Stats theory (SDA)

Alejandro Caceres

2022-11-27

Contents

1 About **13**
1.1 Recommended reading list . . . 13

2 Data description **15**
2.1 Objective . . . 15
2.2 Statistics . . . 15
2.3 Scientific method . . . 15
2.4 Outcome . . . 15
2.5 Types of outcome . . . 16
2.6 Random experiments . . . 16
2.7 Absolute frequencies . . . 16
2.8 Example . . . 17
2.9 Relative frequencies . . . 17
2.10 Example . . . 18
2.11 Bar plot . . . 18
2.12 Pie chart . . . 19
2.13 Categorical and ordered variables . . . 19
2.14 Example . . . 20
2.15 Absolute and relative cumulative frequencies . . . 20
2.16 Frequency table . . . 20
2.17 Cumulative frequency plot . . . 21
2.18 Continuous variables . . . 21
2.19 Bins . . . 22
2.20 Create a categorical variable from a continuous one . . . 22
2.21 Frequency table for a continuous variable . . . 23
2.22 Histogram . . . 23
2.23 Histogram . . . 24
2.24 Cumulative frequency plot: Continous variables . . . 25
2.25 Summary statistics . . . 26
2.26 Average . . . 27
2.27 Average (categorical ordered) . . . 27
2.28 Average (categorical ordered) . . . 27
2.29 Average . . . 28

2.30 Average . . . 28
2.31 Median . . . 28
2.32 Median Vs Average . . . 29
2.33 Dispersion . . . 30
2.34 Dispersion . . . 31
2.35 Sample variance . . . 31
2.36 Sample variance . . . 31
2.37 Standard deviation . . . 32
2.38 IQR . . . 32
2.39 IQR . . . 33
2.40 Box plot . . . 33

3 Probability **35**
3.1 Objective . . . 35
3.2 Random experiments . . . 35
3.3 Probability . . . 36
3.4 Example . . . 36
3.5 Example . . . 36
3.6 Relative frequency . . . 36
3.7 At infinity . . . 37
3.8 Frequentist probability . . . 38
3.9 Classical Probability . . . 38
3.10 Classical and frequentist probabilities . . . 39
3.11 Probability . . . 39
3.12 Sample space . . . 39
3.13 Examples of sample spaces . . . 40
3.14 Discrete and continuous sample spaces . . . 40
3.15 Event . . . 40
3.16 Event operations . . . 41
3.17 Event operations example . . . 41
3.18 Outcomes . . . 41
3.19 Probability definition . . . 42
3.20 Probability properties . . . 42
3.21 Addition Rule . . . 43
3.22 Example Addition Rule . . . 43
3.23 Venn diagram . . . 43
3.24 Probability table . . . 44
3.25 Example probability table . . . 44
3.26 Contingency table . . . 44
3.27 Example contingency table . . . 45
3.28 Misophonia study . . . 45
3.29 Contingency table for frequencies . . . 48
3.30 Heat map . . . 49
3.31 Continous variables . . . 49
3.32 Heat map for continuous variables . . . 52
3.33 Scatter plot . . . 53

4 Conditional Probability **55**
4.1 Objective 55
4.2 Joint Probability 55
4.3 Diagnostics 56
4.4 Diagnostics Test 56
4.5 Observations 56
4.6 Contingency tables 57
4.7 Conditional probability 57
4.8 Conditional probability 58
4.9 Conditional contingency table 58
4.10 Example conditional contingency table 59
4.11 Multiplication rule 59
4.12 Diagnostic performance 59
4.13 Multiplication rule 60
4.14 Contingency table in terms of conditional probabilities 60
4.15 Conditional tree 61
4.16 Contingency table in terms of conditional probabilities 61
4.17 Total probability rule 61
4.18 Conditional tree 62
4.19 Finding reverse probabilities 62
4.20 Recover joint probabilities 62
4.21 Reverse conditionals 62
4.22 Baye's theorem 63
4.23 Example: Bayes' theorem 64
4.24 Example: Bayes' theorem 64
4.25 Statistical independence 65
4.26 Statistical independence 65
4.27 Statistical independence 65
4.28 Statistical independence 66
4.29 Products of marginals products 66
4.30 Example 67

5 Discrete Random Variables **69**
5.1 Objective 69
5.2 How do we assign probability values to outcomes? 69
5.3 Random variable 69
5.4 Random variable 70
5.5 Events of observing a random variable 70
5.6 Probability of random variables 71
5.7 Probability functions 71
5.8 Probability functions 72
5.9 Probability functions 72
5.10 Probability functions 73
5.11 Example: Probability mass function 73
5.12 Probability table for equally likely outcomes 74
5.13 Probability table for X 74

5.14 Example 74
5.15 Example 75
5.16 Probabilities and frequencies 75
5.17 Probabilities and relative frequencies 76
5.18 Mean and Variance 76
5.19 Mean and Variance 77
5.20 Mean 77
5.21 Example: Mean 77
5.22 Mean and Average 78
5.23 Variance 79
5.24 Example: Variance 79
5.25 Functions of X 80
5.26 Example: Variance about the origin 80
5.27 Probability distribution 81
5.28 Example: Probability distribution 81
5.29 Probability distribution 82
5.30 Probability function and Probability distribution 82
5.31 Probability function and Probability distribution 83
5.32 Quantiles 83
5.33 Summary 84

6 Continous Random Variables 85
6.1 Objective 85
6.2 Continuous random variable 85
6.3 Continuous random variable 86
6.4 Continuous random variable 86
6.5 Continuous random variable 87
6.6 Continuous random variable 87
6.7 Total area under the curve 88
6.8 Area under the curve 88
6.9 Area under the curve 89
6.10 Probability distribution 90
6.11 Probability distribution 92
6.12 Probability distribution 92
6.13 Probability distribution 93
6.14 Probability graphics 93
6.15 Probability graphics 94
6.16 Mean 95
6.17 Mean 95
6.18 Variance 96
6.19 Functions of X 96
6.20 Example 97

7 Discrete Probability Models 99
7.1 Objective 99
7.2 Probability mass function 99

7.3 Probability model 100
7.4 Parametric models 100
7.5 Uniform distribution (one parameter) 101
7.6 Uniform distribution 102
7.7 Uniform distribution (two parameters) 102
7.8 Uniform distribution (two parameters) 103
7.9 Uniform distribution 103
7.10 Uniform distribution (two-parameter) 104
7.11 Parameters and Models 105
7.12 Parameters and Models 105
7.13 Bernoulli trial 106
7.14 Bernoulli trial 107
7.15 Bernoulli trial 107
7.16 Bernoulli trial 108
7.17 Binomial distribution 108
7.18 Examples: Binomial distribution 109
7.19 Binomial distribution 109
7.20 Binomial distribution 109
7.21 Binomial distribution: Definition 110
7.22 Binomial distribution: Mean and Variance 111
7.23 Example 1 111
7.24 Example 1 112
7.25 Example 2 112
7.26 Binomial distribution 113
7.27 Negative binomial distribution 113
7.28 Negative binomial distribution 114
7.29 Negative binomial distribution 114
7.30 Mean and Variance 115
7.31 Geometric distribution 115
7.32 Example 115
7.33 Example 116
7.34 Example 116
7.35 Examples 116
7.36 Negative binomial distribution 117

8 Poisson and Exponential Models 119
8.1 Objective 119
8.2 Discrete probability models 119
8.3 Counting events 120
8.4 Counting events 120
8.5 Poisson distribution 120
8.6 Poisson distribution 120
8.7 Poisson distribution: Derivation details 121
8.8 Poisson distribution 121
8.9 Poisson distribution 122
8.10 Poisson distribution 122

8.11 Poisson distribution 123
8.12 Continuous probability models 123
8.13 Exponential density 124
8.14 Exponential density 124
8.15 Exponential density 124
8.16 Exponential density 125
8.17 Exponential density 126
8.18 Exponential Distribution 126
8.19 Exponential Distribution 127
8.20 Exponential Distribution 128

9 Normal Distribution 129
9.1 Objective 129
9.2 Continuous probability models 129
9.3 Normal density 130
9.4 Normal density 130
9.5 Normal density 130
9.6 Normal density 130
9.7 Normal density 131
9.8 Definition 131
9.9 Normal probability density (Gaussian) 132
9.10 Normal distribution 132
9.11 Normal distribution 133
9.12 Normal distribution 133
9.13 Normal distribution 134
9.14 Normal distribution 134
9.15 Standard normal density 135
9.16 Standard normal density 135
9.17 Standard normal density 136
9.18 Normal distribution 136
9.19 Standard distribution 136
9.20 Standard normal density 137
9.21 Standard normal density 137
9.22 Normal and standard distributions 137
9.23 Normal distribution 138
9.24 Summary of probability models 138
9.25 R functions of probability models 139

10 Sampling Distributions 141
10.1 Objective 141
10.2 Normal distribution 141
10.3 Example: When we do not know the parameters 142
10.4 Example 142
10.5 Random sample 143
10.6 Example 144
10.7 Average or sample mean 144

10.8 The average as an estimator . . . 144
10.9 Sample variance . . . 145
10.10 Sample variance . . . 145
10.11 Fitting a model . . . 146
10.12 Prediction . . . 147
10.13 Inference . . . 148
10.14 Example: When we do know the parameters . . . 148
10.15 Density for X and $\bar{X}$ . . . 149
10.16 Sample mean distribution . . . 150
10.17 Inference on the average . . . 150
10.18 Density for $\bar{X}$ . . . 151
10.19 Inference in the sample variance . . . 153
10.20 Probabilities of the sample variance . . . 153
10.21 χ^2-statistic . . . 154
10.22 χ^2-statistic . . . 155

11 Central limit theorem 157
11.1 Objective . . . 157
11.2 Margin of error . . . 157
11.3 Margin of error . . . 157
11.4 Z-statistic . . . 158
11.5 Z-statistic . . . 158
11.6 Z-statistic . . . 159
11.7 Central Limit Theorem . . . 160
11.8 Central Limit Theorem . . . 160
11.9 Central Limit Theorem . . . 161
11.10 Margin of error with CLT . . . 163
11.11 Sample sum and CLT . . . 163
11.12 Unknown σ but large n . . . 164
11.13 T-statistic . . . 164
11.14 T-statistic . . . 165
11.15 T-statistic . . . 165
11.16 Example 1 . . . 166
11.17 Example 2 . . . 167

12 Maximum likelihood 169
12.1 Objective . . . 169
12.2 Statistic . . . 169
12.3 Estimator . . . 169
12.4 Estimator . . . 170
12.5 Examples 1: Average (Sample mean) . . . 170
12.6 Examples 2: Sample Variance . . . 170
12.7 Bias . . . 171
12.8 Consistency . . . 171
12.9 Maximum likelihood . . . 171
12.10 Example . . . 171

12.11 Probability density . . . 172
12.12 Probability density . . . 172
12.13 Example: Maximum likelihood . . . 174
12.14 Maximum likelihood . . . 174
12.15 Method step 1 . . . 175
12.16 Method step 2 . . . 175
12.17 Method step 3 . . . 175
12.18 Method step 3 . . . 176
12.19 Estimation . . . 176
12.20 Estimation . . . 177
12.21 Maximum likelihood: History . . . 177
12.22 Maximum likelihood: History . . . 177
12.23 Maximum likelihood: History . . . 178
12.24 Maximum likelihood: History . . . 178
12.25 Maximum likelihood: History . . . 178
12.26 Normal distribution . . . 179
12.27 Normal distribution . . . 179
12.28 Normal distribution . . . 179
12.29 Normal distribution . . . 180
12.30 Method of Moments . . . 180
12.31 Method of Moments . . . 180
12.32 Method of Moments . . . 181
12.33 Method of Moments . . . 181
12.34 Method of Moments . . . 182
12.35 Method of Moments . . . 182
12.36 Method of Moments . . . 182
12.37 Normal distribution . . . 183
12.38 Normal distribution . . . 183
12.39 Method of Moments . . . 184
12.40 Method of Moments . . . 184
12.41 Method of Moments . . . 185
12.42 Method of Moments . . . 185

13 Interval estimation **187**
13.1 Objective . . . 187
13.2 Average or sample mean . . . 187
13.3 Inference on the average . . . 188
13.4 Margin of error . . . 188
13.5 Outcome probability density Vs sample mean probability density 189
13.6 Real life . . . 191
13.7 Interval estimation . . . 192
13.8 Interval estimation . . . 192
13.9 Interval estimation . . . 193
13.10 Interval estimation . . . 193
13.11 Interval estimation . . . 194
13.12 Interval estimation . . . 194

13.13 Interval estimation . . . 195
13.14 Example . . . 195
13.15 Example . . . 195
13.16 T-statistic . . . 196
13.17 T-statistic . . . 197
13.18 T-statistic . . . 197
13.19 Example . . . 198
13.20 Example . . . 198
13.21 IC with CLT . . . 199
13.22 Central Limit Theorem . . . 200
13.23 Parameter estimation . . . 201
13.24 Interval estimation for proportions . . . 202
13.25 Interval estimation for proportions . . . 203
13.26 Interval estimation for proportions . . . 203
13.27 Interval estimation for proportions . . . 204
13.28 Interval estimation for the variance . . . 204
13.29 Interval estimation for the variance . . . 204
13.30 χ^2-statistic . . . 205
13.31 Interval estimation for the variance . . . 206
13.32 Interval estimation for the variance . . . 206
13.33 Interval estimation . . . 207
13.34 Interval estimation . . . 207

14 Exercises **209**
14.1 Data description . . . 209
14.2 Probability . . . 210
14.3 Conditional Probability . . . 211
14.4 Random variables . . . 214
14.5 Probability Models . . . 217
14.6 Sampling and Central Limit Theorem . . . 218
14.7 Point Estimators . . . 219
14.8 Maximum likelihood . . . 220
14.9 Method of moments . . . 221
14.10 Confidence intervals . . . 222
14.11 Hypothesis testing . . . 223

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

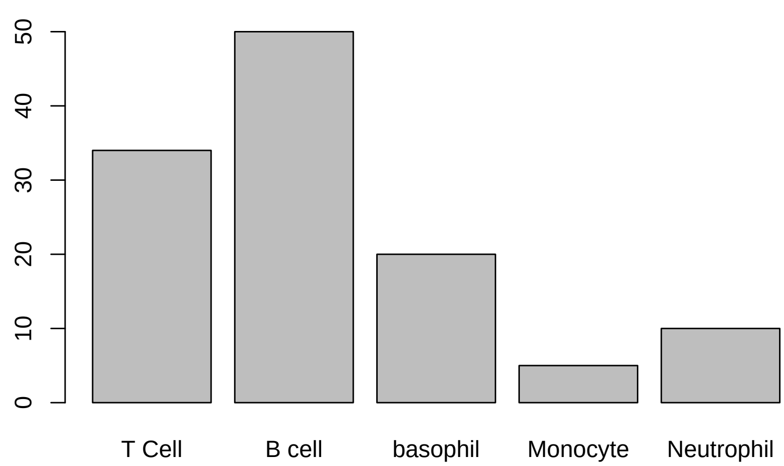
$\sum_{i=1..M} n_i = N$

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

where M is the number of outcomes.

2.11 Bar plot

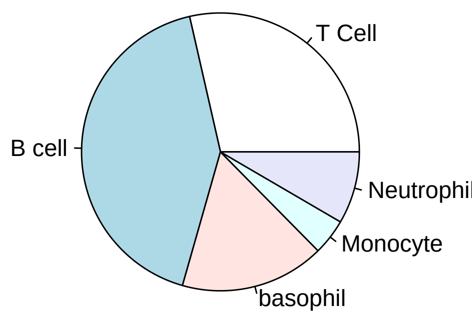
We can plot n_i Vs the outcomes, giving us a bar plot



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

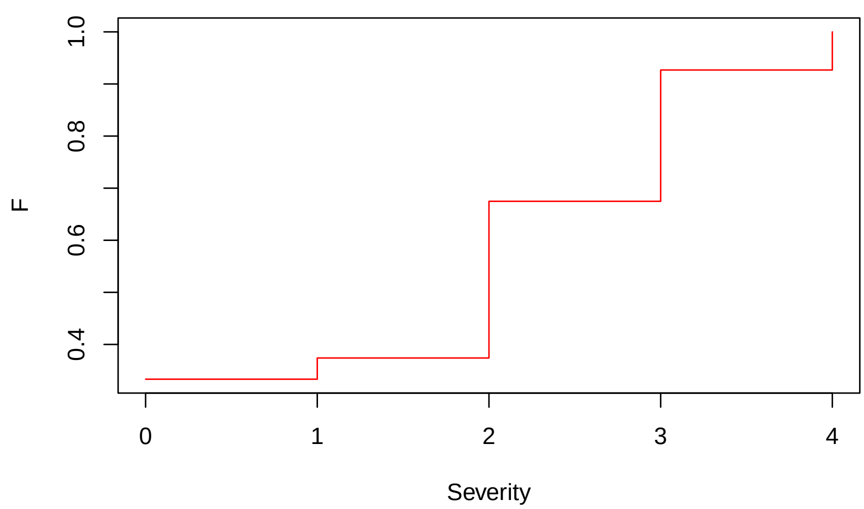
```
##   outcome ni         fi  Ni        Fi
## 0       0 41 0.33333333  41 0.3333333
## 1       1  5 0.04065041  46 0.3739837
```

```
## 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]"
```

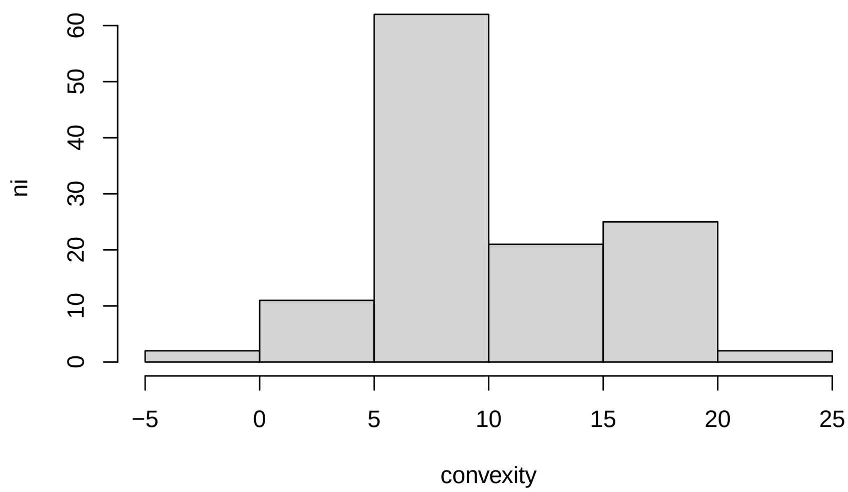
```
##  [41] "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"  "(7.92,12.4]"  "(12.4,16.8]"
##  [46] "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"  "(12.4,16.8]"
##  [51] "(7.92,12.4]"  "[-1.02,3.46]" "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"
##  [56] "(7.92,12.4]"  "(7.92,12.4]"  "(7.92,12.4]"  "(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
```

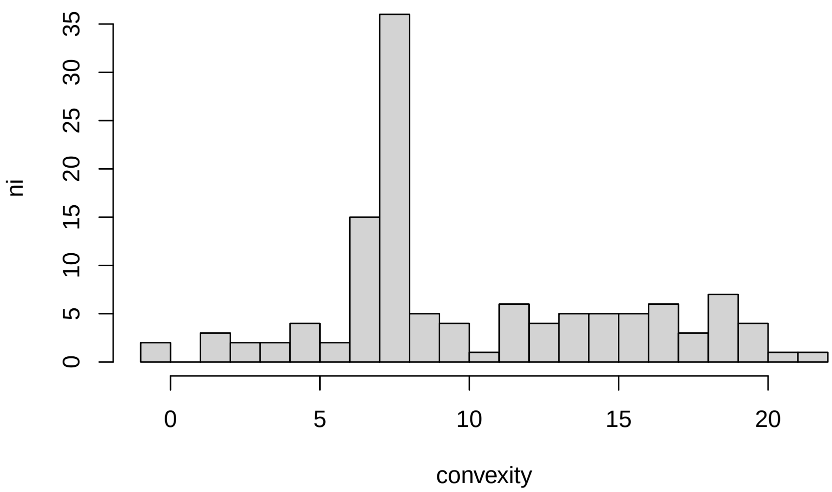
2.22 Histogram

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



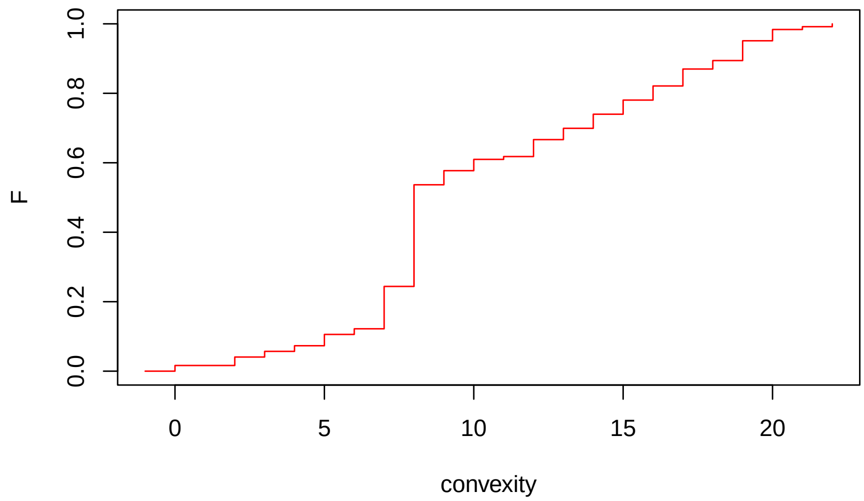
2.23 Histogram

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).

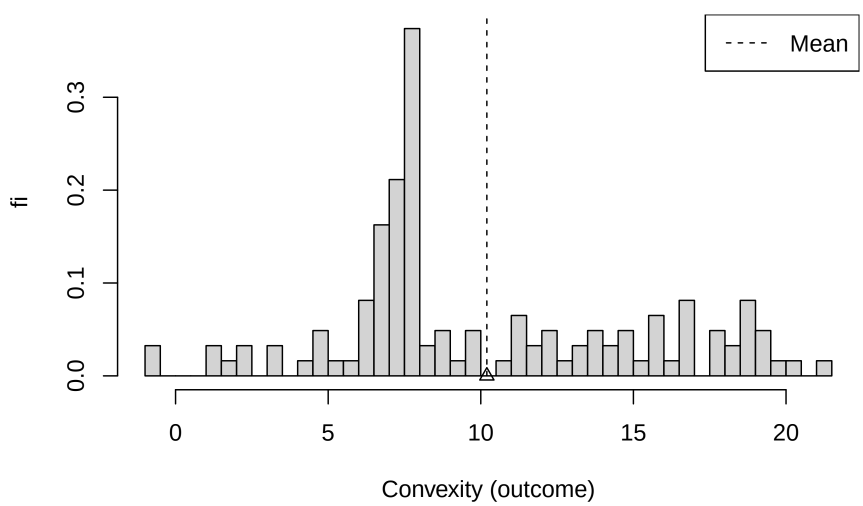
$\bar{x}$ is the **central value** or center of gravity of the outcomes. As if each outcome had a mass density given by f_i.

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

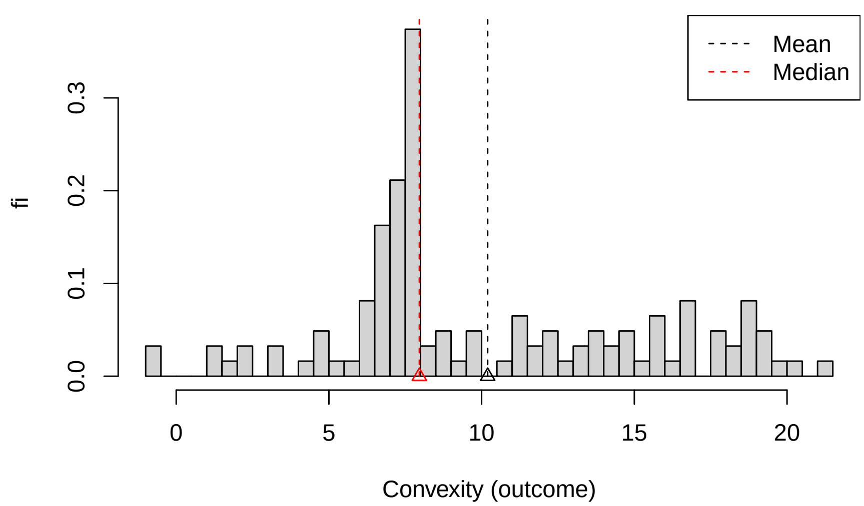
$$\sum_{x \le x_p} 1 = \frac{N}{2}$$

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

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

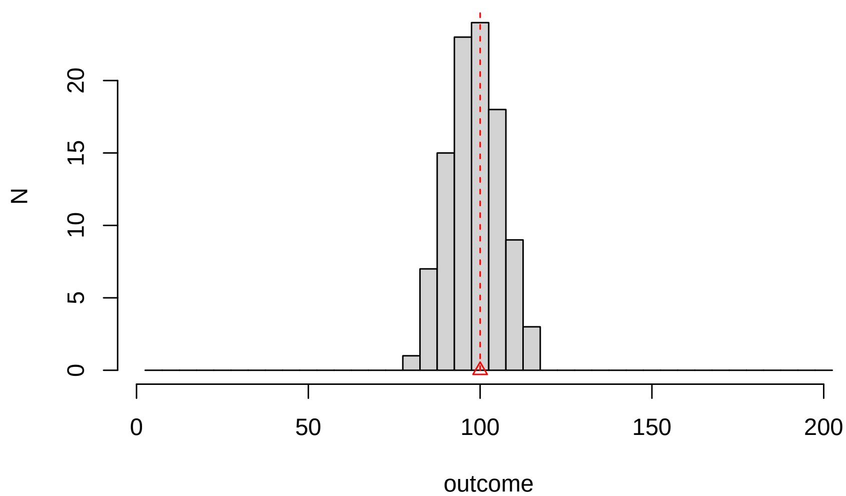
2.32 Median Vs Average

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

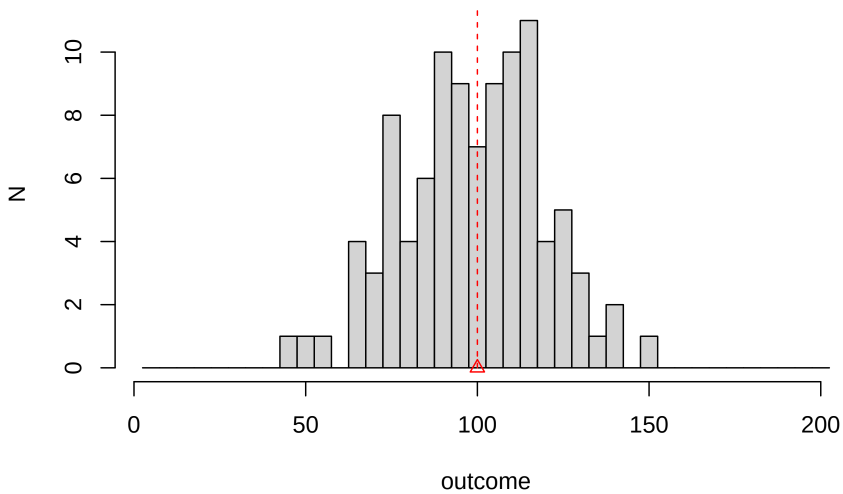


2.33 Dispersion

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

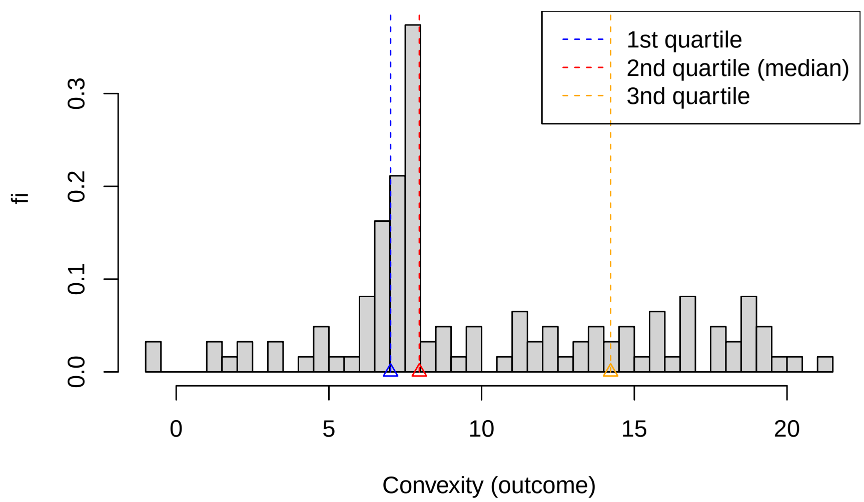
$$q_{0.25} = \sum_{x \leq x_p} f_x = F_p = 0.25$$

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

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

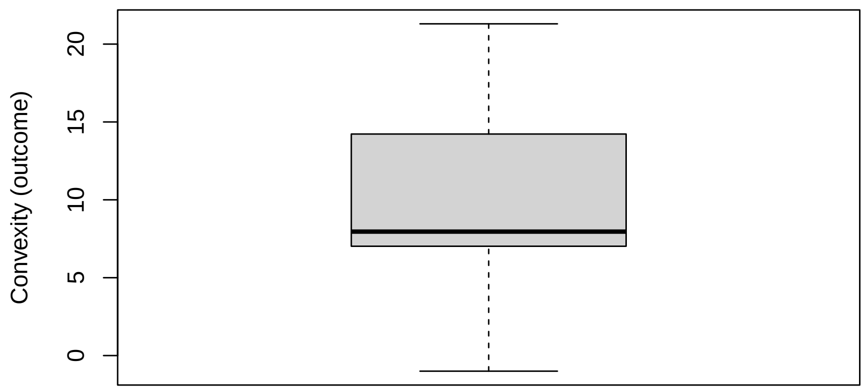
2.39 IQR

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

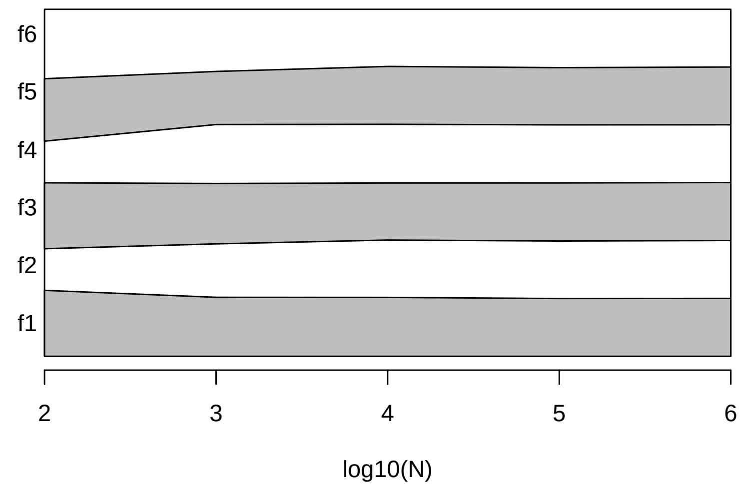
3.7 At infinity

Say we repeated the experiment 1000 times:

```
##   outcome  ni    fi
## 1       1 172 0.172
## 2       2 187 0.187
## 3       3 166 0.166
## 4       4 167 0.167
## 5       5 159 0.159
## 6       6 149 0.149
```

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
```

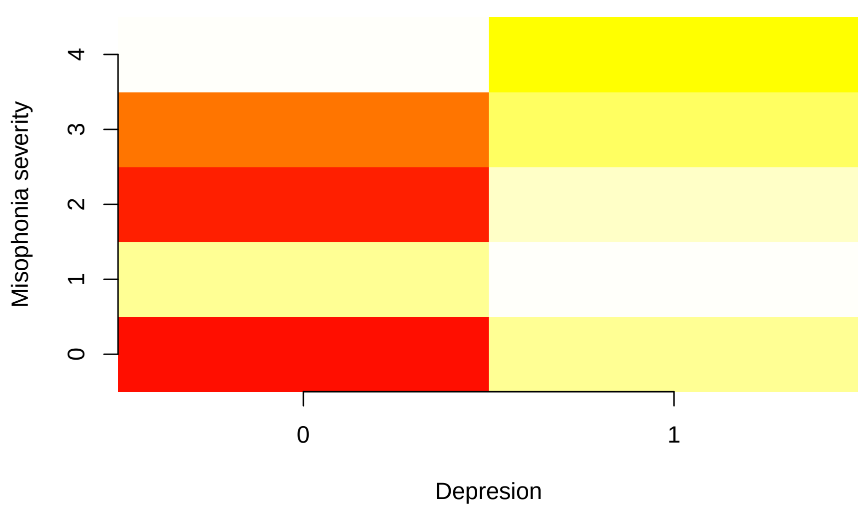
- For the relative frequencies $f_{i,j}$

```
##
```

```
##                 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
```

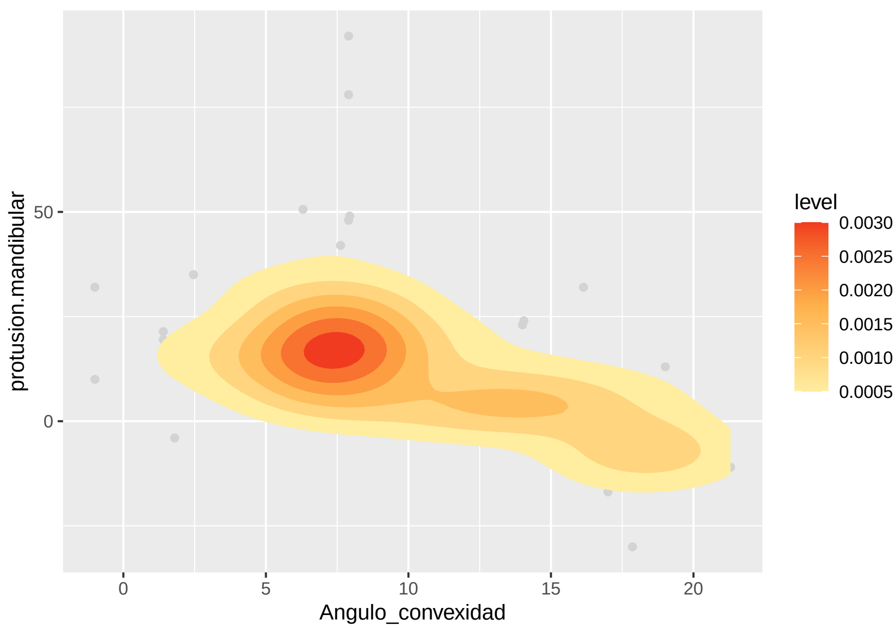
```
## 48              6.91              37.60
## 49              7.10              15.00
## 50             13.40               5.10
## 51             11.60              -2.20
## 52             -1.00              32.00
## 53              6.00              25.00
## 54              7.82              24.00
## 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

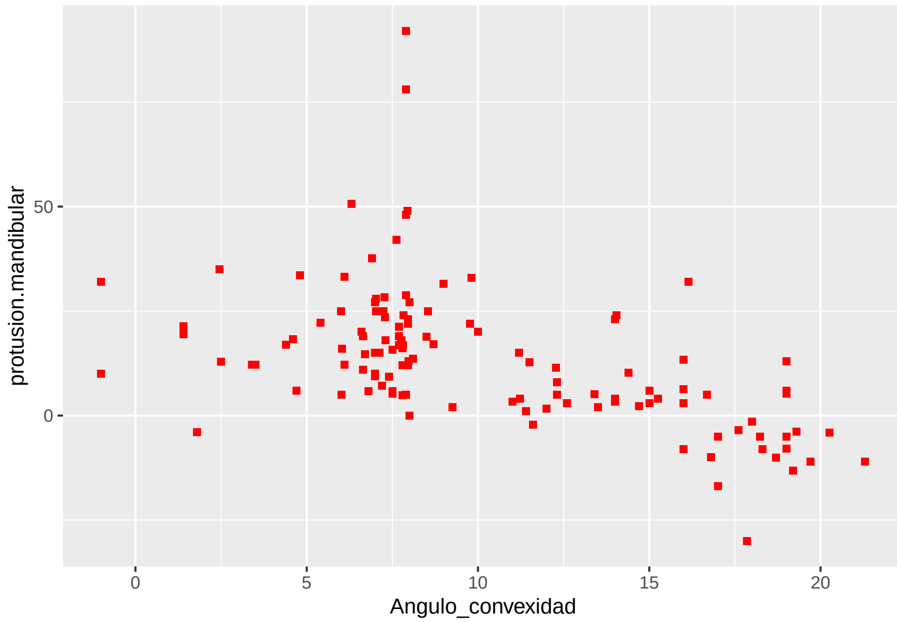


3.33 Scatter plot

- 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:

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

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

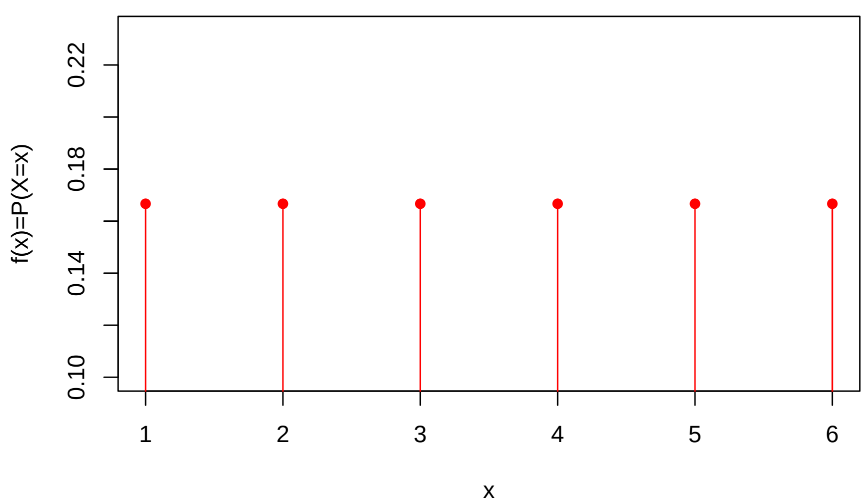
We are interested in assigning probabilities to the values of a random variable.

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

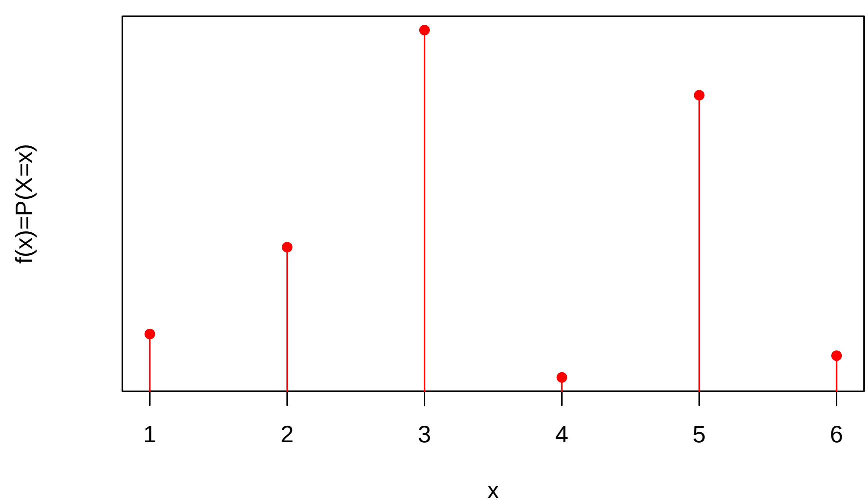


- or write as the function

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

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	$P(X = 2) = 1/6$
3	$P(X = 3) = 1/6$

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

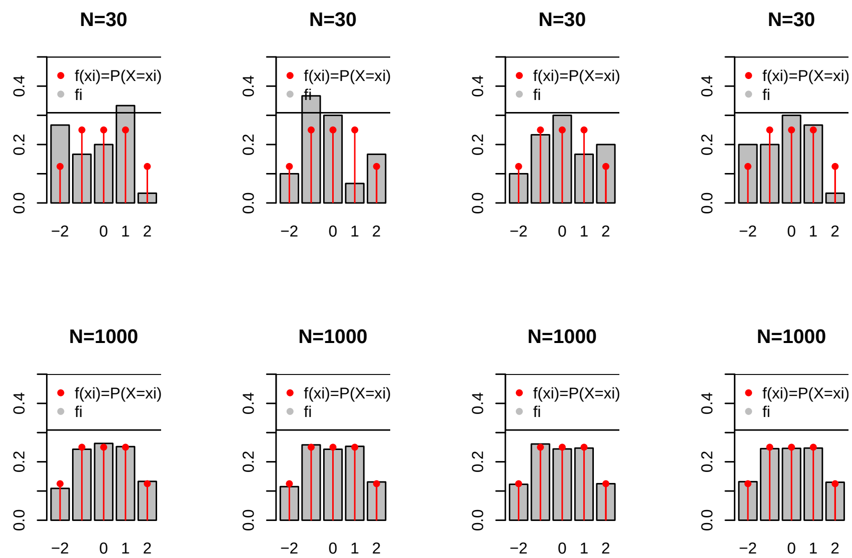
- **repeat** the experiment N times (you have to put the ball back in the urn each time) and at the end compute

$$f_i = n_i/N$$

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

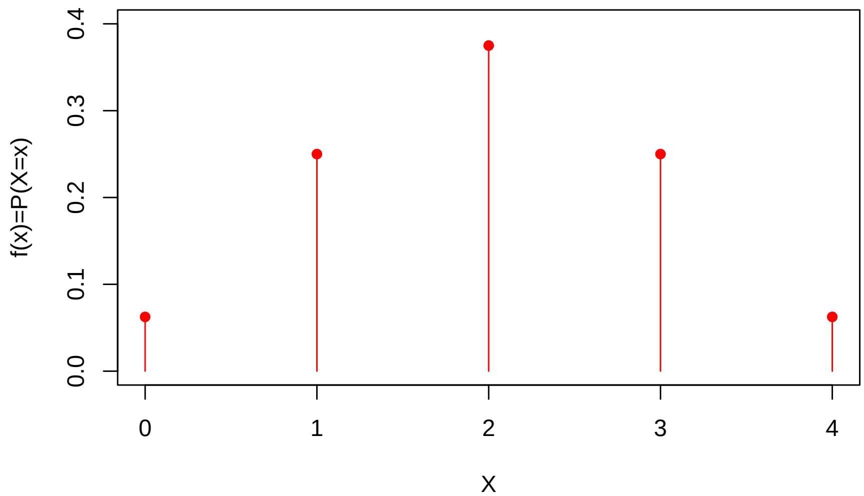
$$\mu = E(X) = \sum_{i=1}^{M} x_i f(x_i)$$

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 Mean and Average

- The mean μ is the centre of grevity of the probability mass function **it does not change**

For instance for

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	$1/8 = 0.125$

- The average $\bar{x}$ is the centre of gravity of the observations (relative frequencies) it **changes** with different data

For instance for

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

5.23 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.

5.24 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

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

$$\sigma = 1$$

5.25 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:

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

5.26 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:

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

$5 = 1 + 2^2$

5.27 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 $F(x) \leq F(y)$

5.28 Example: Probability distribution

For the probability mass function:

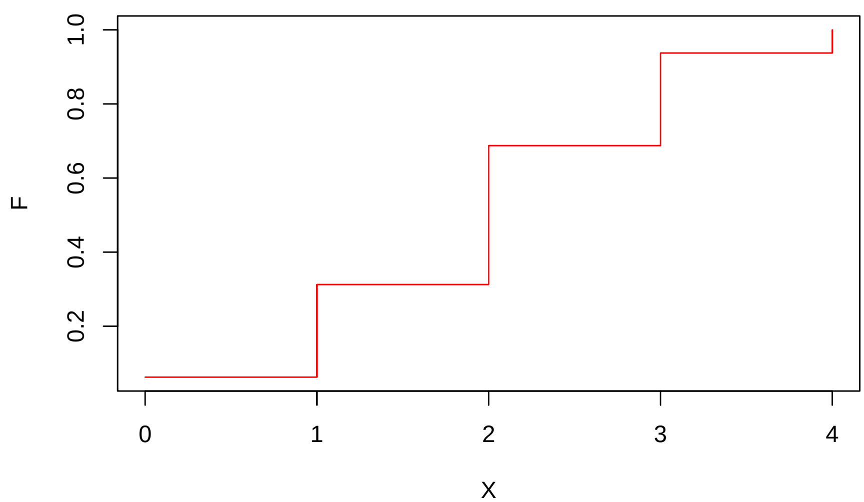
$f(0) = P(X = 0) = 1/16$ $f(1) = P(X = 1) = 4/16$ $f(2) = P(X = 2) = 6/16$ $f(3) = P(X = 3) = 4/16$ $f(4) = P(X = 4) = 1/16$

The probability distribution is:

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

For$X \in \mathbb{Z}$

5.29 Probability distribution



5.30 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.

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

5.31 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 $x_1 \leq x_2 \leq ... \leq x_n$

5.32 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 0.25-quantile is **first quartile** the value x_s such that $q = 0.25$

$$F(x_s) = 0.25$$

5.33 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?*

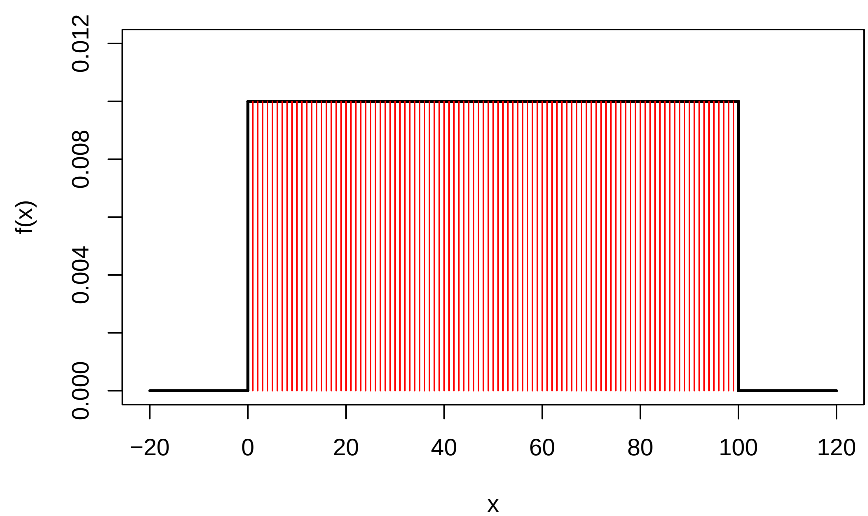
6.7 Total area under the curve

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**

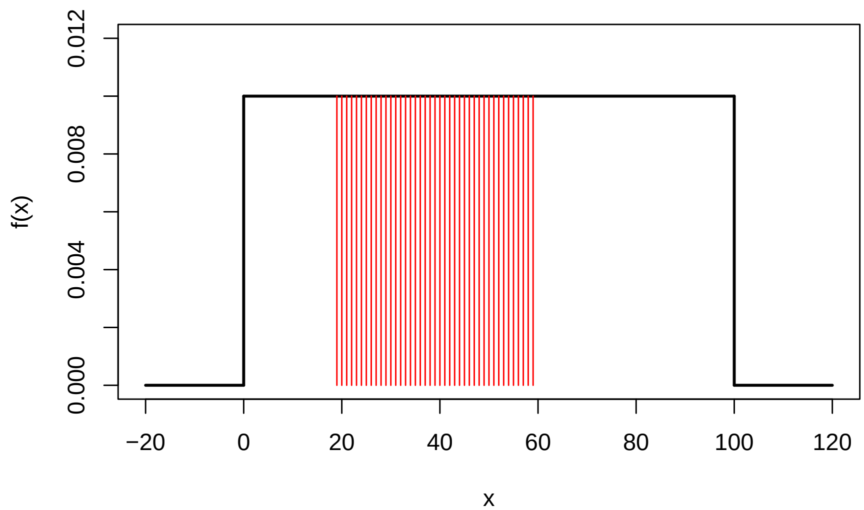
$P(-\infty \leq X \leq \infty) = \int_{-\infty}^{\infty} f(x)dx = 100 * 0.01 = 1$



6.8 Area under the curve

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

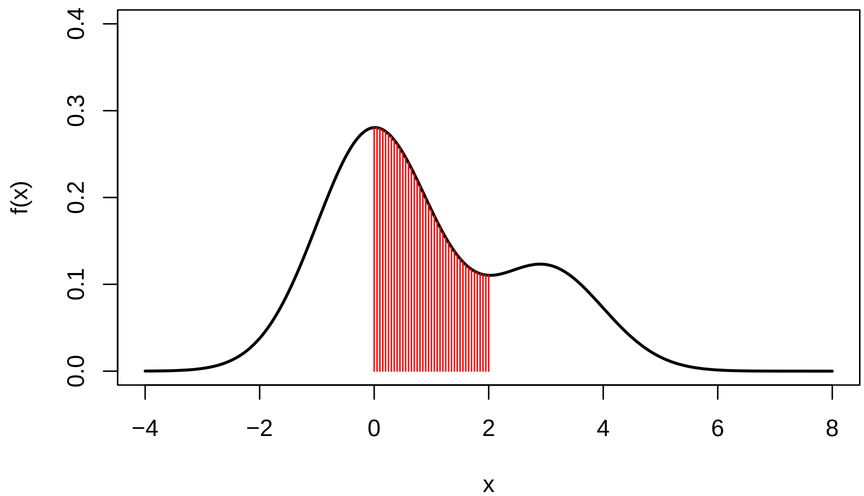
- $P(20 \leq X \leq 60) = \int_{20}^{60} f(x)dx = (60 - 20) * 0.01 = 0.4$



6.9 Area under the curve

In general $f(x)$ should satisfy:

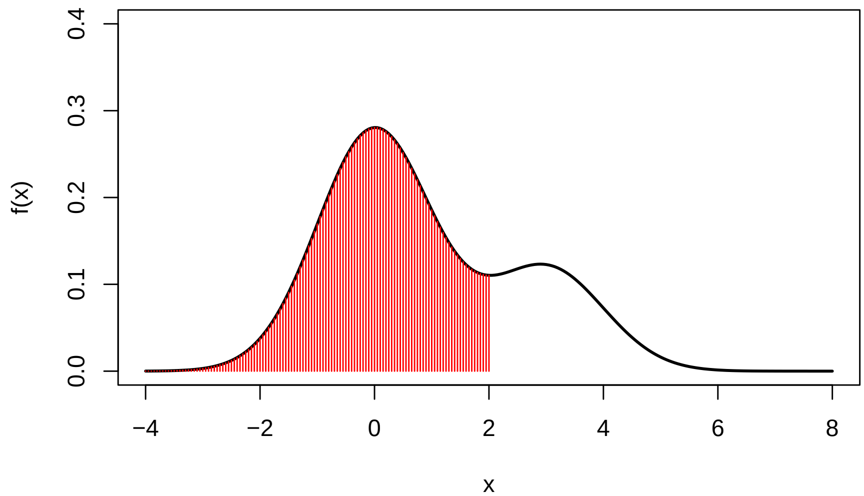
- $0 \leq P(a \leq X \leq b) = \int_a^b f(x)dx \leq 1$



6.10 Probability distribution

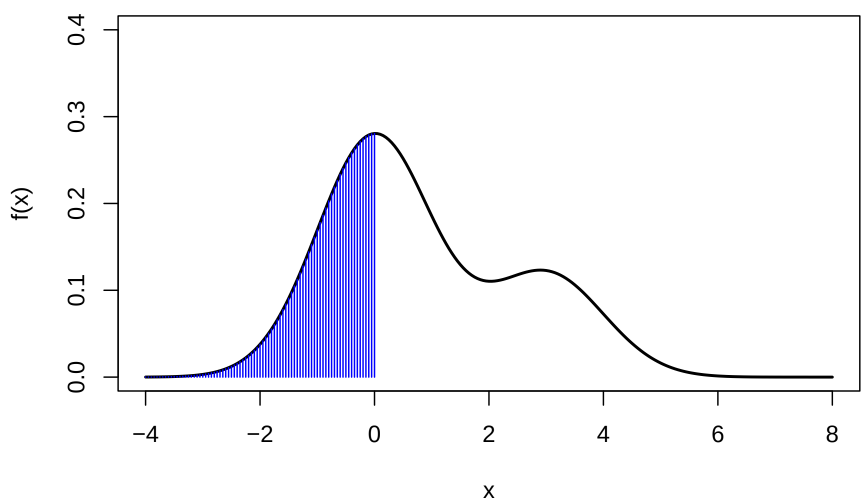
The probability accumulated up to b is defined by the probability distribution F

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



The probability accumulated up to a is

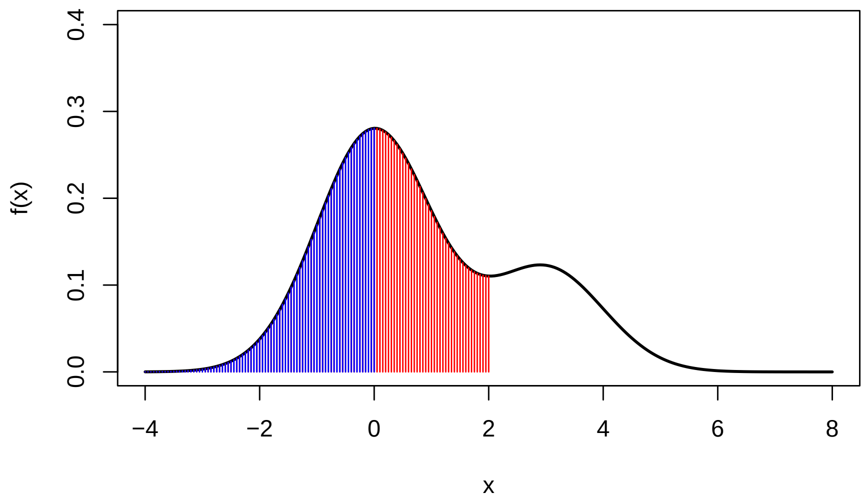
- $F(a) = P(X \leq a)$



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:

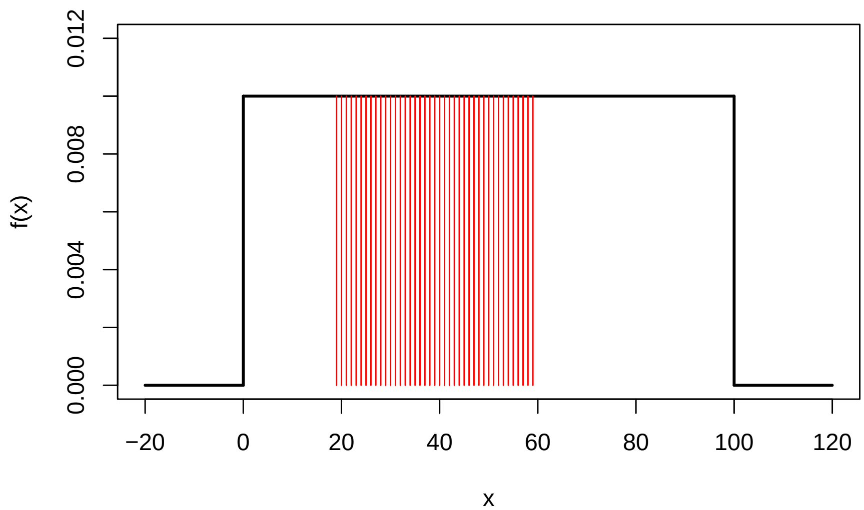
$$f(x) = \begin{cases} \frac{1}{100}, & \text{if } x \in (0, 100) \\ 0, & otherwise \end{cases}$$

The probability distribution is

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

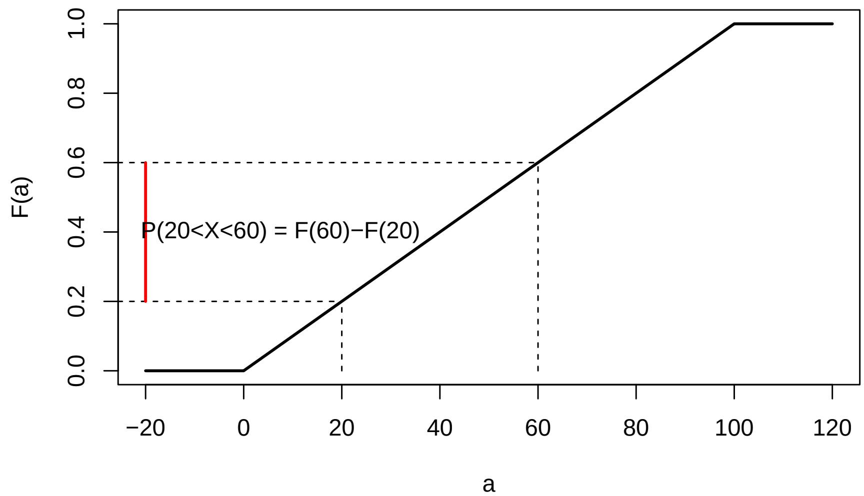
6.14 Probability graphics

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

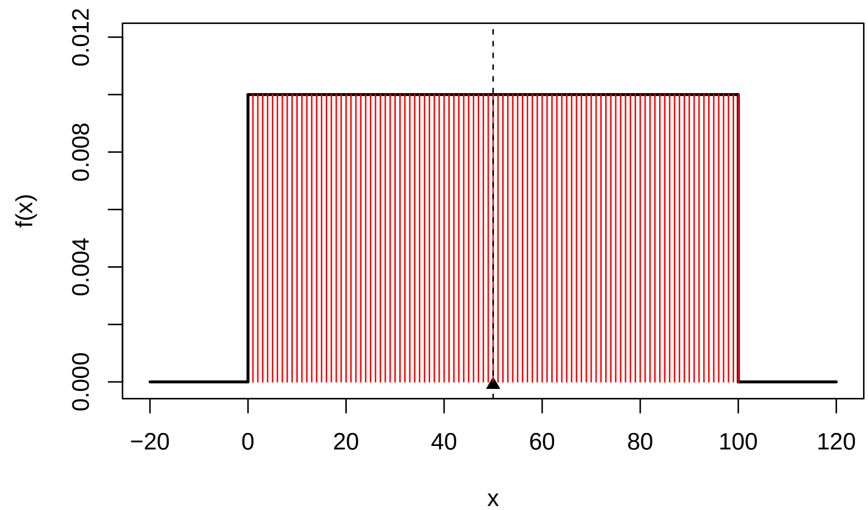
$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

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:

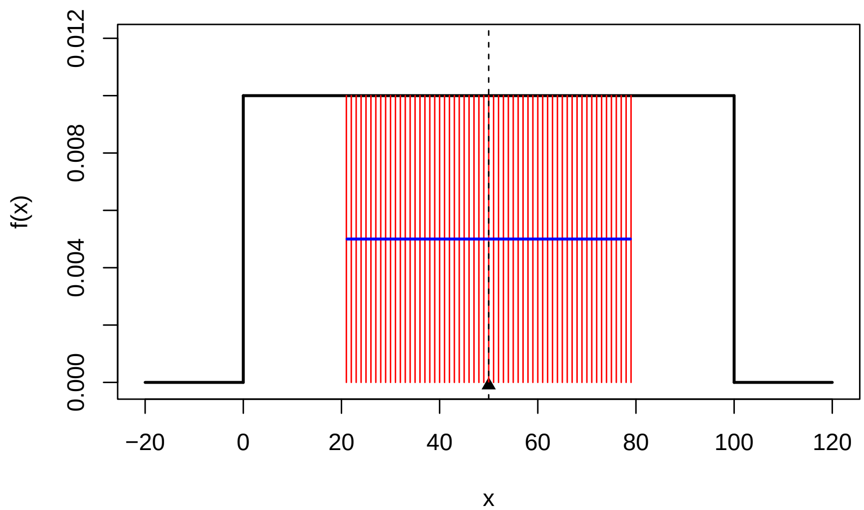
$$E(X^2) = V(X) + E(X)^2$$

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?



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

Chapter 7

Discrete Probability Models

7.1 Objective

Discrete probability models:

- Uniform and Bernoulli probability functions
- Binomial and negative binomial probability functions

7.2 Probability mass function

A probability mass function of a **discrete random variable** X with possible values $x_1, x_2, .., x_M$ is **any function** such that

Positive:

- $f(x_i) \geq 0$

Allow us to compute probabilities:

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

The probability of any outcome is 1

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

Properties:

Central tendency:

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

Dispersion:

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

They are abstract objects with general properties that may or may not **describe** a natural or engineered process.

7.3 Probability model

A **probability model** is a probability mass function that may represent the probabilities of a random experiment.

Examples:

- $f(x) = P(X = x) = 1/6$ represents the probability of the outcomes of **one** throw of a dice.
- The probability mass function

X	$f(x)$
-2	$1/8$
-1	$2/8$
0	$2/8$
1	$2/8$
2	$1/8$

Represents the probability of drawing **one** ball from an urn where there are two balls per label: $-1, 0, 1$ and one ball per label: $-2, 2$.

7.4 Parametric models

When we perform a random experiment and **do not** know the probabilities of the outcomes:

- We can always formulate the model given by the relative frequencies: $\hat{P}(X = x_i) = f_i$ (where $i = 1...M$).

We need to find M numbers each depending on N.

In many cases:

- We can formulate probability functions $f(x)$ that depend on **very few** numbers only.

Example:

A random experiment with M equally likely outcomes has a probability mass function:

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

We only need to know M.

The numbers we **need to know** to fully determine a probability function are called **parameters**.

7.5 Uniform distribution (one parameter)

Definition A random variable X with outcomes $\{1, ...M\}$ has a discrete **uniform distribution** if all its M outcomes have the same probability

$$f(x) = \frac{1}{M}$$

With mean and variance:

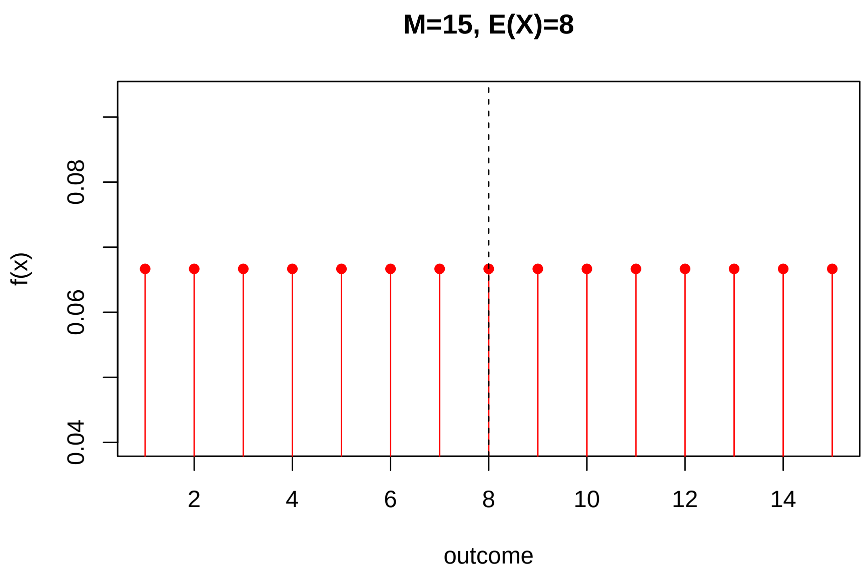
$E(X) = \frac{M+1}{2}$

$V(X) = \frac{M^2-1}{12}$

Note: $E(X)$ and $V(X)$ are also **parameters**. If we know any of them then we can fully determine the distribution.

$$f(x) = \frac{1}{2E(X) - 1}$$

7.6 Uniform distribution



7.7 Uniform distribution (two parameters)

Let's introduce a new uniform probability model with **two parameters**: The minimum and maximum outcomes.

If the random variable takes values in $\{a, a+1, ...b\}$, where a and b are integers and all the outcomes are equally probable then

$$f(x) = \frac{1}{b-a+1}$$

as $M = b - a + 1$.

- We then say that X distributes uniformly between a and b and write

$$X \rightarrow Unif(a, b)$$

7.8 Uniform distribution (two parameters)

Example:

What is the probability of observing a child of a particular age in a primary school (if all classes have the same amount of children)?

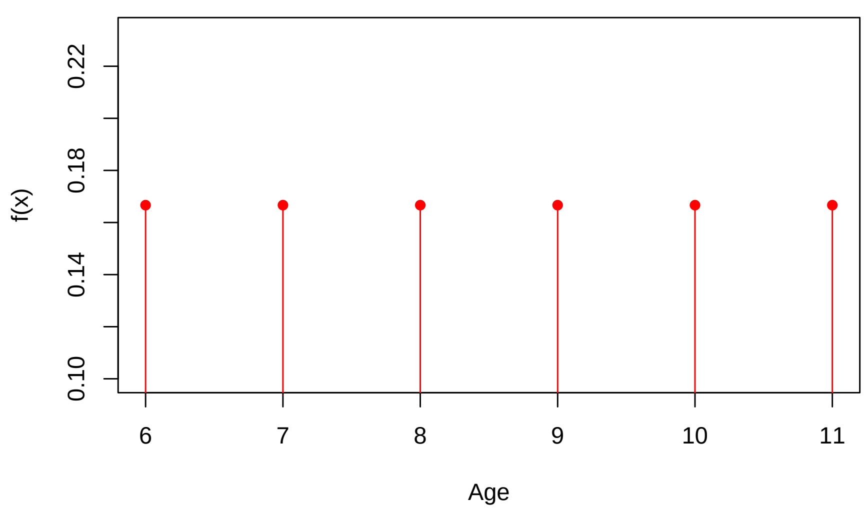
From the experiment we know: $a = 6$ and $b = 11$ then

$$X \to Unif(a = 6, b = 11)$$

that is

$$f(x) = \frac{1}{6}$$

for $x \in \{6, 7, 8, 9, 10, 11\}$, and 0 otherwise



7.9 Uniform distribution

The probability model of a random variable X

$$f(x) = \frac{1}{b-a+1}$$

for $x \in \{a, a+1, ...b\}$

has mean and variance:

- $E(X) = \frac{b+a}{2}$
- $V(X) = \frac{(b-a+1)^2-1}{12}$

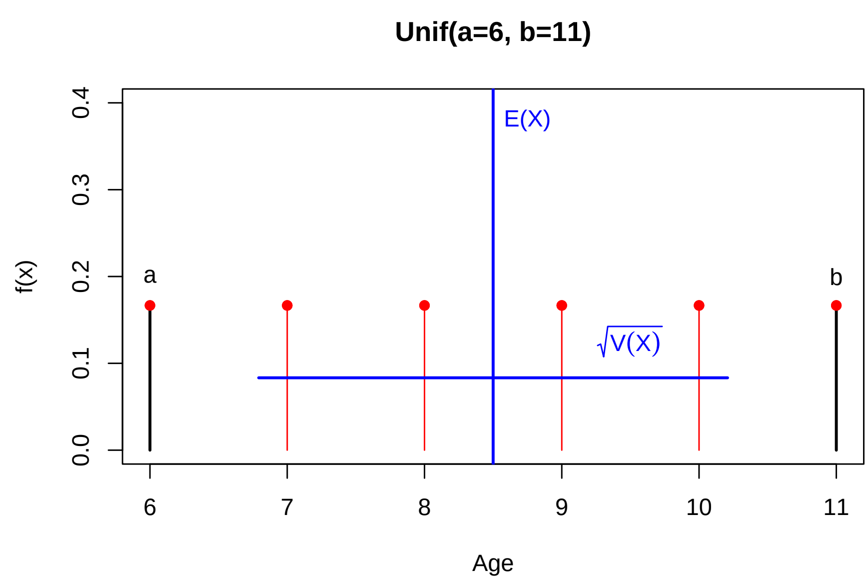
(Change variables $X = Y + a - 1$, $y \in \{1, ...M\}$)

We can either specify a and b or $E(X)$ and $V(X)$.

In our example:

- $E(X) = (11+6)/2 = 8.5$
- $V(X) = (6^2 - 1)/12 = 2.916667$

7.10 Uniform distribution (two-parameter)



7.11 Parameters and Models

- A **model** is a particular function $f(x)$ that **describes** our experiment

- If the model is a **known** function that depends on a few parameters then changing the value of the parameters we produce a **family of models**

- Knowledge of $f(x)$ is reduced to the knowledge of the value of the parameters

- Ideally, the model and the parameters are **interpretable**

Example:

Model: The data of our experiment is produced by a random process in which each age has the **same probability** of being observed.

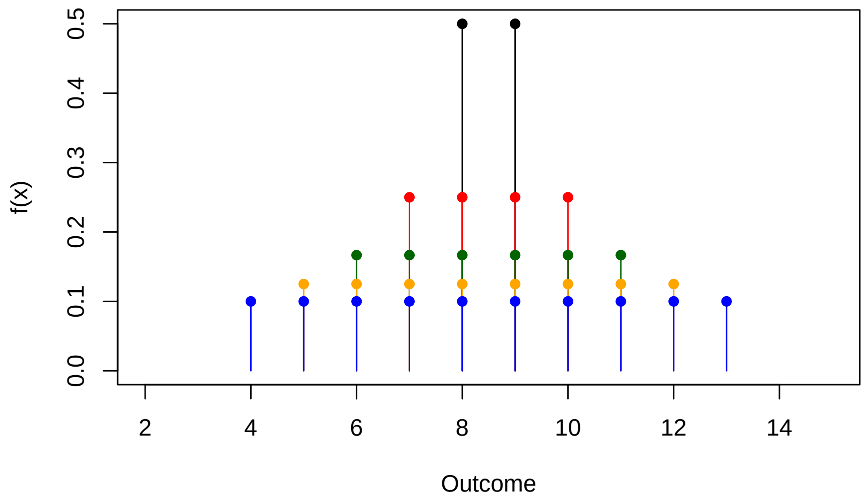
Parameters: a is the minimum age, $E(X)$ is the expected age ... they are **physical properties** of the experiment.

7.12 Parameters and Models

Example:

A **family** of models obtained from two-parameter uniform distributions changing the **variances** and keeping a constant mean ($E(X) = 8.5$). It results on **changing** both **minimum** and **maximum** outcomes.

- Note: Only one model makes sense for our experiment (only one model can represent the ages of children in a school).



- We can think of **families** that change only the **mean**, only the **minimum**, or only the **maximum**

7.13 Bernoulli trial

Let's try to advance from the equal probability case and suppose a model with two outcomes (A and B) that have **unequal** probabilities

Examples:

- Writing down the sex of a patient who goes into an emergency room of a hospital ($A : male$ and $B : female$).
- Recording whether a manufactured machine is defective or not ($A : defective$ and $B : fine$).
- Hitting a target ($A : success$ and $B : failure$).
- Transmitting one pixel correctly ($A : yes$ and $B : no$).

In these examples, the probability of outcome A is usually **unknown**.

7.14 Bernoulli trial

We will introduce the probability of an outcome (A) as the **parameter** of the model:

- outcome A (success): has probability p (parameter)
- outcome B (failure): has a probability $1-p$

Or write, the probability mass function of K taking values $\{0,1\}$ for A and B

$$f(k) = \begin{cases} 1-p, & k = 0 \, (event \, B) \\ p, & k = 1 \, (event \, A) \end{cases}$$

or more shortly

$$f(k;p) = p^k (1-p)^{1-k}$$

for $k = (0,1)$

We only need to know p.

7.15 Bernoulli trial

A Bernoulli variable K with outcomes $\{0,1\}$ has a probability mass function

$$f(k;p) = p^k (1-p)^{1-k}$$

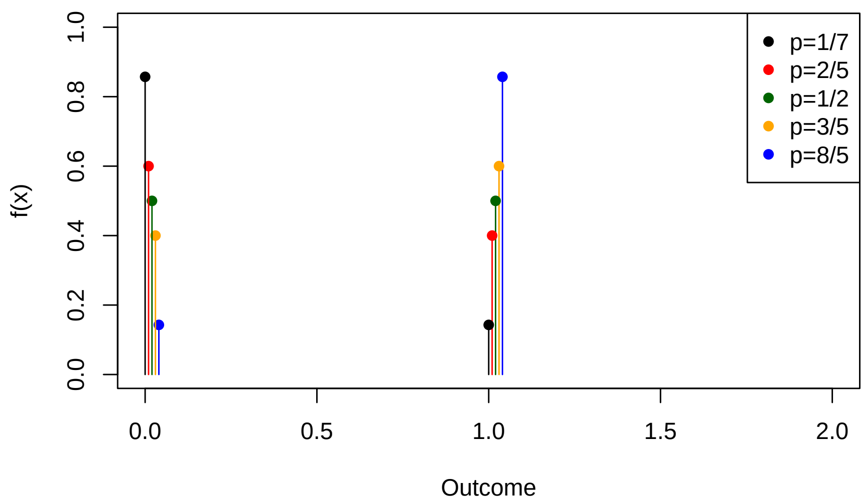
With mean and variance:

- $E(K) = p$
- $V(K) = (1-p)p$

Note:

- The probability of the outcome A is the parameter p which is the same as $f(0) = P(X = 0)$.
- As p is usually **unknown** we typically estimated it by the relative frequency (more on this in the inference sections): $\hat{p} = f_A = \frac{n_A}{N}$

7.16 Bernoulli trial



7.17 Binomial distribution

When we are interested in learning about a particular Bernoulli trial

- We repeat the Bernoulli trial N times and count how many times we obtained A (n_A).
- We define a random variable $X = n_A$ taking values $x \in 0, 1, ...N$

We now ask for the probability of observing x events of type A in the repetition of n independent Bernoulli trials, when the probability of observing A is p.

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

7.18 Examples: Binomial distribution

- Writing down the sex of $n = 10$ patients who go into an emergency room of a hospital. What is the probability that $x = 6$ patients are men when $p = 0.9$?
- Trying $n = 5$ times to hit a target (A : *success* and B : *failure*). What is the probability that I hit the target $x = 5$ times when I usually hit it 25% of the times ($p = 0.25$)?
- Transmitting $n = 100$ pixels correctly (A : *yes* and B : *no*). What is the probability that $x = 2$ pixels are errors, when the probability of error is $p = 0.1$?

7.19 Binomial distribution

What is the probability of observing $X = 4$ errors when transmitting 4 pixels, if the probability of an error is p?

Consider 4 random variables: K_1, K_2, K_3 and K_4 that record whether an error has been made in the 1^{st}, 2^{nd}, 3^{rd} and 4^{th} pixel.

Then

- k_i takes values $\{correct : 0; error : 1\}$
- $X = \sum_{i=1}^{4} K_i$ takes values $\{0, 1, 2, 3, 4\}$

Then the probability of observing 4 errors is:

- $P(X = 4) = P(1, 1, 1, 1) = p * p * p * p = p^4$ because K_i are independent.

The probability of 0 errors is:

- $P(X = 0) = P(0, 0, 0, 0) = (1 - p)(1 - p)(1 - p)(1 - p) = (1 - p)^4$

The probability of 3 errors is:

$P(X = 3) = P(0, 1, 1, 1) + P(1, 0, 1, 1) + P(1, 1, 0, 1) + P(1, 1, 1, 0) = 4p^3(1 - p)^1$

7.20 Binomial distribution

Therefore the probability of x errors is

$$f(x) = \begin{cases} 1 * p^0(1-p)^4, & x = 0 \\ 4 * p^1(1-p)^3, & x = 1 \\ 6 * p^2(1-p)^2, & x = 2 \\ 4 * p^3(1-p)^1, & x = 3 \\ 1 * p^4(1-p)^0, & x = 4 \end{cases}$$

or more shortly

$$f(x) = \binom{4}{x} p^x (1-p)^{4-x}$$

for $x = 0, 1, 2, 3, 4$

where $\binom{4}{x}$ is the number of possible outcomes (transmissions of 4 pixels) with x errors.

7.21 Binomial distribution: Definition

The binomial probability function is the probability mass function of observing x outcomes of type A in n independent Bernoulli trials, where A has the same probability p in each trial.

The function is given by

$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, ...n$

$\binom{n}{x} = \frac{n!}{x!(n-x)!}$ is called **the binomial coefficient** and gives the number of ways one can obtain x events of type A in a set of n.

When a variable X has a binomial probability function we say it distributes binomially and write

$$X \to Bin(n, p)$$

where n and p are parameters.

7.22 Binomial distribution: Mean and Variance

The mean and variance of $X \hookrightarrow Bin(n,p)$ are

- $E(X) = np$
- $V(X) = np(1-p)$
- Since X is the sum of n independent Bernoulli variables

$E(X) = E(\sum_{i=1}^{n} K_i) = np$

and

$V(X) = V(\sum_{i=1}^{n} K_i) = n(1-p)p$

Example:

- The expected value for the number of errors in the transmission of 4 pixels is $np = 4 * 0.1 = 0.4$ when the probability of an error is 0.1.
- The variance is $n(1-p)p = 0.36$

Remember: We can specify either the parameters n and p, or the parameters $E(X)$ and $V(X)$

7.23 Example 1

Now let's answer:

- What is the probability of observing 4 errors when transmitting 4 pixels, if the probability of an error is 0.1?

Since we are repeating a Bernoulli trial $n = 4$ times and counting the number of events of type A (errors), when $P(A) = p = 0.1$ then

$$X \to Bin(n = 4, p = 0.1)$$

That is

$$f(x) = \binom{4}{x} 0.1^x (1-0.1)^{4-x}$$

7.24 Example 1

- We want to compute:

$P(X=4)=f(4)=\binom{4}{4}0.1^4 0.9^0=10^{-4}$

In R dbinom(4,4,0.1)

- We can also compute:

$P(X=2)=\binom{4}{2}0.1^2 0.9^2=0.0486$

In R dbinom(2,4,0.1)

7.25 Example 2

- What is the probability of observing at most 8 voters of the ruling party in an election poll of size 10, if the probability of a positive vote is 0.9

For this case

$$X \rightarrow Bin(n=10, p=0.9)$$

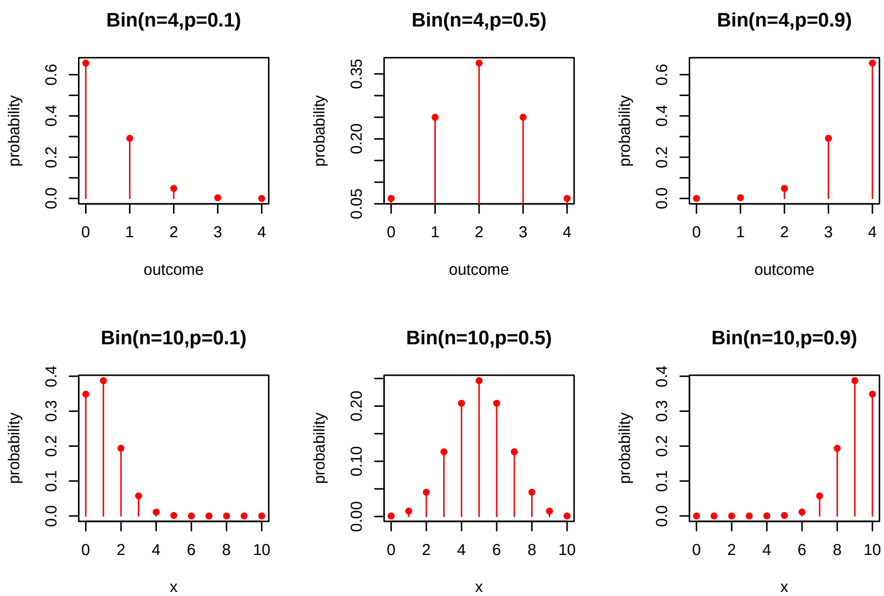
That is

$$f(x)=\binom{10}{x}0.9^x(0.1)^{4-x}$$

We want to compute: $P(X\leq 8)=F(8)=\sum_{i=1..8}f(x_i)=0.2639011$

in R pbinom(8,10, 0.9)

7.26 Binomial distribution



7.27 Negative binomial distribution

Now let us imagine that we are interested in counting the well-transmitted pixels before a **given number** of errors occur. Say we can **tolerate** r errors in transmission.

- Experiment: Suppose performing Bernoulli trials until we observe the outcome A appears r times.
- Random variable: We count the number of events B
- Example: What is the probability of observing y well-transmitted (B) pixels before r errors (A)?

7.28 Negative binomial distribution

Let's first find the probability of one particular transmission with y number of correct pixels (B) and r number of errors (A).

$(0, 0, 1, ., 0, 1, ...0, 1)$ (there are y zeros, and r ones)

We observe y correct pixels in a total of $y + r$ trials.

Then

- $P(0, 0, 1, ., 0, 1, ...0, 1) = p^r(1 - p)^y$ (Remember: p is the probability of error)

How many transmissions can have y correct pixels before r errors?

Note:

- The last bit is fixed (marks the end of transmission)
- The total number of transmissions with y number of correct pixels (B) that we can obtain in $y + r - 1$ trials is: $\binom{y+r-1}{y}$

7.29 Negative binomial distribution

Therefore, the probability of observing y events of type B before r events of type A (with probability p) is

$$P(Y = y) = f(y) = \binom{y + r - 1}{y} p^r(1 - p)^y$$

for $y = 0, 1, ...$

We then say that Y follows a negative binomial distribution and we write

$$Y \to NB(r, p)$$

where r and p are parameters representing the tolerance and the probability of a single error.

7.30 Mean and Variance

A random variable with $Y \to NB(r,p)$ has

- mean: $E(Y) = r\frac{1-p}{p}$
- variance: $V(Y) = r\frac{1-p}{p^2}$

7.31 Geometric distribution

We call **geometric distribution** to the negative binomial distribution with $r = 1$

The probability of observing B events before observing the **first** event of type A is

$$P(Y = y) = f(y) = p(1-p)^y$$

$$Y \to Geom(p)$$

with mean

- mean: $E(Y) = \frac{1-p}{p}$
- variance: $V(Y) = \frac{1-p}{p^2}$

7.32 Example

- A website has three servers.
- One server operates at a time and only when a request fails another server is used.
- If the probability of failure for a request is known to be $p = 0.0005$ then
- what is the expected number of successful requests before the three computers fail?

7.33 Example

Since we are repeating a Bernoulli trial until $r = 3$ events of type A (failure) are observed (each with $P(A) = p = 0.0005$) and are counting the number of events of type B (successful requests) then

$$Y \to NB(r = 3, p = 0.0005)$$

Therefore, the expected number of requests before the system fails is:

$E(Y) = r\frac{1-p}{p} = 3\frac{1-0.0005}{0.0005} = 5997$

- Note that there are actually 6000 trials

7.34 Example

What is the probability of dealing with at most 5 successful requests before the system fails?

Recall the cumulative function distribution $F(y) = P(Y \le 5)$

$F(5) = P(Y \le 5) = \Sigma_{y=0}^{5} f(y)$

$= \sum_{y=0}^{5} \binom{y+2}{y} 0.0005^r 0.9995^y$

$= \binom{2}{0} 0.0005^3 0.9995^0 + \binom{3}{1} 0.0005^3 0.9995^1$

$+ \binom{4}{2} 0.0005^3 0.9995^2 + \binom{5}{3} 0.0005^3 0.9995^3$

$+ \binom{6}{4} 0.0005^3 0.9995^4 + \binom{7}{5} 0.0005^3 0.9995^5$

$= 6.9 \times 10^{-9}$

In R pnbinom(5,3,0.0005)

7.35 Examples

With the negative binomial probability function:

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

We can now answer questions like:

- What is the probability of observing 10 correct pixels before 2 errors, if the probability of an error is 0.1?

$f(10; r = 2, p = 0.1) = 0.03835463$

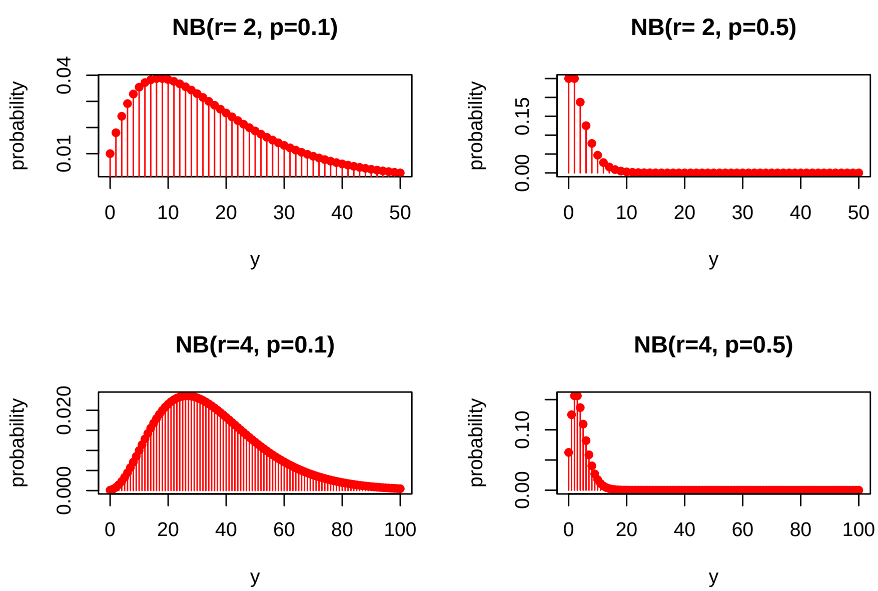
in R dnbinom(10, 2, 0.1)

- What is the probability that 2 girls enter the class before 4 boys if the probability that a girl enters is 0.5?

$f(2; r = 4, p = 0.5) = 0.15625$

in R dnbinom(2, 4, 0.5)

7.36 Negative binomial distribution



Chapter 8

Poisson and Exponential Models

8.1 Objective

Discrete probability model:

- Poisson

Continuous probability model:

- Exponential

8.2 Discrete probability models

We are building up more complex models from simple ones:

Uniform: Classical interpretation of probability ↓ **Bernoulli**: Introduction of a **parameter** p (family of models) ↓ **Binomial**: **Repetition** of a random experiment (n-times Bernoulli trials) ↓ **Poisson**: Repetition of random experiment within a continuous interval, having **no control** on when/where the Bernouilli trial occurs.

8.3 Counting events

Imagine that we are observing events that **depend** on time or distance **intervals**.

- cars arriving at a traffic light
- getting messages on your mobile phone
- impurities occurring at random in a copper wire

Suppose that the events are outcomes of **independent** Bernoulli trials each appearing randomly on a continuous interval, and we want to **count** them.

8.4 Counting events

What is the probability of observing X events in an interval's unit (time or distance)?

Imagine that some impurities in a copper wire deposit randomly along a wire

- at each centimeter, you would count an average of $\lambda = 10/cm$.
- divide the centimeter into micrometers ($0.0001cm$)

8.5 Poisson distribution

micrometers are small enough so

- either there is or there is not an impurity in each micrometer
- each micrometer can be considered a **Bernoulli trial**

8.6 Poisson distribution

The probability of observing X impurities in $n = 10,000\mu$ (1cm) approximately follows a binomial distribution

$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$

where p is the probability of finding an impurity in a micrometer.

Remember that $E(X) = np$ so for $\lambda = np$ (average number of impurities per 1cm), we can write

$$P(X = x) = \binom{n}{x}(\frac{\lambda}{n})^x(1 - \frac{\lambda}{n})^{n-x}$$

- There **could** still be two impurities in a micrometer so we need to increase the partition of the wire and $n \to \infty$.

Then in the limit:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Where λ is constant because it is the density of impurities per centimeter, a **physical property** of the system.

8.7 Poisson distribution: Derivation details

For $P(X = x) = \binom{n}{x}(\frac{\lambda}{n})^x(1 - \frac{\lambda}{n})^{n-x}$

in the limit ($n \to \infty$)

- $\frac{1}{n^x}\binom{n}{x} = \frac{1}{n^x}\frac{n!}{x!(n-x)!} = \frac{(n-x)!(n-x+1)...(n-1)n}{n^x x!(n-x)!} = \frac{n(n-1)..(n-x+1)}{n^x x!} \to \frac{1}{x!}$
- $(1 - \frac{\lambda}{n})^n \to e^{-\lambda}$ (definition of exponential)
- $(1 - \frac{\lambda}{n})^{-x} \to 1$

Therefore $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$

8.8 Poisson distribution

Definition

Given

- an interval in the real numbers
- counts occur at random in the interval
- the average number of counts on the interval is known (λ)
- if one can find a small regular partition of the interval such that each of them can be considered Bernoulli trials

Then…

8.9 Poisson distribution

Definition

The random variable X that counts events across the interval is a **Poisson** variable with probability mass function

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \lambda > 0$$

Properties:

- mean $E(X) = \lambda$
- variance $V(X) = \lambda$

8.10 Poisson distribution

With the Poisson probability function:

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

for $x \in \{0, 1, ...\}$

We can now answer questions like:

- What is the probability of receiving 4 emails in an hour, when the average number of emails in **one** hour is 1?

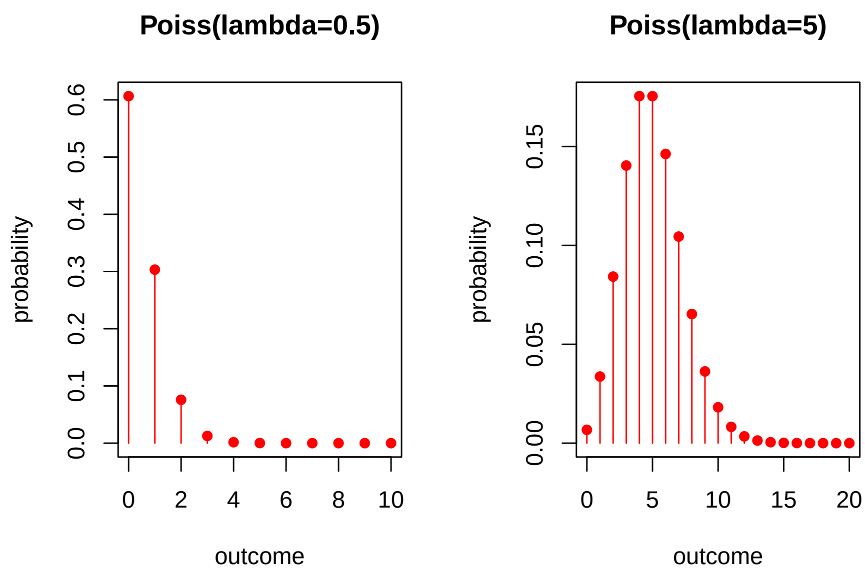
$f(4; \lambda = 1) = 0.18$

in R dpois(2,1)

- What is the probability of counting at least 10 cars arriving at a road toll in a minute, when the average number of cars that arrive at the toll in a minute is 5; $P(X \leq 10) = F(10; \lambda = 5) = 0.98$?

in R ppois(10,5)

8.11 Poisson distribution



8.12 Continuous probability models

Continuous probability models are probability density functions $f(x)$ of a continuous random variables that we **believe** describe real random experiments.

Definition:

Positive:

- $f(x) \geq 0$

Allows us to compute probabilities using the area under the curve:

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

The probability of any value is 1:

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

8.13 Exponential density

Let's go back to the Poisson probability for the number of events (k) in an interval

$$f(k) = \frac{e^{-\lambda}\lambda^k}{k!}, \lambda > 0$$

- Let's now consider only the first event
- the distance/time we have to wait until the fisrt event is a **continuous** random variable.

We can ask for the probability that the first event is at distance X.

8.14 Exponential density

The probability of observing 0 events **if** an interval has unit x is

$$f(0|x) = \frac{e^{-x\lambda}x\lambda^0}{0!}$$

or

$$f(0|x) = e^{-x\lambda}$$

We can treat this as the conditional probability of 0 events in a distance x: $f(K = 0|X = x)$ and apply the Bayes theorem to reverse it:

$$f(x|0) = Cf(0|x) = Ce^{-x\lambda}$$

So we can calculate the **probability of observing a distance** x with 0 events (this is the distance until the first event or the distance between any two events).

8.15 Exponential density

In a Poisson process with parameter λ the probability of waiting a distance/time X until the first event has a **probability density**

$$f(x) = Ce^{-x\lambda}$$

- C is a constant that ensures: $\int_{-\infty}^{\infty} f(x)dx = 1$
- by integration $C = \lambda$

Therefore

$$f(x) = \lambda e^{-\lambda x}$$

8.16 Exponential density

An exponential random variable X has a probability density

$$f(x) = \lambda e^{-\lambda x}, x \geq 0$$

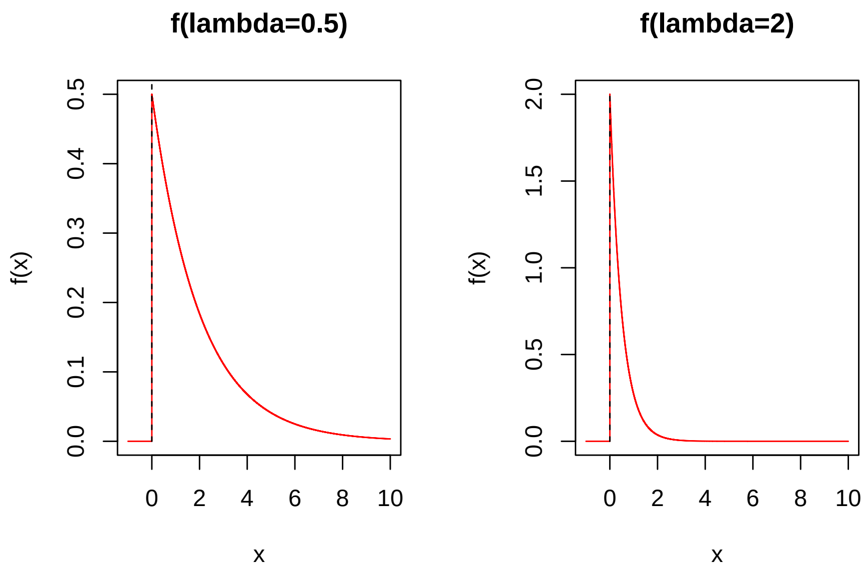
Properties:

- Mean: $E(X) = \frac{1}{\lambda}$
- Variance: $V(Y) = \frac{1}{\lambda^2}$

Where λ is its single parameter, known as a **decay rate**.

Note: The exponential model is a general model. It can describe the time/length until the first count in a Poisson process of the size of a whole made by a drill.

8.17 Exponential density



8.18 Exponential Distribution

In a Poisson process: ¿What is the probability of observing distance **smaller** than a until the first event?

Remember that this probability $F(a) = P(X \leq a)$ is the probability density

$$F(a) = \lambda \int_{\infty}^{a} e^{-x\lambda} dx = 1 - e^{-a\lambda}$$

- ¿What is the probability of observing a distance **larger** than a until the first event?

$$P(X > a) = 1 - P(X \leq a) = 1 - F(a) = e^{-a\lambda}$$

8.19 Exponential Distribution

With the exponential density function:

$$f(x) = \lambda e^{-\lambda x}$$

We can answer questions like:

- What is the probability that we have to wait for a bus for more than 1 hour when on average there are two buses per hour?

$$P(X > 1) = 1 - P(X \leq 1) = 1 - F(1, \lambda = 2) = 0.1353$$

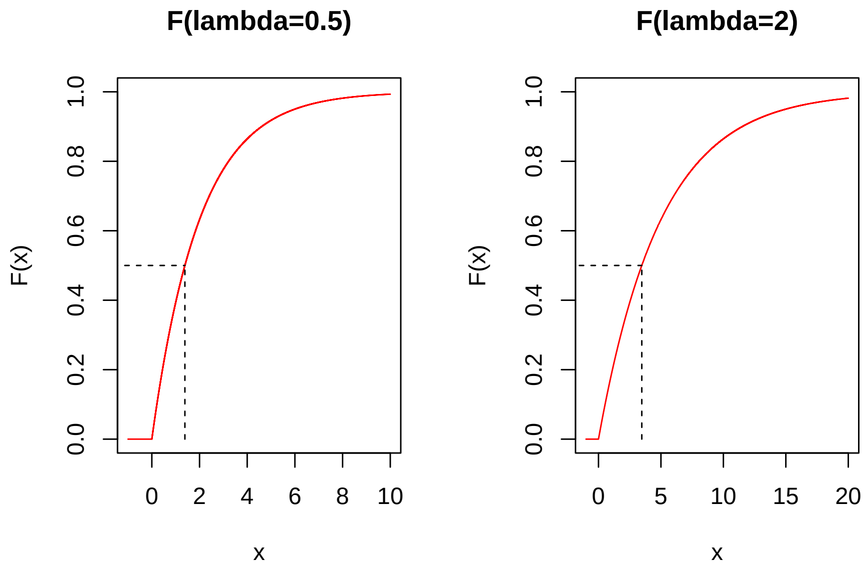
in R 1-pexp(1,2)

- What is the probability of having to wait less than 2 seconds to detect one particle when the radioactive decay rate is 2 particles each second; $F(2, \lambda = 2)$

$$P(X \leq 2) = F(2, \lambda = 2) = 0.981$$

in R pexp(2,2)

8.20 Exponential Distribution



The median x_m is such that $F(x_m) = 0.5$. That is $x_m = \frac{\log(2)}{\lambda}$

Chapter 9

Normal Distribution

9.1 Objective

Continuous probability model:

- Normal distribution

9.2 Continuous probability models

Continuous probability models are probability density functions $f(x)$ of a continuous random variables that we **believe** describe real random experiments.

Definition:

Positive:

- $f(x) \geq 0$

Allows us to compute probabilities using the area under the curve:

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

The probability of any value is 1:

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

9.3 Normal density

In 1801 Gauss analyzed the orbit of Ceres (large asteroid between Mars and Jupiter).

- People suspected it was a new planet.
- The measurements had errors.
- He was interested in finding how the observations were distributed so he could find the most probable orbit.
- He wanted to predict where astronomers should point their telescopes to find it a few months after it had passed behind the Sun.

9.4 Normal density

Errors due to measurement.

9.5 Normal density

He assumed that

- small errors were more likely than large errors
- error at a distance $-\epsilon$ or ϵ from the most likely measurement were equally likely
- the most **likely** altitude of Ceres at a given time in the sky was the **average** of multiple altitude measurements at that latitude.

9.6 Normal density

That was enough to show that the random deviations y **from the orbit** distributed like

$$f(y) = \frac{h}{\sqrt{\pi}} e^{-h^2 y^2}$$

*The evolution of the Normal distribution, Saul Stahl, Mathematics Magazine, 2006.

9.7 Normal density

Let's write the distribution of errors

$$f(y) = \frac{h}{\sqrt{\pi}} e^{-h^2 y^2}$$

for the errors of measurements from the horizon X then $y = x - x_0$

$$f(x) = \frac{h}{\sqrt{\pi}} e^{-h^2 (x - x_0)^2}$$

- The **mean** of this probability density is:

$E(X) = \mu = x_0$, that represents the **true** position of Ceres from the horizon (property of the physical system).

- The **variance** is:

$V(X) = \sigma^2 = \frac{1}{2h^2}$, that represents the dispersion of the error in the observations (property of the measurement system).

9.8 Definition

A random variable X defined in the real numbers has a **Normal** density if it takes the form

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$$

with mean and variance:

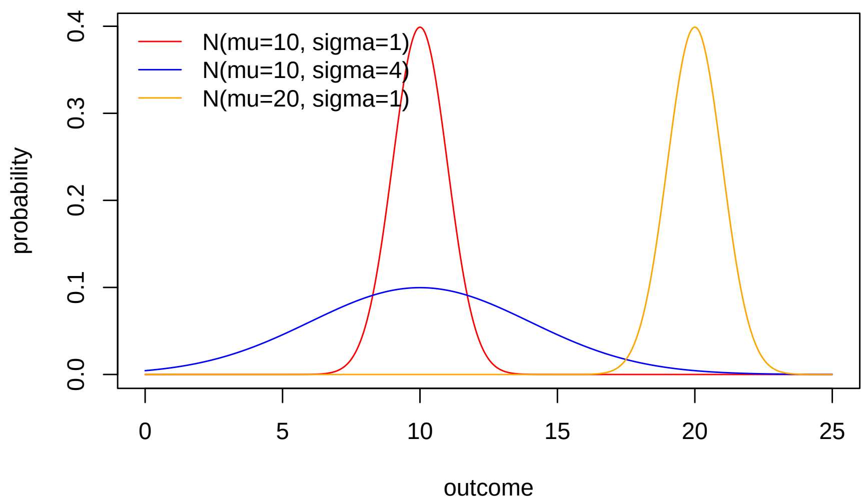
- $E(X) = \mu$
- $V(X) = \sigma^2$

μ and σ are the **two parameters** that fully describe the normal density function and their **interpretation** depends on the random experiment.

When X follows a Normal density, i.e. distributes normally, we write

$$X \to N(\mu, \sigma^2)$$

9.9 Normal probability density (Gaussian)



9.10 Normal distribution

The probability distribution of the Normal density:

$$F_{normal}(a) = P(Z \leq a)$$

is the **error** function defined by the area under the curve from $-\infty$ to a

$$F_{normal}(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

The function is found in most computer programs.

9.11 Normal distribution

When

$$X \to N(\mu, \sigma^2)$$

We can ask questions like:

- What is the probability that a woman in the population is at most $150cm$ tall if women have a mean height of $165cm$ with standard deviation of $8cm$?

$P(X \leq 150) = F(150, \mu = 165, \sigma = 8) = 0.03039636$

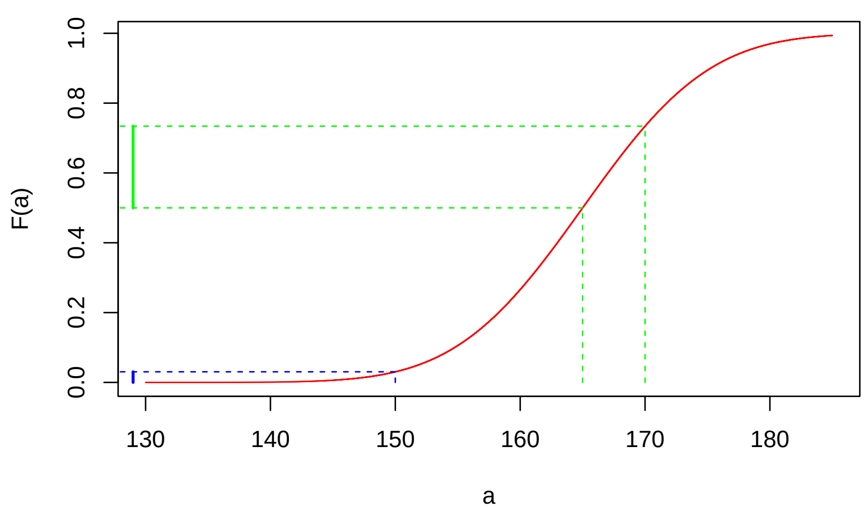
in R pnorm(150, 165, 8)

- What is the probability that a woman's height in the population is between $165cm$ and $170cm$?

$P(165 \leq X \leq 170) = F(170, \mu = 165, \sigma = 8) - F(165, \mu = 165, \sigma = 8) = 0.2340145$

in R pnorm(170, 165, 8)-pnorm(165, 165, 8)

9.12 Normal distribution



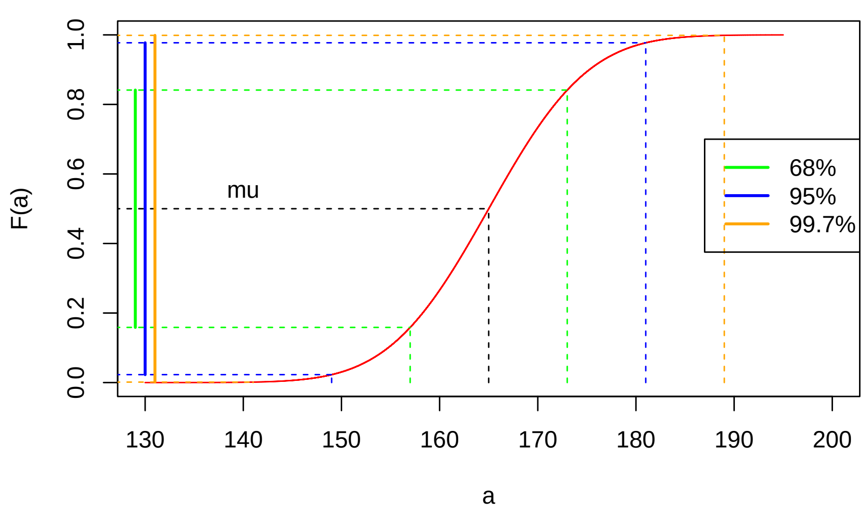
9.13 Normal distribution

- the mean μ is also the median as it splits the measurements in two
- x values that fall farther than 2σ are considered **rare** 5%
- x values that fall farther than 3σ are considered **extremely rare** 0.2%

9.14 Normal distribution

We can define the limits of **common observations** for the distribution of women's height in the population.

- $P(165 - 8 \leq X \leq 165 + 8) = P(157 \leq X \leq 173) = 0.68$
- $P(165 - 2 \times 8 \leq X \leq 165 + 2 \times 8) = P(149 \leq X \leq 181) = 0.95$
- $P(165 - 3 \times 8 \leq X \leq 165 + 3 \times 8) = P(141 \leq X \leq 189) = 0.997$



9.15 Standard normal density

Let's change variables to a **standardized variable**

$$Z = \frac{X - \mu}{\sigma}$$

in the density

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$$

replacing $x = \sigma z + \mu$ and $dx = \sigma dz$ in the probability expression we have

$$\begin{aligned} P(x \leq X \leq x + dx) = P(z \leq Z \leq z + dz) \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \end{aligned}$$

we obtain the **standardized** form of the normal density.

9.16 Standard normal density

Definition

A random variable Z defined in the real numbers has a **standard** density if it takes the form

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz, z \in \mathbb{R}$$

with mean and variance

- $E(X) = 0$
- $V(X) = 1$.

9.17 Standard normal density

The standard density:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz, z \in \mathbb{R}$$

- is the normal density $N(\mu = 0, \sigma^2 = 1)$
- any normally distributed variable X can be transformed to a variable Z

$$Z = \frac{x - \mu}{\sigma}$$

that follows a standard distribution:

$$Z \to N(0, 1)$$

9.18 Normal distribution

All normal densities can be obtained from the standard density with the values of μ and σ

9.19 Standard distribution

The probability distribution of the standard density:

$$\phi(a) = F_{standard}(a) = P(Z \leq a)$$

is the **error** function defined by

$$\phi(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

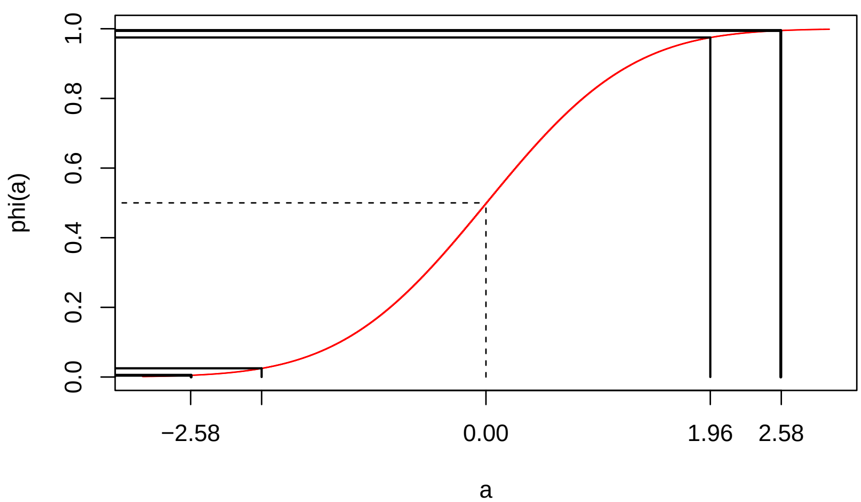
You can find it in most computer programs

9.20 Standard normal density

9.21 Standard normal density

We define the limits of the **most common observations** for the standard variable

- $P(-0.67 \leq X \leq 0.67) = 0.50$
- $P(-1.96 \leq X \leq 1.96) = 0.95$
- $P(-2.58 \leq X \leq 2.58) = 0.99$



9.22 Normal and standard distributions

For any normally distributed variable X, such that

$$X \rightarrow N(\mu, \sigma^2)$$

its distribution $F(a) = P(X \leq a)$ can be computed from

$$F(a) = \phi(\frac{a-\mu}{\sigma})$$

9.23 Normal distribution

For computing $P(a \leq X \leq b)$, we use the property of the probability distributions

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

Let's standardize

$= P(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}) - P(\frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma})$

$= P(Z \leq \frac{b-\mu}{\sigma}) - P(Z \leq \frac{a-\mu}{\sigma})$

$= \phi(\frac{b-\mu}{\sigma}) - \phi(\frac{a-\mu}{\sigma})$

Then

$$F(b) - F(a) = \phi(\frac{b-\mu}{\sigma}) - \phi(\frac{a-\mu}{\sigma})$$

The probabilities of **any normal variable** can be obtained from the **standard distribution**, after standardization (subtract the mean and divide by the standard deviation).

9.24 Summary of probability models

Model	X	range of x	f(x)	E(X)	V(X)
Uniform	integer or real number	$[a, b]$	$\frac{1}{n}$	$\frac{b+a}{2}$	$\frac{(b-a+1)^2-1}{12}$
Bernoulli	event A	0,1	$(1-p)^{1-x}p^x$	p	$p(1-p)$

Model	X	range of x	f(x)	E(X)	V(X)
Binomial	# of A events in n repetitions of Bernoulli trials	0,1,...	$\binom{n}{x}(1-p)^{n-x}p^x$	np	$np(1-p)$
Negative Binomial for events	# of B events in Bernoulli repetitions before r As are observed	0,1,..	$\binom{x+r-1}{x}(1-p)^x p^r$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$
Hypergeometric	# A events in a sample n from population N with K As	$\max(0, n+K-N)$, ... $\min(K, n)$	$\frac{1}{\binom{N}{n}}\binom{K}{x}\binom{N-K}{n-x}$	$n * \frac{N}{K}$	$n\frac{N}{K}(1-\frac{N}{K})\frac{N-n}{N-1}$
Poisson	# of events A in an interval	0,1, ..	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ
Exponential	Interval between two events A	$[0, \infty)$	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normal	measurement with symmetric errors whose most likely value is the average	$(-\infty, \infty)$	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2

9.25 R functions of probability models

Model	R
Uniform (continuous)	dunif(x, a, b)

Model	R
Binomial	dbimon(x,n,p)
Negative Binomial for events	dnbinom(x,r,p)
Hypergeometric	dhyper(x, K, N-K, n)
Poisson	dpois(x, lambda)
Exponential	dexp(x, lambda)
Normal	dnomr(x, mu, sigma)

Chapter 10

Sampling Distributions

10.1 Objective

Distributions for

- Sample mean (average)
- sample sum
- Sample variance

10.2 Normal distribution

When we have a normal random variable

$$X \to N(x; \mu, \sigma^2)$$

How do we estimate μ and σ^2?

- we need to take a **random sample**
- we need to **estimate** each parameter

10.3 Example: When we do not know the parameters

Imagine a client asking your metallurgical company to sell them 8 cables that can carry up to 96 Tons; that is 12 Tons each.

- You have in **stock** a set of cables that could do the job.

Can you use the cables in stock or do you need to produce new ones?

10.4 Example

you take a sample of 8 random experiments, each of which consists of loading a cable until it breaks and recording the breaking load.

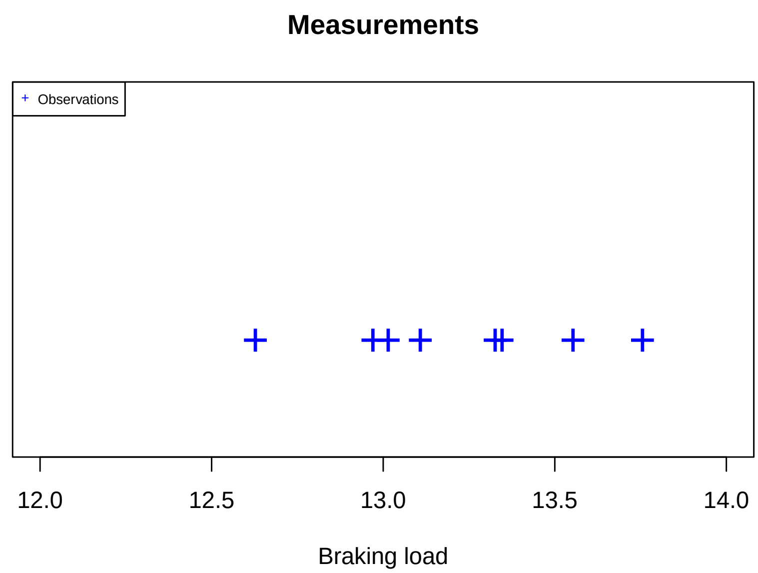
These are the results: The observation of a **sample** of size 8

```
## [1] 13.34642 13.32620 13.01459 13.10811 12.96999 13.55309 13.75557 12.62747
```

- None of them broke at 12 Tons.

- There was one that broke at 12.62747 Tons.

Do you take the risk and sell a random sample of 8 cables from your stock?



10.5 Random sample

A **random sample** of size n is the **repetition** of a random experiment n **independent** times.

- A random sample is a n-dimensional **random variable**

$$(X_1, X_2, ...X_n)$$

where X_i is the *i-th* repetition of the random experiment with comon distribution $f(x;\theta)$ for any i

- **One observation** of a random sample is the set of n values obtained from the experiments

$$(x_1, x_2, ...x_n)$$

Our **observation** of the sample of 8 cables was

```
## [1] 13.34642 13.32620 13.01459 13.10811 12.96999 13.55309 13.75557 12.62747
```

10.6 Example

We would like to compute $P(X \leq 12)$.

We are going to **assume** that the braking point is **normally distributed**.

$$X \to N(x; \mu, \sigma^2)$$

- For computing $P(X \leq 12)$ we need the parameters μ and σ^2.
- How do we estimate the parameters from the observed sample?

10.7 Average or sample mean

Definition

The sample mean (or average) of a **random sample** of size n is defined as

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

The average is a **random variable** that in our 8-size sample took the value

$$\bar{x}_{stock} = 13.21$$

10.8 The average as an estimator

This number can be used to **estimate** the unknown parameter μ because:

- $E(\bar{X}) = E(X) = \mu$
- $V(\bar{X}) = \frac{V(X)}{n} = \frac{\sigma^2}{n}$

(since each random experiment in the sample is independent)

as

- $n \to \infty$, $V(\bar{X}) \to 0$

then

- $\bar{x}$ **concentrates closer and closer** to μ as n increases.

We can take one value of $\bar{x}$ as estimation for μ or

$$\bar{x} = \hat{\mu}$$

10.9 Sample variance

Definition

The **sample variance** S^2 of a random sample of size n

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

is the dispersion of the measurements about $\bar{X}$. In our 8-size sample S^2 took the value

$$s^2_{stock} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0.1275608$$

The expected value of S^2 is

- $E(S^2) = V(X) = \sigma^2$ (unbiased)

and therefore S^2 is

- an estimator of $V(X)$
- it also concentrates around σ^2 because as $n \to \infty$, $V(\bar{S^2}) \to 0$ (consistent)

We can take one value of s^2 as estimation for σ^2 or

$$s^2 = \hat{\sigma}^2$$

10.10 Sample variance

S^2 aims to estimate the dispersion of the outcomes about μ (the variance)

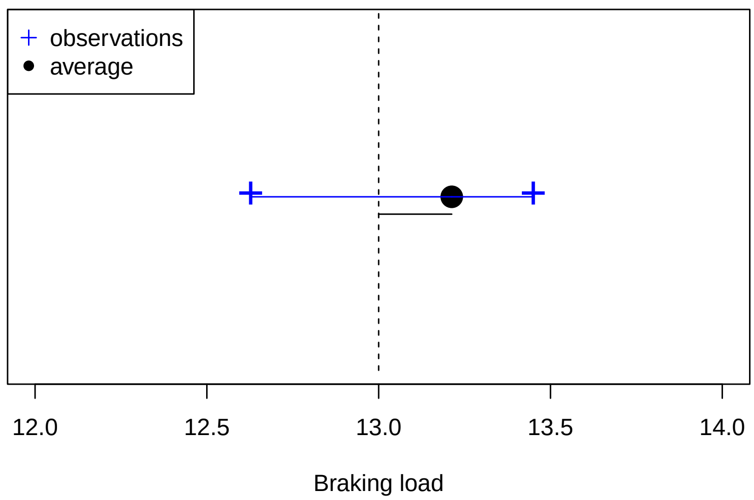
If we use $\bar{X}$ as an estimator of μ we need to correct for its dispersion (i.e. mean squared error of $\bar{X}$).

The correction is achieved by dividing by $n-1$ and not n in the definition of S^2

For:

$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$

$E(S_n^2) = \sigma^2 - \frac{\sigma^2}{n} \neq \sigma^2$ (we say that S_n^2 is **biased** estimator of σ)



10.11 Fitting a model

We **fit a model** when we

- **estimate** the parameters of the model

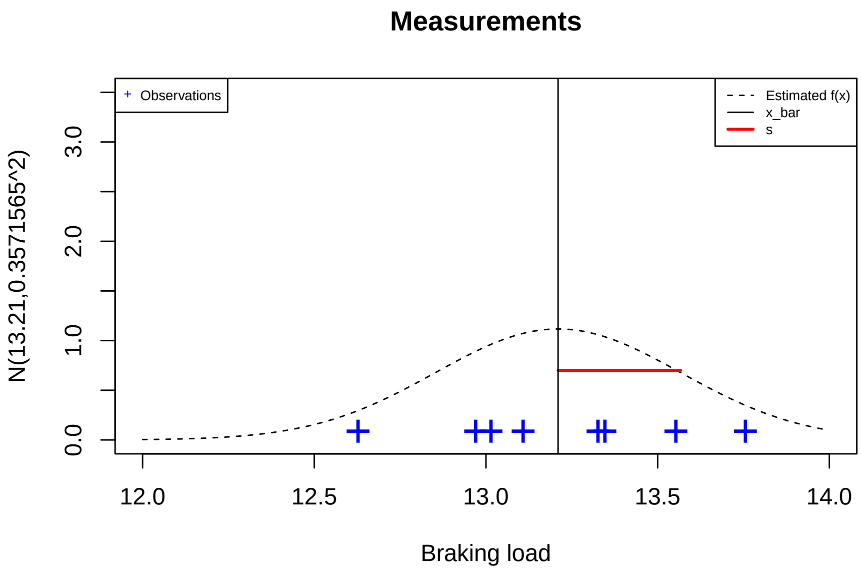
We also say we **train** a model (machine learning)

Assuming that

$$X \to N(x; \mu, \sigma^2)$$

Since we do not know the parameters, we **plugin** the estimates $\bar{x}$ and s^2 as the values of μ and σ^2

$$X \to N(x; \mu = 13.21, \sigma^2 = 0.3571565^2)$$



10.12 Prediction

We **predict** the value of an **outcome** when we compute its **probability**

What is the probability that the cable breaks at 12 Tons?

If we assumed the random variable

$$X \to N(x; \mu, \sigma^2)$$

We plug in the estimates $\bar{x}$ and s^2 into the probability distribution

$$P(X \leq 12) = F_{normal}(12; \mu = 13.21, \sigma^2 = 0.1275608)$$

In R pnorm(12,13.21, 0.3571565)= 0.000352188

Given the **observed** sample, there is an estimated probability of 0.03% that a single cable will break at 12 Tons.

10.13 Inference

When we have a normal random variable

$$X \to N(x; \mu, \sigma^2)$$

and we know μ and σ^2.

We can make inferences about $\bar{X}$, that is, calculate probabilities of the random variable $\bar{X}$.

When we make **inferences**, we usually ask the question:

How **sure** are we that the value of the estimator **is close** to the **true parameter**?

10.14 Example: When we do know the parameters

Let us imagine that our cables are certified to break with an average load of $\mu = 13$ Tons with variance $\sigma^2 = 0.35^2$.

We take a random sample of 8 cables

```
## [1] 13.34642 13.32620 13.01459 13.10811 12.96999 13.55309 13.75557 12.62747
```

Can we say that we actually produced stronger cables because we got $\bar{x} = 13.21$ from this sample of 8 cables?

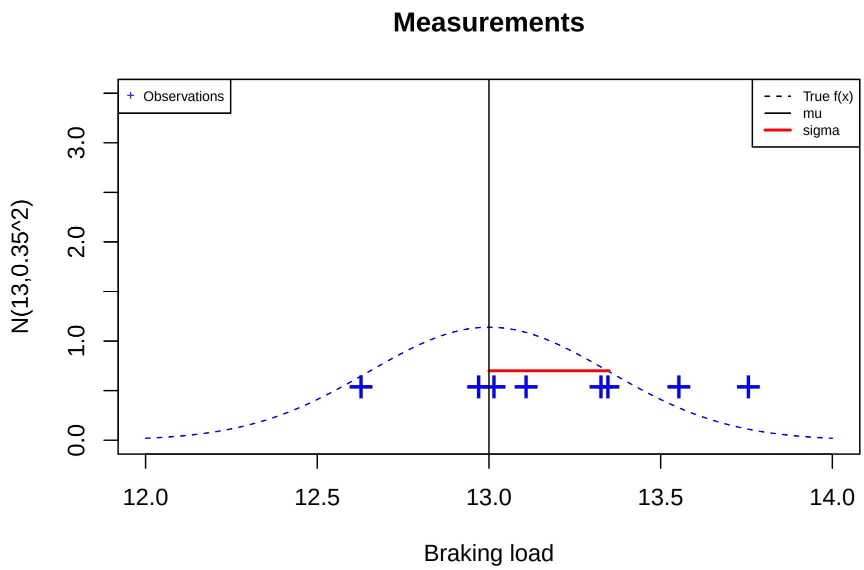
We need to calculate probabilities of $\bar{X}$.

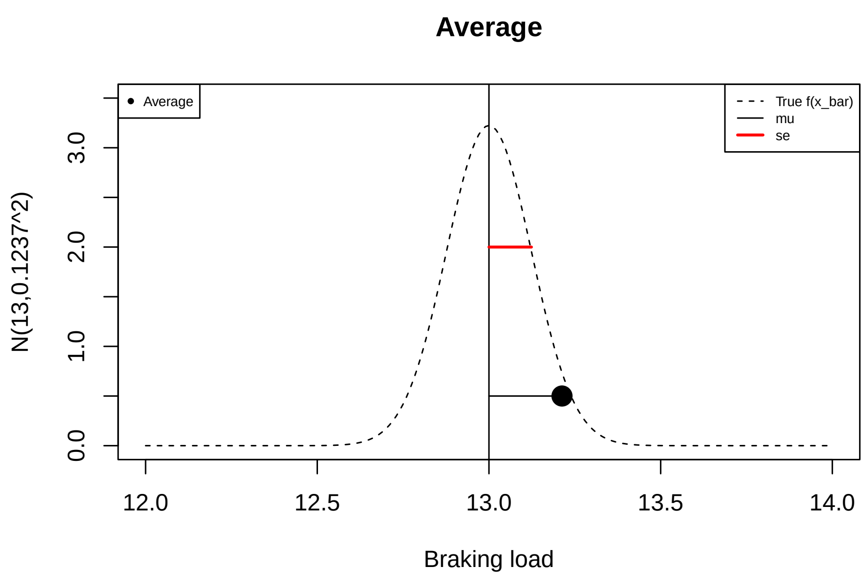
- What is the probability that the distance between $\bar{X}$ and μ is less than $\bar{x}_{stock} - \mu = 0.21$?

$$P(-0.21 \leq \bar{X} - \mu \leq 0.21)$$

10.15 Density for X and $\bar{X}$

If we **know** that the **true** parameters are $\mu = 13$ and $\sigma = 0.35$ this is what we would see





10.16 Sample mean distribution

Theorem: When X follows a normal distribution $X \rightarrow N(\mu, \sigma^2)$

$\bar{X}$ is normal:

$$\bar{X} \rightarrow N(\mu, \frac{\sigma^2}{n})$$

Then, if we **know** μ and σ we can compute the true **probabilities of** $\bar{X}$ using the normal distribution.

The mean and variance of $\bar{X}$ are

- $E(\bar{X}) = \mu$
- $V(\bar{X}) = \frac{\sigma^2}{n}$

10.17 Inference on the average

Example:

If we **know** that the breaking load of our cables truly distribute as

$$X \rightarrow N(\mu = 13, \sigma^2 = 0.35^2)$$

then

$$\bar{X} \rightarrow N(13, \frac{0.35^2}{8})$$

- $E(\bar{X}) = 13$
- $V(\bar{X}) = \frac{0.35^2}{8} = 0.01530169$

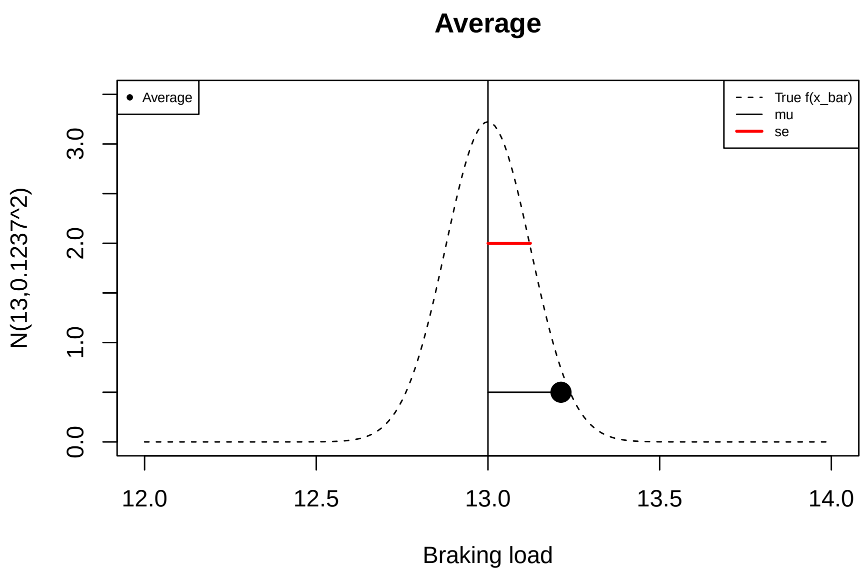
Our **observed error** in the estimation of the mean is the difference

$$\bar{x}_{stock} - \mu = 13.21 - 13 = 0.21$$

We ask: Is this a **typical** error?

10.18 Density for $\bar{X}$

If we **knew** that the **true** parameters were $\mu = 13$ and $\sigma = 0.35$ this is the error we would see



10.18.1 Probabilities of $\bar{X}$

If we **know** that the braking load of our cables **truly** distribute as

$$\bar{X} \to N(\mu = 13, \frac{\sigma^2}{n} = 0.1237^2)$$

What is the probability of observing an **error in estimation** of μ (distance between $\bar{X}$ and μ) smaller than 0.21?

We want to compute

$$P(-0.21 \leq \bar{X} - 13 \leq 0.21) = P(12.79 \leq \bar{X} \leq 13.21)$$

$= F_{normal}(13.21; \mu, se^2) - F_{normal}(12.79; \mu, se^2)$

In R we can compute it as:

pnorm(13.21, 13, 0.1237)-pnorm(12.79, 13, 0.1237)=0.9104.

91.0% of the errors are less than 0.21, therefore the **observed** error does not seem to be too typical (only 9% of the errors are higher).

Maybe we have stronger cables than we thought.

10.18.2 Sample sum

If we are interested in using all the 8 cables at the same time to carry a total of 96 Tons, then we should consider adding their individual contributions.

The **sample sum** is the **statistic**:

$$Y = n\bar{X} = \sum_{i=1}^{n} X_i$$

Theorem: if $X \to N(\mu, \sigma^2)$ then

$$Y \to N(n\mu, n\sigma^2)$$

With mean and variance:

- $E(Y) = n\mu$
- $V(Y) = n\sigma^2$

10.18.3 Inference on the sample sum

If we **know** that for our cables

$$X \to N(\mu = 13, \sigma^2 = 0.35^2)$$

then

$$Y \to N(n\mu = 104, n\sigma^2 = 8 \times 0.35^2)$$

- $E(Y) = 104$
- $V(Y) = 8 \times 0.35^2 = 0.98$

For our 8-sample, we observed

- $y_{stock} = 105.7014$

and, therefore, the **observed error** in the estimation of the mean of the **true** total braking load ($n\mu$) of 8 cables was

- $y_{stock} - n\mu = 1.7014$

Is this a **typical** error?

10.18.4 Probabilities of the sample sum: Propagation of error

What is the probability of observing a difference $Y - E(Y)$ smaller than 1.7014?

We want to compute the probability

$$P(-1.7014 \leq \bar{Y} - 104 \leq 1.7014) = P(102.2986 \leq Y \leq 105.7014)$$

$= F_{normal}(105.7014; n\mu, n\sigma^2) - F_{normal}(102.2986; n\mu, n\sigma^2)$

In R we can compute it as:

pnorm(105.7014, 104, sqrt(0.98)) - pnorm(102.2986, 104, sqrt(0.98))=0.914.

91.4% of the times we obtain sample sums that are smaller than 1.7014

This proportion is higher than the proportion for individual cables.

10.19 Inference in the sample variance

Consider a quality control process that requires that the cables are produced close to the specified value μ.

If a sample of 8 cables is too dispersed ($S^2 > 0.3$), we stop production: the process is out of control.

What is the probability that the sample variance of a sample of 8 cables is greater than the required 0.3?

10.20 Probabilities of the sample variance

Theorem: When X follows a normal distribution

$$X \to N(\mu, \sigma^2)$$

The **statistic**:

$$W = \frac{(n-1)S^2}{\sigma^2} \to \chi^2(n-1)$$

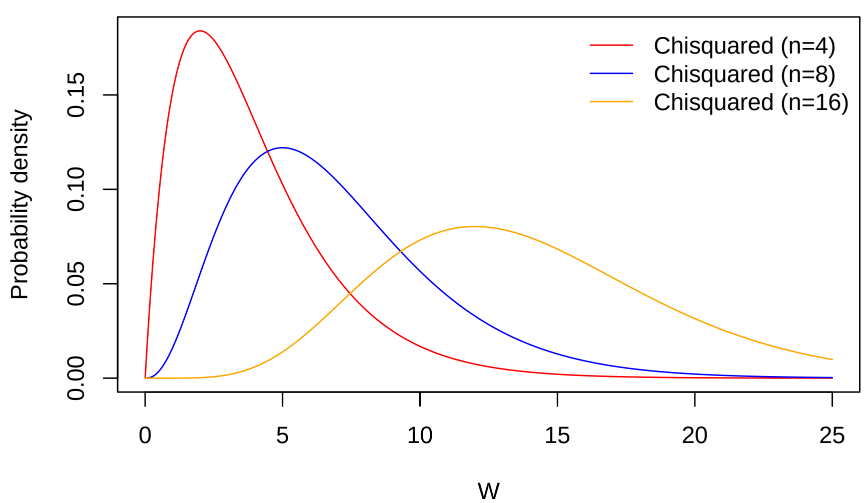
has a χ^2 (chi-squared) distribution with $df = n-1$ degrees of freedom given by

$$f(w) = C_n w^{\frac{n-3}{2}} e^{-\frac{w}{2}}$$

where:

- $C_n = \frac{1}{2^{(n-1)/2}\sqrt{\pi(n-1)}}$ ensures $\int_{-\infty}^{\infty} f(t)dt = 1$
- $\Gamma(x)$ is Euler's factorial for real numbers
- If we **know** the true values of μ and σ we can compute probabilities of S^2 using the χ^2 distribution for W.

10.21 χ^2-statistic



10.22 χ^2-statistic

If we **know** that our cables trully distribute as

$$X \to N(\mu = 13, \sigma^2 = 0.35^2)$$

then we can compute

$$P(S^2 > 0.2) = P(\frac{(n-1)S^2}{\sigma^2} > \frac{(n-1)0.3}{\sigma^2})$$

$= P(W > \frac{(n-1)0.3}{\sigma^2})$

$= 1 - P(W \leq \frac{(n-1)0.3}{\sigma^2}) = 1 - P(W \leq \frac{(8-1)0.3}{0.1225})$

$= 1 - F_{\chi^2, df=7}(17.14286) = 0.016$

In R 1-pchisq(17.14286, df=7)=0.016

There is only a probability of 1% of obtaining a value greater than $s^2 = 0.3$.

- $s^2 > 0.3$ seems to be a good criterion to stop production and revise the process.
- our observed value was $s^2_{stock} = 0.1275608$
- the sample is not too dispersed and we believe that the production is under control.

Chapter 11

Central limit theorem

11.1 Objective

- Margin of errors
- Central limit theorem
- t-statistic

11.2 Margin of error

When deciding whether an **observed error** is large or not we usually compare it with a **predefined** tolerance.

- The **margin of error** at 5% level is the distance m such that distribution of $\bar{X}$ captures 95% of the estimations:

$$P(-m \leq \bar{X} - \mu \leq m) = P(\mu - m \leq \bar{X} \leq \mu + m) = 0.95$$

- or that 95% of the values of $\bar{X}$ are a distance m from μ

11.3 Margin of error

Let's continue with the breaking load example.

for the 8-sample

```
## [1] 13.34642 13.32620 13.01459 13.10811 12.96999 13.55309 13.75557 12.62747
```

the **observed error** is the difference

$$\bar{x}_{stock} - \mu = 13.21 - 13 = 0.21$$

Is this value below the margin of error at 5%?

11.4 Z-statistic

If we **know** that for our cables truly distribute as

$$X \rightarrow N(\mu = 13, \sigma^2 = 0.35^2)$$

then,

$$\bar{X} \rightarrow N(\mu, \frac{\sigma^2}{n})$$

and the 5% margin of error for the average in our 8-sample can be computed from the **standardized statistic**:

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0, 1)$$

11.5 Z-statistic

to compute the margin of error m at 5% level we standardize (subtract μ and divide by $\sigma/\sqrt{n}$)

$$P(\mu - m \leq \bar{X} \leq \mu + m) = P(-\frac{m}{\sigma/\sqrt{n}} \leq \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{m}{\sigma/\sqrt{n}})$$

$$= P(-\frac{m}{\sigma/\sqrt{n}} \leq Z \leq \frac{m}{\sigma/\sqrt{n}}) = 0.95$$

(compare it with the plot) we have

$$m = z_{0.025}\frac{\sigma}{\sqrt{n}} = 1.96 \times se = 1.96\frac{0.35}{\sqrt{8}} = 0.24$$

where $z_{0.025} = 1.96$ is the value Z that leaves 2.5% at each side of standard normal density (0.025-quantile)

Our observed error 0.21

- is less than the margin of error 0.24 at level 5%.
- and, therefore, it is expected within the 95% of errors.

If an observation of $\bar{x}$ distance more than ~ 2 times the se we say that the error is **unusually** large.

11.6 Z-statistic

Definition

For a normal random variable X

$$X \to N(\mu, \sigma^2)$$

with **known** σ

The Z statistic:

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}}$$

is a standard random variable whose $1 - \alpha/2$-quantiles ($z_{1-\alpha/2}$) give a measure of the margin of error of $\bar{X}$ at $1 - \alpha$ level

$$m = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

A common situation:

- What happens when X is not normally distributed?

11.7 Central Limit Theorem

For any random variable X with **unknown** (any type of) distribution

$$X \to f(x;\theta)$$

the standardized statistic

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}}$$

approximates to a standard distribution

$$Z \to_d N(0,1)$$

when $n \to \infty$

Therefore:

- We can compute probabilities for $\bar{X}$ if n is large, using the normal distribution:

$$\bar{X} \sim_{aprox} N(E(X), \frac{V(X)}{n})$$

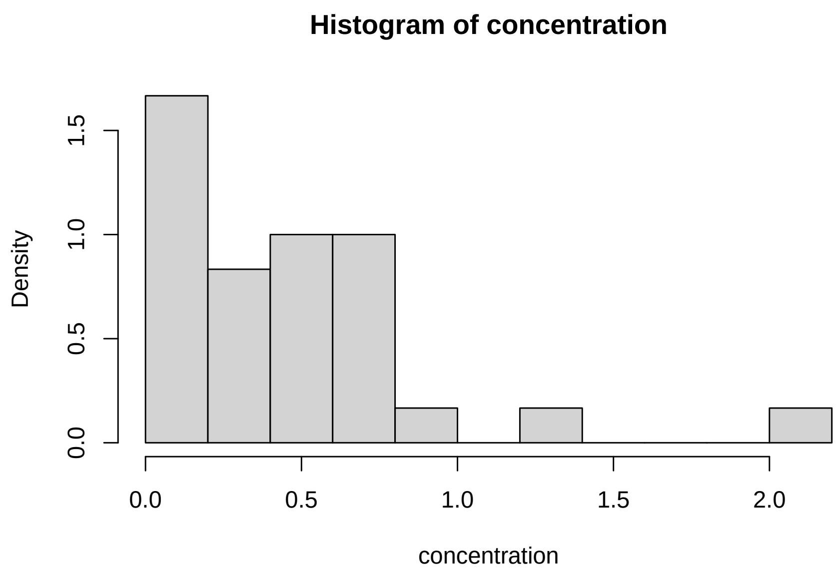
11.8 Central Limit Theorem

Example:

Consider an experiment where we measure the concentration in blood of a drug after 10-hour administration in 30 patients. We obtain the following results:

```
##  [1] 0.42172863 0.28830514 0.66452743 0.01578868 0.02810549 0.15825061
##  [7] 0.15711365 0.07263340 1.36311823 0.01457672 0.50241503 0.24010736
## [13] 0.14050681 0.18855892 0.09414202 0.42489306 0.78160177 0.23938021
## [19] 0.29546742 2.02050586 0.42157487 0.48293561 0.74263790 0.67402224
## [25] 0.58426449 0.80292617 0.74837143 0.78532627 0.01588387 0.29892485
```

- the average is $\bar{x} = 0.56$
- the histogram of the results is:



11.9 Central Limit Theorem

If we **know** that levels follow an exponential distribution

$$X \to exp(\lambda = 2)$$

The mean and variance are:

- $E(X) = \frac{1}{\lambda} = 0.5$
- $V(X) = \frac{1}{\lambda^2} = 0.25$

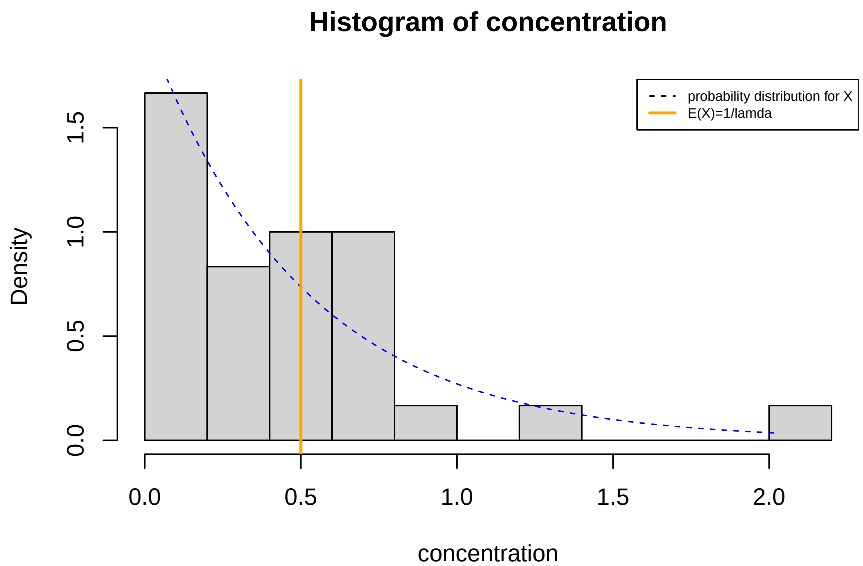
Therefore the mean and variance of $\bar{X}$ are:

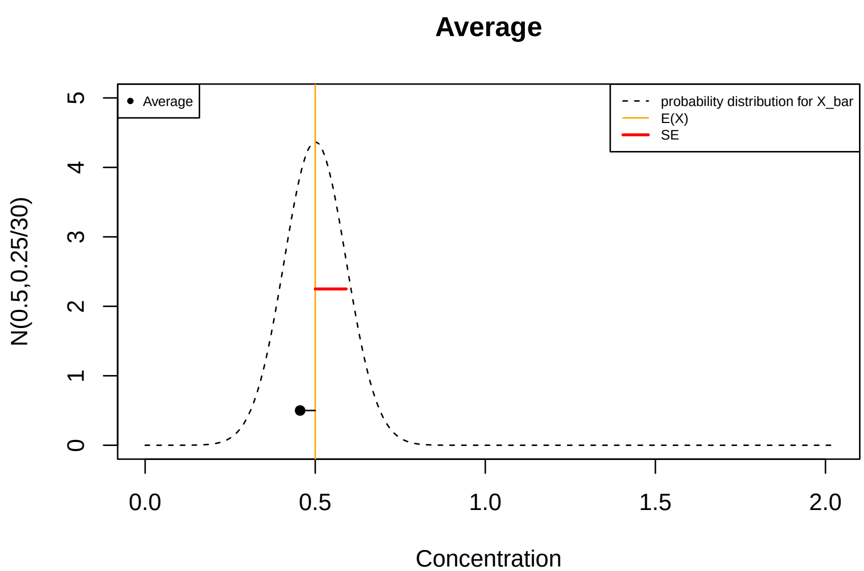
- $E(\bar{X}) = \frac{1}{\lambda} = 0.5$
- $V(\bar{X}) = \frac{V(X)}{n} = \frac{1}{n\lambda^2} = 0.25/30$

As $n \geq 30$

$$Z = \frac{\bar{X} - \lambda}{\sqrt{\frac{1}{n\lambda^2}}}$$

is a standard normal variable and: $\bar{X} \sim_{aprox} N(\lambda, \frac{1}{n\lambda^2})$


Histogram of concentration
probability distribution for X
E(X)=1/lamda
Density
0.0
0.5
1.0
1.5
0.0
0.5
1.0
1.5
2.0
concentration


Average
Average
probability distribution for X_bar
E(X)
SE
N(0.5,0.25/30)
0
1
2
3
4
5
0.0
0.5
1.0
1.5
2.0
Concentration

11.10 Margin of error with CLT

Since

$$\bar{X} \sim_{aprox} N(E(X), \frac{V(X)}{n})$$

The margin of error at 5% level

$$P(E(X) - m \leq \bar{X} \leq E(X) + m) = 0.95$$

can be computed again with the standard distribution

$$m = z_{0.025}\sqrt{\frac{V(X)}{n}} = 1.96\sqrt{\frac{0.25}{30}} = 0.1789227$$

We **observed** $\bar{x} = 0.5638725$ therefore the **observed error** in estimation is

$$\bar{x} - E(X) = 0.5638725 - 0.5 = 0.063$$

which is within the margin of error.

The error that we observed is common and within the 95% of errors.

11.11 Sample sum and CLT

For any random variable X with **unknown** (any type of) distribution

$$X \to f(x; \theta)$$

the standardized statistic

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{n\bar{X} - nE(\bar{X})}{\sqrt{nV(\bar{X})}}$$

approximates to a standard distribution

$$Z \to_d N(0, 1)$$

when $n \to \infty$

Therefore:

- We can compute probabilities for the sample sum $Y = n\bar{X}$ if n is large, using the normal distribution:

$$\bar{Y} \sim_{aprox} N(nE(X), nV(X))$$

11.12 Unknown σ but large n

For any random variable X with **unknown** (any type of) distribution

$$X \to f(x;\theta)$$

with **unknown** variance $V(X)$, we can estimate the standard error ($se = \sqrt{V(X)/n}$) by the sample standard deviation

$$\hat{se} = \frac{s}{\sqrt{n}}$$

and write the standardized statistic

$$Z = \frac{\bar{X} - E(\bar{X})}{\frac{s}{\sqrt{n}}}$$

$$Z \to_d N(0,1)$$

to recover the CLT when $n \to \infty$ (a good aproximation is when $n > 30$)

11.13 T-statistic

When

- σ is **unknown**

and

- n is small (cannot apply CLT)

However, if X is normal

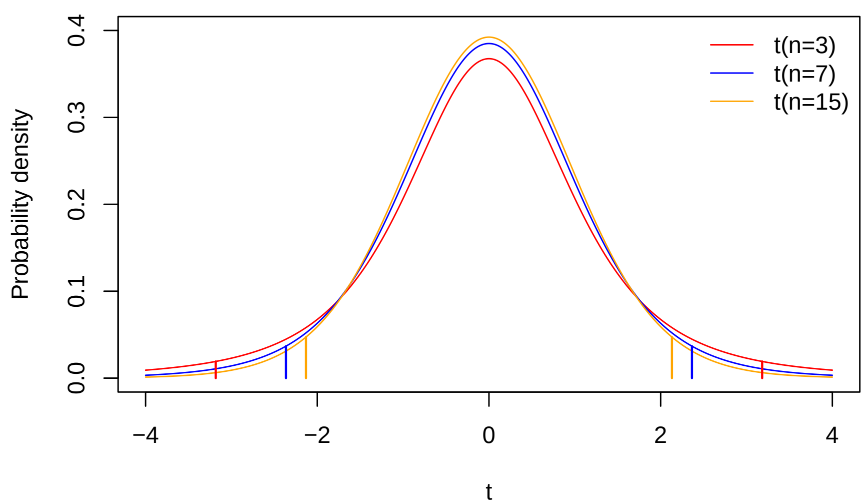
$$X \to N(\mu, \sigma^2)$$

then the standardized statistic

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

Follows a t-distribution with $n-1$ degrees of freedom, and we can compute probabilities on $\bar{X}$.

11.14 T-statistic



11.15 T-statistic

To compute the margin of error m at 5% level when n is small, σ unknown but X normal

$P(\mu - m \leq \bar{X} \leq \mu + m) = P(-\frac{m}{s/\sqrt{n}} \leq \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \leq \frac{m}{s/\sqrt{n}})$

$$= P(-\frac{m}{s/\sqrt{n}} \leq T \leq \frac{m}{s/\sqrt{n}}) = 0.95$$

We use the t-distribution

$$m = t_{0.025,n-1}\frac{s}{\sqrt{n}}$$

where $t_{0.025,n-1}$ is the value T that leaves 2.5% at each side of t-distribution with $n-1$ degrees of freedom (0.025-quantile)

11.16 Example 1

Going back to the breaking load example, we computed the margin of error with **known** $\sigma^2 = 0.35^2$.

$$m = z_{0.025}\frac{\sigma}{\sqrt{n}} = 1.96 \times se = 1.96\frac{0.35}{\sqrt{8}} = 0.24$$

- In most applications we **do not know** the parameters

If we only assumed that the breaking load is a normal random variable

$$X \rightarrow N(\mu, \sigma^2)$$

with **unknown** μ and σ^2 then from the data

- $s_{stock} = \sqrt{0.1275608}$

and the margin of error is

$$m = t_{0.025,n-1}\frac{s}{\sqrt{n}} = 2.36 \times \hat{se} = 2.36\frac{0.3571565}{\sqrt{8}} = 0.29$$

where $t_{0.025,n-1} = 2.36$

in R is qt(1-0.025, 7)

It increased from the value we obtained with **known** σ

11.17 Example 2

We can also ask for the probability of observing an error in the estimation of μ (distance between $\bar{X}$ and μ) smaller than the observed value 0.21?

We thus want to compute

$$P(-0.21 \leq \bar{X} - \mu \leq 0.21) = P(\frac{-0.21}{s/\sqrt{n}} \leq T \leq \frac{0.21}{s/\sqrt{n}})$$

$= P(\frac{-0.21}{0.3571565/\sqrt{8}} \leq T \leq \frac{0.21}{0.3571565/\sqrt{8}})$

$= F_{t,n-1}(0.21) - F_{t,n-1}(-0.21)$

In R we can compute it as:

pt(1.663052, 7)-pt(-1.663052, 7)=0.859.

85.9% of the errors are less than 0.21, therefore the **observed** error seems more typical than the 91% that we obtain with $\sigma^2 = 0.35^2$.

Note that in the calculations we have substituted $\sigma = 0.35$ by a higher estimate $s = 0.3571565$ **obtained from data**.

Chapter 12

Maximum likelihood

12.1 Objective

- Maximum likelihood
- Method of Moments

12.2 Statistic

Definition

Given a random sample $X_1, ...X_n$ a **statistic** is any real value function of the random variables that define the random sample: $f(X_1, ...X_n)$

- $\bar{X} = \frac{1}{N}\sum_{j=1..N} X_j$
- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
- $\max X_1, X_n$

are statistics

12.3 Estimator

Definition

An **estimator** is a statistic Θ whose values $\hat{\theta}$ are measures of a parameter θ of the population distribution on which the sample is defined: $E(\Theta) \sim \theta$

$$X \to f(x;\theta)$$

Then

- θ is a **parameter** of the population distribution $f(x;\theta)$
- Θ is an **estimator** of θ: A random variable
- $\hat{\theta}$ is the **estimate** of θ: A realized value of Θ

12.4 Estimator

12.5 Examples 1: Average (Sample mean)

When

$$X \to N(\mu, \sigma^2)$$

For the mean:

- μ is a **parameter** of the **population** distribution: distribution of X, $N(\mu, \sigma^2)$
- $\bar{X}$ is an **estimator** of μ
- $\bar{x} = \hat{\mu} = 13.21\,Tons$ is the **estimate** of μ

12.6 Examples 2: Sample Variance

When

$$X \to N(\mu, \sigma^2)$$

For the variance:

- σ^2 is a **parameter** of the population distribution $N(\mu, \sigma^2)$
- S^2 is an **estimator** of σ^2
- $s^2 = \widehat{\sigma^2} = 0.127\,Tons^2$ is the **estimate** of σ^2

12.7 Bias

An estimator is unbiased if $E(\Theta) = \theta$

- $\bar{X}$ is an **unbiased** estimator of μ because $E(\bar{X}) = \mu$
- S^2 is an **unbiased** estimator of σ^2 because $E(S^2) = \sigma^2$

12.8 Consistency

An estimator is consistent if $V(\Theta) \to 0$ when $n \to \infty$

- $\bar{X}$ is **consistent** because $V(\bar{X}) = \frac{\sigma^2}{n} \to 0$ when $n \to \infty$.
- S^2 is also **consistent** (we will not show).

12.9 Maximum likelihood

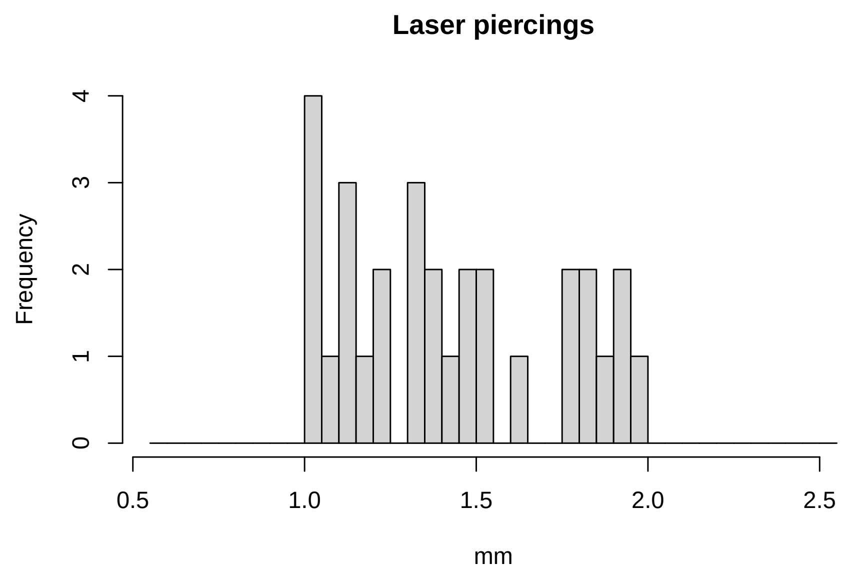
How can we **estimate** the parameter of **any** parametric model?

- Imagine we design a laser with a diameter of $1mm$ that we want to use for clinical applications.
- We want to characterize the diameter of a piercing in a tissue made with the laser
- and take a random sample of 30 cuts made with the laser

```
##  [1] 1.11 1.64 1.20 1.79 1.89 1.01 1.31 1.81 1.34 1.25 1.92 1.24 1.49 1.36 1.03
## [16] 1.82 1.09 1.01 1.14 1.91 1.80 1.51 1.44 1.98 1.46 1.53 1.33 1.39 1.12 1.04
```

12.10 Example

with histogram



12.11 Probability density

We consider that maximum probability should be given to diameters of $x = 1mm$, and that the diameters should decrease as the inverse power of some **unknown** parameter α, with a limit of $2mm$ beyond which the probability is 0.

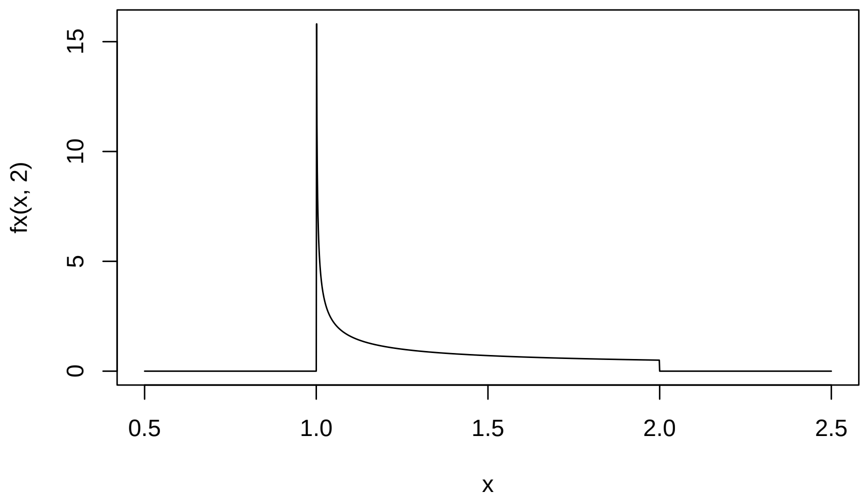
A suitable probability density distribution is

$$f(x) = \begin{cases} \frac{1}{\alpha}(x-1)^{\frac{1}{\alpha}-1}, & \text{if } x \in (1,2) \\ 0, & x \notin (1,2) \end{cases}$$

Where α is a parameter. This is a probability density.

12.12 Probability density

In particular, for $\alpha = 2$ we can plot it

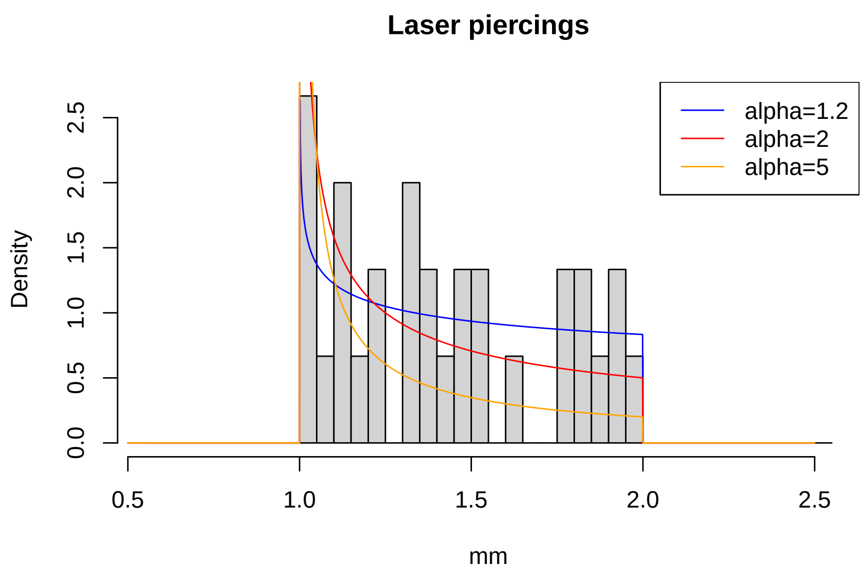


If we were to perform a n-sample: $X_1, ...X_n$, how should we combine the data for obtaining the best value of α?

- The **maximum likelihood** method **gives us the estimator** for α

$$\hat{\alpha}_{ml}$$

12.13 Example: Maximum likelihood



12.14 Maximum likelihood

The objective is to find the value of the parameter that we **believe** can **best** represent the data.

We search for the parameter that makes the **observation** of the sample the most **probable**.

Note:

- **Probabilities** are assigned **to observations**.
- Probabilities are **not** assigned to **parameters** (we assign beliefs, and likelihoods).

Parameters are not supposed to change, they are properties of the system.

12.15 Method step 1

1. We calculate the probability of having observed the n-sample: $x_1, ...x_n$. It is the product of probabilities because observations are independent of one another:

$$P(M = x_1, ...x_n) = P(X = x_1)P(X = x_2)...P(X = x_n)$$

$$= f(x_1; \alpha)f(x_2; \alpha)...f(x_n; \alpha)$$

- Once the data is observed they are **fixed**.
- The unknown is α
- This probability as a function of the α we call it the **likelihood function**

$$L(\alpha) = \Pi_{i=1..n} f(x_i; \alpha)$$

then in our case

$$L(\alpha; x_1, ..x_n) = \frac{1}{\alpha^n}\Pi_{i=1..n}(x_i - 1)^{\frac{1-\alpha}{\alpha}} = \frac{1}{\alpha^n}\{(x_1 - 1)(x_2 - 1)...(x_n - 1)\}^{\frac{1-\alpha}{\alpha}}$$

12.16 Method step 2

We ask: what is the value of α that makes the observations the most probable? We thus want to maximize $L(\alpha)$ with respect to α.

Since we have the multiplication of many factors is easier to maximize the logarithm of $L(\alpha)$

2. Take the logarithm, obtain the **Log-likelihood**

$$\ln L(\alpha; x_1, ..x_n) = -n\ln(\alpha) + \frac{1-\alpha}{\alpha}\Sigma_{i=1...n}\ln(x_i - 1)$$

12.17 Method step 3

3. Maximize the log-likelihood with respect to the parameter

Therefore,

- we differentiate with respect to α

$$\frac{d \ln L(\alpha)}{d\alpha} = -\frac{n}{\alpha} - \frac{1}{\alpha^2}\Sigma_{i=1\ldots n}\ln(x_i)$$

- The maximum is where the derivative is 0. This maximum is the value of our estimator $\hat{\alpha}_{ml}$.

$$\hat{\alpha}_{ml} = -\frac{1}{n}\Sigma_{i=1\ldots n}\ln(x_i - 1)$$

12.18 Method step 3

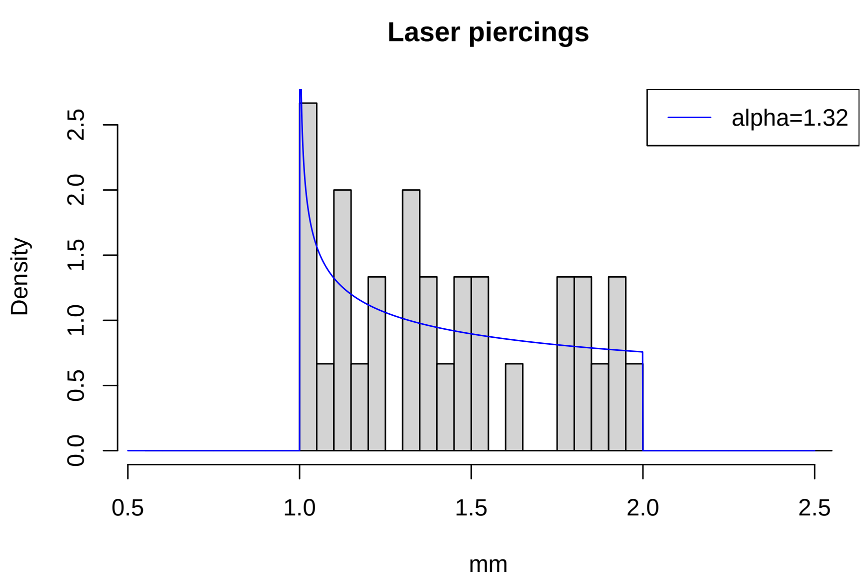
$\hat{\alpha}_{ml} = -\frac{1}{n}\Sigma_{i=1\ldots n}\ln(x_i - 1)$

is the **statistic** that estimates the parameter.

In our example we thus compute:

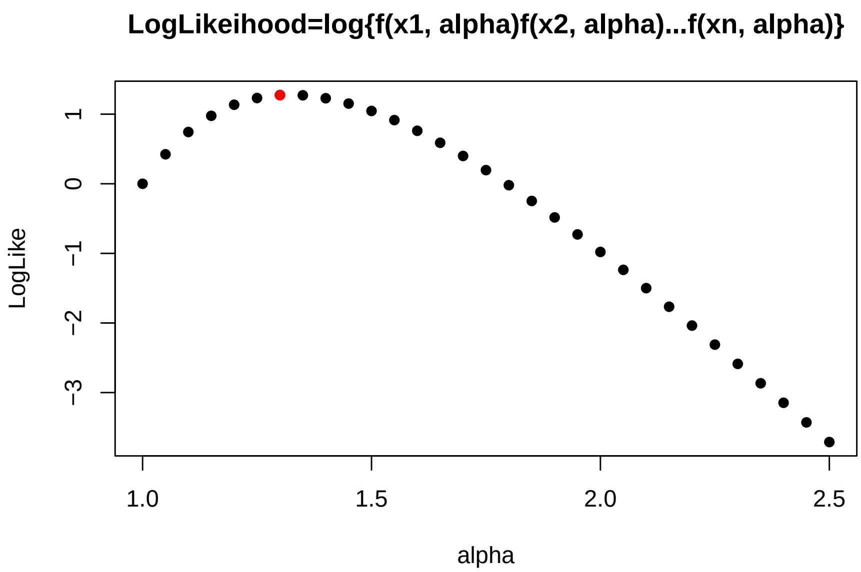
$\hat{\alpha}_{ml} = -\frac{1}{n}\{\ln(1.11 - 1) + \ln(1.64 - 1) + \ldots \ln(1.04 - 1)\} = 1.320$

12.19 Estimation



12.20 Estimation

Let's look at the log-likelihood for our 30 laser cuts. Remember, data is fixed by our experiment and α varies



Note: If we take another sample this function changes and so does its maximum.

12.21 Maximum likelihood: History

12.22 Maximum likelihood: History

- Ceres was thought to be a planet
- It disappeared behind the Sun

- Predictions were needed to know where in the sky to look for it after it passed behind the sun
- The trajectory (parallel to the planets) would determine if it was likely a planet
- With several observations with errors, what would be the best representative of the true position of Ceres at a given time?

12.23 Maximum likelihood: History

What is the statistic that best represents the true position of Ceres?

12.24 Maximum likelihood: History

Gauss proposed that at a **given** time

- the **true** position of Ceres was the mean μ
- the probabilities around the mean were symmetrical.

12.25 Maximum likelihood: History

Gauss discovered that if the average ($\bar{x}$) is the **most likely** value for the real position of Ceres (μ), then the probability density for the errors is

$$\frac{h}{\sqrt{\pi}} e^{-h^2(x-\mu)^2}$$

which we call the Gaussian and Pearson (1920) baptized it as the normal curve.

Note: We assume that the **true** position of Ceres exists μ.

Can we say the same about the height of men? is there a **true** mean height? (Galton)

12.26 Normal distribution

Imagine that we take a 8-sample for the breaking load of cables

```
## [1] 13.34642 13.32620 13.01459 13.10811 12.96999 13.55309 13.75557 12.62747
```

and

- **assume** that
$$X \to N(\mu, \sigma^2)$$
.
- What are the estimators of μ and σ^2 that maximize the probability of the observed data?

12.27 Normal distribution

1. The likelihood function, the probability of having observed $(x_1, ...x_n)$ is

$L(\mu, \sigma^2) = \Pi_{i=1..n} N(x_i; \mu, \sigma)$

$$= (\frac{1}{\sigma\sqrt{2\pi}})^n e^{-\frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2}$$

2. We can take the log of L, and compute the **log-likelihood**

$$\ln L(\mu, \sigma^2) = -n\ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\Sigma_i(x_i - \mu)^2$$

12.28 Normal distribution

The estimates of μ, σ^2 are where the likelihood is maximum, and give the highest probability for the data.

3. we differentiate with respect to μ and σ^2 (i.e. making a substitution like $t = \sigma^2$)

- $\frac{d\ln L(\mu,\sigma^2)}{d\mu} = \frac{1}{\sigma^2}\sum_i(x_i - \mu)$
- $\frac{d\ln L(\mu,\sigma^2)}{d\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_i(x_i - \mu)^2$

12.29 Normal distribution

The derivatives are 0 at the maxima

- $\frac{1}{\hat{\sigma}^2}\sum_i(x_i - \hat{\mu}) = 0$
- $-\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4}\sum_i(x_i - \hat{\mu})^2 = 0$

solving for the parameters we find

- $\hat{\mu}_{ml} = \frac{1}{n}\sum_i x_i = \bar{x}$ (the **average**)
- $\hat{\sigma^2}_{ml} = \frac{1}{n}\sum_i(x_i - \bar{x})^2$ (the **uncorrected** sample variance)

The maximum likelihood estimator of σ^2 is a **biased** estimator as:

$$E(\hat{\sigma}^2_{ml}) = \sigma^2 - \frac{\sigma^2}{n} \neq \sigma^2$$

12.30 Method of Moments

The method of maximum likelihood aims to produce the estimators of probability distributions from data.

- Is there another way to produce those estimators? would they be equal?

12.31 Method of Moments

Let's re-write the estimator $\hat{\mu} = \bar{x}$ for a normal randon variable in terms of the outcomes of X

For instance:

$$\hat{\mu} = \frac{1}{n}\sum_i x_i = \sum_x x\frac{n_x}{n}$$

and remember that in the limit $n \to \infty$ the frequentist interpretation requires $\frac{n_x}{n} \to P(X = x)$ and therefore in the limit

$$\hat{\mu} = \frac{1}{n}\sum_i x_i \to E(X) = \mu$$

12.32 Method of Moments

The method os moments says that we can take the **observed** value of the average $\bar{X}$ as an estimator of $E(X) = \mu$

$$E(X) \sim \bar{x}$$

$\bar{X} = \frac{1}{n}\sum_i X_i$ is called the first **sample moment**

if $X \rightarrow f(x, \theta)$ the estimator of the parameter θ is then obtained from the equation:

$$E(X; \hat{\theta}) = \bar{x}$$

Example: If

$X \hookrightarrow exp(\lambda)$ then

1. $E(X; \mu, \sigma^2) = \mu$
2. We write down the equation where we make the expected value equal to the first sample moment $\frac{1}{\lambda} = \bar{x} = \frac{1}{n}\sum_i x_i$
3. We solve for the parameter

$$\hat{\lambda} = \frac{1}{\bar{x}}$$

12.33 Method of Moments

Suppose that we have several batteries (new and old) that we charge over the period of 1 hour. We measure the state of charge of the battery, being 1 a 100% charge.

The state of charge of a battery is a random variable that may have a uniform distribution, where we do not know the minimum value that x can take, but we know that the maximum is 1 (100% of charge)

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

What is the estimator of a (the minimum charge after one hour)?

- We run an experiment and obtain $x_1, ...x_n$ how can we estimate a form the data?

12.34 Method of Moments

1. We compute the expected value of the random variable

$$E(X) = \frac{a+1}{2}$$

2. We obtain the equation for $\hat{a}$ where we make the expected value equal to the first sample moment

$$\frac{\hat{a}+1}{2} = \bar{x}$$

3. We solve for the estimator $\hat{a}$

This is the estimator of the minimum charge we may observe.

12.35 Method of Moments

Note that taking the minimum of the measurements is clearly suboptimal.

The method gave us a clever answer:

- we can compute $\bar{x}$ with increasing precision given by n
- We know that no measurement surpasses $b = 1$
- Then compute the distance between $\bar{x}$ and b: $1 - \bar{x}$
- Subtract it from $\bar{x}$: $\bar{x} - (1 - \bar{x}) = 2\bar{x} - 1$

12.36 Method of Moments

The method says that an estimator for the parameter θ of $f(x;\theta)$ can be found from the equation:

$$E(X) = \frac{1}{n}\sum_i x_i$$

If there are more parameters, we use the higher **sample moments**

- The second sample moment is

$$\frac{1}{n}\sum_i X_i^2$$

as such, an observation of this moment is

$E(X^2)\ \frac{1}{n}\sum_i x_i^2$.

The method says that an estimation for the the parameters θ_1 and θ_2 of $f(x;\theta_1,\theta_2)$ can be found from the equations:

a. $E(X) = \frac{1}{n}\sum_i x_i$

b. $E(X^2) = \frac{1}{n}\sum_i x_i^2$

We can have as many equations as parameters we need to compute, incrementing the degree of the moments.

12.37 Normal distribution

If X distributes normally, we have two parameters to estimate

$$X \to N(\mu, \sigma^2)$$

1. We compute the mean and expected value of the second moment $E(X^2)$:

$E(X) = \mu$ and $E(X^2) = \sigma^2 - \mu^2$

2. We obtain the equations for the parametes where we make the expected value equal to the first sample moment, and the second moment equal to the expected value of the second moment

a. $\hat{\mu} = \frac{1}{n}\sum_i x_i$
b. $\hat{\sigma}^2 - \hat{\mu}^2 = \frac{1}{n}\sum_i x_i^2$

12.38 Normal distribution

3. We solve for the parameters

The first equation gives the estimator for the mean μ.

a. $\hat{\mu} = \frac{1}{n}\sum_i x_i$

$E(X^2)$ follows from the property: $E(X^2) = \hat{\mu}^2 + V(X) = \hat{\mu}^2 + \hat{\sigma}^2$

From the second equation we obtain

b. $\hat{\sigma}^2 = \frac{1}{n}\sum_i x_i^2 - \hat{\mu}^2$

which can also be written as:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_i (x_i - \hat{\mu})^2$$

12.39 Method of Moments

What is the estimator of parameter α for the laser cut given by the method of moments?

$$f(x;\alpha) = \begin{cases} \frac{1}{\alpha}(x-1)^{\frac{1}{\alpha}-1}, & \text{if } x \in (1,2) \\ 0, & x \notin (1,2) \end{cases}$$

Where α is a parameter.

12.40 Method of Moments

The method says that an estimator for the parameter α of $f(x;\alpha)$ can be found from the equation:

$$E(X) = \frac{1}{n}\sum_i x_i$$

for $\hat{\alpha}$

1. We compute the expected value $E(X)$

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

12.41 Method of Moments

Consider a change of variables $Z = X - 1$ then $E(X) = E(Z) + 1$ and

$E(Z) = \frac{1}{\alpha}\int_0^1 z z^{\frac{1-\alpha}{\alpha}} dz = \frac{1}{\alpha}\int_0^1 z^{1+\frac{1-\alpha}{\alpha}} dz$

$= \frac{1}{\alpha}\frac{z^{2+\frac{1-\alpha}{\alpha}}}{2+\frac{1-\alpha}{\alpha}}|_0^1 = \frac{1}{1+\alpha}$

Therefore,

$$E(X) = E(Z+1) = \frac{1}{1+\alpha} + 1$$

Substituting for $\hat{\alpha}$, the method of moments gives us the equation

$$\frac{1}{1+\hat{\alpha}} + 1 = \bar{x}$$

2. We obtain the equation for $\hat{\alpha}$ where we make the expected value equal to the first sample moment
3. We solve for $\hat{\alpha}$
$$\hat{\alpha}_m = \frac{1}{\bar{x}-1} - 1$$
4. We compute the value for our data

$\hat{\alpha}_m = 1.314$

12.42 Method of Moments

Note that this is an example for which the estimates by maximum likelihood and the method of moments are **different**

- $\hat{\alpha}_{ml} = -\frac{1}{n}\sum_{i=1}^{n}\ln(x_i - 1) = 1.320$
- $\hat{\alpha}_m = \hat{\alpha}_m = \frac{1}{\bar{x}-1} - 1 = 1.314$

We need **simulation** studies, where **we know** the true value of the parameter α, to find which of these statistics have less mean squared error.

Note: the data for 30 laser piercings were simulated with $\alpha = 2$, therefore we should prefer the maximum likelihood estimate.

To obtain better estimates of α we need to increase the size of the sample.

Chapter 13

Interval estimation

13.1 Objective

- Interval estimation for the mean and the proportion
- Interval estimation for the variance

13.2 Average or sample mean

Definition

The sample mean (or average) of a **random sample** of size n is

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

Because each random experiment is independent, the mean and variance of $\bar{X}$ are

- $E(\bar{X}) = E(X)$
- $V(\bar{X}) = \frac{V(X)}{n}$
- We call $se = \sqrt{V(\bar{X})}$ the standard error

$\bar{X}$ is therefore

- an **estimator** of $E(X)$, that is μ.
- a random variable

13.3 Inference on the average

Example:

You perform 8 random experiments: Load a cable until it breaks and record the breaking load. These are the results.

```
## [1] 13.34642 13.32620 13.01459 13.10811 12.96999 13.55309 13.75557 12.62747
```

If we **know** that our cables truly distribute as

$$X \to N(\mu = 13, \sigma^2 = 0.35^2)$$

then

$$\bar{X} \to N(13, \frac{0.35^2}{8})$$

- $E(\bar{X}) = 13$
- $V(\bar{X}) = \frac{0.35^2}{8} = 0.01530169$
- $se = \frac{0.35}{\sqrt{8}} = 0.1237$

then the **observed error** in the estimation is the difference

$$\bar{x}_{stock} - \mu = 13.21 - 13 = 0.21$$

13.4 Margin of error

When deciding whether the **error** in estimation: $\bar{X} - \mu$ is large or not we usually compare it with a predefined tolerance.

- The **margin of error** at 5% level is the distance m such that distribution of $\bar{X}$ captures 95% of the estimations:

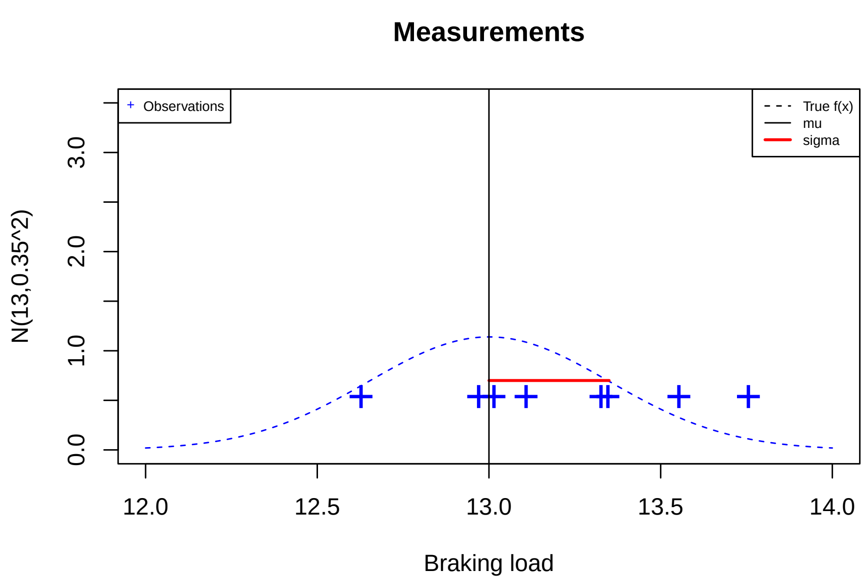
$$P(-m \leq \bar{X} - \mu \leq m) = P(\mu - m \leq \bar{X} \leq \mu + m) = 0.95$$

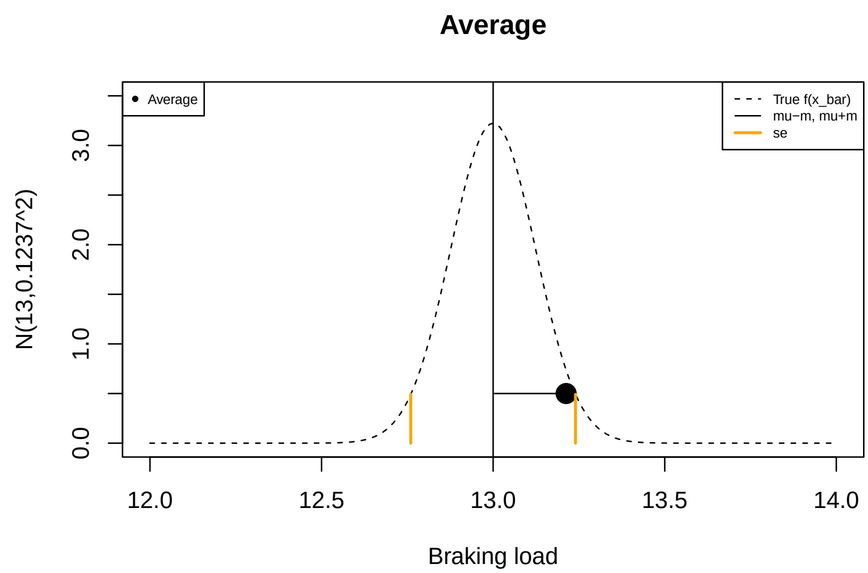
- or that 95% of the values of $\bar{X}$ are a distance m from μ.

In our example, we assume that $\bar{X}$ is normally distributed then

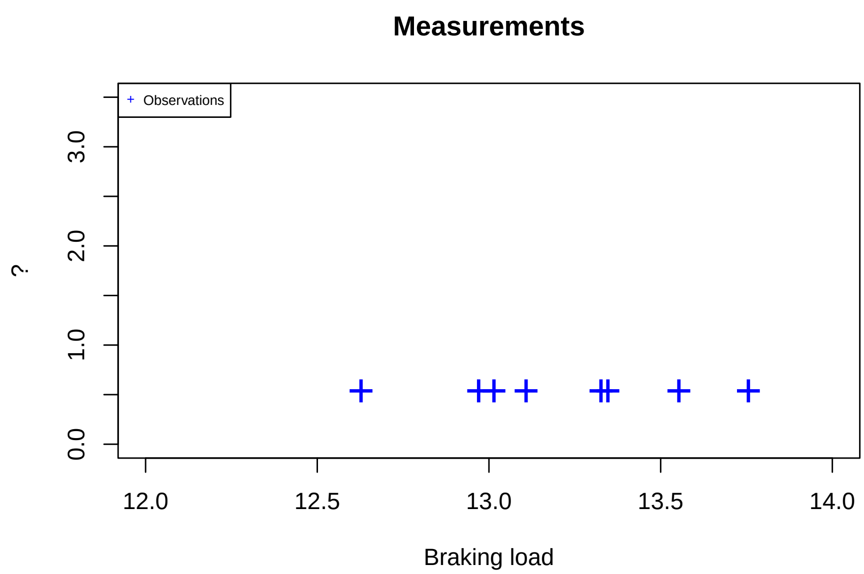
$$m = z_{0.025}\frac{\sigma}{\sqrt{n}} = 1.96 \times se = 1.96\frac{0.35}{\sqrt{8}} = 0.24$$

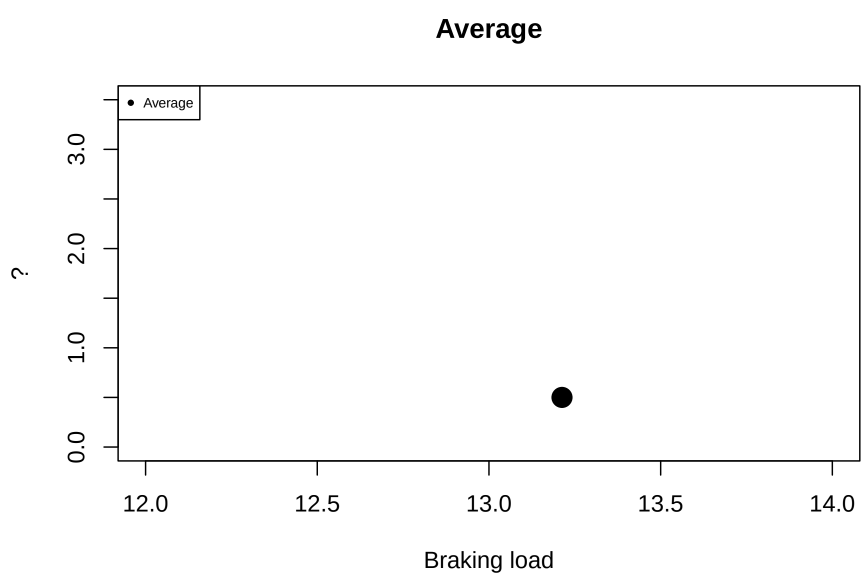
13.5 Outcome probability density Vs sample mean probability density




Average
Average
True f(x_bar)
mu−m, mu+m
se
N(13,0.1237^2)
0.0
1.0
2.0
3.0
12.0
12.5
13.0
13.5
14.0
Braking load

13.6 Real life





13.7 Interval estimation

From the margin of error equation:

$$P(-m \leq \bar{X} - \mu \leq m) = 0.95$$

let's solve for μ (the real unknown)

$$P(\bar{X} - m \leq \mu \leq \bar{X} + m) = 0.95$$

The left and right limits of the inequality are random variables which motivate the definition for the **random confidence interval at** 95%

$$(L, U) = (\bar{X} - m, \bar{X} + m)$$

This interval is a **random variable** and it has by definition a probability of 0.95 to contain μ.

13.8 Interval estimation

When we perform n-random experiments (n-sample) we can calculate m **if**

- X is normal
- we know σ^2.

The interval that we obtain from the experiment is (script size)

$$(l, u) = (\bar{x} - m, \bar{x} + m)$$

- this interval either contains or does not the parameter μ: we will **never know**!
- We say that we have a confidence of 95% that the interval (l, u) will capture the true unknown parameter μ. Think of buying a lottery ticket for which you do not know the result.

13.9 Interval estimation

In our example, we assume that $\bar{X}$ is normally distributed then

$$m = z_{0.025}\frac{\sigma}{\sqrt{n}}$$

and the 95% confidence interval is

$$(l, u) = (\bar{x} - z_{0.025}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{0.025}\frac{\sigma}{\sqrt{n}}) = (12.97, 13.45)$$

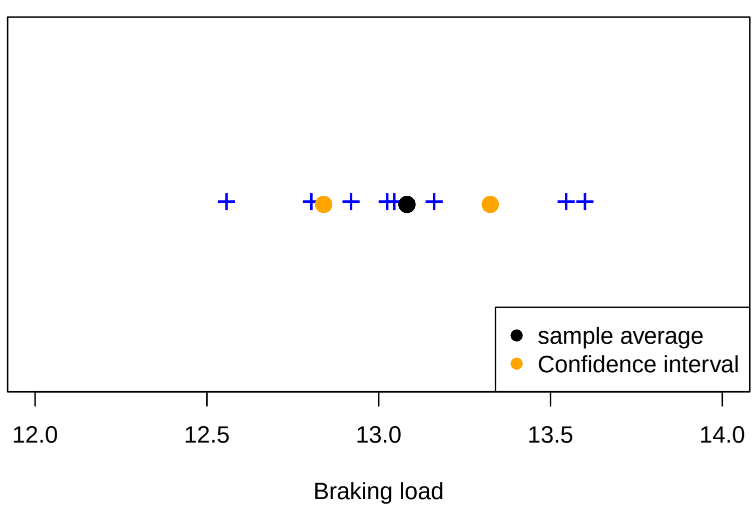
or

$$\hat{\mu} = 13.21 \pm 0.24$$

It also means that, in the estimation, we are confident about the units but not so much about the decimal places.

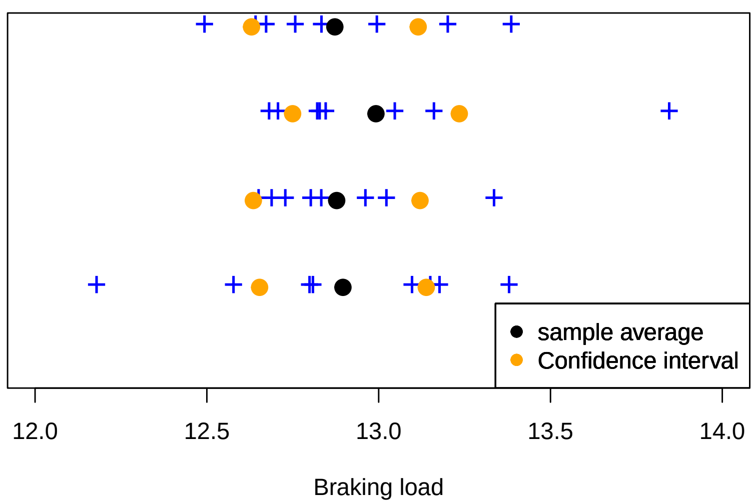
13.10 Interval estimation

For a sample of 8 observations, we have one estimate of the mean and one confidence interval



13.11 Interval estimation

Every time that we obtain a new sample then the estimates change. If we perform 100 samples of size n then 95 of the confidence intervals will contain μ (we do not know which!)



13.12 Interval estimation

We can change our confidence from 95% to 99%

- We had left out $\alpha = 0.05$ probability, 0.025 on each side.
- Now, we can leave out $\alpha = 0.01$ probability, 0.005 on each side.

Therefore the 99% confidence interval is

$(l, u) = (\bar{x} - z_{0.005}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{0.005}\frac{\sigma}{\sqrt{n}})$

$$= (\bar{x} - 2.58\frac{\sigma}{\sqrt{n}}, \bar{x} + 2.58\frac{\sigma}{\sqrt{n}})$$

or

$$\hat{\mu} = \bar{x} \pm 2.58\frac{\sigma}{\sqrt{n}}$$

If we want to be more confident then we need larger confidence intervals!

For our cables:

$$\hat{\mu} = 13.21 \pm 0.31$$

13.13 Interval estimation

13.14 Example

A metallic material is tested for impact to measure the energy required to cut it at a given temperature.

- Ten specimens of A238 steel were cut at 60ºC at the following impact energies (J)
- $64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, 64.3$
- If we know that the impact energy is normally distributed with $\sigma = 1J$ what is the 95% CI for the mean of these data?

13.15 Example

We know

- $x_i = \{64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, 64.3\}$
- $X \to N(\mu, \sigma^2)$
- $\sigma = 1J$

- $\alpha = 0.05$

The confidence interval is then

$$CI = (\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}})$$

$$= (64.46 - 1.96\frac{1}{\sqrt{10}}, 64.46 + 1.96\frac{1}{\sqrt{10}}) = (63.84, 65.08)$$

or

$$\hat{\mu} = 64.46 \pm 0.61$$

this tells us that we can be sure on the first digit (6), somewhat confident on the second (4), and unsure on the decimals (46).

What if σ^2 is **unknown**?

13.16 T-statistic

When

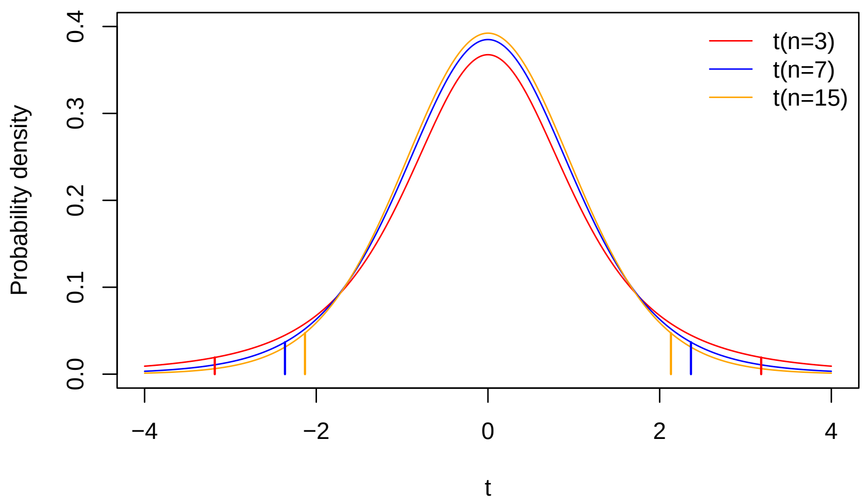
- X is normal, and
- σ is **unknown**

then the standardized statistic

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

Follows a t-distribution with $n - 1$ degrees of freedom, and we can compute probabilities for $\bar{X}$.

13.17 T-statistic



13.18 T-statistic

To compute the margin of error m at 5% level when

- X is normal and
- σ^2 is unknown

We use the t-distribution

$P(\mu - m \leq \bar{X} \leq \mu + m)$

$$= P(-\frac{m}{s/\sqrt{n}} \leq T \leq \frac{m}{s/\sqrt{n}}) = 0.95$$

$$m = t_{0.025,n-1}\frac{s}{\sqrt{n}}$$

where $t_{0.025,n-1}$ is the value T that leaves 2.5% of probability at the right hand side of the t-distribution with $n-1$ degrees of freedom (0.025-quantile)

the 95% confidence interval is then

$$(l,u) = (\bar{x} - t_{0.025,n-1}\frac{s}{\sqrt{n}}, \bar{x} + t_{0.025,n-1}\frac{s}{\sqrt{n}})$$

in R: $t_{0.025,n-1}$=qt(1-0.025, n-1)

13.19 Example

A metallic material is tested for impact to measure the energy required to cut it at a given temperature.

- Ten specimens of A238 steel were cut at 60ºC at the following impact energies (J)
- $64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, 64.3$
- If we know that the impact energy is normally distributed but we **do not know** the variance what is the 95% CI for the mean of these data?

13.20 Example

- $\bar{x} = 64.46$
- $s = 0.227$
- $\alpha = 0.05$
- $t_{0.025,9} = 2.26$ obtained from $P(T \leq t_{0.025,9}) = 0.975$; qt(1-0.025, 9)

The confidence interval is then

$CI = (\bar{x} - t_{0.025,9}\frac{s}{\sqrt{n}}, \bar{x} + t_{0.025,9}\frac{s}{\sqrt{n}})$

$$= (64.46 - 2.26\frac{0.227}{\sqrt{10}}, 64.46 + 2.26\frac{0.227}{\sqrt{10}})$$
$$= (64.29, 64.62)$$

We have seen a larger $CI = (63.84, 65.08)$ when $\sigma = 1$. Data then suggests $\sigma < 1$.

R: t.test(c(64.1,64.7,64.5,64.6,64.5,64.3,64.6,64.8,64.2,64.3))

13.21 IC with CLT

If

- we do not know how X distributes but
- we take a large sample $n \geq 30$

Then we can use the CLT to find the confidence intervals.

the 95% confidence interval is then

$$(l, u) = (\bar{x} - z_{0.025}\frac{s}{\sqrt{n}}, \bar{x} + z_{0.025}\frac{s}{\sqrt{n}})$$

since $t_{0.025,n-1} \to z_{0.025}$ for $n \to \infty$ then it is also ok to use the T distribution in this case.

Note: This is why R only implements t.test and not z.test in the base functions to compute CI.

Example:

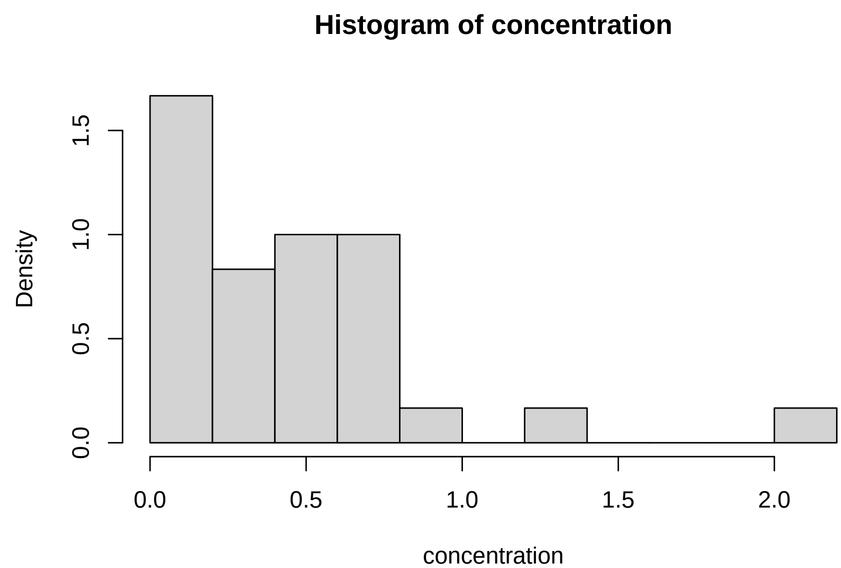
Consider an experiment where we measure the concentration in blood of a drug after 10-hour administration in 30 patients. We obtain the following results:

```
##  [1] 0.42172863 0.28830514 0.66452743 0.01578868 0.02810549 0.15825061
##  [7] 0.15711365 0.07263340 1.36311823 0.01457672 0.50241503 0.24010736
## [13] 0.14050681 0.18855892 0.09414202 0.42489306 0.78160177 0.23938021
## [19] 0.29546742 2.02050586 0.42157487 0.48293561 0.74263790 0.67402224
## [25] 0.58426449 0.80292617 0.74837143 0.78532627 0.01588387 0.29892485
```

- the average is $\bar{x} = 0.4556198$

- the standard deviation is $s = 0.4335571$

- the histogram of the results is:



13.22 Central Limit Theorem

We **assumed** that $X \to exp(\lambda = 2)$

With mean and variance:

- $E(X) = \frac{1}{\lambda}$
- $V(X) = \frac{1}{\lambda^2}$

What is the confidence interval for the mean $E(X) = \mu$?

- We use a 95% CI to estimate it

Since $n \geq 30$ we can use the CLT

$$\bar{X} \sim_{aprox} N(\lambda, \frac{1}{n\lambda^2})$$

and the 95% confidence interval is then

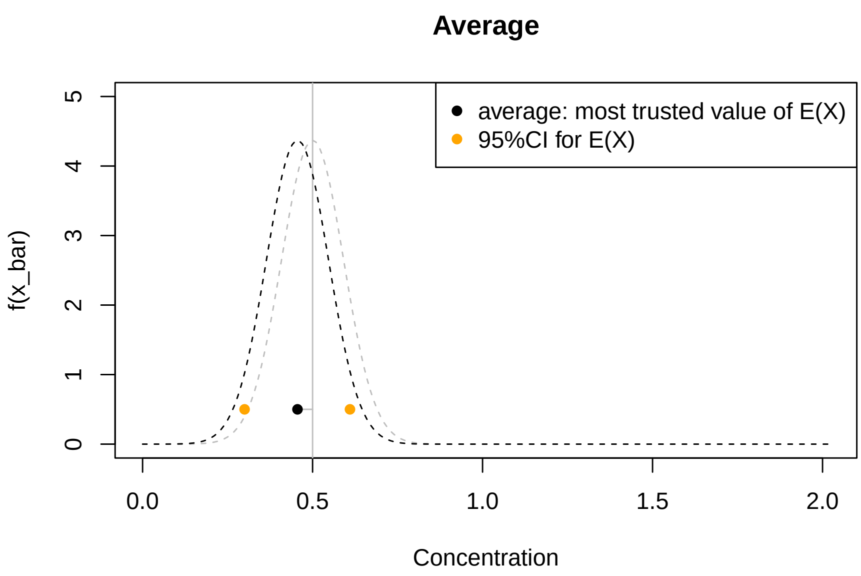
$$(l, u) = (\bar{x} - z_{0.025}\frac{s}{\sqrt{n}}, \bar{x} + z_{0.025}\frac{s}{\sqrt{n}})$$

$(l, u) = (0.4556198 - 1.96\frac{0.4335571}{\sqrt{30}}, 0.4556198 + 1.96\frac{0.4335571}{\sqrt{30}})$

$$= (0.300, 0.610)$$

or

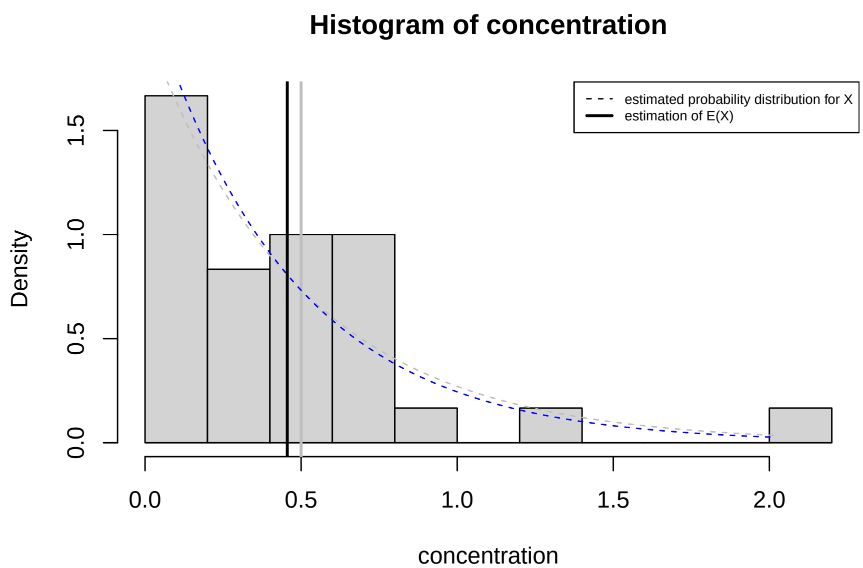
$$\hat{\mu} = 0.45 \pm 0.15$$



13.23 Parameter estimation

Since $E(X) = \mu = \frac{1}{\lambda}$ then

$$\hat{\lambda} = \frac{1}{\hat{\mu}} = 2.194812$$



or its 95% CI:

$$\hat{\lambda} = (1.66, 3.33)$$

13.24 Interval estimation for proportions

A random sample of 400 patients was selected for testing a new vaccine for the influenza virus, after 6 months of vaccination 136 were ill.

- What is the expected efficacy of the vaccine?

We have 136 failures in 400 trials, each trial is a Bernoulli trial

$$X \to Bernoulli(p)$$

with:

- the probability p of failure for one person $(x = 1)$
- mean $E(X) = p$
- variance $V(X) = p(1 - p)$

We want to have a 95% CI for p.

13.25 Interval estimation for proportions

If the distribution of a random experiment is

$$X \to Bernoulli(p)$$

Then $\bar{X}$ has

- mean $E(\bar{X}) = E(X) = p$ (unbiased estimator of p)
- variance $V(\bar{X}) = \frac{V(X)}{n} = \frac{p(p-1)}{n}$ (consistent estimator of p)

$$\hat{p} = \bar{x}$$

13.26 Interval estimation for proportions

When $\hat{p}n > 5$ and $(\hat{p} - 1)n > 5$

- The **standardized statistic** of $\bar{X}$ can be approximated by a standard distribution

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{\bar{X} - p}{\left[\frac{p(1-p)}{n}\right]^{1/2}} \to N(0,1)$$

- The 95% CI interval of p is:

$$CI = (l, u) = (\bar{x} - z_{0.025}\left[\frac{\bar{x}(1-\bar{x})}{n}\right]^{1/2}, \bar{x} + z_{0.025}\left[\frac{\bar{x}(1-\bar{x})}{n}\right]^{1/2})$$

Where we estimate the Bernoulli variance $p(1-p)$ by $\bar{x}(1-\bar{x})$.

13.27 Interval estimation for proportions

In our case, we are counting failures on vaccinations 136 in 400 trials

we know

- $\bar{x} = 134/400 = 0.34$
- $z_{0.025} = 1.96$

$CI = (l, u) = (\bar{x} - 1.96[\frac{\bar{x}(1-\bar{x})}{n}]^{1/2}, \bar{x} + 1.96[\frac{\bar{x}(1-\bar{x})}{n}]^{1/2})$

$$= (0.29, 0.39)$$

The probability of failure of the vaccine is

$$\hat{p} = 0.34 \pm 0.05$$

Note: Polls for the intention to vote (Bernoulli trial) in a sample of n individuals report this type of estimate with its margin of error. It does not mean that the **true value** of p is within this interval with probability 95%.

13.28 Interval estimation for the variance

A metallic material is tested for impact to measure the energy required to cut it at a given temperature.

- Ten specimens of A238 steel were cut at 60ºC at the following impact energies (J)
- $64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, 64.3$

We know that the estimate for $s^2 = 0.227^2 = 0.051$, but what is its confidence interval?

13.29 Interval estimation for the variance

If

- X is a **normal** variable

then

$$W = \frac{S^2(n-1)}{\sigma^2}$$

Captures the proportion in the error of σ^2 and follows a χ^2 distribution with $n-1$ degrees of freedom

$$\frac{S^2}{\sigma^2}(n-1) \rightarrow \chi^2_{n-1}$$

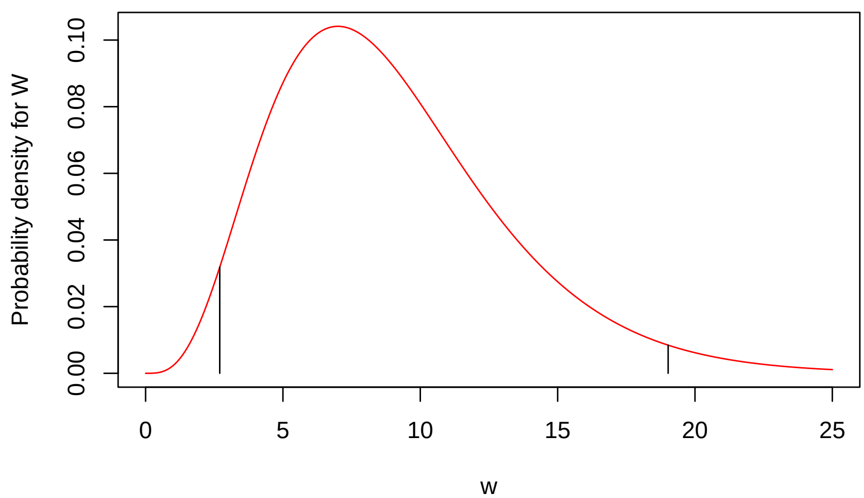
- We look for confidence interval of σ^2 at confidence 95% (L, U) such that

$$P(L \leq \sigma^2 \leq U) = 0.95$$

We can use the χ^2 to determine the 95% of the distribution about W

$$P(\chi^2_{0.975,n-1} \leq W \leq \chi^2_{0.025,n-1}) = 0.95$$

13.30 χ^2-statistic



13.31 Interval estimation for the variance

replacing the value of W

$$P(\chi^2_{0.975,n-1} \leq \frac{S^2}{\sigma^2}(n-1) \leq \chi^2_{0.025,n-1}) = 0.95$$

and solving for σ^2

$$P(\frac{S^2(n-1)}{\chi^2_{0.025,n-1}} \leq \sigma^2 \leq \frac{S^2(n-1)}{\chi^2_{0.975,n-1}}) = 0.95$$

The **random** interval at 95% confidence is

$$(L, U) = (\frac{S^2(n-1)}{\chi^2_{0.025,n-1}}, \frac{S^2(n-1)}{\chi^2_{0.975,n-1}})$$

and the **observed** 95% confidence interval (script size)

$$(l, u) = (\frac{s^2(n-1)}{\chi^2_{0.025,n-1}}, \frac{s^2(n-1)}{\chi^2_{0.975,n-1}})$$

13.32 Interval estimation for the variance

$\chi^2_{0.975,n-1} = F^{-1}(0.025)$ for $n = 10$ or $df = n - 1 = 9$

```
chi0.975 <- qchisq(0.025, df=9)
chi0.975
```

```
[1] 2.700389
```

```
chi0.025 <- qchisq(0.975, df=9)
chi0.025
```

```
[1] 19.02277
```

13.33 Interval estimation

In our example

- $s = 0.227$
- $n = 10$

$\hat{\sigma}^2 = (l, u) = (\frac{s^2(n-1)}{\chi^2_{0.025,n-1}}, \frac{s^2(n-1)}{\chi^2_{0.975,n-1}})$

$$= (\frac{0.227^2(10-1)}{19.02277}, \frac{0.227^2(10-1)}{2.700389}) = (0.02, 0.17)$$

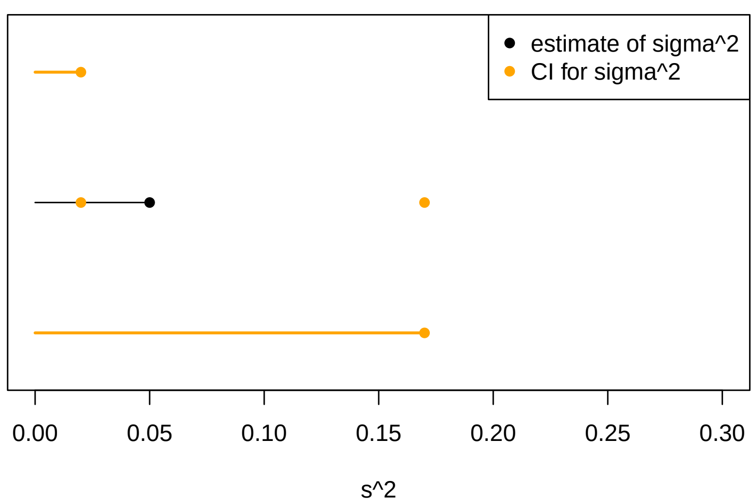
According to the data $\sigma^2 \neq 1$ at 95% confidence.

- Had we made an error considering $\sigma = 1$ when we calculated the first CI for this data?

in R: library(Ecfun); confint.var(0.05, 9)

13.34 Interval estimation

The interval for the variance is **not symmetric** and we cannot formulate it as an estimate $\pm$ margin of error.



Chapter 14

Exercises

14.1 Data description

14.1.0.1 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?

14.1.0.2 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.

14.2 Probability

14.2.0.1 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? (R/0.3)
- What is the probability of not being misophonic **and** not depressed? (R/0.293)
- What is the probability of being misophonic **or** depressed? (R/0.293)
- What is the probability of being misophonic **and** depressed? (R/0.707)
- Describe in English the outcomes with probability 0.

14.2.0.2 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? (R/0.6)
- What is the marginal frequency of being alive? (R/0.5)
- What is the frequency of being alive **or** female? (R/0.6)

14.3 Conditional Probability

14.3.0.1 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? (R: 2/207)
- What is the estimated probability that the machine produces a rod that does not satisfy at least one quality control?(R: 7/207)
- What is the estimated probability that the machine produces rounded and smoothed surfaced rods? (R: 200/207)
- what is the estimated probability that the rod is rounded if the rod is smooth? (R: 201/201)
- what is the estimated probability that the rod is smooth if it is rounded? (R: 201/204)
- what is the estimated probability that the rod is neither smooth nor rounded if it does not satisfy at least one quality control? (R: 2/7)

- Are smoothness and roundness independent events? (no)

14.3.0.2 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? (R: 0.30)

14.3.0.3 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? (R: 357/370)
- What is the probability that the rod has been produced by machine 1? (R: 207/370)
- what is the probability that the rod is not smooth? (R: 20/370)
- What is the probability that the rod is smooth or rounded or produced by machine 1? (R: 364/370)
- What is the probability that the rod is rounded if it is smoothed and from machine 1? (R: 200/201)
- What is the probability that the rod is not rounded if it is not smoothed and is from machine 2? (R: 6/8)
- what is the probability that the rod has come from machine 1 if it it is smoothed and rounded? (R: 200/345)
- what is the probability that the rod has come from machine 2 if it does not pass at least one of the quality controls? (R:0.72)

14.3.0.4 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? (R:0.09)
- What is the probability of having to stop at at least one traffic light?(R:0.7)
- What is the probability of having to stop at only one traffic light? (R:0.61)
- If I stopped at the second traffic light, what is the probability that I had to stop at the first one? (R: 0.62)
- If I had to stop at any traffic light, what is the probability that I had to do it twice? (R: 0.12)
- Is stopping at the first traffic light an independent event from stopping at the second traffic light? (no)

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? (R:0.013)
- What is the probability of having to stop at at least one traffic light? (R:0.775)
- What is the probability of having to stop at only one traffic light? (R:0.5425)

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: $P(R_1, R_2, R_3) = P(R_3|R_2)P(R_2|R_1)P(R_1)$

14.3.0.5 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 $P(D)$? (R:0.1)
- What is the probability that a brick that has passed the test is really defective? (R:0.022)
- The probability that a brick is not defective **and** that it does not pass the test (R:0.009)
- Are D and $\bar{E}$ statistical independent? (no)

14.4 Random variables

14.4.0.1 Exercise 1

Given the probability mass function

x	$f(x) = P(X = x)$
10	0.1
12	0.3
14	0.25
15	0.15
17	?
20	0.15

- what is its expected value and standard deviation? (R: 14.2; 2.95)

14.4.0.2 Exercise 2

Given the probability distribution for a discrete variable X

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

- find $f(x)$
- find $E(X)$ and $V(X)$ (R:1; 1.5)
- what is the expected value and variance of $Y = 2X + 3$ (R: 6)
- what is the median and the first and third quartiles of X? (R:2,0,2)

14.4.0.3 Exercise 3

We are testing a system to transmit digital pictures. We first consider the experiment of sending 3 pixels and having as **possible** outcomes events such like $(0,1,1)$. This is the event of receiving the first pixel with no error, the second with error and third with error.

- List in one column the sample space of the random experiment.

- In the a second column assign the random variable that counts the number of errors transmitted for each outcome

Consider that we have a totally noisy channel, that is any outcome of three pixels is equally likely.

- What is the probability of receiving 0, 1, 2, or 3 errors in the transmission of 3 pixels? (R: 1/8; 3/8; 3/8; 1/8)

- Sketch the probability mass function for the number of errors

- What is the expected value for the number of errors? (R:1.5)

- What is its variance? (R: 0.75)

- Sketch the probability distribution

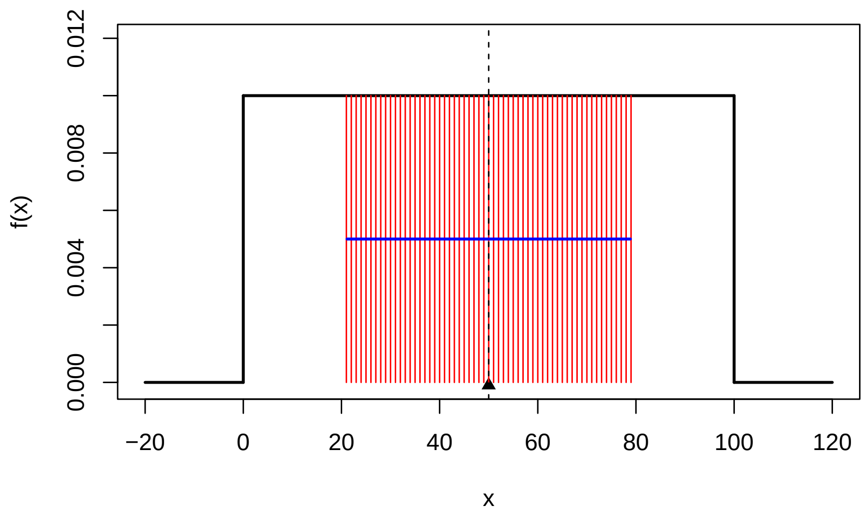
- What is the probability of transmitting at least 1 error? (R:7/8)

14.4.0.4 Exercise 4

- for the probability density

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

- compute the mean (R:50)
- compute variance using $E(X^2) = V(X) + E(X)^2$ (R:100^2/12)
- compute $P(\mu - \sigma \leq X \leq \mu + \sigma)$ (R: 0.57)
- What are the first and third quartiles? (R: 25; 75)



14.4.0.5 Exercise 5

Given

$$f(x) = \begin{cases} 0, & x < 0 \\ ax, & x \in [0,3] \\ b, & x \in (3,5) \\ \frac{b}{3}(8-x), & x \in [5,8] \\ 0, & x > 8 \end{cases}$$

- What are the values of a and b such that $f(x)$ is a continous probability density function? (R: 1/15; 1/5)
- what is the mean of X? (R:4)

14.4.0.6 Exercise 6

For the probability density

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

- Confirm that this is a probability density
- Compute the mean (R: $1/\lambda$)
- Compute the expected value of X^2 (R: $2/\lambda^2$)
- Compute variance (R: $1/\lambda^2$)

- Find the probability distribution $F(a)$ (R: $1 - exp(-\lambda a)$)
- Find the median (R: $\log 2/\lambda$)

14.4.0.7 Exercise 7

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:

- $P(X > 0)$ (R:63/80)
- $E(X)$ (R:1.93)
- $P(X > 0|X < 2)$ (R:28/45)

14.5 Probability Models

14.5.0.1 Exercise 1

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?(R: 0.240)
- What is the probability that a family has only one girl?(R: 0.259)
- What is the probability that a family has only one boy or only one girl?(R: 0.4999)
- What is the probability that the family has at least two boys?(R: 0.7023)
- What is the number of children that a family should have such that the probability of having at least one girl is more than 0.75?(R:$n = 3 > \log(0.25)/\log(0.51)$)

14.5.0.2 Exercise 2

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?(R:0.1385651)

- What is the probability that the engine successfully completes 15 searches before the first failure?(R:0.020)

- We consider that a search engine works sufficiently well when it is able to find information for moe than 10 requests for every 2 failures. What is the probability that in a reliability trial our search engine is satisfactory?(R: 0.697)

14.5.0.3 Exercise 3

The average number of radioactive particles hitting a Geiger counter in a nuclear energy plant under control is 2.3 per minute.

- What is the probability of counting exactly 2 particles in a minute? (R:0.265)
- What is the probability of detecting exactly 10 particles in 5 minute? (R:0.112)
- What is the probability of at least one count in two minutes? (R:0.9899)
- What is the probability of having to wait less than 1 second to detect a radioactive particle, after we switch on the detector? (R:0.037)
- We suspect that a nuclear plant has a radioactive leak if we wait less than 1 second to detect a radioactive particle, after we switch on the detector. What is the probability that when we visit in 5 plants that are under control, we suspect that at least one has a leak? (R:0.1744).

14.5.0.4 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? (R:0.841)
- What is the probability that a man's height is between 165cm and 185cm? (R:0.682)
- What is the height that defines the 5% of the smallest men? (R:158.55)

14.6 Sampling and Central Limit Theorem

14.6.0.1 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 batteries, what is the probability that the difference in charge between the sample average and the mean charge is at most 5%? (R:0.9044)
- If we charge 100, what is that probability? (R:0.9991)
- If, instead we only charge 9 batteries, which value c is surpassed by the sample mean with probability of 0.015? (R:85.850)

14.6.0.2 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? (R:0.1863)
- How many components do we need such that the telescope is operational 2750 consecutive hours with at least 0.95 probability? (R:31)

14.6.0.3 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?

14.6.0.4 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?

14.6.0.5 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.

14.7 Point Estimators

14.7.0.1 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 $N \to \infty$

14.7.0.2 Exercise 2

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

14.8 Maximum likelihood

14.8.0.1 Exercise 1

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?

- Compute $E(X) = \mu$ as a function of θ. What is the maximum likelihood estimate for μ?

14.8.0.2 Exercise 2

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?

14.8.0.3 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?

- What is the maximum likelihood estimate of the parameter $\alpha = \frac{n}{\lambda}$
- Is α an unbiased and consistent estimator of the mean of the sample sum $E(Y)$, where $Y = \sum_1^n X_i$?

14.9 Method of moments

14.9.0.1 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

14.9.0.2 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 $x_1 = 0.92;\quad x_2 = 0.79;\quad x_3 = 0.90;\quad x_4 = 0.65;\quad x_5 = 0.86$

What is the estimated value of the parameter θ?

14.9.0.3 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

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

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:

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

14.10 Confidence intervals

14.10.0.1 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

14.10.0.2 Exercise 2

compute 95% CI the mean of a normal variable with known variance $\sigma^2 = 9$ and $\bar{x} = 22$, using a sample of size 36.

14.10.0.3 Exercise 3

This year, 17 from 1000 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 significant drop in influenza complications?

14.11 Hypothesis testing

14.11.0.1 Exercise 1

Imagine we take a random sample of size $n = 41$ of a normal random variable X, and find that the sample average is 10 and the sample variance is 1.5.

- What is then the confidence interval for the mean of X at 95% confidence level?

Consider that $t_{0.025,40}$ = qt(0.975, 40) $\sim$ 2.

- Test the hypothesis that the mean of X is **different** than 10.5, using a 5% significance threshold.
- Write the code to calculate the P-value to test the hypothesis that the mean of μ is **lower** than 10.5, using a 5% significance threshold.

Consider that the code for the T probability distribution with $n - 1$ degrees of freedom is pt(tobs, n-1).

14.11.0.2 Exercise 2

10 gas condensates showed the following concentrations of mercury (in ng/ml):

23.3, 22.5, 21.9, 21.5, 19.9, 21.3, 21.7, 23.8, 22.6, 24.7

Assuming that the mercury concentration is distributed normally a across gas condensates, test the hypothesis that a condensate does not surpass the toxicity limit established at $24 ng/ml$.

14.11.0.3 Exercise 3

The manufacturer of gene expression microarrays guarantees that at least 97% of the microarrys they produce have high quality signals. A customer receives a batch of 200 pieces and finds that 8 unperformed.

Should the costumer return the lot due to poor quality?