Assignment 2

EMATM0061: Statistical Computing and Empirical Methods, TB1, 2024

Introduction

Load packages

Then we need to load two packages, namely Stat2Data and tidyverse, before answering the questions. If they haven't been installed in your computer, please use install.packages() to install them first.

1. Load the tidyverse package:

```
library(tidyverse)
```

2. Load the Stat2Data package and then the dataset Hawks:

```
library(Stat2Data)
data("Hawks")
```

1. Data Wrangling

This part is mainly about data wrangling. Basic concepts of data wrangling can be found in lecture 4.

1.1 Select and filter

(Q1). Use a combination of the **select()** and **filter()** functions to generate a data frame called "hSF" which is a sub-table of the original Hawks data frame, such that

1. Your data frame should include the columns:
   a) "Wing"
   b) "Weight"
   c) "Tail"
2. Your data frame should contain a row for every hawk such that:
   a) They belong to the species of Red-Tailed hawks
   b) They have weight at least 1kg.
3. Use the pipe operator "%>%" to simplify your code.

```
hSF <- Hawks %>%
  filter(Species=="RT" & Weight>=1000) %>%
  select(Wing,Weight,Tail)
head(hSF,5)

##   Wing Weight Tail
## 1  412   1090  230
## 2  412   1210  210
```

```
## 3   405   1120  238
## 4   393   1010  222
## 5   371   1010  217
```

(Q2) How many variables does the data frame hSF have? What would you say to communicate this information to a Machine Learning practitioner?

Answer There are 3 variables, or equivalently 3 features.

How many examples does the data frame hSF have? How many observations? How many cases?

Answer The "examples", "observations", and "cases" have the same meaning here.

```
hSF%>%nrow()

## [1] 398
```

1.2 The arrange function

(Q1) Use the arrange() function to sort the hSF data frame created in the previous section so that the rows appear in order of increasing wing span.

Then use the head command to print out the top five rows of your sorted data frame.

```
head(arrange(hSF,Wing),5)

##    Wing Weight Tail
## 1  37.2   1180  210
## 2 111.0   1340  226
## 3 199.0   1290  222
## 4 241.0   1320  235
## 5 262.0   1020  200
```

1.3 Join and rename functions

The species of Hawks within the data frame have been indicated via a two-letter code (e.g., RT, CH, SS). The correspondence between these codes and the full names is given by the following data frame:

```
##   species_code species_name_full
## 1           CH           Cooper's
## 2           RT         Red-tailed
## 3           SS      Sharp-shinned
```

(Q1). Use **data.frame()** to create a data frame that is called **hawkSpeciesNameCodes** and is the same as the above data frame (i.e., containing the correspondence between codes and the full species names).

Answer

```r
species_code<-c("CH","RT","SS")
species_name_full<-c("Cooper's","Red-tailed","Sharp-shinned")
hawkSpeciesNameCodes<-data.frame(species_code,species_name_full)
hawkSpeciesNameCodes

##   species_code species_name_full
## 1           CH          Cooper's
## 2           RT        Red-tailed
## 3           SS     Sharp-shinned
```

(Q2). Use a combination of the functions **left_join()**, the **rename()** and the **select()** functions to create a new data frame called "hawksFullName" which is the same as the "Hawks" data frame except that the Species column contains the full names rather than the two-letter codes.

Answer

```r
hawksFullName <- Hawks %>%
  left_join(hawkSpeciesNameCodes %>%
            rename(Species=species_code))%>%
  select(-Species)%>%
  rename(Species=species_name_full)
```

(Q3). Use a combination of the **head()** and **select()** functions to print out the top seven rows of the columns "Species", "Wing" and "Weight" of the data frame called **"hawksFullName"**. Do this without modifying the data frame you just created

Answer

```r
hawksFullName %>%
  select(Species,Wing,Weight) %>%
  head(7)

##         Species Wing Weight
## 1    Red-tailed  385    920
## 2    Red-tailed  376    930
## 3    Red-tailed  381    990
## 4      Cooper's  265    470
## 5 Sharp-shinned  205    170
## 6    Red-tailed  412   1090
## 7    Red-tailed  370    960
```

Does it matter what type of join function you use here?

Answer The results obtained with any one of left_join(), right_join(), inner_join(), and full_join() are the same, because the two data frames share the same set of species codes.

In what situations would it make a difference?

Answer It would matter if there were some unmatched entries in either data frame.

1.4 The mutate function

Suppose that the fictitious "Healthy Hawks Society"[1] has proposed a new measure called the "bird BMI" which attempts to measure the mass of a hawk standardized by their wing span. The "bird BMI" is equal to the weight of the hawk (in grams) divided by their wing span (in millimeters) squared. That is,

$$\text{Bird-BMI} := 1000 \times \text{Weight}/\text{Wing-pan}^2.$$

(Q1). Use the **mutate()**, **select()** and **arrange()** functions to create a new data frame called "hawksWithBMI" which has the same number of rows as the original Hawks data frame but only two columns - one with their Species and one with their "bird BMI". Also, arrange the rows in descending order of "bird BMI".

Answer

```
hawksWithBMI <- Hawks %>%
  mutate(bird_BMI=1000*Weight/Wing^2)%>%
  select(Species, bird_BMI)%>%
  arrange(desc(bird_BMI))

hawksWithBMI %>% head(8)

##   Species  bird_BMI
## 1      RT 852.69973
## 2      RT 108.75741
## 3      RT  32.57493
## 4      RT  22.72688
## 5      CH  22.40818
## 6      RT  19.54932
## 7      CH  15.21998
## 8      RT  14.85927
```

1.5 Summarize and group-by functions

Using the data frame "hawksFullName", from Section 1.3 above, to do the following tasks:

(Q1). In combination with the summarize() and the group_by functions, create a summary table, broken down by Hawk species, which contains the following summary quantities:

1. The number of rows (num_rows);
2. The average wing span in centimeters (mn_wing);
3. The median wing span in centimeters (nd_wing);

---

[1] Both the "Healthy Hawks Society" and the concept of "bird BMI" were made up purely for this assignment.

4. The trimmed average wing span in centimeters with trim=0.1, i.e., the mean of the numbers after the 10% largest and the 10% smallest values being removed (t_mn_wing);
5. The biggest ratio between wing span and tail length (b_wt_ratio).

Type ?summarize to see a list of useful functions (mean, sum, etc) that can be used to compute the summary quantities.

Answer

```
hawksFullName %>%
  group_by(Species)%>%
  summarize(num_rows=n(),
            mn_wing=mean(Wing, na.rm=TRUE),
            md_wing=median(Wing,na.rm=TRUE),
            t_mn_wing=mean(Wing,na.rm=TRUE, trim=0.1),
            b_wt_ratio=max(Wing/Tail,na.rm=TRUE))

## # A tibble: 3 × 6
##   Species       num_rows mn_wing md_wing t_mn_wing b_wt_ratio
##   <chr>            <int>   <dbl>   <dbl>     <dbl>      <dbl>
## 1 Cooper's            70    244.     240      243.       1.67
## 2 Red-tailed         577    383.     384      385.       3.16
## 3 Sharp-shinned      261    185.     191      184.       1.67
```

(Q2). Next create a summary table of the following form: Your summary table will show the number of missing values, broken down by species, for the columns Wing, Weight, Culmen, Hallux, Tail, StandardTail, Tarsus, and Crop. You can complete this task by combining the **select()**, **group_by()**, **summarize()**, **across()**, **everything()**, **sum()** and **is.na()** functions.

Answer

```
hawksFullName %>%
  select(Species,Wing,Weight,Culmen,
Hallux,Tail,StandardTail,Tarsus,Crop) %>%
  group_by(Species) %>%
  summarize(across(everything(),~sum(is.na(.x)))) %>%
  head()

## # A tibble: 3 × 9
##   Species        Wing Weight Culmen Hallux  Tail StandardTail Tarsus
Crop
##   <chr>         <int>  <int>  <int>  <int> <int>        <int>  <int>
<int>
## 1 Cooper's          1      0      0      0     0           19     62
21
## 2 Red-tailed        0      5      4      3     0          250    538
254
## 3 Sharp-shinned     0      5      3      3     0           68    233
68
```

2. Random experiments, events and sample spaces, and the set theory

In this exercise, we will learn about random experiments, events and sample spaces and set theory that were introduced in Lecture 4.

In this section, you are not required to compute your results using R codes. If you want to write math formulas in R-markdown, the document called "Assignment_R MarkdownMathformulasandSymbolsExamples.rmd" (available under the "resource list" tab at Blackboard course webpage) provides a list of examples for your reference.

2.1 Random experiments, events and sample spaces

(Q1) Firstly, write down the definition of a random experiment, event and sample space. This question aims to help you recall the basic concepts before completing the subsequent tasks.

Answer: A random experiment is a procedure (real or imagined) which:

1. has a well-defined set of possible outcomes;
2. could (at least in principle) be repeated arbitrarily many times.

An event is a set (i.e. a collection) of possible outcomes of an experiment A sample space is the set of all possible outcomes of interest for a random experiment

(Q2) Consider a random experiment of rolling a dice twice. Give an example of what is an event in this random experiment. Also, can you write down the sample space as a set? What is the total number of different events in this experiment? Is the empty set considered as an event?

Answer:

1. {(1,2), (2,3)} is an event. Here (1,2) means the number on the top of the first die is 1 and the number on the top of the second die is 2.
2. The sample space is $\{(a, b)|a, b \in \{1,2,3,4,5,6\}\}$
3. There are 36 different outcomes, so the total number of events is 2^{36}.
4. The empty set is an event.

2.2 Set theory

Remember that a set is just a collection of objects. All that matters for the identity of a set is the objects it contains. In particular, the elements within the set are unordered, so for example the set {1, 2, 3} is exactly the same as the set {3, 2, 1}. In addition, since sets are just collections of objects, each object can only be either included or excluded and multiplicities do not change the nature of the set. In particular, the set {1, 2, 2, 2, 3, 3} is exactly the same as the set A = {1, 2, 3}. In general there is no concept of "position" within a set, unlike a vector or matrix.

(Q1) Set operations:

Let the sets A, B, C be defined by A := {1, 2, 3}, B := {2, 4, 6}, C := {4, 5, 6}.

1. What are the unions $A \cup B$ and $A \cup C$?
2. What are the intersections $A \cap B$ and $A \cap C$?
3. What are the complements $A \setminus B$ and $A \setminus C$?
4. Are A and B disjoint? Are A and C disjoint?
5. Are B and $A \setminus B$ disjoint?
6. Write down an arbitrary partition of {1,2,3,4,5,6} consisting of two sets. Also, write down another partition of {1,2,3,4,5,6} consisting of three sets.

Answers:

1. $A \cup B = \{1,2,3,4,6\}$. $A \cup C = \{1,2,3,4,5,6\}$.
2. $A \cap B = \{2\}$. $A \cap C = \emptyset$
3. $A \setminus B = \{1,3\}$. $A \setminus C = \{1,2,3\}$.
4. A and B are not disjoint, since $A \cap B = \{2\}$. A and C are disjoint, since $A \cap C = \emptyset$.
5. Yes.
6. A partition consisting of two sets: {1,2,3} and {4,5,6}. A partition consisting of three sets: {1,2}, {3,4,5}, {6}.

(Q2) Complements, subsets and De Morgan's laws

Let Ω be a sample space. Recall that for an event $A \subseteq \Omega$ the complement $A^c := \Omega \setminus A := \{w \in \Omega : w \notin A\}$. Take a pair of events $A \subseteq \Omega$ and $B \subseteq \Omega$.

1. Can you give an expression for $(A^c)^c$ without using the notion of a complement?
2. What is Ω^c?
3. (Subsets) Show that if $A \subseteq B$, then $B^c \subseteq A^c$.
4. (De Morgan's laws) Show that $(A \cap B)^c = A^c \cup B^c$. Let's suppose we have a sequence of events $A_1, A_2, \cdots, A_K \subseteq \Omega$. Can you write out an expression for $(\cap_{k=1}^{K} A_k)^c$?
5. (De Morgan's laws) Show that $(A \cup B)^c = A^c \cap B^c$.
6. Let's suppose we have a sequence of events $A_1, A_2, \cdots, A_K \subseteq \Omega$. Can you write out an expression for $(\cup_{k=1}^{K} A_k)^c$?

Answers:

1. A
2. The empty set.
3. Suppose that $A \subseteq B$. Now suppose $x \in B^c$. Then $x \notin B$. Hence, $x \notin A$ since if it were the case that $x \in A$ we would also have $x \in B$. Hence, $x \in A^c$. Since this holds for all elements of B^c, we must have $B^c \subseteq A^c$.

4. First, we show that $x \in (A \cap B)^c$ implies $x \in A^c \cup B^c$. Suppose that $x \in (A \cap B)^c$. Then $x \notin A \cap B$. So either $x \notin A$, in which case $x \in A^c$, or $x \notin B$, in which case $x \in B^c$ (or both holds). Hence, if $x \in (A \cap B)^c$ we mush have $x \in A^c \cup B^c$. Second, we show that $x \in A^c \cup B^c$ implies $x \in (A \cap B)^c$. Suppose that $x \in A^c \cup B^c$. Then either $x \in A^c$, so $x \notin A$, so $x \notin A \cap B$, or $x \in B^c$, so $x \notin B$ so $x \notin A \cap B$. Either way $x \notin A \cap B$, and hence $x \in (A \cap B)^c$.
5. Let's apply $(A \cap B)^c = A^c \cup B^c$. with A^c in place of A and B^c in place of B. This gives the result $(A^c \cap B^c)^c = (A^c)^c \cup (B^c)^c$. By taking complements, and applying the result $(S^c)^c = S$ with A, B, in place of S yields
$$A^c \cap B^c = ((A^c \cap B^c)^c)^c = ((A^c)^c \cup (B^c)^c)^c = (A \cup B)^c.$$
6. $(\cup_{k=1}^K A_k)^c = \cap_{k=1}^K A_k^c$

(Q3) Cardinality and the set of all subsets:

Suppose that $\Omega = \{w_1, w_2, \cdots, w_K\}$ contains K elements for some natural number K. Here Ω has cardinality K.

Let E be a set of all subsets of Ω, i.e., $E := \{A | A \subset \Omega\}$. Note that here E is a set. Give a formula for the cardinality of E in terms of K.

Answer: The cardinality of E is 2^K.

(Q4) Disjointness and partitions.

Suppose we have a sample space Ω, and events A_1, A_2, A_3, A_4 are subsets of Ω.

1. Can you think of a set which is disjoint from every other set? That is, find a set $A \subseteq \Omega$ such that $A \cap B = \emptyset$ for all $B \subseteq \Omega$.
2. Define events $S_1 := A_1, S_2 = A_2 \setminus A_1, S_3 = A_3 \setminus (A_1 \cup A2), S_4 = A_4 \setminus (A_1 \cup A2 \cup A_3)$. Show that S_1, S_2, S_3, S_4 form a partition of $A_1 \cup A_2 \cup A_3 \cup A_4$.

Answers:

1. The empty set satisfies $\emptyset \cap B = \emptyset$ for all $B \subseteq \Omega$.
2. We must show (1) that $S_1 \cup S_2 \cup S_3 \cup S_4 = A_1 \cup A_2 \cup A_3 \cup A_4$ and (2) that S_1, S_2, S_3, S_4 are disjoint.

(1). Take $x \in S_1 \cup S_2 \cup S_3 \cup S_4$. Then $x \in S_i = A_i \setminus_{j<i} A_j$ for some $i \in \{1,2,3,4\}$ and so $x \in A_i$ and hence $x \in A_1 \cup A_2 \cup A_3 \cup A_4$. On the other hand, if $x \in A_1 \cup A_2 \cup A_3 \cup A_4$, we may choose $i \in \{1,2,3,4\}$ as small as possible with $x \in A_i$. It follows that $x \in S_i = A_i \setminus_{j<i} A_j$ and so $x \in S_1 \cup S_2 \cup S_3 \cup S_4$. Hence, the sets $S_1 \cup S_2 \cup S_3 \cup S_4$ and $A_1 \cup A_2 \cup A_3 \cup A_4$ have exactly the same elements and so are equal.

(2). To show $S_1 \cup S_2 \cup S_3 \cup S_4$ are disjoint, let's show that there cannot be any $x \in S_{i_0} \cap S_{i_1}$ with $i_0 < i_1$. Indeed, if $x \in S_{i_1} = A_{i_1} \setminus \left(\cup_{j<i_i} A_j\right)$ so $x \notin A_{i_0}$ and so $x \notin S_{i_0}$.

(Q5) Indicator function.

Suppose we have a sample space Ω, and the event A is a subset of Ω. Let $\mathbf{1}_A$ be the indicator function of A.

1. Write down the indicator function $\mathbf{1}_{A^c}$of A^c (use $\mathbf{1}_A$ in your formula).
2. Can you find a set B whose indicator function is $\mathbf{1}_{A^c} + \mathbf{1}_A$?
3. Recall that $\mathbf{1}_{A\cap B} = \mathbf{1}_A \cdot \mathbf{1}_B$ and $\mathbf{1}_{A\cup B} = \max(\mathbf{1}_A, \mathbf{1}_B) = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_A \cdot \mathbf{1}_B$ for any $A \subseteq \Omega$ and $B \subseteq \Omega$. Combining this with the conclusion from Question (Q5) 1, use indicator functions to prove $(A \cap B)^c = A^c \cup B^c$ (De Morgan's laws).

Answer:

1. $\mathbf{1}_{A^c} = 1 - \mathbf{1}_A$.
2. $\mathbf{1}_{A^c} + \mathbf{1}_A = 1$, so this is the indicator function of $\Omega = A \cup A^c$.
3. $\mathbf{1}_{A^c\cup B^c} = (1 - \mathbf{1}_A) + (1 - \mathbf{1}_B) - (1 - \mathbf{1}_A) \cdot (1 - \mathbf{1}_B) = 1 - \mathbf{1}_A\mathbf{1}_B = 1 - \mathbf{1}_{A\cap B} = \mathbf{1}_{(A\cap B)^c}$. Since $(A \cap B)^c$ and $A^c \cup B^c$ have the same indicator function, they must be the same.

(Q6) Uncountable infinities (this is an optional extra).

This is a challenging optional extra. You may want to return to this question once you have completed all other questions.

Show that the set of numbers $\Omega := [0,1]$ is uncountably infinite.

Answer:

We use a proof due to Cantor known as the diagonalisation argument. We shall suppose that $[0,1]$ is countable and then deduce a contradiction. So suppose that $[0,1]$ is countable and that there is an enumeration $(a_n)_{n\in\mathbb{N}}$ of $[0,1]$. For each $n \in \mathbb{N}$, let $a_{n,j}$ be the corresponding decimal expansion, so each $a_{n,j} \in \{0,1,\cdots,9\}$ and $a_n = 0.\,a_{n,1}a_{n,2}a_{n,3}a_{n,4}\cdots$. Now choose $x \in [0,1]$ with decimal expansion $\left(x_j\right)_{j\in\mathbb{N}}$ by setting $x_j \in \{1,2,\cdots,9\} \setminus a_{j,j}$ for all $j \in \mathbb{N}$. It follows that $x \neq a_n$ for all $n \in \mathbb{N}$, and hence $(a_n)_{n\in\mathbb{N}}$ is not an enumeration of $[0,1]$. This is a contradiction, hence there does not exist such an enumeration, so $[0,1]$ is uncountable.

3. Probability theory

In this section we consider some of the concepts introduced in Lecture 6.

Recall that we have introduced the three key rules of probability. Given a sample space Ω along with a well-behaved collection of events $\mathcal{E}$, a probability $\mathbb{P}$ is a function which assigns a number $\mathbb{P}(A)$ to each event $A \in \mathcal{E}$, and satisfies rules 1, 2, and 3:

: $\mathbb{P}(A) \geq 0$ for any event $A \in \mathcal{E}$

: $\mathbb{P}(\Omega) = 1$ for sample space Ω

: For pairwise disjoint events $A_1, A_2, \cdots$ in $\mathcal{E}$, we have

$$\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

3.1 Rules of probability

(Q1) Construct a probability function based on the Rules of probability

Consider a sample space $\Omega = \{a, b, c\}$ and a set of events $\mathcal{E} = \{A \subseteq \Omega\}$ (i.e., $\mathcal{E}$ consists of all subsets of Ω). Based on the rules of probability, find a probability function $\mathbb{P}: \mathcal{E} \to [0,1]$ that satisfies

$$\mathbb{P}(\{a, b\}) = 0.6 \quad \text{ and } \mathbb{P}(\{b, c\}) = 0.5.$$

In your example, you need to define a function called $\mathbb{P}$. The function maps each event in $\mathcal{E}$ to a number. Make sure that your function $\mathbb{P}$ satisfies the three rules, but you don't need to write down the proof (that it satisfies the three rules).

Answer:

$\mathbb{P}(\{a\}) = 0.5$, $\mathbb{P}(\{b\}) = 0.1$, $\mathbb{P}(\{c\}) = 0.4$, $\mathbb{P}(\{a, b\}) = 0.6$, $\mathbb{P}(\{b, c\}) = 0.5$, $\mathbb{P}(\{a, c\}) = 0.9$, $\mathbb{P}(\{a, b, c\}) = 1$.

(Q2) Verify that the following probability space satisfies the rules of probability.

Consider a setting in which the sample space $\Omega = \{0,1\}$, and $\mathcal{E} = \{A \subseteq \Omega\} = \{\emptyset, \{0\}, \{1\}, \{0,1\}\}$. For a fixed $q \in [0,1]$, define a function $\mathbb{P}: \mathcal{E} \to [0,1]$ by

$$\mathbb{P}(\emptyset) = 0,\ \mathbb{P}(\{0\}) = 1 - q,\ \mathbb{P}(\{1\}) = q,\ \mathbb{P}(\{0,1\}) = 1.$$

Show that the probability space $(\Omega, \mathcal{E}, \mathbb{P})$ satisfies the three rules of probability.

Answer:

Rule 1: Since by definition $\mathbb{P}(\emptyset) = 0 \geq 0,\ \mathbb{P}(\{0\}) = 1 - q \geq 0,\ \mathbb{P}(\{1\}) = q \geq 0,\ \mathbb{P}(\{0,1\}) = 1 \geq 0$, so $\mathbb{P}(A) \geq 0$ for any $A \in \mathcal{E}$.

Rule 2: $\mathbb{P}(\Omega) = \mathbb{P}(\{0,1\}) = 1$

Rule 3: To show that the third rule holds we consider three cases.

1. If $A_n = \emptyset$ for all $n \in \mathbb{N}$, then $0 = \mathbb{P}(\emptyset) = \mathbb{P}(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mathbb{P}(A_k)$.
2. If there is exactly one $n \in \mathbb{N}$ such that $A_n \neq 0$, then $\cup_{k=1}^{\infty} A_k = A_n$, and hence $\mathbb{P}(\cup_{k=1}^{\infty} A_k) = \mathbb{P}(A_n) = \sum_{k=1}^{\infty} \mathbb{P}(A_k)$.
3. If there are exactly two of the sets in $\{A_j\}$ that are non-empty, then there must be exactly one $n_0 \in \mathbb{N}$ such that $A_{n_0} = \{0\}$ and exactly one $n_1 \in \mathbb{N}$ such that $A_{n_1} = \{1\}$. So

$$\sum_{k=1}^{\infty} \mathbb{P}\,(A_k) = \mathbb{P}\big(A_{n_0}\big) + \mathbb{P}\big(A_{n_0}\big) + \sum_{k\in\mathbb{N}\setminus\{n_0,n_1\}} \mathbb{P}\,(A_k) = \mathbb{P}\big(A_{n_0}\big) + \mathbb{P}\big(A_{n_0}\big)$$
$$= (1-q) + q = 1 = \mathbb{P}(\Omega) = \mathbb{P}(\cup_{k=1}^{\infty} A_k).$$

Moreover, it is not possible to have more than two of the sets in $\{A_j\}$ that are non-empty, since they are disjoint sets.

3.2 Deriving new properties from the rules of probability

(Q1) Union of a finite sequence of disjoint events.

Recall that in Rule 3, we have

$$\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}\,(A_i)$$

for an infinite sequence of pairwise disjoint events $A_1, A_2, \cdots$. Show that for a finite sequence of disjoint events $A_1, A_2, \cdots A_n$, for any integer n bigger than 1, the below equality holds as a consequence of Rule 3:

$$\mathbb{P}(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbb{P}\,(A_i)$$

Please note that in left hand side of the equation above we have the union of a finite sequence instead of an infinite sequence.

Answer:

Let's define a sequence of events $A_{n+1} = \emptyset, A_{n+2} = \emptyset, \cdots$.

So $A_1, A_2, \cdots, A_n, A_{n+1}, \cdots$ are pairwise disjoint. If not, then there exist i, j such that A_i and A_j are not disjoint. As $A_1, A_2, \cdots A_n$ are pairwise disjoint, at least one of $\{i, j\}$ must be bigger than n, which corresponds to an empty set, hence contradicting to the fact that an empty set is disjoint from any sets.

Therefore

$$\mathbb{P}(\cup_{i=1}^{n} A_i) = \mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}\,(A_i) = \sum_{i=1}^{n} \mathbb{P}\,(A_i),$$

where the second equality follows from Rule 3.

(Q2) Probability of a complement.

Prove that if Ω is a sample space, $S \subseteq \Omega$ is an event and $S^c := \Omega \setminus S$ is its complement, then we have

$$\mathbb{P}(S^c) = 1 - \mathbb{P}(S).$$

Answer:

By Rule 3, we have

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(S^c \cup S) = \mathbb{P}(S^c) + \mathbb{P}(S)$$

Therefore, $\mathbb{P}(S^c) = 1 - \mathbb{P}(S)$.

(Q3) The union bound

In Rule 3, for pairwise disjoint events $A_1, A_2, \cdots$, we have

$$\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}\,(A_i)$$

Recall that in the lecture we have also shown the union bound as a consequence of the rules of probability: for a sequence of events $S_1, S_2, \cdots$, we have $\mathbb{P}(\cup_{i=1}^{\infty} S_i) \leq \sum_{i=1}^{\infty} \mathbb{P}\,(S_i)$.

Give an example of a probability space and a sequence of sets $S_1, S_2, \cdots$, such that $\mathbb{P}(\cup_{i=1}^{\infty} S_i) \neq \sum_{i=1}^{\infty} \mathbb{P}\,(S_i)$.

Answer: Consider the example in 1.1 (Q2). Assuming $q > 0$, let $S_1 = \{1\}, S_2 = \{1\}, S_3 = \{1\}$ and $S_i = \emptyset$ for $i > 3$. Then

$$\mathbb{P}(\cup_{i=1}^{\infty} S_i) = q \neq 3q = \sum_{i=1}^{\infty} \mathbb{P}\,(S_i).$$

(Q4) Probability of union and intersection of events.

Show that for events $A \subseteq \Omega$ and $B \subseteq \Omega$, we have

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Answer:

Since $A \setminus B$ and B are disjoint and $(A \setminus B) \cup B = A \cup B$, we have

$$\mathbb{P}(A \cup B) = \mathbb{P}(A \setminus B) + \mathbb{P}(B).$$

Let $S = A \cap B$, so $A \setminus S = A \setminus B$. Since $A \setminus S$ and S are disjoint and $(A \setminus S) \cup S = A \cup S = A$, we have

$$\mathbb{P}(A) = \mathbb{P}\big((A \setminus S) \cup S\big) = \mathbb{P}(A \setminus S) + \mathbb{P}(S) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B).$$

Therefore, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.