = x) = x^2$ if $x \geq 7$, $f(g(X) = x^2) = f(X = x) = \binom{10}{x}0.15^x * 0.85^{(10-k)}$, so $E[g(x)] = \sum_{i=0}^{1} 0g(x)f(x)$.

$g(X = x) = -5 * x$ if $x \leq 6$.

So $49 * \binom{10}{7}0.15^7 * 0.85^3 + 64 * \binom{10}{8}0.15^8 * 0.85^2 + 81 * \binom{10}{9}0.15^9 * 0.85 + 100 * (0.15)^10 - 20 * \binom{10}{6}0.15^6 * 0.85^4 - 25 * \binom{10}{5}0.15^5 * 0.85^5 - 30 * \binom{10}{4}0.15^4 * 0.85^6 - 35 * \binom{10}{3}0.15^3 * 0.85^7 - 40 * \binom{10}{2}0.15^2 * 0.85^8 - 45 * \binom{10}{1}0.15^1 * 0.85^9 - 50 * 0.85^10 = -42.5TL$.

```
rewards <- (7:10)^2
probs_rewards <- dbinom(7:10,10,0.15)
losses <- -5*(10-(6:0))
probs_losses <- dbinom(6:0,10,0.15)

expectation <- sum(rewards*probs_rewards) + sum(losses*probs_losses)
expectation
```

```
## [1] -42.4913
```

c) $V(X) = E[X^2] - (E[X])^2$. We know that $E[X] = -42.5$, then $(E[X])^2 = 1805.51$. Also $E[X^2] = 1838.434$ and $V(X) = 32.92$.

```
ex2<-sum(rewards^2*probs_rewards) + sum(losses^2*probs_losses)
ex2 - expectation^2
```

```
## [1] 32.92423
```

Bibliography

Walpole, R. E., Myers, R. H., Myers, S. L., and Ye, K. (2012). *Probability and Statistics for Engineers and Scientists*. Pearson, Boston, MA, 9th edition. ISBN 978-0-321-74823-2.