Stats theory (SDA)

Alejandro Caceres

2022-10-01

Contents

1 About **9**
1.1 Recommended reading list 9

2 Data description **11**
2.1 Objective 11
2.2 Statistics 11
2.3 Scientific method 11
2.4 Outcome 11
2.5 Types of outcome 12
2.6 Random experiments 12
2.7 Absolute frequencies 12
2.8 Example 13
2.9 Relative frequencies 13
2.10 Example 14
2.11 Bar plot 14
2.12 Pie chart 15
2.13 Categorical and ordered variables 15
2.14 Example 16
2.15 Absolute and relative cumulative frequencies 16
2.16 Frequency table 16
2.17 Cumulative frequency plot 17
2.18 Continuous variables 17
2.19 Bins 18
2.20 Create a categorical variable from a continuous one 18
2.21 Frequency table for a continuous variable 19
2.22 Histogram 19
2.23 Histogram 20
2.24 Cumulative frequency plot: Continous variables 21
2.25 Summary statistics 22
2.26 Average 23
2.27 Average (categorical ordered) 23
2.28 Average (categorical ordered) 23
2.29 Average 24

2.30 Average 24
2.31 Median 24
2.32 Median Vs Average 25
2.33 Dispersion 26
2.34 Dispersion 27
2.35 Sample variance 27
2.36 Sample variance 27
2.37 Standard deviation 28
2.38 IQR 28
2.39 IQR 29
2.40 Box plot 29

3 Probability **31**
3.1 Objective 31
3.2 Random experiments 31
3.3 Probability 32
3.4 Example 32
3.5 Example 32
3.6 Relative frequency 32
3.7 At infinity 33
3.8 Frequentist probability 34
3.9 Classical Probability 34
3.10 Classical and frequentist probabilities 35
3.11 Probability 35
3.12 Sample space 35
3.13 Examples of sample spaces 36
3.14 Discrete and continuous sample spaces 36
3.15 Event 36
3.16 Event operations 37
3.17 Event operations example 37
3.18 Outcomes 37
3.19 Probability definition 38
3.20 Probability properties 38
3.21 Addition Rule 39
3.22 Example Addition Rule 39
3.23 Venn diagram 39
3.24 Probability table 40
3.25 Example probability table 40
3.26 Contingency table 40
3.27 Example contingency table 41
3.28 Misophonia study 41
3.29 Contingency table for frequencies 44
3.30 Heat map 45
3.31 Continous variables 45
3.32 Heat map for continuous variables 48
3.33 Scatter plot 49

4 Conditional Probability **51**
4.1 Objective 51
4.2 Joint Probability 51
4.3 Diagnostics 52
4.4 Diagnostics Test 52
4.5 Observations 52
4.6 Contingency tables 53
4.7 Conditional probability 53
4.8 Conditional probability 54
4.9 Conditional contingency table 54
4.10 Example conditional contingency table 55
4.11 Multiplication rule 55
4.12 Diagnostic performance 55
4.13 Multiplication rule 56
4.14 Contingency table in terms of conditional probabilities 56
4.15 Conditional tree 57
4.16 Contingency table in terms of conditional probabilities 57
4.17 Total probability rule 57
4.18 Conditional tree 58
4.19 Finding reverse probabilities 58
4.20 Recover joint probabilities 58
4.21 Reverse conditionals 58
4.22 Baye's theorem 59
4.23 Example: Bayes' theorem 60
4.24 Example: Bayes' theorem 60
4.25 Statistical independence 61
4.26 Statistical independence 61
4.27 Statistical independence 61
4.28 Statistical independence 62
4.29 Products of marginals products 62
4.30 Example 63

5 Discrete Random Variables **65**
5.1 Objective 65
5.2 How do we assign probability values to outcomes? 65
5.3 Random variable 65
5.4 Random variable 66
5.5 Events of observing a random variable 66
5.6 Probability of random variables 67
5.7 Probability functions 67
5.8 Probability functions 68
5.9 Probability functions 68
5.10 Probability functions 69
5.11 Example: Probability mass function 69
5.12 Probability table for equally likely outcomes 70
5.13 Probability table for X 70

5.14 Example . . . 70
5.15 Example . . . 71
5.16 Probabilities and frequencies . . . 71
5.17 Probabilities and relative frequencies . . . 72
5.18 Mean and Variance . . . 72
5.19 Mean and Variance . . . 73
5.20 Mean . . . 73
5.21 Example: Mean . . . 73
5.22 Variance . . . 74
5.23 Example: Variance . . . 75
5.24 Functions of X . . . 75
5.25 Example: Variance about the origin . . . 76
5.26 Probability distribution . . . 76
5.27 Example: Probability distribution . . . 77
5.28 Probability distribution . . . 77
5.29 Probability function and Probability distribution . . . 78
5.30 Probability function and Probability distribution . . . 78
5.31 Quantiles . . . 78
5.32 Summary . . . 79

6 Continous Random Variables 81
6.1 Objective . . . 81
6.2 Continuous random variable . . . 81
6.3 Continuous random variable . . . 82
6.4 Continuous random variable . . . 82
6.5 Continuous random variable . . . 83
6.6 Continuous random variable . . . 83
6.7 Total area under the curve . . . 84
6.8 Area under the curve . . . 84
6.9 Area under the curve . . . 85
6.10 Probability distribution . . . 86
6.11 Probability distribution . . . 88
6.12 Probability distribution . . . 88
6.13 Probability distribution . . . 89
6.14 Probability graphics . . . 89
6.15 Probability graphics . . . 90
6.16 Mean . . . 91
6.17 Mean . . . 91
6.18 Variance . . . 92
6.19 Functions of X . . . 92
6.20 Example . . . 93

7 Exercises 95
7.1 Data description . . . 95
7.2 Probability . . . 96
7.3 Conditional Probability . . . 97

7.4 Random variables 100
7.5 Probability Models 102
7.6 Point Estimators 103
7.7 Sampling and Central Limit Theorem 103
7.8 Maximum likelihood 105
7.9 Method of moments 106
7.10 Confidence intervals 107

Chapter 1

About

- This is the introduction to statistics course from EEBE (UPC).
- Exam dates and additional study material can be found in ATENEA

1.1 Recommended reading list

- Douglas C. Montgomery and George C. Runger. "Applied Statistics and Probability for Engineers" 4th Edition. Wiley 2007.

Chapter 2

Data description

2.1 Objective

- Data: discrete, continuous
- Summarizing data in tables and figures

2.2 Statistics

- Solve problems in a systematic way (science, engineering and technology)
- Modern humans use a general **method** historically developed for thousands of years! ... and still under development.
- It has three main components: observation, logic, and generation of new knowledge

2.3 Scientific method

2.4 Outcome

Observation or *Realization*

- an **observation** is the acquisition of a number or a characteristic from an experiment

… 1 0 0 1 0 **1** 0 1 1 … (the number in bold is an observation in a repetition of the experiment)

Outcome

- An **outcome** is a possible observation that is the result of an experiment.

1 is an outcome, **0** is the other outcome

2.5 Types of outcome

- **Categorical**: If the result of an experiment can only take discrete values (number of car pieces produced per hour, number of leukocytes in blood)
- **Continuous**: If the result of an experiment can only take continuous values (battery state of charge, engine temperature).

2.6 Random experiments

Definition:

A **random experiment** is an experiment that gives different outcomes when repeated in the same manner.

Examples:

- on the same object (person): temperature, sugar levels.
- on different objects but the same measurement: the weight of an animal.
- on events: a number of emails received in an hour.

2.7 Absolute frequencies

When we repeat a random experiment, we record a list of outcomes.

We summarize the **categorical** observations by counting how many times we saw a particular outcome.

Absolute frequency:

$$n_i$$

is the number of times we observed the outcome i

2.8 Example

Random experiment: Extract a leukocyte from **one** donor and write down its type. Repeat experiment $N = 119$ times.

`(T cell, Tcell, Neutrophil, ..., B cell)`

```
##      outcome ni
## 1     T Cell 34
## 2     B cell 50
## 3   basophil 20
## 4   Monocyte  5
## 5 Neutrophil 10
```

- For instance: $n_1 = 34$ is total number of T cells
- $N = \sum_i n_i = 119$

2.9 Relative frequencies

We can also summarize the observations by computing the **proportion** of how many times we saw a particular outcome.

$$f_i = n_i/N$$

where N is the total number of observations

In our example there are recorded $n_1 = 34$ T cells, so we ask for the proportion of T cells from the total of 119.

2.10 Example

```
##      outcome ni         fi
## 1     T Cell 34 0.28571429
## 2     B cell 50 0.42016807
## 3   basophil 20 0.16806723
## 4   Monocyte  5 0.04201681
## 5 Neutrophil 10 0.08403361
```

We have

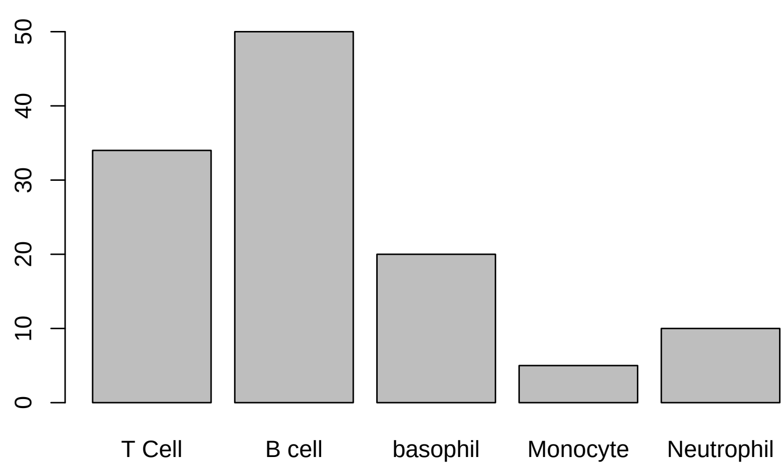
$\sum_{i=1..M} n_i = N$

$\sum_{i=1..M} f_i = 1$

where M is the number of outcomes.

2.11 Bar plot

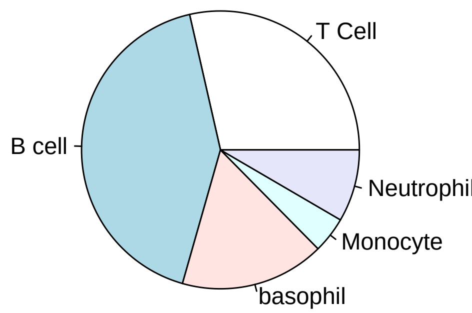
We can plot n_i Vs the outcomes, giving us a bar plot



2.12 Pie chart

We can visualize the relative frequencies with a pie chart

- Where the area of the circle represents 100% of observations (proportion = 1) and the sections the relative frequencies of all the outcomes.



2.13 Categorical and ordered variables

Cell types are not meaningfully ordered concerning the outcomes. However, sometimes **categorical** variables can be **ordered**.

Misophonia study:

- 123 patients were examined for misophonia: anxiety/anger produced by certain sounds
- They were categorized into 4 different groups according to severity.

2.14 Example

The results of the study are:

```
##   [1] 4 2 0 3 0 0 2 3 0 3 0 2 2 0 2 0 0 3 3 0 3 3 2 0 0 0 4 2 2 0 2 0 0 0 3 0 2
##  [38] 3 2 2 0 2 3 0 0 2 2 3 3 0 0 4 3 3 2 0 2 0 0 0 2 2 0 0 2 3 0 1 3 2 4 3 2 3
##  [75] 0 2 3 2 4 1 2 0 2 0 2 0 2 2 4 3 0 3 0 0 0 2 2 1 3 0 0 3 2 1 3 0 4 4 2 3 3
## [112] 3 0 3 2 1 2 3 3 4 2 3 2
```

And its frequency table

```
##   outcome ni         fi
## 1       0 41 0.33333333
## 2       1  5 0.04065041
## 3       2 37 0.30081301
## 4       3 31 0.25203252
## 5       4  9 0.07317073
```

2.15 Absolute and relative cumulative frequencies

Misophonia severity is **categorical** and **ordered**.

When outcomes can be ordered then it is useful to ask how many observations were obtained up to a given outcome we call this number the absolute cumulative frequency up to the outcome i:

$$N_i = \sum_{k=1..i} n_k$$

It is also useful to compute the **proportion** of the observations that was obtained up to a given outcome

$$F_i = \sum_{k=1..i} f_k$$

2.16 Frequency table

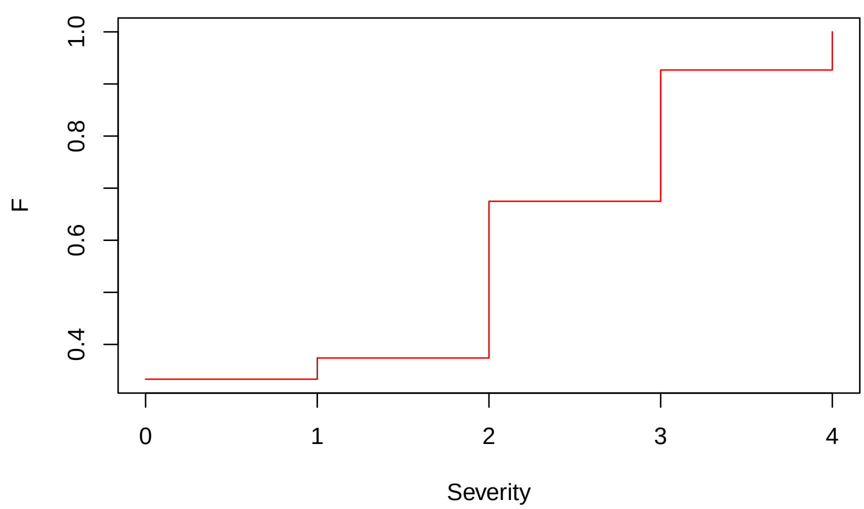
```
##   outcome ni         fi  Ni        Fi
## 0       0 41 0.33333333  41 0.3333333
## 1       1  5 0.04065041  46 0.3739837
```

```
## 2        2 37 0.30081301  83 0.6747967
## 3        3 31 0.25203252 114 0.9268293
## 4        4  9 0.07317073 123 1.0000000
```

- **67%** of patients had misophonia up to severity **2**
- **37%** of patients have severity less or equal than **1**

2.17 Cumulative frequency plot

We can also plot the cumulative frequency Vs the outcomes



2.18 Continuous variables

The result of a random experiment can also give continuous outcomes.

In the misophonia study, the researchers asked whether the convexity of the jaw would affect the misophonia severity (the scientific hypothesis is that the convexity angle of the jaw can influence the ear and its sensitivity). These are the results for the convexity of the jaw (degrees)

```
##   [1]  7.97 18.23 12.27  7.81  9.81 13.50 19.30  7.70 12.30  7.90 12.60 19.00
##  [13]  7.27 14.00  5.40  8.00 11.20  7.75  7.94 16.69  7.62  7.02  7.00 19.20
##  [25]  7.96 14.70  7.24  7.80  7.90  4.70  4.40 14.00 14.40 16.00  1.40  9.76
##  [37]  7.90  7.90  7.40  6.30  7.76  7.30  7.00 11.23 16.00  7.90  7.29  6.91
##  [49]  7.10 13.40 11.60 -1.00  6.00  7.82  4.80 11.00  9.00 11.50 16.00 15.00
##  [61]  1.40 16.80  7.70 16.14  7.12 -1.00 17.00  9.26 18.70  3.40 21.30  7.50
##  [73]  6.03  7.50 19.00 19.01  8.10  7.80  6.10 15.26  7.95 18.00  4.60 15.00
##  [85]  7.50  8.00 16.80  8.54  7.00 18.30  7.80 16.00 14.00 12.30 11.40  8.50
##  [97]  7.00  7.96 17.60 10.00  3.50  6.70 17.00 20.26  6.64  1.80  7.02  2.46
## [109] 19.00 17.86  6.10  6.64 12.00  6.60  8.70 14.05  7.20 19.70  7.70  6.02
## [121]  2.50 19.00  6.80
```

2.19 Bins

Continuous outcomes cannot be counted!

We transform them into ordered categorical variables

- We cover the range of the observations into regular intervals of the same size (bins)

```
## [1] "[-1.02,3.46]" "(3.46,7.92]"  "(7.92,12.4]"  "(12.4,16.8]"  "(16.8,21.3]"
```

2.20 Create a categorical variable from a continuous one

- We map each observation to its interval: creating an **ordered categorical** variable; in this case with 5 possible outcomes

```
##   [1] "(7.92,12.4]"  "(16.8,21.3]"  "(7.92,12.4]"  "(3.46,7.92]"  "(7.92,12.4]"
##   [6] "(12.4,16.8]"  "(16.8,21.3]"  "(3.46,7.92]"  "(7.92,12.4]"  "(3.46,7.92]"
##  [11] "(12.4,16.8]"  "(16.8,21.3]"  "(3.46,7.92]"  "(12.4,16.8]"  "(3.46,7.92]"
##  [16] "(7.92,12.4]"  "(7.92,12.4]"  "(3.46,7.92]"  "(7.92,12.4]"  "(12.4,16.8]"
##  [21] "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"  "(16.8,21.3]"  "(7.92,12.4]"
##  [26] "(12.4,16.8]"  "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"
##  [31] "(3.46,7.92]"  "(12.4,16.8]"  "(12.4,16.8]"  "(12.4,16.8]"  "[-1.02,3.46]"
##  [36] "(7.92,12.4]"  "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"
```

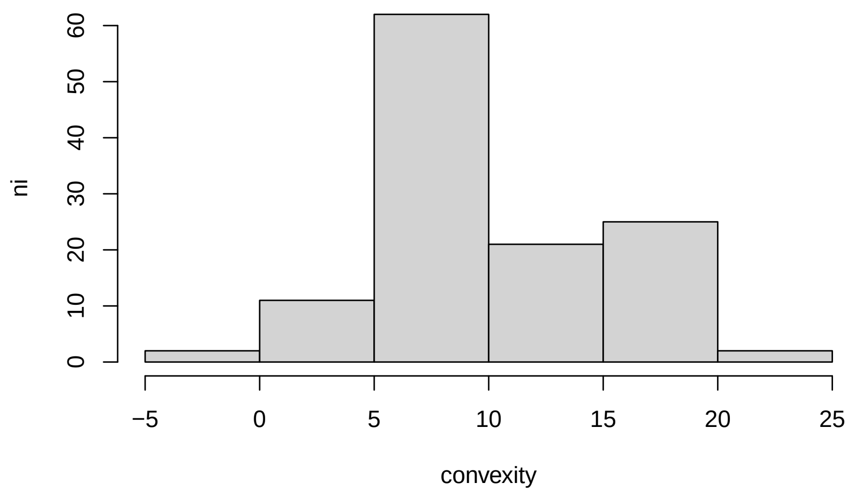
```
##  [41] "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"  "(7.92,12.4]"  "(12.4,16.8]"
##  [46] "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"  "(12.4,16.8]"
##  [51] "(7.92,12.4]"  "[-1.02,3.46]" "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"
##  [56] "(7.92,12.4]"  "(7.92,12.4]"  "(7.92,12.4]"  "(12.4,16.8]"  "(12.4,16.8]"
##  [61] "[-1.02,3.46]" "(12.4,16.8]"  "(3.46,7.92]"  "(12.4,16.8]"  "(3.46,7.92]"
##  [66] "[-1.02,3.46]" "(16.8,21.3]"  "(7.92,12.4]"  "(16.8,21.3]"  "[-1.02,3.46]"
##  [71] "(16.8,21.3]"  "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"  "(16.8,21.3]"
##  [76] "(16.8,21.3]"  "(7.92,12.4]"  "(3.46,7.92]"  "(3.46,7.92]"  "(12.4,16.8]"
##  [81] "(7.92,12.4]"  "(16.8,21.3]"  "(3.46,7.92]"  "(12.4,16.8]"  "(3.46,7.92]"
##  [86] "(7.92,12.4]"  "(12.4,16.8]"  "(7.92,12.4]"  "(3.46,7.92]"  "(16.8,21.3]"
##  [91] "(3.46,7.92]"  "(12.4,16.8]"  "(12.4,16.8]"  "(7.92,12.4]"  "(7.92,12.4]"
##  [96] "(7.92,12.4]"  "(3.46,7.92]"  "(7.92,12.4]"  "(16.8,21.3]"  "(7.92,12.4]"
## [101] "(3.46,7.92]"  "(3.46,7.92]"  "(16.8,21.3]"  "(16.8,21.3]"  "(3.46,7.92]"
## [106] "[-1.02,3.46]" "(3.46,7.92]"  "[-1.02,3.46]" "(16.8,21.3]"  "(16.8,21.3]"
## [111] "(3.46,7.92]"  "(3.46,7.92]"  "(7.92,12.4]"  "(3.46,7.92]"  "(7.92,12.4]"
## [116] "(12.4,16.8]"  "(3.46,7.92]"  "(16.8,21.3]"  "(3.46,7.92]"  "(3.46,7.92]"
## [121] "[-1.02,3.46]" "(16.8,21.3]"  "(3.46,7.92]"
```

2.21 Frequency table for a continuous variable

```
##        outcome ni         fi  Ni         Fi
## 1 [-1.02,3.46]  8 0.06504065   8 0.06504065
## 2  (3.46,7.92] 51 0.41463415  59 0.47967480
## 3  (7.92,12.4] 26 0.21138211  85 0.69105691
## 4  (12.4,16.8] 20 0.16260163 105 0.85365854
## 5  (16.8,21.3] 18 0.14634146 123 1.00000000
```

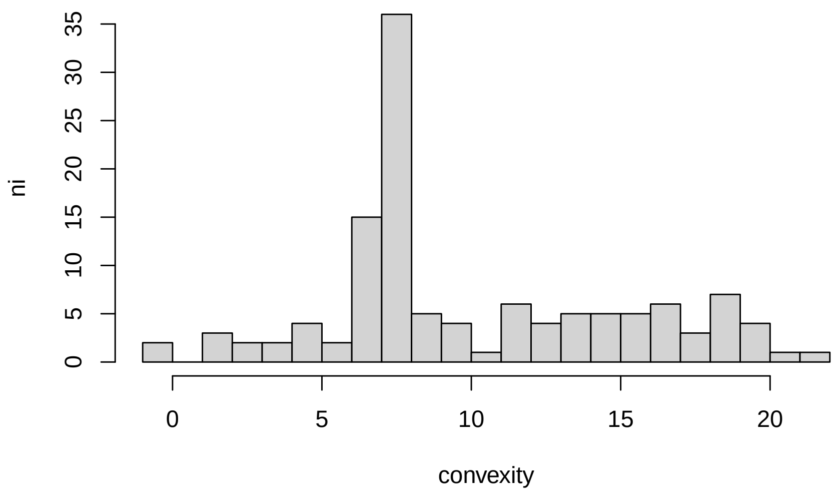
2.22 Histogram

The histogram is the plot of n_i or f_i Vs the outcomes (bins). The histogram depends on the size of the bins



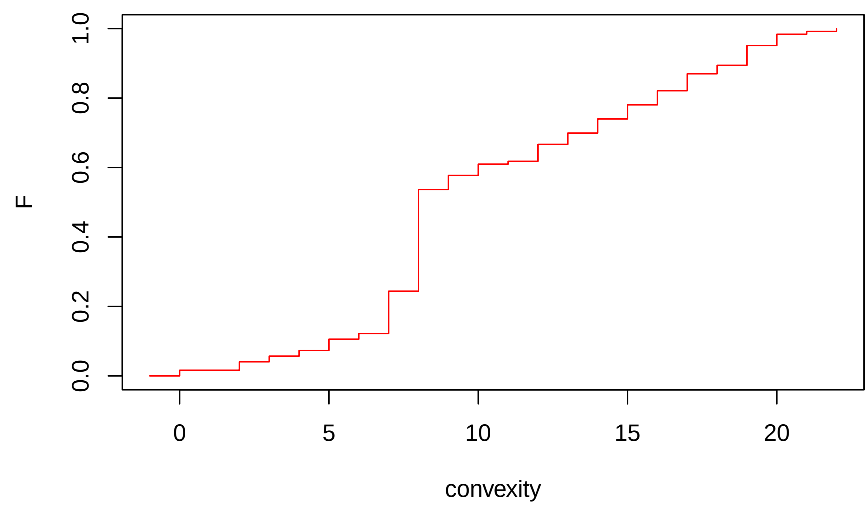
2.23 Histogram

The histogram is the plot of n_i or f_i Vs the outcomes (bins). The histogram depends on the size of the bins



2.24 Cumulative frequency plot: Continous variables

We can also plot the cumulative frequency Vs the outcomes



2.25 Summary statistics

The summary statistics are numbers computed from the data that tell us important features of numerical variables (categorical or continuous).

Limiting values:

- minimum: the minimum outcome observed
- maximum: the maximum outcome observed

Central value for the outcomes

- The average is defined as

$$\bar{x} = \frac{1}{N} \sum_{j=1..N} x_j$$

where x_j is the **observation** j (convexity) from a total of N.

2.26 Average

The average convexity can be computed directly from the **observations**

$\bar{x} = \frac{1}{N} \sum_j x_j$

$= \frac{1}{N}(7.97 + 18.23 + 12.27... + 6.80) = 10.19894$

2.27 Average (categorical ordered)

For **categorical ordered** variables we can use the frequency table to compute the average

```
##   outcome ni         fi
## 1       0 41 0.33333333
## 2       1  5 0.04065041
## 3       2 37 0.30081301
## 4       3 31 0.25203252
## 5       4  9 0.07317073
```

The average **severity** of misophonia in the study can **also** be computed from the relative frequencies of the **outcomes**

$\bar{x} = \frac{1}{N} \sum_{i=1...N} x_j = \frac{1}{N} \sum_{i=1...M} x_i * n_i = \sum_{i=1...M} x_i * f_i$

$= 0 * f_0 + 1 * f_1 + 2 * f_2 + 3 * f_3 + 4 * f_4 = 1.691057$

(note the change from N to M in the second summation)

2.28 Average (categorical ordered)

In terms of the **outcomes** of categorical ordered variables, the **average** can be written as

$$\bar{x} = \sum_{i=1...M} x_i f_i$$

from a total of M possible outcomes (number of severity levels).

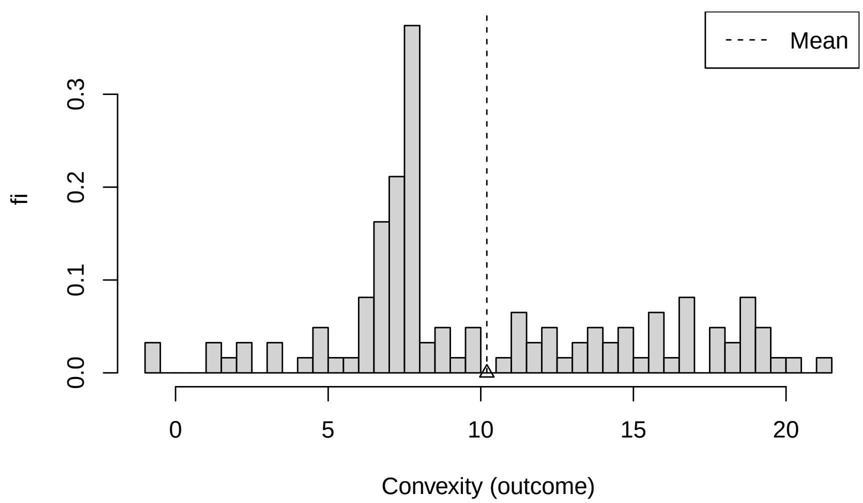
$\bar{x}$ is the **central value** or center of gravity of the outcomes. As if each outcome had a mass density given by f_i.

2.29 Average

- The average is not the result of one observation (random experiment).
- It is the result of a series of observations (sample).
- It describes the number where the observed values balance.

That is why we hear, for instance, that a patient with an infection can infect an average of 2.5 people.

2.30 Average



2.31 Median

Another measure of centrality is the median. The median $q_{0.5}$ is the value x_p

$$median(x) = q_{0.5} = x_p$$

below which we find half of the observations

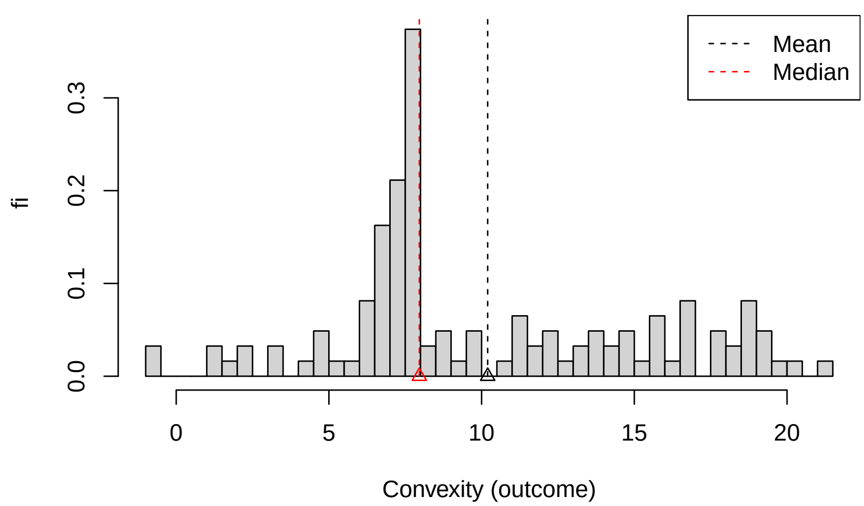
$$\sum_{x \le x_p} 1 = \frac{N}{2}$$

or in terms of the frequencies, is the value x_p that makes the cumulative frequency F_p equal to 0.5

$$q_{0.5} = \sum_{x \le x_p} f_x = F_p = 0.5$$

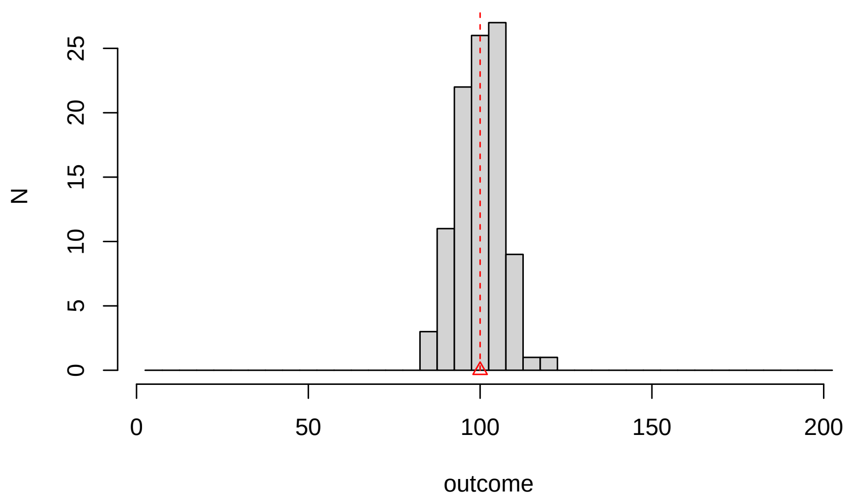
2.32 Median Vs Average

- Average: Center of mass (compensates distant values)
- Median: Half of the mass

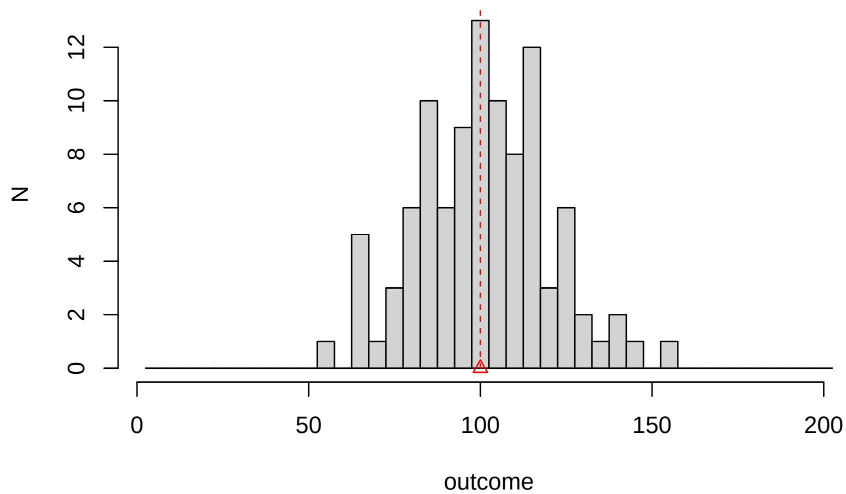


2.33 Dispersion

An important measure of the outcomes is their **dispersion**. Many experiments can share their mean but differ on how dispersed the values are.



2.34 Dispersion



2.35 Sample variance

Dispersion about the mean is measured with the

- The sample variance:

$$s^2 = \frac{1}{N-1} \sum_{j=1..N} (x_j - \bar{x})^2$$

It measures the average square distance of the **observations** to the average. The reason for $N-1$ will be explained when we talk about inference.

2.36 Sample variance

- In terms of the frequencies of **categorical and ordered** variables

$$s^2 = \frac{N}{N-1}\sum_x (x-\bar{x})^2 f_x$$

s^2 can be thought of as the moment of inertia of the observations.

2.37 Standard deviation

The squared root of the sample variance is called the **standard deviation** s.

The standard deviation of the convexity angle is

$s = [\frac{1}{123-1}((7.97 - 10.19894)^2 + (18.23 - 10.19894)^2$
$+(12.27 - 10.19894)^2 + ...)]^{1/2} = 5.086707$

The jaw convexity deviates from its mean by 5.086707.

2.38 IQR

- Dispersion of data can also be measured with respect to the median by the **interquartile range**
- We define the **first** quartile as the value x_p that makes the cumulative frequency F_p equal to 0.25

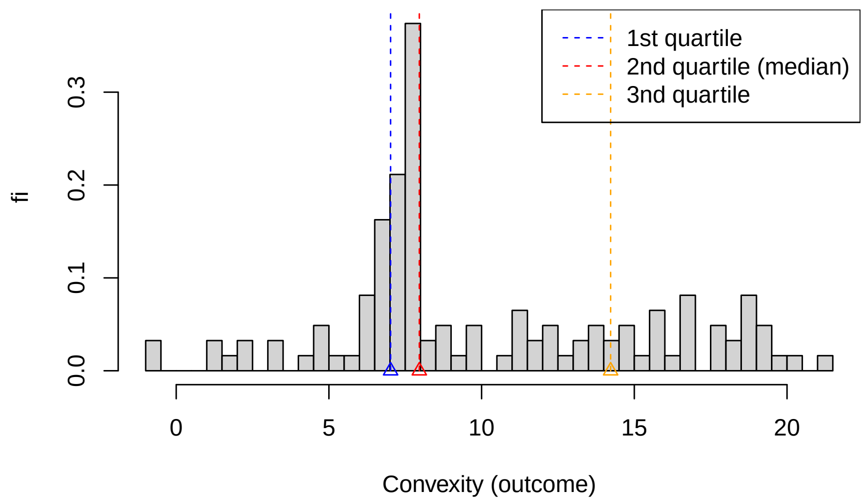
$$q_{0.25} = \sum_{x \le x_p} f_x = F_p = 0.25$$

- We also define the **third** quartile as the value x_p that makes the cumulative frequency F_p equal to 0.75

$$q_{0.75} = \sum_{x \le x_p} f_x = F_p = 0.75$$

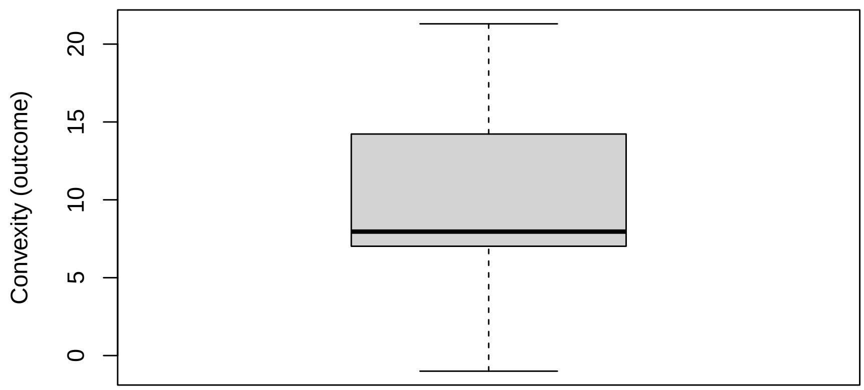
2.39 IQR

The distance between the third quartile and the first quartile is called the **interquartile range** (IQR) and captures the central 50% of the observations



2.40 Box plot

The interquartile range, the median, and the 5% and 95% of the data can be visualized in a **boxplot**, here the values of the outcomes are on the y-axis. The IQR is the box, the median is the line in the middle and the whiskers mark the 5% and 95% of the data.


Convexity (outcome)
0
5
10
15
20

Chapter 3

Probability

3.1 Objective

- Definition of probability
- Probability algebra
- Joint probability

3.2 Random experiments

Observation

- An **observation** is the acquisition of a number or a characteristic from an experiment

Outcome

- An **outcome** is a possible observation that is the result of an experiment.

Random experiment

- An experiment that gives **different** outcomes when repeated in the same manner.

3.3 Probability

The **probability** of an outcome is a measure of how sure we are to observe that outcome when performing a random experiment.

- 0: We are sure that the observation will **not** happen.
- 1: We are sure that the observation will happen.

3.4 Example

- Consider the following observations of a random experiment:

1 5 1 2 2 1 2 2

- How sure we are to obtain 2 in the following observation?

3.5 Example

The frequency table is

```
##   outcome ni    fi
## 1       1  3 0.375
## 2       2  4 0.500
## 3       5  1 0.125
```

The **relative frequency** f_i

- is a number between 0 and 1.
- measures the proportion of total observations that we observed a particular outcome.
- seems a reasonable probability measure.

As $f_2 = 0.5$ then we would be half certain to obtain a 2 in the next repetition of the experiment.

3.6 Relative frequency

As a measure of certainty is f_i enough?

Say we repeated the experiment 12 times more:

1 5 1 2 2 1 2 2 **3 1 1 3 3 1 6 3 5 6 4 4**

The frequency table is now

```
##   outcome ni  fi
## 1       1  6 0.3
## 2       2  4 0.2
## 3       3  4 0.2
## 4       4  2 0.1
## 5       5  2 0.1
## 6       6  2 0.1
```

New outcomes appeared and f_2 is now 0.2, we are now a fifth certain of obtaining 2 in the next experiment… probability should not depend on N

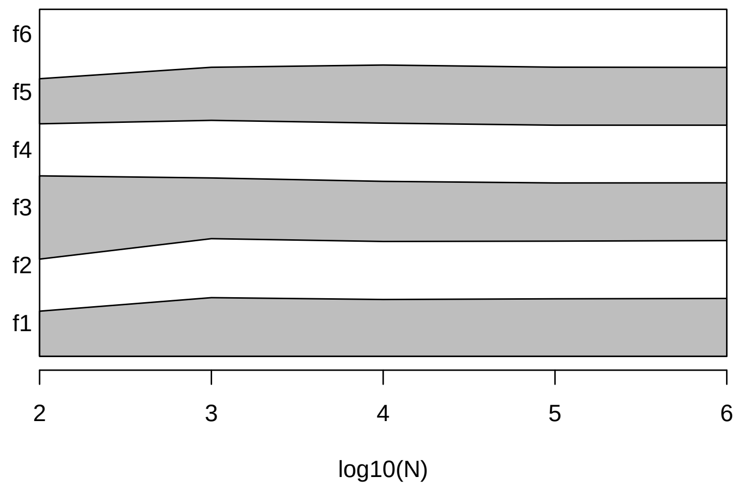
3.7 At infinity

Say we repeated the experiment 1000 times:

```
##   outcome  ni    fi
## 1       1 156 0.156
## 2       2 173 0.173
## 3       3 147 0.147
## 4       4 178 0.178
## 5       5 189 0.189
## 6       6 157 0.157
```

We find that f_i is converging to a constant value

$$lim_{N\to\infty} f_i = P_i$$



3.8 Frequentist probability

We call **Probability** P_i to the limit when $N \to \infty$ of the **relative frequency** of observing the outcome i in a random experiment.

Championed by Venn (1876)

The frequentist interpretation of probabilities is derived from data/experience (empirical).

- We do not observe P_i, we observe f_i
- When we **estimate** P_i with f_i (typically when N is large), we write:

$$\hat{P}_i = f_i$$

3.9 Classical Probability

Whenever a random experiment has M possible outcomes that are all **equally likely**, the probability of each outcome is $\frac{1}{M}$.

Championed by Laplace (1814).

Since each outcome is **equally probable** we declare complete ignorance and the best we can do is to fairly distribute the same probability to each outcome.

What if I told you that our experiment was the throw of the dice? then

$P_2 = 1/6 = 0.166666$.

$$P_i = lim_{N\to\infty}\frac{n_i}{N} = \frac{1}{M}$$

3.10 Classical and frequentist probabilities

3.11 Probability

Probability is a number between 0 and 1 that is assigned to each member E of a collection of **events** of a **sample space** (S) from a random experiment.

$$P(E) \in (0, 1)$$

where $E \in S$

3.12 Sample space

We start by reasoning what are all the possible values (outcomes) that a random experiment could give.

Note that we do not have to observe them in a particular experiment: We are using **reason/logic** and not observation.

Definition:

- The set of all possible outcomes of a random experiment is called the **sample space** of the experiment.
- The sample space is denoted as S.

3.13 Examples of sample spaces

- temperature 35 and 42 degrees Celcius
- sugar levels: 70-80mg/dL
- the size of one screw from a production line: 70mm-72mm
- number of emails received in an hour: 0-100
- a dice throw: 1, 2, 3, 4, 5, 6

3.14 Discrete and continuous sample spaces

- A sample space is discrete if it consists of a finite or countable infinite set of outcomes.
- A sample space is continuous if it contains an interval (either finite or infinite in length) of real numbers.

3.15 Event

Definition:

An **event** is a **subset** of the sample space of a random experiment. It is a **collection** of outcomes.

Examples of events:

- The event of a healthy temperature: temperature 37-38 degrees Celsius
- The event of producing a screw with a size: of 71.5mm
- The event of receiving more than 4 emails in an hour.
- The event of obtaining a number less than 3 in the throw of a dice

One event refers to a possible set of **outcomes**.

3.16 Event operations

For two events A and B, we can construct the following derived events:

- Complement A': the event of **not** A
- Union $A \cup B$: the event of A **or** B
- Intersection $A \cap B$: the event of A **and** B

3.17 Event operations example

Take

- Event $A : \{1, 2, 3\}$ a number less or equal to three in the throw of a dice
- Event $B : \{2, 4, 6\}$ an even number in the throw of a dice

New events:

- Not less than three: $A' : \{4, 5, 6\}$
- Less or equal to three **or** even: $A \cup B : \{1, 2, 3, 4, 6\}$
- Less or equal to three **and** even $A \cap B : \{2\}$

3.18 Outcomes

Outcomes are events that are **mutually exclusive**

Definition:

Two events denoted as E_1 and E_2, such that

$$E_1 \cap E_2 = \emptyset$$

They cannot occur at the same time.

Example:

- The outcome of obtaining 1 **and** the outcome of obtaining 5 in the throw of one dice are mutually exclusive:
- The event of obtaining 1 and 5 is empty:

$$\{1\} \cap \{5\} = \emptyset$$

3.19 Probability definition

A probability is a number that is assigned to each possible event (E) of a sample space (S) of a random experiment that satisfies the following properties:

- $P(S) = 1$
- $0 \leq P(E) \leq 1$
- when $E_1 \cap E_2 = \emptyset$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

Proposed by Kolmogorov's (1933)

3.20 Probability properties

Kolmogorov says that we can build a probability table (likewise the relative frequency table)

outcome	Probability
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6
$P(1 \cup 2 \cup ... \cup 6)$	1

As $\{1, 2, 3, 4, 5, 6\}$ are mutually exclusive then

$$P(S) = P(1 \cup 2 \cup ... \cup 6) = P(1) + P(2) + ... + P(n) = 1$$

3.21 Addition Rule

When A and B are not mutually exclusive then:

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

Where $P(A)$ and $P(B)$ are called the **marginal probabilities**

3.22 Example Addition Rule

Take

- Event $A : \{1, 2, 3\}$ a number less or equal to three in the throw of a dice

- Event $B : \{2, 4, 6\}$ an even number in the throw of a dice

then:

- $P(A) : P(1) + P(2) + P(3) = 3/6$
- $P(B) : P(2) + P(4) + P(6) = 3/6$
- $P(A \cap B) : P(2) = 1/6$

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 3/6 + 3/6 - 1/6 = 5/6$

Note: $P(2)$ appears in $P(A)$ and $P(B)$ that's why we subtract it with the intersection

---

3.23 Venn diagram

Note that can always break down the sample space in **mutually exclusive** sets involving the intersections:

$S = \{A \cap B, A \cap B', A' \cap B, A' \cap B'\}$

Marginals:

- $P(A) = P(A \cap B') + P(A \cap B) = 2/6 + 1/6 = 3/6$
- $P(B) = P(A' \cap B) + P(A \cap B) = 2/6 + 1/6 = 3/6$

3.24 Probability table

Let's look at the probability table

outcome	Probability
$A \cap B$	$P(A \cap B)$
$A \cap B'$	$P(A \cap B')$
$A' \cap B$	$P(A' \cap B)$
$A' \cap B'$	$P(A' \cap B')$
sum	1

3.25 Example probability table

We also write $A \cap B$ as (A, B) and call it the **joint probability** of A and B

In our example:

outcome	Probability
(A, B)	$P(A, B) = 1/6$
(A, B')	$P(A, B') = 2/6$
(A', B)	$P(A', B) = 2/6$
(A', B')	$P(A', B') = 1/6$
sum	1

Note: each outcome has *two* values (one for the characteristic of type A and another for type B)

3.26 Contingency table

We can organize the probability of **joint outcomes** in a **contingency table**

	B	B'	sum
A	$P(A, B)$	$P(A, B')$	$P(A)$
A'	$P(A', B)$	$P(A', B')$	$P(A')$
sum	$P(B)$	$P(B')$	1

Marginals:

- $P(A) = P(A, B') + P(A, B)$
- $P(B) = P(A', B) + P(A, B)$

3.27 Example contingency table

- Event $A : \{1, 2, 3\}$ a number less or equal to three in the throw of a dice

- Event $B : \{2, 4, 6\}$ an even number in the throw of a dice

	B	B'	sum
A	1/6	2/6	3/6
A'	2/6	1/6	3/6
sum	3/6	3/6	1

Three forms of the **addition rule**:

$P(A \cup B)$

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

$$= P(A \cap B) + P(A \cap B') + P(A' \cap B)$$

$$= 1 - P(A' \cap B')$$

3.28 Misophonia study

In the misophonia study, the patients were assessed for their misophonia severity **and** if they were depressed.

The outcome of one random experiment is to measure the misophonia severity **and** depression status of one patient. The repetition of the random experiment was to perform the same two measurements on another patient.

```
##     Misofonia.dic depresion.dic
## 1               4             1
## 2               2             0
## 3               0             0
## 4               3             0
## 5               0             0
```

```
## 6                  0              0
## 7                  2              0
## 8                  3              0
## 9                  0              1
## 10                 3              0
## 11                 0              0
## 12                 2              0
## 13                 2              1
## 14                 0              0
## 15                 2              0
## 16                 0              0
## 17                 0              0
## 18                 3              0
## 19                 3              0
## 20                 0              0
## 21                 3              0
## 22                 3              0
## 23                 2              0
## 24                 0              0
## 25                 0              0
## 26                 0              0
## 27                 4              1
## 28                 2              0
## 29                 2              0
## 30                 0              0
## 31                 2              0
## 32                 0              0
## 33                 0              0
## 34                 0              0
## 35                 3              0
## 36                 0              0
## 37                 2              0
## 38                 3              1
## 39                 2              0
## 40                 2              0
## 41                 0              0
## 42                 2              0
## 43                 3              0
## 44                 0              0
## 45                 0              0
## 46                 2              0
## 47                 2              0
## 48                 3              0
## 49                 3              0
## 50                 0              0
## 51                 0              0
```

```
## 52                    4             1
## 53                    3             0
## 54                    3             1
## 55                    2             1
## 56                    0             1
## 57                    2             0
## 58                    0             0
## 59                    0             0
## 60                    0             0
## 61                    2             0
## 62                    2             0
## 63                    0             0
## 64                    0             0
## 65                    2             0
## 66                    3             1
## 67                    0             0
## 68                    1             0
## 69                    3             0
## 70                    2             0
## 71                    4             1
## 72                    3             0
## 73                    2             1
## 74                    3             0
## 75                    0             1
## 76                    2             0
## 77                    3             0
## 78                    2             0
## 79                    4             1
## 80                    1             0
## 81                    2             0
## 82                    0             0
## 83                    2             0
## 84                    0             0
## 85                    2             0
## 86                    0             1
## 87                    2             0
## 88                    2             0
## 89                    4             1
## 90                    3             0
## 91                    0             1
## 92                    3             0
## 93                    0             0
## 94                    0             0
## 95                    0             0
## 96                    2             0
## 97                    2             0
```

```
## 98              1              0
## 99              3              0
## 100             0              0
## 101             0              0
## 102             3              1
## 103             2              0
## 104             1              0
## 105             3              0
## 106             0              0
## 107             4              1
## 108             4              1
## 109             2              0
## 110             3              0
## 111             3              0
## 112             3              1
## 113             0              0
## 114             3              0
## 115             2              0
## 116             1              0
## 117             2              0
## 118             3              1
## 119             3              0
## 120             4              1
## 121             2              0
## 122             3              0
## 123             2              0
```

3.29 Contingency table for frequencies

- For the number of observations $n_{i,j}$ of each outcome (x_i, y_i), misophonia: $x \in \{0, 1, 2, 3, 4\}$ and depression $y \in \{0, 1\}$ (no:0, yes:1)

```
##
##                Depression:0 Depression:1
##   Misophonia:4            0            9
##   Misophonia:3           25            6
##   Misophonia:2           34            3
##   Misophonia:1            5            0
##   Misophonia:0           36            5
```

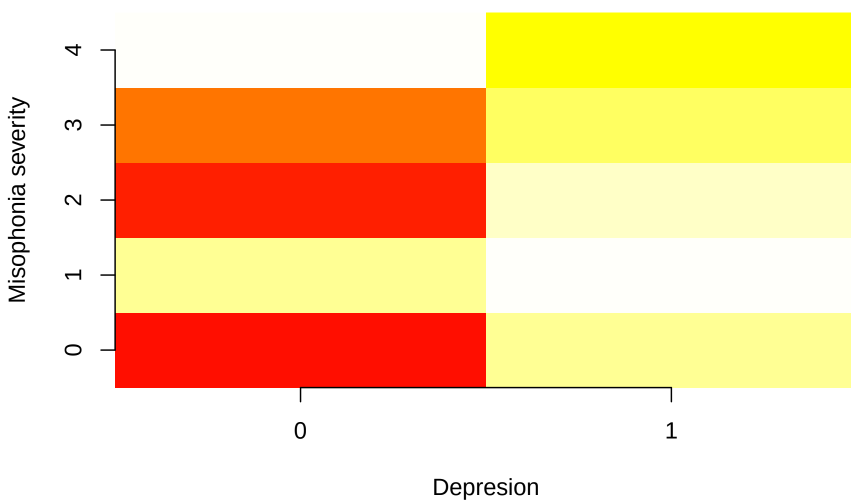
- For the relative frequencies $f_{i,j}$

```
##
```

```
##                Depression:0 Depression:1
##    Misophonia:4   0.00000000   0.07317073
##    Misophonia:3   0.20325203   0.04878049
##    Misophonia:2   0.27642276   0.02439024
##    Misophonia:1   0.04065041   0.00000000
##    Misophonia:0   0.29268293   0.04065041
```

3.30 Heat map

The contingency table can be plotted as a **heat map**



3.31 Continous variables

In the misophonia study, the jaw protrusion was also measured as a possible cephalometric factor for de disease.

```
##     Angulo_convexidad protusion.mandibular
## 1                7.97                13.00
```

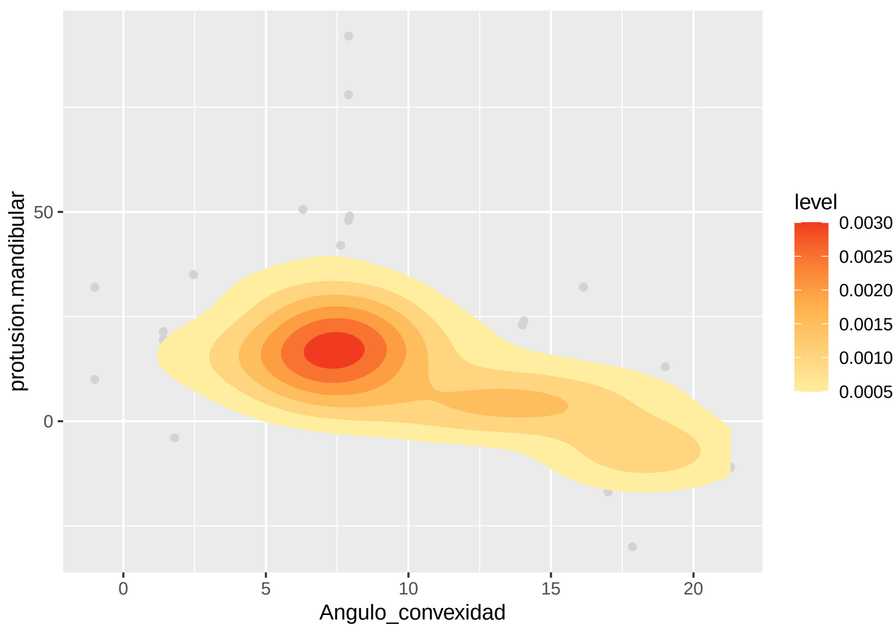
```
## 2                18.23              -5.00
## 3                12.27              11.50
## 4                 7.81              16.80
## 5                 9.81              33.00
## 6                13.50               2.00
## 7                19.30              -3.90
## 8                 7.70              16.80
## 9                12.30               8.00
## 10                7.90              28.80
## 11               12.60               3.00
## 12               19.00              -7.90
## 13                7.27              28.30
## 14               14.00               4.00
## 15                5.40              22.20
## 16                8.00               0.00
## 17               11.20              15.00
## 18                7.75              17.00
## 19                7.94              49.00
## 20               16.69               5.00
## 21                7.62              42.00
## 22                7.02              28.00
## 23                7.00               9.40
## 24               19.20             -13.20
## 25                7.96              23.00
## 26               14.70               2.30
## 27                7.24              25.00
## 28                7.80               4.90
## 29                7.90              92.00
## 30                4.70               6.00
## 31                4.40              17.00
## 32               14.00               3.30
## 33               14.40              10.30
## 34               16.00               6.30
## 35                1.40              19.50
## 36                9.76              22.00
## 37                7.90               5.00
## 38                7.90              78.00
## 39                7.40               9.30
## 40                6.30              50.60
## 41                7.76              18.00
## 42                7.30              18.00
## 43                7.00              10.00
## 44               11.23               4.00
## 45               16.00              13.30
## 46                7.90              48.00
## 47                7.29              23.50
```

```
## 48                6.91                37.60
## 49                7.10                15.00
## 50               13.40                 5.10
## 51               11.60                -2.20
## 52               -1.00                32.00
## 53                6.00                25.00
## 54                7.82                24.00
## 55                4.80                33.60
## 56               11.00                 3.30
## 57                9.00                31.50
## 58               11.50                12.80
## 59               16.00                 3.00
## 60               15.00                 6.00
## 61                1.40                21.40
## 62               16.80               -10.00
## 63                7.70                19.00
## 64               16.14                32.00
## 65                7.12                15.00
## 66               -1.00                10.00
## 67               17.00               -16.90
## 68                9.26                 2.00
## 69               18.70               -10.10
## 70                3.40                12.20
## 71               21.30               -11.00
## 72                7.50                 5.20
## 73                6.03                16.00
## 74                7.50                 5.80
## 75               19.00                 5.20
## 76               19.01                13.00
## 77                8.10                13.60
## 78                7.80                16.10
## 79                6.10                33.20
## 80               15.26                 4.00
## 81                7.95                12.00
## 82               18.00                -1.50
## 83                4.60                18.30
## 84               15.00                 3.00
## 85                7.50                15.80
## 86                8.00                27.10
## 87               16.80               -10.00
## 88                8.54                25.00
## 89                7.00                27.10
## 90               18.30                -8.00
## 91                7.80                12.00
## 92               16.00                -8.00
## 93               14.00                23.00
```

```
## 94              12.30               5.00
## 95              11.40               1.00
## 96               8.50              18.90
## 97               7.00              15.00
## 98               7.96              22.00
## 99              17.60              -3.50
## 100             10.00              20.00
## 101              3.50              12.20
## 102              6.70              14.70
## 103             17.00              -5.00
## 104             20.26              -4.15
## 105              6.64              11.00
## 106              1.80              -4.00
## 107              7.02              25.00
## 108              2.46              35.00
## 109             19.00              -5.00
## 110             17.86             -30.00
## 111              6.10              12.20
## 112              6.64              19.00
## 113             12.00               1.60
## 114              6.60              20.00
## 115              8.70              17.10
## 116             14.05              24.00
## 117              7.20               7.10
## 118             19.70             -11.00
## 119              7.70              21.30
## 120              6.02               5.00
## 121              2.50              12.90
## 122             19.00               5.90
## 123              6.80               5.80
```

3.32 Heat map for continuous variables

- Two dimensional **histogram**.
- It illustrates the "continuous contingency" table for continuous variables

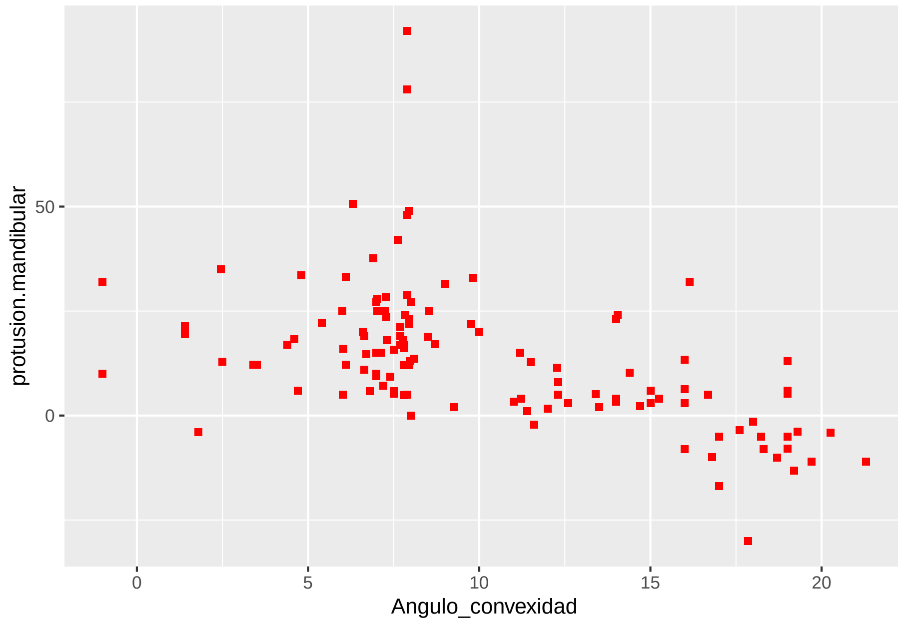


3.33 Scatter plot

- The **histogram** depends on the size of the bin (pixel).

- If the pixel is small enough to contain a single observation then the heat map results in a **scatter plot**

The scatter plot is the illustration of a "contingency table" for continuous variables when the bin (pixel) is small enough to contain one single observation (consisting of a pair of values).



protusion.mandibular
50
0
0
5
10
15
20
Angulo_convexidad

Chapter 4

Conditional Probability

4.1 Objective

- Conditional probability
- Independence
- Bayes' theorem

4.2 Joint Probability

The joint probability of two events A and B is

$$P(A, B) = P(A \cap B)$$

Let's imagine a random experiment that measures two different types of outcomes.

- height and weight of an individual: (h, w)
- time and place of an electric charge: (p, t)
- a throw of two dice: (n_1,n_2)
- cross two traffic lights in green: $(\bar{R_1}, \bar{R_2})$

In many cases, we are interested in finding out whether the values of one outcome **condition** the values of the other.

4.3 Diagnostics

Let's consider a **diagnostic tool**

We want to find the state of a system (s):

- inadequate (yes)
- adequate (no)

with a test (t):

- positive
- negative

We test a battery to find how long it can live. We stress a cable to find if it resists carrying a certain load. We perform a PCR to see if someone is infected.

4.4 Diagnostics Test

Let's consider diagnosing infection with a new test.

Infection status:

- yes (infected)
- no (not infected)

Test:

- positive
- negative

4.5 Observations

Each individual is a random experiment with two measurements: (Infection, Test)

Subject	Infection	Test
s_1	yes	positive
s_2	no	negative
s_3	yes	positive
...	...	...
s_i	no	positive*

Subject	Infection	Test
...	...	...
...	...	...
s_n	yes	negative*

4.6 Contingency tables

- For the number of observations of each outcome

	Infection: yes	Infection: no	sum
Test: positive	18	12	30
Test: negative	30	300	330
sum	48	312	360

- For the relative frequencies, if $N >> 0$ we will take $f_{i,j} = \hat{P}(x_i, y_j)$

	Infection: yes	Infection: no	sum
Test: positive	0.05	0.0333	0.0833
Test: negative	0.0833	0.833	0.9166
sum	0.133	0.866	1

4.7 Conditional probability

Let's think first in terms of those who are **infected**

Within those who are infected (**yes**), what is the probability of those who tested positive?

- Sensitivity (true positive rate)

$$\hat{P}(positive|yes) = \frac{n_{positive,yes}}{n_{yes}}$$

$$= \frac{\frac{n_{positive,yes}}{N}}{\frac{n_{yes}}{N}} = \frac{f_{positive,yes}}{f_{yes}}$$

Therefore, in the limit, we expect to have a probability of the type

$$P(positive|yes) = \frac{P(positive, yes)}{P(yes)} = \frac{P(positive \cap yes)}{P(yes)}$$

4.8 Conditional probability

Definition: The conditional probability of an event B given an event A, denoted as $P(A|B)$, is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- you can prove that the conditional probability satisfies the axioms of probability.
- it is the probability with the sampling space given by B: S_B.

4.9 Conditional contingency table

	Infection: Yes	Infection: No
Test: positive	P(positive | yes)	P(positive | no)
Test: negative	P(negative | yes)	P(negative | no)
sum	1	1

- True positive rate (Sensitivity): The probability of testing positive **if** having the disease $P(positive|yes)$
- True negative rate (Specificity): The probability of testing negative **if** not having the disease $P(negative|no)$
- False-positive rate: The probability of testing positive **if** not having the disease $P(positive|no)$
- False-negative rate: The probability of testing negative **if** having the disease $P(negative|yes)$

4.10 Example conditional contingency table

Taking the frequencies as estimates of the probabilities then

	Infection: Yes	Infection: No
Test: positive	$18/48 = 0.375$	$12/312 = 0.038$
Test: negative	$30/48 = 0.625$	$300/312 = 0.962$
sum	1	1

Our diagnostic tool has low sensitivity (0.375) but high specificity (0.962).

4.11 Multiplication rule

Now let's imagine the real situation where we want to compute **joint** probabilities from conditional **probabilities**

- PCRs for coronavirus were (performed)[https://www.nejm.org/doi/full/10.1056/NEJMp2015897] in people in the hospital who we are sure to be infected. They have a sensitivity of 70%. They have also been tested in the lab in conditions of no infection with 96% specificity
- A prevalence study in Spain showed that $P(yes) = 0.05$, $P(no) = 0.95$ before summer.

With this data, what was the probability that a randomly selected person in the population tested positive **and** was infected: $P(yes \cap positive) = P(yes, positive)$?

4.12 Diagnostic performance

To study the performance of a new diagnostic test:

- you select specimens that are inadequate (disease: **yes**) and apply the test, trying to find its sensitivity: $P(positive|yes)$ (0.70 for PCRs)
- you select specimens that are adequate (disease: **no**) and apply the test, trying to find its specificity: $P(negative|no)$ (0.96 for PCRs)

	Infection: Yes	Infection: No
Test: positive	P(positive|yes)=0.7	P(positive|no)=0.06
Test: negative	P(negative|yes)=0.3	P(negative|no)=0.94
sum	1	1

From this matrix, can we obtain $P(yes, positive)$?

4.13 Multiplication rule

How do you recover the joint probability from the conditional probability?

For two events A and B we have the multiplication rule

$$P(A, B) = P(A|B)P(B)$$

that follows from the definition of the conditional probability.

4.14 Contingency table in terms of conditional probabilities

	Infection: Yes	Infection: No	sum
Test: positive	P(positive | yes)P(yes)	P(positive | no)P(no)	P(positive)
Test: negative	P(negative | yes)P(yes)	P(negative | no) P(no)	P(negative)
sum	P(yes)	P(no)	1

For instance the probability of testing *positive* and being infected *yes*:

- $P(positive, yes) = P(positive \cap yes) = P(positive|yes)P(yes)$

4.15 Conditional tree

4.16 Contingency table in terms of conditional probabilities

	Infection: yes	Infection: no	sum
Test: positive	0.035	0.057	0.092
Test: negative	0.015	0.893	0.908
sum	0.05	0.95	1

- $P(positive, yes) = 0.035$

But we also found the marginal of being positive:

- $P(positive) = 0.092$

4.17 Total probability rule

	Infection: Yes	Infection: No	sum
Test: positive	P(positive | yes)P(yes)	P(positive | no)P(no)	P(positive)
Test: negative	P(negative | yes)P(yes)	P(negative | no) P(no)	P(negative)
sum	P(yes)	P(no)	1

When we write the unknown marginals in terms of their conditional probabilities we call it the **total probability rule**

- $P(positive) = P(positive|yes)P(yes) + P(positive|no)P(no)$
- $P(negative) = P(negative|yes)P(yes) + P(negative|no)P(no)$

4.18 Conditional tree

Total probability rule for the marginal of B: In how many ways I can obtain the outcome B?

$P(B) = P(B|A)P(A) + P(B|A')P(A')$

4.19 Finding reverse probabilities

From the conditional contingency table

	Infection: Yes	Infection: No
Test: positive	P(positive | yes)	P(positive | no)
Test: negative	P(negative | yes)	P(negative | no)
sum	1	1

How can we calculate the probability of being infected if tested positive: $P(yes|positive)$?

4.20 Recover joint probabilities

1. We recover the contingency table for joint probabilities

	Infection: Yes	Infection: No	sum
Test: positive	P(positive | yes)P(yes)	P(positive | no)P(no)	P(positive)
Test: negative	P(negative | yes)P(yes)	P(negative | no) P(no)	P(negative)
sum	P(yes)	P(no)	1

4.21 Reverse conditionals

2. We compute the conditional probabilities for the test:

$$P(infection|test) = \frac{P(test|infection)P(infection)}{P(test)}$$

	Infection: Yes	Infection: No	sum
Test: positive	P(yes|positive)	P(no|positive)	1
Test: negative	P(yes|negative)	P(no|negative)	1

For instance:

$$P(yes|positive) = \frac{P(positive|yes)P(yes)}{P(positive)}$$

since we usually don't have $P(positive)$ we use the **total probability** rule in the denominator

$$P(yes|positive) = \frac{P(positive|yes)P(yes)}{P(positive|yes)P(yes) + P(positive|no)P(no)}$$

4.22 Baye's theorem

The expression:

$$P(yes|positive) = \frac{P(positive|yes)P(yes)}{P(positive|yes)P(yes) + P(positive|no)P(no)}$$

is called the **Bayes theorem**

Theorem

If $E1, E2, ..., Ek$ are k mutually exclusive and exhaustive events and B is any event,

$$P(Ei|B) = \frac{P(B|Ei)P(Ei)}{P(B|E1)P(E1) + ... + P(B|Ek)P(Ek)}$$

It allows to reverse the conditionals:

$$P(B|A) \rightarrow P(A|B)$$

Or **design** a test B in controlled condition A and then use it to **infer** the probability of the condition when the test is positive.

4.23 Example: Bayes' theorem

Baye's theorem:

$$P(yes|positive) = \frac{P(positive|yes)P(yes)}{P(possitive|yes)P(yes) + P(positive|no)P(no)}$$

we know:

- $P(positive|yes) = 0.70$
- $P(positive|no) = 1 - P(negative|no) = 0.06$
- the probability of infection and not infection in the population: $P(yes) = 0.05$ and $P(no) = 1 - P(yes) = 0.95$.

Therefore:

$$P(yes|positive) = 0.47$$

Tests are not so good to **confirm** infections.

4.24 Example: Bayes' theorem

Let's now apply it to the probability of not being infected if the test is negative

$$P(no|negative) = \frac{P(negative|no)P(no)}{P(negative|no)P(no) + P(negative|yes)P(yes)}$$

Substitution of all the values gives

$$P(no|negative) = 0.98$$

Tests are good to **rule out** infections.

4.25 Statistical independence

In many applications, we want to know if the knowledge of one event conditions the outcome of another event.

- there are cases where we want to know if the events are not conditioned

4.26 Statistical independence

Consider conductors for which we measure their surface flaws and if their conduction capacity is defective

The estimated **joint probabilities** are

	flaws (F)	no flaws (F')	sum
defective (D)	0.005	0.045	0.05
no defective (D')	0.095	0.855	0.95
sum	0.1	0.9	1

where, for instance, the joint probability of F and D is

- $P(D, F) = 0.005$

The marginal probabilities are

- $P(D) = P(D, F) + P(D, F') = 0.05$
- $P(F) = P(D, F) + P(D', F) = 0.1.$

4.27 Statistical independence

What is the **conditional probability** of observing a defective conductor if they have a flaw?

	F	F'
D	P(D|F) = 0.05	P(D|F')=0.05
D'	P(D'|F)=0.95	P(D'|F')=0.95
sum	1	1

The marginals and the conditional probabilities are the same!

- $P(D|F) = P(D|F') = P(D)$
- $P(D'|F) = P(D'|F') = P(D')$

The probability of observing a defective conductor **does not** depend on having observed or not a flaw.

$$P(D) = P(D|F)$$

4.28 Statistical independence

Two events A and B are statistically independent if

- $P(A|B) = P(A)$; A is independent of B
- $P(B|A) = P(B)$; B is independent of A

and by the multiplication rule, their joint probability is

- $P(A \cap B) = P(A|B)P(B) = P(A)P(B)$

the multiplication of their marginal probabilities.

4.29 Products of marginals products

	F	F'	sum
D	0.005	0.045	0.05
D'	0.095	0.855	0.95
sum	0.1	0.9	1

Confirm that all the entries of the matrix are the product of the marginals.

For example:

- $P(F)P(D) = P(D \cap F)$
- $P(D')P(F') = P(D' \cap F')$

4.30 Example

Outcomes of throwing two coins: $S = (H, H), (H, T), (T, H), (T, T)$

	H	T	sum
H	1/4	1/4	1/2
T	1/4	1/4	1/2
sum	1/2	1/2	1

- Obtaining a head in the first coin does not condition obtaining a tail in the result of the second coin $P(T|H) = P(T) = 1/2$
- the probability of obtaining a head and then a tail is the product of each independent outcome $P(H, T) = P(H) * P(T) = 1/4$

Chapter 5

Discrete Random Variables

5.1 Objective

- Random variables
- Probability mass function
- Mean and variance
- Probability distribution

5.2 How do we assign probability values to outcomes?

5.3 Random variable

Definition:

A **random variable** is a function that assigns a real **number** to each **outcome** in the sample space of a random experiment.

- Most commonly a random variable is the value of the **measurement** of interest that is made in a random experiment.

A random variable can be:

- Discrete (nominal, ordinal)

- Continuous (interval, ratio)

5.4 Random variable

A **value** (or **outcome**) of a random variable is one of the possible numbers that the variable can take in a random experiment.

We write the random variable in **capitals**.

Example:

If $X \in \{0, 1\}$, we then say X is a random variable that can take the values 0 or 1.

Observation of a random variable

- An observation is the **acquisition** of the value of a random variable in a random experiment

Example:

1 0 0 1 0 **1** 0 1 1

The number in bold is an observation of X

5.5 Events of observing a random variable

- $X = 1$ is the **event** of observing the random variable X with value 1
- $X = 2$ is the **event** of observing the random variable X with value 2

...

In general:

- $X = x$ is the **event** of observing the random variable X with value x (little x)
- Any two values of a random variable define two **mutually exclusive** events.

5.6 Probability of random variables

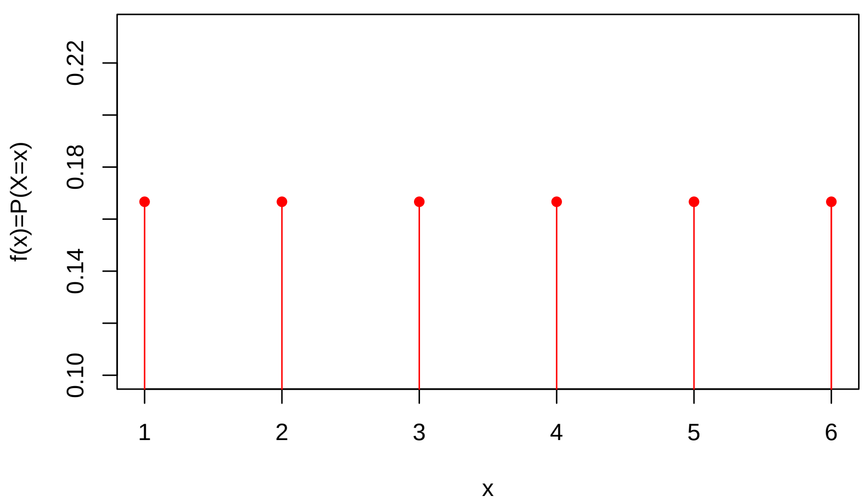
We are interested in assigning probabilities to the values of a random variable.

We have already done this for the dice: $X \in \{1, 2, 3, 4, 5, 6\}$ (classical interpretation of pribability)

X	Probability
1	$P(X = 1) = 1/6$
2	$P(X = 2) = 1/6$
3	$P(X = 3) = 1/6$
4	$P(X = 4) = 1/6$
5	$P(X = 5) = 1/6$
6	$P(X = 6) = 1/6$

5.7 Probability functions

- We can write the probability table
- plot it

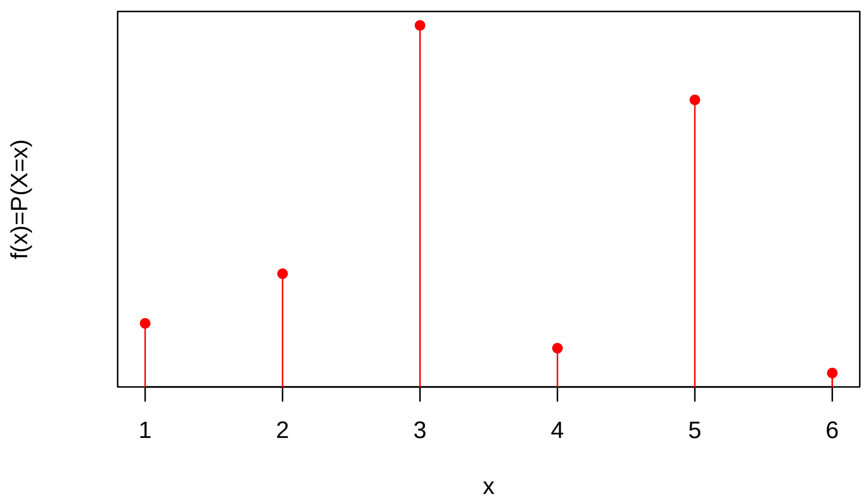


- or write as the function

$$f(x) = P(X = x) = 1/6$$

5.8 Probability functions

We can **create** any type of probability function if we respect the probability rules:



5.9 Probability functions

For a discrete random variable $X \in \{x_1, x_2, .., x_M\}$, a **probability mass function**

is always positive

- $f(x_i) \geq 0$

is used to compute probabilities

- $f(x_i) = P(X = x_i)$

and its sum over all the values of the variable is 1:

- $\sum_{i=1}^{M} f(x_i) = 1$

5.10 Probability functions

- Note that the definition of X and its probability mass function is general **without reference** to any experiment. The functions live in the model (abstract) space.
- X and $f(x)$ are abstract objects that may or may not map to an experiment
- We have the freedom to construct them as we want as long as we respect their definition.
- They have some **properties** that are derived exclusively from their definition.

5.11 Example: Probability mass function

Consider the following random variable X over the outcomes

outcome	X
a	0
b	0
c	1.5
d	1.5
e	2
f	3

If each outcome is equally probable then what is the probability mass function of x?

5.12 Probability table for equally likely outcomes

outcome	Probability(outcome)
a	$1/6$
b	$1/6$
c	$1/6$
d	$1/6$
e	$1/6$
f	$1/6$

5.13 Probability table for X

X	$f(x) = P(X = x)$
0	$P(X = 0) = 2/6$
1.5	$P(X = 1.5) = 2/6$
2	$P(X = 2) = 1/3$
3	$P(X = 3) = 1/3$

We can compute, for instance, the following probabilities for events on the values of X

- $P(X > 3)$
- $P(X = 0 \cup X = 2)$
- $P(X \leq 2)$

5.14 Example

Probability model:

Consider the following experiment: In one urn put 8 balls and:

- mark 1 ball with -2
- mark 2 balls with -1
- mark 2 balls with 0

- mark 2 balls with 1
- mark 1 ball with 2

experiment: Take one ball and read the number.

X	$P(X = x)$
-2	$1/8 = 0.125$
-1	$2/8 = 0.25$
0	$2/8 = 0.25$
1	$2/8 = 0.25$
2	$1/8 = 0.125$

5.15 Example

Consider another experiment where we do not know what is in the previous urn. We draw a ball 30 times, write its nuber and put it back in the urn.

- we do not know what the primary events with equal probabilities are.
- we then **estimate** the probability mass function from the relative frequencies observed for a random variable

X	f_i
-2	0.132
-1	0.262
0	0.240
1	0.248
2	0.118

5.16 Probabilities and frequencies

For computing the relative frequencies f_i you have to

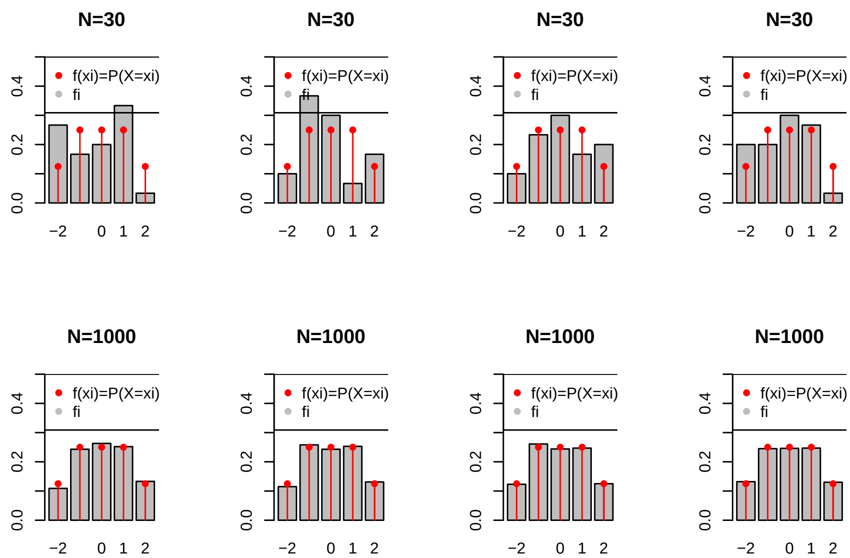
- **repeat** the experiment N times (you have to put the ball back in the urn each time) and at the end compute

$$f_i = n_i/N$$

We are assuming that:

$$lim_{N \to \infty} f_i = f(x_i) = P(X = x_i)$$

5.17 Probabilities and relative frequencies



- In this example we **know** the probability **model** $f(x) = P(X = x)$ by design.
- We never observe $f(x)$
- We can use relative frequencies to estimate the probabilities

$$f_i = \hat{f}(x_i) = \hat{P}(X = x_i)$$

(f_i depends on N)

5.18 Mean and Variance

The probability mass functions $f(x)$ have two main properties

- its center
- its spread

We can ask,

- around which values of X the probability concentrated?
- How dispersed are the values of X in relation to their probabilities?

5.19 Mean and Variance

5.20 Mean

Remember that the **average** in terms of the relative frequencies of the values of x_i (categorical ordered outcomes) can be written as

$$\bar{x} = \sum_{i=1}^{M} x_i \frac{n_i}{N} = \sum_{i=1}^{M} x_i f_i$$

Definition

The **mean** (μ) or expected value of a discrete random variable X, $E(X)$, with mass function $f(x)$ is given by

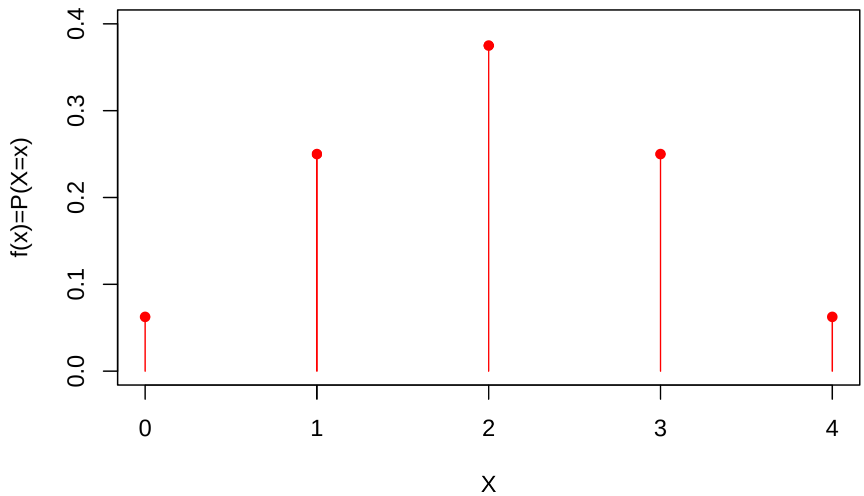
$$\mu = E(X) = \sum_{i=1}^{M} x_i f(x_i)$$

It is the center of gravity of the **probabilities**: The point where probability loadings on a road are balanced

5.21 Example: Mean

What is the mean of X if its probability mass function $f(x)$ is given by

$P(X = 0) = 1/16$ $P(X = 1) = 4/16$ $P(X = 2) = 6/16$ $P(X = 3) = 4/16$ $P(X = 4) = 1/16$



$$\mu = E(X) = \sum_{i=1}^{m} x_i f(x_i)$$

$E(X)$ =**0** * 1/16 + **1** * 4/16 + **2** * 6/16 + **3** * 4/16 + **4** * 1/16 =2

5.22 Variance

In similar terms we define the mean squared distance from the mean:

Definition

The variance, written as σ^2 or $V(X)$, of a discrete random variable X with mass function $f(x)$ is given by

$$\sigma^2 = V(X) = \sum_{i=1}^{M} (x_i - \mu)^2 f(x_i)$$

- $\sigma = \sqrt{V(X)}$ is called the **standard deviation** of the random variable
- Think of it as the moment of inertia of probabilities about the mean.

5.23 Example: Variance

What is the variance of X if its probability mass function $f(x)$ is given by

$P(X = 0) = 1/16$ $P(X = 1) = 4/16$ $P(X = 2) = 6/16$ $P(X = 3) = 4/16$ $P(X = 4) = 1/16$

$$\sigma^2 = V(X) = \sum_{i=1}^{m} (x_i - \mu)^2 f(x_i)$$

$V(X)$ =**(0-2)**^{2}* 1/16 + **(1-2)**^{2}* 4/16 + **(2-2)**^{2}* 6/16 + **(3-2)**^{2}* 4/16 + **(4-2)**^{2}* 1/16 =1

$$V(X) = \sigma^2 = 1$$

$$\sigma = 1$$

5.24 Functions of X

Definition

For any function h of a random variable X, with mass function $f(x)$, its expected value is given by

$$E[h(X)] = \sum_{i=1}^{M} h(x_i) f(x_i)$$

This is an important definition that allows us to prove three important properties of the median and variance:

- The mean of a linear function is the linear function fo the mean:

  $$E(a \times X + b) = a \times E(X) + b$$

  for a and b scalars (numbers).

- The variance of a linear function of X is:

  $$V(a \times X + b) = a^2 \times V(X)$$

- The variance **about the origin** is the variance **about the mean** plus the mean squared:
$$E(X^2) = V(X) + E(X)^2$$

5.25 Example: Variance about the origin

What is the variance X about the origin, $E(X^2)$, if its probability mass function $f(x)$ is given by

$P(X = 0) = 1/16$ $P(X = 1) = 4/16$ $P(X = 2) = 6/16$ $P(X = 3) = 4/16$ $P(X = 4) = 1/16$

$$E(X^2) = \sum_{i=1}^{m} x_i^2 f(x_i)$$

$E(X^2) =$**(0)**^{2}* 1/16 + **(1)**^{2}* 4/16 + **(2)**^{2}* 6/16 + **(3)**^{2}* 4/16 + **(4)**^{2}* 1/16 =5

We can also verify:

$$E(X^2) = V(X) + E(X)^2$$

$5 = 1 + 2^2$

5.26 Probability distribution

Definition:

The **probability distribution** function is defined as

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

That is the accumulated probability up to a given value x

$F(x)$ satisfies:

- $0 \leq F(x) \leq 1$
- If $x \leq y$, then $F(x) \leq F(y)$

5.27 Example: Probability distribution

For the probability mass function:

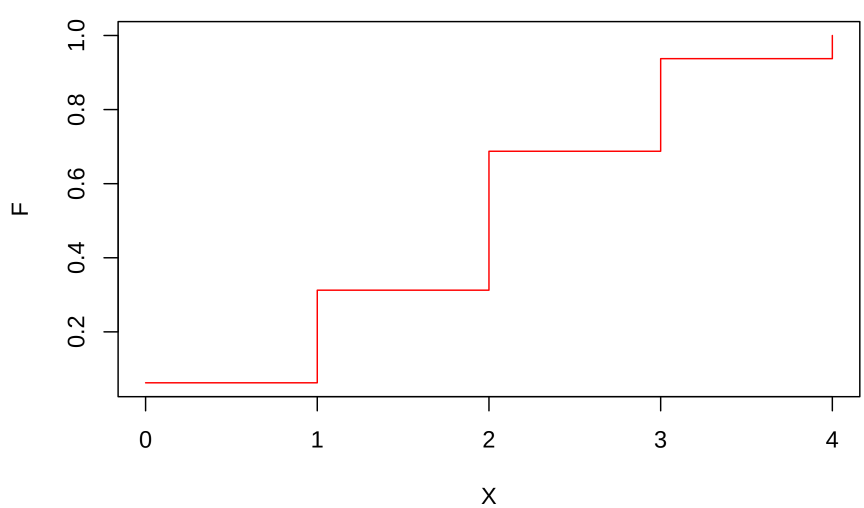
$f(0) = P(X = 0) = 1/16$ $f(1) = P(X = 1) = 4/16$ $f(2) = P(X = 2) = 6/16$ $f(3) = P(X = 3) = 4/16$ $f(4) = P(X = 4) = 1/16$

The probability distribution is:

$$F(x) = \begin{cases} 1/16, & \text{if } x < 1 \\ 5/16, & 1 \leq x < 2 \\ 11/16, & 2 \leq x < 3 \\ 15/16, & 3 \leq x < 4 \\ 16/16, & x \leq 5 \end{cases}$$

For$X \in \mathbb{Z}$

5.28 Probability distribution



5.29 Probability function and Probability distribution

Compute the mass probability function of the following probability distribution:

$F(0) = 1/16$, $F(1) = 5/16$, $F(2) = 11/16$, $F(3) = 15/16$, $F(4) = 16/16$,

Let's work backward.

$f(0) = F(0) = 1/16$ $f(1) = F(1) - f(0) = 5/32 - 1/32 = 4/16$ $f(2) = F(2) - f(1) - f(0) = F(2) - F(1) = 6/16$ $f(3) = F(3) - f(2) - f(1) - f(0) = F(3) - F(2) = 4/16$ $f(4) = F(4) - F(3) = 1/16$

5.30 Probability function and Probability distribution

The Probability distribution is another way to specify the probability of a random variable

$$f(x_i) = F(x_i) - F(x_{i-1})$$

with

$$f(x_1) = F(x_1)$$

for X taking values in $x_1 \leq x_2 \leq ... \leq x_n$

5.31 Quantiles

We define the **q-quantile** as the value x_p **under** which we have accumulated q*100% of the probability

$$q = \sum_{i=1}^{p} f(x_i) = F(x_p)$$

- The **median** is value x_m such that $q = 0.5$

$$F(x_m) = 0.5$$

- The 0.05-quantile is the value x_r such that $q = 0.05$

$$F(x_r) = 0.05$$

- The 0.95-quantile is the value x_s such that $q = 0.95$

$$F(x_s) = 0.95$$

5.32 Summary

quantity names	model (unobserved)	data (observed)
probability mass function // relative frequency	$f(x_i) = P(X = x_i)$	$f_i = \frac{n_i}{N}$
probability distribution // cumulative relative frequency	$F(x_i) = P(X \leq x_i)$	$F_i = \sum_{k\leq i} f_k$
mean // average	$\mu = E(X) = \sum_{i=1}^{M} x_i f(x_i)$	$\bar{x} = \sum_{j=1}^{N} x_j/N$
variance // sample variance	$\sigma^2 = V(X) = \sum_{i=1}^{M}(x_i - \mu)^2 f(x_i)$	$s^2 = \sum_{j=1}^{N}(x_j - \bar{x})^2/(N-1)$
standard deviation // sample sd	$\sigma = \sqrt{V(X)}$	s
variance about the origin // 2nd sample moment	$E(X^2) = \sum_{i=1}^{M} x_i^2 f(x_i)$	$m_2 = \sum_{j=1}^{N} x_j^2/n$

Note that:

- $i = 1...M$ is an **outcome** of the random variable X.
- $j = 1...N$ is an **observation** of the random variable X.

Properties:

- $\sum_{i=1...N} f(x_i) = 1$
- $f(x_i) = F(x_i) - F(x_{i-1})$
- $E(a \times X + b) = a \times E(X) + b$; for a and b scalars.
- $V(a \times X + b) = a^2 \times V(X)$
- $E(X^2) = V(X) + E(X)^2$

Chapter 6

Continous Random Variables

6.1 Objective

- Probability density function
- Mean and variance
- Probability distribution

6.2 Continuous random variable

What happens with continuous random variables?

Let's reconsider the convexity angle of misophonia patients (Section 2.21).

- We redefined the outcomes as little regular intervals (bins) and computed the relative frequency for each of them as we did in the discrete case.

```
##         outcome ni         fi
## 1 [-1.02,3.46]  8 0.06504065
## 2  (3.46,7.92] 51 0.41463415
## 3  (7.92,12.4] 26 0.21138211
## 4  (12.4,16.8] 20 0.16260163
## 5  (16.8,21.3] 18 0.14634146
```

6.3 Continuous random variable

Let's consider again that their relative frequencies are the probabilities when $N \to \infty$

$$f_i = \frac{n_i}{N} \to f(x_i) = P(X = x_i)$$

The probability depends now on the length of the bins Δx. If we make the bins smaller and smaller then the frequencies get smaller and therefore

$P(X = x_i) \to 0$ when $\Delta x \to 0$, because $n_i \to 0$

```
##          outcome ni         fi
## 1  [-1.02,0.115]  2 0.01626016
## 2   (0.115,1.23]  0 0.00000000
## 3    (1.23,2.34]  3 0.02439024
## 4    (2.34,3.46]  3 0.02439024
## 5    (3.46,4.58]  2 0.01626016
## 6    (4.58,5.69]  4 0.03252033
## 7     (5.69,6.8] 11 0.08943089
## 8     (6.8,7.92] 34 0.27642276
## 9    (7.92,9.04] 12 0.09756098
## 10   (9.04,10.2]  4 0.03252033
## 11   (10.2,11.3]  3 0.02439024
## 12   (11.3,12.4]  7 0.05691057
## 13   (12.4,13.5]  2 0.01626016
## 14   (13.5,14.6]  6 0.04878049
## 15   (14.6,15.7]  4 0.03252033
## 16   (15.7,16.8]  8 0.06504065
## 17     (16.8,18]  4 0.03252033
## 18     (18,19.1]  9 0.07317073
## 19   (19.1,20.2]  3 0.02439024
## 20   (20.2,21.3]  2 0.01626016
```

6.4 Continuous random variable

We define a quantity at a point x that is the amount of probability per unit distance that we would find in an **infinitesimal** bin dx at x

$$f(x) = \frac{P(x \leq X \leq x + dx)}{dx}$$

$f(x)$ is called the probability **density** function.

Therefore, the probability of observing x between x and $x + dx$ is given by

$$P(x \leq X \leq x + dx) = f(x)dx$$

6.5 Continuous random variable

Definition

For a continuous random variable X, a **probability density** function is such that

The function is positive:

- $f(x) \geq 0$

The probability of observing a value within an interval is the **area under the curve**:

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

The probability of observing **any** value is 1:

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

6.6 Continuous random variable

- The probability density function is a step forward in the abstraction of probabilities: we add the continuous limit ($dx \to 0$).
- All the properties of probabilities are translated in terms of densities ($\sum \to \int$).
- Assignment of probabilities to a random variable can be done with equiprobability (classical) arguments.
- Densities are mathematical quantities some will map to experiments some will not. *Which density will map best to my experiment?*

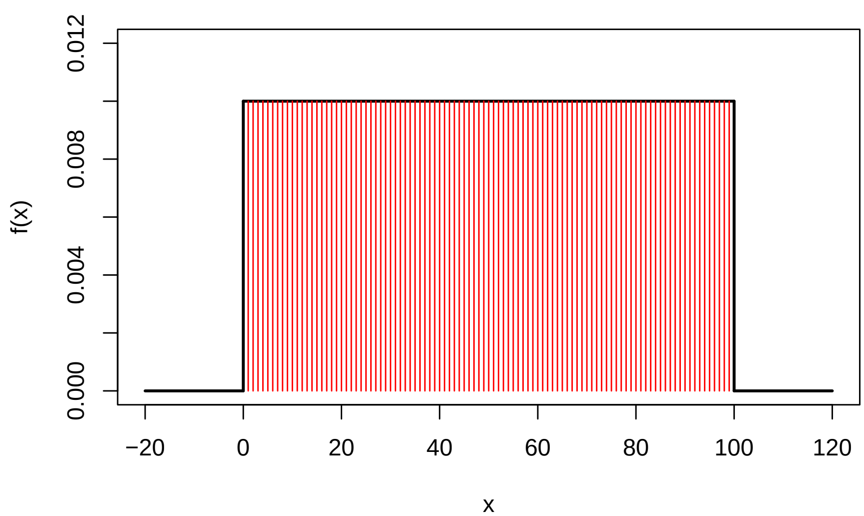
6.7 Total area under the curve

Example: take the **probability density** that may describe the random variable that measures where a raindrop falls in a rain gutter of length $100cm$.

$$f(x) = \begin{cases} \frac{1}{100}, & \text{if } x \in (0, 100) \\ 0, & otherwise \end{cases}$$

Then the probability of **any** observation is the total **area under the curve**

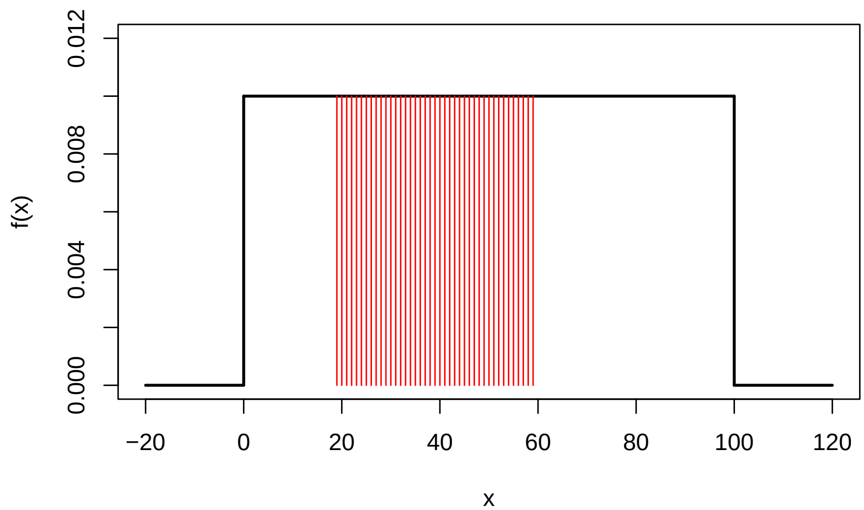
$P(-\infty \leq X \leq \infty) = \int_{-\infty}^{\infty} f(x)dx = 100 * 0.01 = 1$



6.8 Area under the curve

The probability of observing x in an interval is the **area under the curve** within the interval

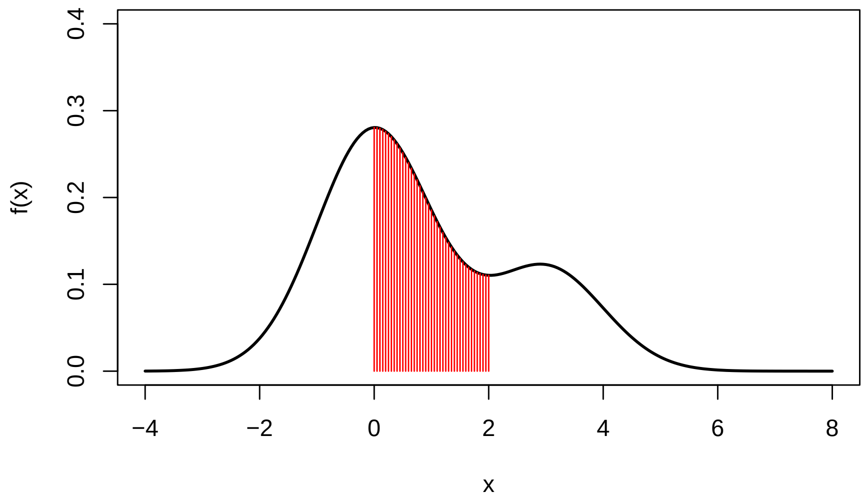
- $P(20 \leq X \leq 60) = \int_{20}^{60} f(x)dx = (60 - 20) * 0.01 = 0.4$



6.9 Area under the curve

In general $f(x)$ should satisfy:

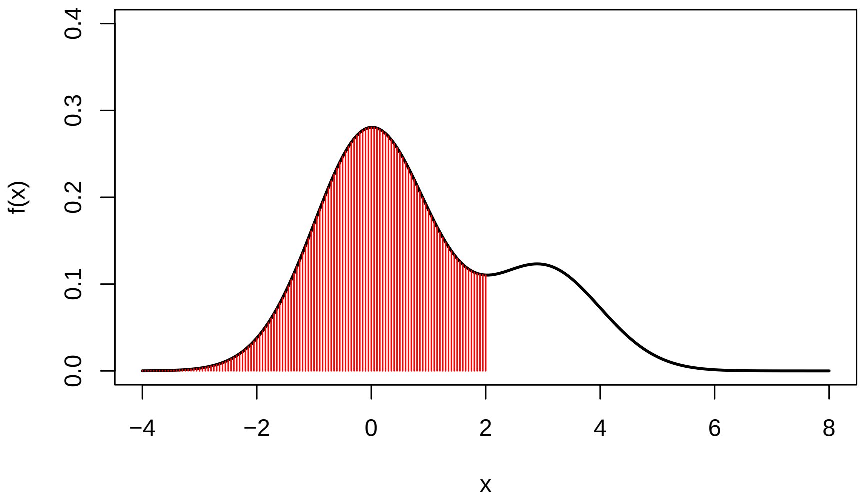
- $0 \leq P(a \leq X \leq b) = \int_a^b f(x)dx \leq 1$



6.10 Probability distribution

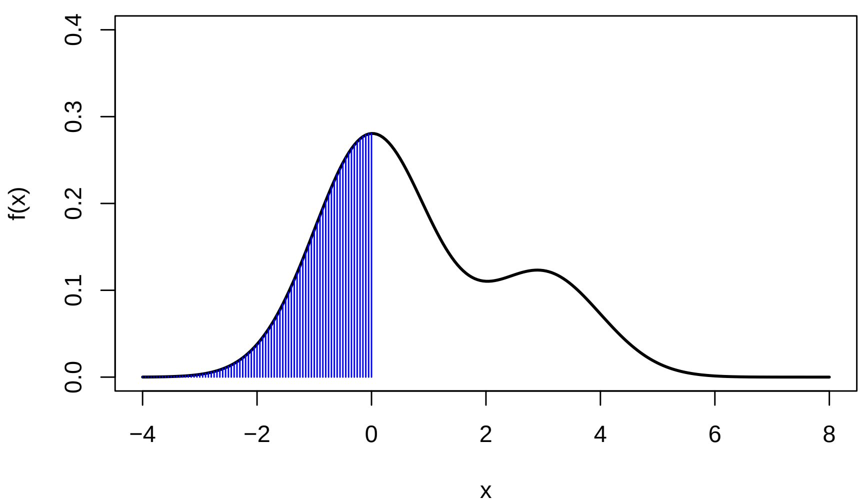
The probability accumulated up to b is defined by the probability distribution F

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



The probability accumulated up to a is

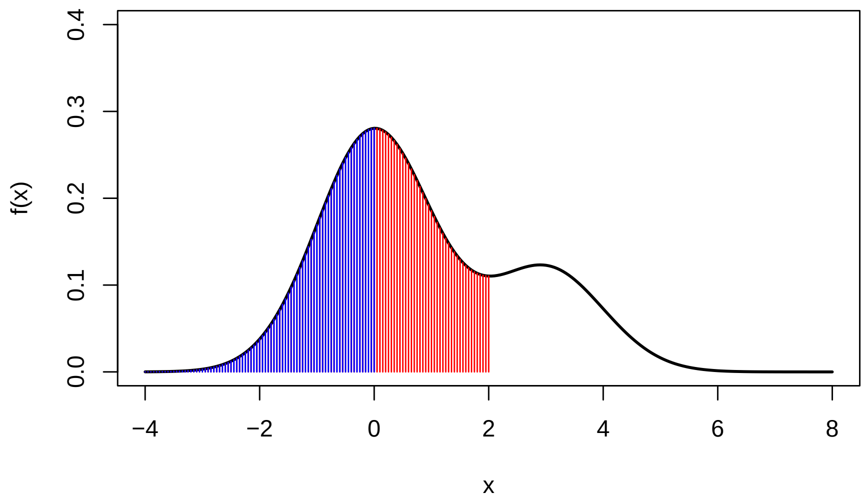
- $F(a) = P(X \leq a)$



6.11 Probability distribution

The probability between a and b is defined by the probability distribution F

- $P(a \leq X \leq b) = \int_a^b f(x)dx = F(b) - F(a)$



6.12 Probability distribution

The probability distribution of a continuous random variable is defined as $F(a) = P(X \leq a) = \int_{-\infty}^a f(x)dx$

with the properties that:

It is between 0 and 1:

- $F(-\infty) = 0$ and $F(\infty) = 1$

It always increases:

- if $a \leq b$ then $F(a) \leq F(b)$

It can be used to compute probabilities:

- $P(a \leq X \leq b) = F(b) - F(a)$

It recovers the probability density:

- $f(x) = \frac{dF(x)}{dx}$

We use **probability distributions** to **compute probabilities** of a random variable with intervals

---

6.13 Probability distribution

For the uniform density function:

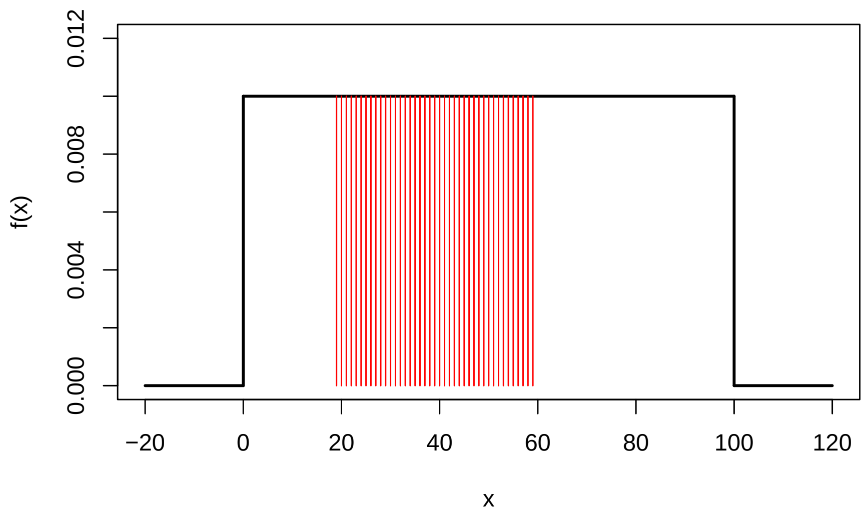
$$f(x) = \begin{cases} \frac{1}{100}, & \text{if } x \in (0, 100) \\ 0, & otherwise \end{cases}$$

The probability distribution is

$$F(a) = \begin{cases} 0, & a \leq 0 \\ \frac{a}{100}, & \text{if } a \in (0, 100) \\ 1, & 10 \leq a \\ \end{cases}$$

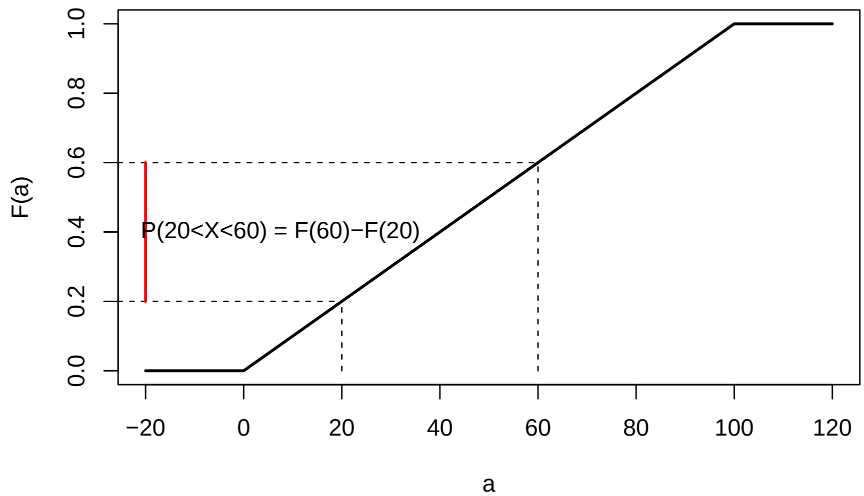
6.14 Probability graphics

The probability $P(20 < X < 60)$ is the *area* under the **density** curve



6.15 Probability graphics

The probability $P(20 < X < 60)$ is the *difference* in **distribution** values



6.16 Mean

As in the discrete case, the **mean** measures the center of the distribution

Definition

Suppose X is a continuous random variable with probability **density** function $f(x)$. The mean or expected value of X, denoted as μ or $E(X)$, is

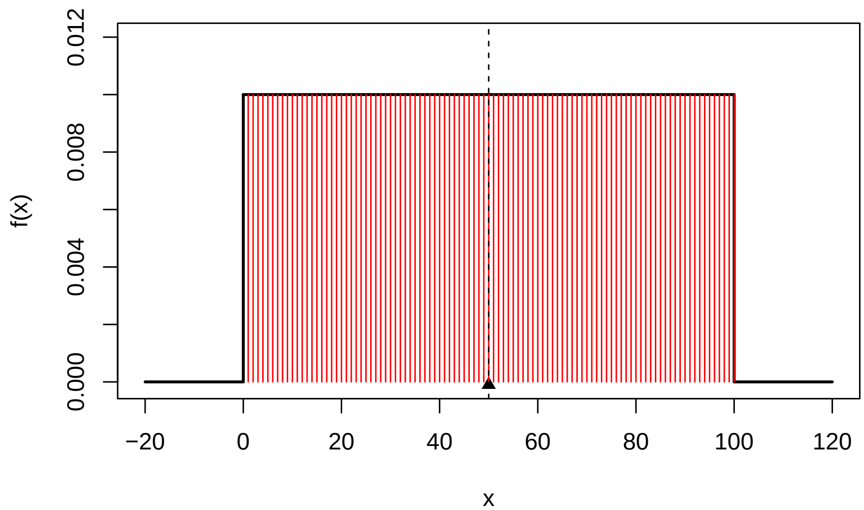
$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

It is the continuous version of the center of mass.

6.17 Mean

$$f(x) = \begin{cases} \frac{1}{100}, & \text{if } x \in (0, 100) \\ 0, & otherwise \end{cases}$$

$E(X) = 50$



6.18 Variance

As in the discrete case, the variance measures the dispersion about the mean

Definition

Suppose X is a continuous random variable with probability density function $f(x)$. The variance of X, denoted as σ^2 or $V(X)$, is

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

6.19 Functions of X

Definition

For any function h of a random variable X, with mass function $f(x)$, its expected value is given by

$$E[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)dx$$

And we have the same properties as in the discrete case

- The mean of a linear function is the linear function fo the mean:

$$E(a \times X + b) = a \times E(X) + b$$

  for a and b scalars.

- The variance of a linear function of X is:

$$V(a \times X + b) = a^2 \times V(X)$$

- The variance about the origin is the variance about the mean plus the mean squared:

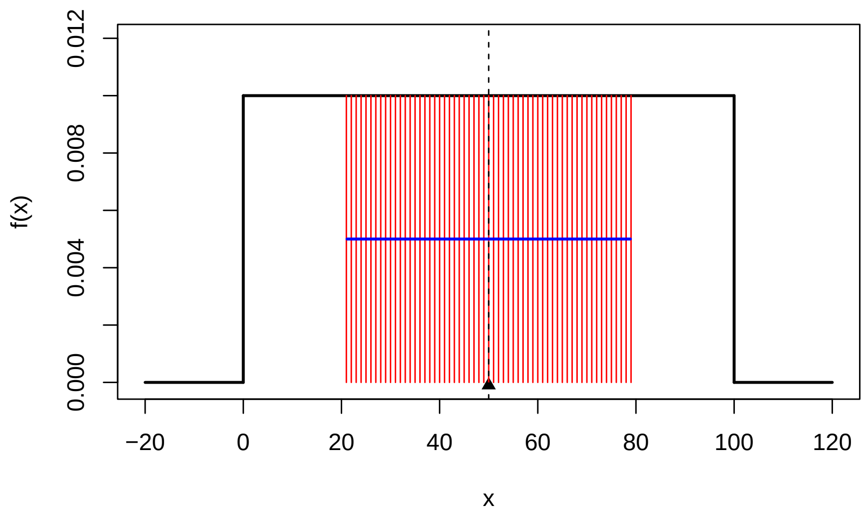
$$E(X^2) = V(X) + E(X)^2$$

6.20 Example

- for the probability density

$$f(x) = \begin{cases} \frac{1}{100}, & \text{if } x \in (0, 100) \\ 0, & otherwise \end{cases}$$

- compute the mean
- compute variance using $E(X^2) = V(X) + E(X)^2$
- compute $P(\mu - \sigma \leq X \leq \mu + \sigma)$
- What are the first and third quartiles?


0.000
0.004
0.008
0.012
f(x)
−20
0
20
40
60
80
100
120
x

Chapter 7

Exercises

7.1 Data description

7.1.0.1 Exercise 1

We have performed an experiment 8 times with the following results

```
## [1]  3  3 10  2  6 11  5  4
```

Answer the following questions:

- Compute the relative frequencies of each outcome.
- Compute the cumulative frequencies of each outcome.
- What is the average of the observations?
- What is the median?
- What is the third quartile?
- What is the first quartile?

7.1.0.2 Exercise 2

We have performed an experiment 10 times with the following results

```
##  [1] 2.875775 7.883051 4.089769 8.830174 9.404673 0.455565 5.281055 8.924190
##  [9] 5.514350 4.566147
```

Consider 10 bins of size 1: [0,1], (1,2]…(9,10].

Answer the following questions:

- Compute the relative frequencies of each outcome and draw the histogram
- Compute the cumulative frequencies of each outcome and sketch the cumulative plot.
- Sketch a boxplot.

7.2 Probability

7.2.0.1 Exercise 1

The outcome of one random experiment is to measure the misophonia severity **and** depression status of one patient.

- Misophonia severity: $x \in \{0, 1, 2, 3, 4\}$
- Depression: $y \in \{0, 1\}$ (no:0, yes:1)

```
##   Misofonia.dic depresion.dic
## 1             4             1
## 2             2             0
## 3             0             0
## 4             3             0
## 5             0             0
## 6             0             0
```

A large study on 123 patients showed the frequencies $n_{x,y}$ given in the contingency table:

```
##
##                Depression:0 Depression:1
##   Misophonia:4            0            9
##   Misophonia:3           25            6
##   Misophonia:2           34            3
##   Misophonia:1            5            0
##   Misophonia:0           36            5
```

Let's assume that $N >> 0$ and that the frequencies **estimate** the probabilities $f_{x,y} = \hat{P}(X, Y)$

```
##
##                Depression:0 Depression:1
##   Misophonia:4   0.00000000   0.07317073
##   Misophonia:3   0.20325203   0.04878049
##   Misophonia:2   0.27642276   0.02439024
##   Misophonia:1   0.04065041   0.00000000
##   Misophonia:0   0.29268293   0.04065041
```

- What is the marginal probability of misophonia severity 3?
- What is the probability of not being misophonic **and** not depressed?
- What is the probability of being misophonic **or** depressed?
- What is the probability of being misophonic **and** depressed?
- Describe in English the outcomes with probability 0.

7.2.0.2 Exercise 2

We have performed an experiment 10 times with the following results

```
##          A     B
## 1     male  dead
## 2     male  dead
## 3     male  dead
## 4   female alive
## 5     male  dead
## 6   female alive
## 7   female  dead
## 8   female alive
## 9     male alive
## 10    male alive
```

- Create the contingency table for the number ($n_{i,j}$) of observations of each outcome (A, B)
- Create the contingency table for the relative frequency ($f_{i,j}$) of the outcomes
- What is the marginal frequency of being male?
- What is the marginal frequency of being alive?
- What is the frequency of being alive **or** female?

7.3 Conditional Probability

7.3.0.1 Exercise 1

A machine is tested for its performance to produce high-quality turning rods. These are the results of the testing

	Rounded: Yes	Rounded: No
smooth surface: yes	200	1
smooth surface: no	4	2

- What is the estimated probability that the machine produces a rod that does not satisfy any quality control?
- What is the estimated probability that the machine produces a rod that does not satisfy at least one quality control?
- What is the estimated probability that the machine produces rounded and smoothed surfaced rods?
- what is the estimated probability that the rod is rounded if the rod is smooth?
- what is the estimated probability that the rod is smooth if it is rounded?
- what is the estimated probability that the rod is neither smooth nor rounded if it does not satisfy at least one quality control?

- Are smoothness and roundness independent events?

7.3.0.2 Exercise 2

We develop a test to detect the presence of bacteria in a lake. We find that if the lake contains the bacteria the test is positive 70% of the time. If there are no bacteria then the test is negative 60% of the time. We deploy the test in a region where we know that 20% of the lakes have bacteria.

- What is the probability that one lake that tests positive is contaminated with bacteria?

7.3.0.3 Exercise 3

Two machines are tested for their performance to produce high-quality turning rods. These are the results of the testing

Machine 1

	Rounded: Yes	Rounded: No
smooth surface: yes	200	1
smooth surface: no	4	2

Machine 2

	Rounded: Yes	Rounded: No
smooth surface: yes	145	4
smooth surface: no	8	6

- what is the probability that the rod is rounded?
- What is the probability that the rod has been produced by machine 1?
- what is the probability that the rod is not smooth?
- What is the probability that the rod is smooth or rounded or produced by machine 1?
- What is the probability that the rod is rounded if it is smoothed and from machine 1?
- What is the probability that the rod is not rounded if it is not smoothed and is from machine 2?
- what is the probability that the rod has come from machine 1 if it it is smoothed and rounded?
- what is the probability that the rod has come from machine 2 if it does not pass at least one of the quality controls?

7.3.0.4 Exercise 4

We want to cross an avenue with two traffic lights. The probability of finding the first traffic light in red is 0.6. If we stopped at the first traffic light, the probability of stopping at the second one is 0.15. Whereas the probability of stopping on the second one if we do not stop on the first one is 0.25.

When we try to cross both traffic lights:

- what is the probability of having to stop at each traffic light?
- What is the probability of having to stop at at least one traffic light?
- What is the probability of having to stop at only one traffic light?
- If I stopped at the second traffic light, what is the probability that I had to stop at the first one?

- If I had to stop at any traffic light, what is the probability that I had to do it twice?
- Is stopping at the first traffic light an independent event from stopping at the second traffic light?

Now, we want to cross an avenue with three traffic lights. The probability of finding a traffic light in red only depends on the previous one. In particular, the probability of finding one traffic light in red given that the previous one was in red is 0.15. Whereas, the probability of finding one traffic right in red given that the previous one was in green is 0.25. Also, the probability of finding the first traffic light in red is 0.6.

- What is the probability of having to stop at each traffic light?
- What is the probability of having to stop at at least one traffic light?
- What is the probability of having to stop at only one traffic light?

hints:

- If the probability that one traffic light is red depends only on the previous one then $P(R_3|R_2, R_1) = P(R_3|R_2, \bar{R}_1) = P(R_3|R_2)$ and $P(R_3|\bar{R}_2, R_1) = P(R_3|\bar{R}_2, \bar{R}_1) = P(R_3|\bar{R}_2)$

- The joint probability of finding three traffic lights in red can be written as: $P(R_1, R_2, R_3) = P(R_3|R_2)P(R_2|R_1)P(R_1)$

7.3.0.5 Exercise 5

A quality test on a random brick is defined by the events:

- Pass quality test: E, do no pass quality test: $\bar{E}$
- Defective: D, non-defective: $\bar{D}$

If the diagnostic test has sensitivity $P(E|\bar{D}) = 0.99$ and specificity $P(\bar{E}|D) = 0.98$, and the probability of passing a test is $P(E) = 0.893$ then

- what is the probability that a brick chosen at random is defective $P(D)$?

- What is the probability that a brick that has passed the test is really defective?
- The probability that a brick is not defective **and** that it does not pass the test
- Are D and $\bar{E}$ statistical independent?

7.4 Random variables

7.4.0.1 Exercise 1

Given the probability distribution for a discrete variable X

$$F(x) = \begin{cases} 0, & x \leq -1 \\ 0.2, & x \in [-1, 0) \\ 0.35, & x \in [0, 1) \\ 0.45, & x \in [1, 2) \\ 1, & x \geq 2 \end{cases}$$

a) find $f(X)$
b) find $E(X)$ and $V(X)$
c) what is the expected value and variance of $Y = 2X + 3$
d) what is the median of X?

7.4.0.2 Exercise 2

We have a system of transmission of pixels that is totally noisy. We are testing the system and have designed an experiment to transmit 3 pixels.

- What is the probability of receiving 0, 1, 2, or 3 errors in the transmission of 3 pixels?
- Sketch the probability mass function
- What is the expected value of the error?
- What is its variance?
- Sketch the probability distribution
- What is the probability of transmitting at least 1 error?

hints:

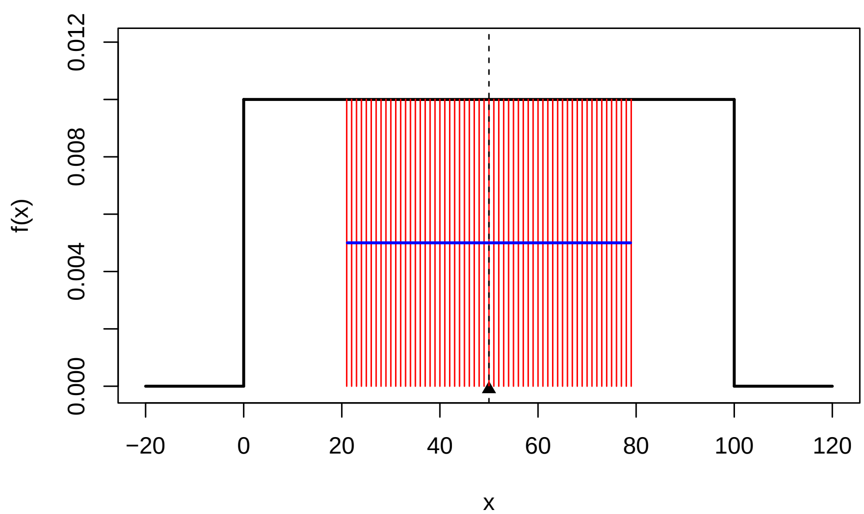
- Sample space: $\{(0,0,0),(1,0,0),(0,1,0),(0,0,1),(0,1,1),(1,0,1),(1,1,0),(1,1,1)\}$
- where, for example, the event $(0,1,1)$ is the event of receiving the first pixel with no error and the second and third pixels with errors.
- All events are equally probable.

7.4.0.3 Exercise 3

- for the probability density

$$f(x) = \begin{cases} \frac{1}{100}, & \text{if } x \in (0, 100) \\ 0, & otherwise \end{cases}$$

- compute the mean
- compute variance using $E(X^2) = V(X) + E(X)^2$
- compute $P(\mu - \sigma \leq X \leq \mu + \sigma)$
- What are the first and third quartiles?



7.4.0.4 Exercise 4

For the probability density

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } 0 \leq \\ 0, & otherwise \end{cases}$$

- Confirm that this is a probability density
- Find the probability distribution $F(a)$
- Compute the mean
- Compute variance using $E(X^2) = V(X) + E(X)^2$

7.4.0.5 Exercise 5

Given the cumulative distribution for a random variable X

$$F(x) = \begin{cases} 0, & x < -1 \\ \frac{1}{80}(17 + 16x - x^2), & x \in [-1, 7) \\ 1, & x \geq 7 \end{cases}$$

compute:

- $P(X > 0)$
- $E(X)$
- $P(X > 0|X < 2)$

7.5 Probability Models

7.5.0.1 Exercise 1

A search engine fails to retrieve information with a probability 0.1

- If we system receives 50 search requests, what is the probability that the system fails to answer three of them?
- What is the probability that the engine successfully completes 15 searches before the first failure?
- We consider that a search engine works sufficiently well when it is able to find information for 10 requests for every 2 failures. What is the probability that in a reliability trial our search engine is satisfactory?

7.5.0.2 Exercise 2

In a population, the probability that a baby boy is born is $p = 0.51$. Consider a family of 4 children

- What is the probability that a family has only one boy?
- What is the probability that a family has only one girl?
- What is the probability that a family has only one boy or only one girl?
- What is the probability that the family has at least two boys?
- What is the number of children that a family should have such that the probability of having at least a girl is more than 0.75?

7.5.0.3 Exercise 3

The average number of radioactive particles hitting a Geiger counter is 2.3 seconds.

- What is the probability of counting exactly 2 particles in a second?

- What is the probability of detecting exactly 10 particles in 5 seconds?
- What is the probability of at least one count in two seconds?
- What is the probability of having to wait 2.5 seconds after we switch on the detector?

7.5.0.4 Exercise 4

- What is the probability that a man's height is at least 165cm if the population mean is 175cm y the standard deviation is 10cm?
- What is the probability that a man's height is between 165cm and 180cm.
- What is the height that defines the 5% of the smallest men?

7.6 Point Estimators

7.6.0.1 Exercise 1

Consider the probability model

$$f(x) = \begin{cases} 1/2 - a, & \text{if } x = -1 \\ 1/2, & \text{if } x = 0 \\ a, & 1\text{if } x = 1 \end{cases}$$

where a is a parameter.

Compute the mean and variance of the statistic:

$$T = \frac{\bar{X}}{2} + \frac{1}{4}$$

where $\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$

- is T a biased estimator of a?
- is T consistent? i.e. $V(T) \to 0$ when $N \to \infty$

7.6.0.2 Exercise 2

- Is $\bar{X}^2 = (\frac{1}{N} \sum_{i=1}^{N} X_i)^2$ an unbiased estimator of $E(X)^2$?

7.7 Sampling and Central Limit Theorem

7.7.0.1 Exercise 1

A battery model charges up to 75% of its capacity within an hour with a standard deviation of 15%.

- If we charge 25, what is the probability that the sample average is within a distance of 5% charge from the mean?
- If we charge 100, what is that probability?
- If, instead we only charge 9 batteries, what is the charge that is surpassed by the sample average with only 0.015 probability?

7.7.0.2 Exercise 2

An electronic component is needed for the correct functioning of a telescope. It needs to be replaced immediately when it wears out.

The mean life of the component (μ) is 100 hours and its standard deviation σ is 30 hours.

- what is the probability that the average of the mean life of 50 components is within 1 hour from the mean life of a single component?
- How many components do we need such that the telescope is operational 2750 consecutive hours with 0.95 probability?

7.7.0.3 Exercise 3

An automated machine fills test tubes with biological samples with mean $\mu = 130$mg and a standard deviation of $\sigma = 5$mg.

- for a random sample of size 50. What is the probability that the sample mean (average) is between 128 and 132gr?
- what should be the size of the sample (n) such that the sample mean $\bar{X}$ is higher than 131gr with a probability less or equal than 0.025?

7.7.0.4 Exercise 4

In the Caribbean, there appears to be an average of 6 hurricanes per year. Considering that hurricane formation is a Poisson process, meteorologists plan to estimate the mean time between the formation of two hurricanes. They plan to collect a sample of size 36 for the times between two hurricanes.

- What is the probability that their sample average is between 45 and 60 days?
- Which should be the sample size such that they have a probability of 0.025 that the sample mean is greater than 70 days?

7.7.0.5 Exercise 5

The probability that a particular mutation is found in the population is 0.4. If we test 2000 people for the mutation:

- What is the probability that the total number of people with the mutation is between 791 and 809?

hint: Use the CLT with a sample of 2000 Bernoulli trials. This is known as the normal approximation of the binomial distribution.

7.8 Maximum likelihood

7.8.0.1 Exercise 1

For a random variable with a binomial probability function

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

- What is the maximum-likelihood estimator of p for a sample of size 1 of this random variable?
- In **one** exam of 100 students we observed $x_1 = 68$ students that passed the exam. What is the estimate of the p?

7.8.0.2 Exercise 2

Take a random variable with the following probability density function

$$f(x) = \begin{cases} (1+\theta)x^{\theta}, & \text{if } x \in (0,1) \\ 0, & x \notin (0,1) \end{cases}$$

- What is the maximum likelihood estimate for θ?
- If we take a 5-sample with observations $x_1 = 0.92; \quad x_2 = 0.79; \quad x_3 = 0.90; \quad x_4 = 0.65; \quad x_5 = 0.86$

What is the estimated value of the parameter θ?

7.8.0.3 Exercise 3

Take a random variable with the following probability density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } 0 \leq \\ 0, & otherwise \end{cases}$$

- What is the maximum likelihood estimate for λ?
- If we take a 5-sample with observations $x_1 = 0.223 \quad x_2 = 0.681; \quad x_3 = 0.117; \quad x_4 = 0.150; \quad x_5 = 0.520$

What is the estimated value of the parameter λ?

7.9 Method of moments

7.9.0.1 Exercise 1

What are the estimators of the following parametric models given by the method of moments?

Model	f(x)	E(X)
Bernoulli	$p^x(1-p)^{1-x}$	p
Binomial	$\binom{n}{x}p^x(1-p)^{n-x}$	np
Shifted geometric	$p(1-p)^{x-1}$	$\frac{1}{p}$
Negative Binomial	$\binom{x+r-1}{x}p^r(1-p)^x$	$r\frac{1-p}{p}$
Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ
Exponential	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$
Normal	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ

7.9.0.2 Exercise 2

Take a random variable with the following probability density function

$$f(x) = \begin{cases} (1+\theta)x^\theta, & \text{if } x \in (0,1) \\ 0, & x \notin (0,1) \end{cases}$$

- Compute $E(X)$ as a function of θ
- What is the estimate for θ using the method of moments?
- If we take a 5-sample with observations $x_1 = 0.92$; $x_2 = 0.79$; $x_3 = 0.90$; $x_4 = 0.65$; $x_5 = 0.86$

What is the estimated value of the parameter θ?

7.9.0.3 Exercise 3

Consider a discrete random variable X that follows a negative binomial distribution with probability mass function:

$$f(x) = \binom{x+r-1}{x}p^r(1-p)^x$$

Given that

- $E(X) = \dfrac{r(1-p)}{p}$
- $V(X) = \dfrac{r(1-p)}{p^2}$

compute:

- An estimate for the parameter r and an estimate for the parameter p obtained from a random sample of size n using the method of moments.
- The values of the estimates of r y p for the folowing random sample:

$$x_1 = 27; \qquad x_2 = 8; \qquad x_3 = 22; \qquad x_4 = 29; \qquad x_5 = 19; \qquad x_5 = 32$$

7.10 Confidence intervals

7.10.0.1 Exercise 1

In a scientific paper the authors report a 95% confidence interval of $(228, 232)$ for the natural frequency (Hz) of metallic beam. They useda sample of size 25 and considered that the measurements distributed normally.

- What is the mean and the standard deviation of the measurements?
- Compute the 99% confidence interval.

hints:

- in R $t_{0.025,24}$ = qt(0.975, 24)$\sim$ 2
- in R $t_{0.005,24}$ =qt(0.995, 24)$\sim$ 2.8

7.10.0.2 Exercise 2

compute 95% CI for the mean of a normal variable with known variance $\sigma^2 = 9$ and $\mu = 22$

7.10.0.3 Exercise 3

This year a 1000 sample of patients with influenza developed complications.

- Compute the 99% confidence interval for the proportion of complications.
- The previous year 2% showed complications. Can we say with 99% confidence that this year there is a significnat drop in influenza complications?