

Introduction to Statistics for Astronomers and Physicists

Section 2a: Fundamentals of Probability

Dr Angus H Wright

2022-02-09

Section 2: Introduction

Section 2: Probability & Decision Making (Weeks 3-5)

For all aspects of modern science, an understanding of probability is required. We cover a range of topics in probability, from decision theory and the fundamentals of probability theory, to standard probabilistic distributions and their origin. From this module, students will gain an insight into different statistical distributions that govern modern observational sciences, the interpretation of these distributions, and how one accurately models distributions of data in an unbiased manner.

Topics include:

- Decision theory
- Fundamentals of probability
- Statistical distributions and their origins

Reminder of Set Notation

In this section, we will explore the fundamentals of probability. For this, we will use set notation quite heavily. The important things to remember about set notation are:

- A set X is defined as a collection of items, grouped by curly brackets:

$$X = \{x_1, x_2, \dots, x_n\}$$

- The size (or cardinality) of a set is given by $|X|$.
- A set with cardinality of 0 is the empty set, which has the symbol \emptyset .

Events and Outcomes

Given an experiment, there can be a set of possible results. These results are known as **outcomes**. Every run of an experiment will produce one, *and only one*, outcome.

Sample Space

When we toss a coin, there are two possible outcomes: Heads (H) or Tails (T). This is therefore the total set of available outcomes, known as the **sample space** (Ω).

$$\Omega = \{H, T\}$$

If we were to toss our coin twice, the set of available outcomes increases:

$$\Omega = \{HH, HT, TH, TT\}$$

Now: consider a game which only ends when you throw a head *and then* a tail. What is the sample space of outcomes of this game?

•

$$\Omega = \{HT, HHT, THT, HHHT, THHT, TTHT, HHHHT, \dots\}$$

This sample space is infinite, but is still totally valid.

Probability derives from Outcomes

The available outcomes of our single coin-toss experiment are:

$$\Omega = \{H, T\}$$

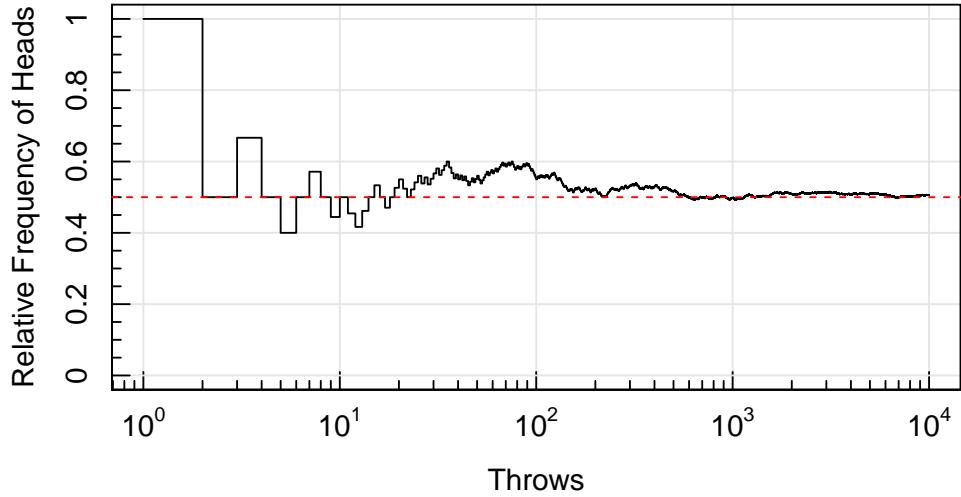
If we were to perform this experiment N times (i.e. run N trials), we could record the number of occurrences of each outcome A . We can then observe the relative frequency of each outcome:

$$f_A = \frac{\#(A)}{N}$$

We can perform this experiment using a simulated coin toss in **R**:

If we plot this as a running “Relative Frequency of Heads”:

```
vals<-rbinom(1E4,1,0.5)
magplot(1:length(vals),cumsum(vals)/1:length(vals),
        type='s',ylab='Relative Frequency of Heads',xlab='Throws',
        ylim=c(0,1),log='x')
abline(h=0.5,lty=2,col='red')
```



We can see that, as the number of throws becomes large, the relative frequency tends towards 50%. With this knowledge, we can define the probability of an outcome A as the relative frequency of the outcome after infinitely many trials:

$$P(A) = \lim_{N \rightarrow \infty} \frac{\#(A)}{N}.$$

This gives us some information about probability already.

- Because probability derives from relative frequency of observations:

$$0 \leq P(A) \leq 1$$

Furthermore, when we look at the probability of all possible outcomes (i.e. $A_i \in \Omega$):

- Ω is the sample space of all outcomes, so the sum of all relative frequencies must be 1:
 $\sum_{A_i \in \Omega} P(A_i) = 1$.

Probability Trees

Outcomes of experiments can be visualised using probability trees, that summarise all possible outcomes (and strings of outcomes).

```
library(DiagrammeR)
mermaid(``
graph LR
    Start --> H[H, P=0.5]
    Start --> T[T, P=0.5]
    H --> HH[H, P=0.5]
    H --> HT[T, P=0.5]
    T --> TH[H, P=0.5]
    T --> TT[T, P=0.5]
    HH --> HHout[HH, P=0.25]
    HT --> HTout[HT, P=0.25]
    TH --> THout[TH, P=0.25]
    TT --> TTout[TT, P=0.25]
``)
```

Rolling a Die

Lets increase the complexity of our experiment, by using a standard 6-sided die instead of a coin. The sample space of observations for the die is:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

We can simulate a roll of a fair die in **R** in a few ways, but using `sample` is the most intuitive:

```
#Simulating a dice roll
sample(1:6, size=1)
```

```
## [1] 5
```

We can simulate many rolls of the die by increasing the value of `size`, and specifying that `replace=TRUE`:

```
#Simulating many dice rolls
sample(1:6, size=30, replace=TRUE)
```

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

The `replace` function is best understood using a different analogy: Consider that, instead of a die, we had a sack with marbles numbered 1 – 6. Our “roll” in this situation is to pull one marble from the sack. The `replace` option specifies that, after we’ve read the number on our marble, we want to put it back in the bag prior to performing our next “roll”. Said another way: each roll of the dice *doesn’t know* about the outcomes of the previous rolls. The rolls are **independent**, and the outcomes of each roll are **disjoint**.

The probability of each number appearing on our fair die is 1/6:

```
#Probability with Dice
table(sample(1:6, size=1E4, replace=TRUE))/1E4
```

```
##
##      1      2      3      4      5      6
## 0.1649 0.1681 0.1671 0.1642 0.1639 0.1718
```

But what is the probability, say, of rolling a prime number?

Events

An **event** is defined as a set of outcomes. Take our example of the prime number roll on a dice. Let’s call our event \mathcal{E} :

$$\mathcal{E} = \{2, 3, 5\}$$

The probability of observing event \mathcal{E} is the sum of the probabilities of observing each of the independent outcomes within the event:

$$P(\mathcal{E}) = \sum_{A_i \in \mathcal{E}} P(A_i)$$

If the event contains all possible outcomes (that is, the event space is the sample space: $\mathcal{E} = \Omega$), then we recover our earlier summation

$$P(\Omega) = 1.$$

Given the above two properties, observing any event that is not A , which is denoted as the compliment of A :

$$P(A^\perp) = 1 - P(A).$$

Given the complementarity rule, if the event space contains no outcomes (that is, it is the empty set: $\mathcal{E} = \Omega^\perp = \emptyset$), then the probability of the event is 0:

$$P(\emptyset) = 1 - P(\Omega) \quad (1)$$

$$= 0. \quad (2)$$

So, what is the probability of observing a prime number when we roll a fair die:

$$P(\mathcal{E} \in \{2, 3, 5\}) = \sum_{A_i \in \mathcal{E}} P(A_i) \quad (3)$$

$$= P(2) + P(3) + P(5) \quad (4)$$

$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \quad (5)$$

$$\therefore P(\mathcal{E}) = 0.5 \quad (6)$$

Rolling two Dice

Lets now complicate the sample space further by looking at the outcome of rolling two dice at the same time, and summing together the results.

Firstly, let's look at the sample space of this experiment.

As before, each die has the outcomes $A = \{1, 2, 3, 4, 5, 6\}$. The sample space of the two-dice roll is the set of all possible **ordered combinations** or **permutations** of two draws from these values.

We can construct this sample space by hand:

$$\Omega = \{1 + 1, 1 + 2, 1 + 3, 1 + 4, 1 + 5, 1 + 6, \quad (7)$$

$$2 + 1, 2 + 2, 2 + 3, 2 + 4, 2 + 5, 2 + 6, \quad (8)$$

$$\vdots \quad (9)$$

$$6 + 1, 6 + 2, 6 + 3, 6 + 4, 6 + 5, 6 + 6\} \quad (10)$$

We can also construct the sample space in **R** using the `expand.grid` function, which produces all possible permutations of two vectors:

```
expand.grid(1:6,1:6)
```

```
##      Var1 Var2
## 1      1    1
## 2      2    1
## 3      3    1
## 4      4    1
## 5      5    1
## 6      6    1
## 7      1    2
## 8      2    2
## 9      3    2
## 10     4    2
## 11     5    2
## 12     6    2
## 13     1    3
## 14     2    3
## 15     3    3
## 16     4    3
## 17     5    3
## 18     6    3
```

```

## 19   1   4
## 20   2   4
## 21   3   4
## 22   4   4
## 23   5   4
## 24   6   4
## 25   1   5
## 26   2   5
## 27   3   5
## 28   4   5
## 29   5   5
## 30   6   5
## 31   1   6
## 32   2   6
## 33   3   6
## 34   4   6
## 35   5   6
## 36   6   6

```

With two fair die, the probability of observing each of these outcomes is equal:

$$P(A_i) = \frac{1}{|\Omega|} \quad (11)$$

$$= \frac{1}{36} \quad (12)$$

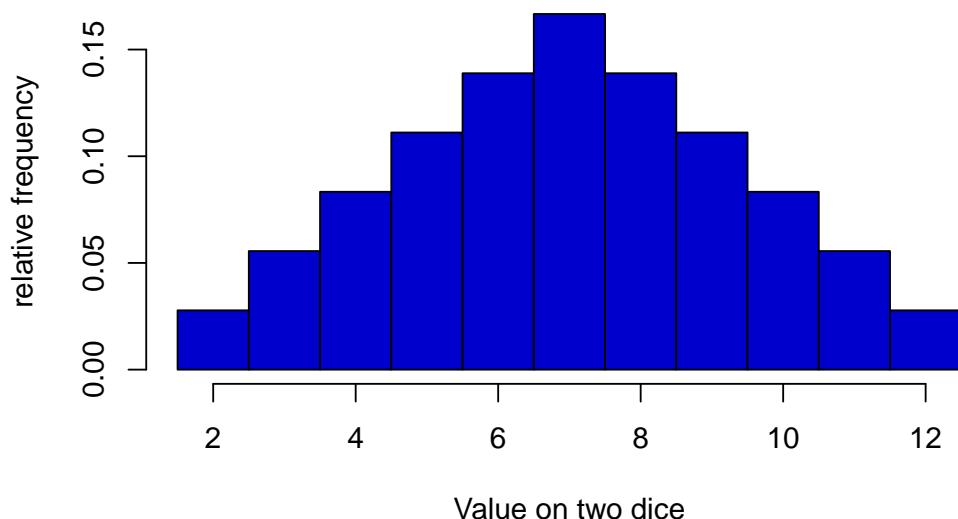
However, our game specified that we wanted to calculate the *sum* of the dice, which changes the event space we are interested in. The summation obviously doesn't care whether we observe $1 + 4$ or $4 + 1$ on our dice, it only cares that we have observed the *event*

$$\mathcal{E}_j = 5$$

What is the probability, then, of all distinct *events* in our two-dice roll?

```
hist(rowSums(expand.grid(1:6,1:6)),breaks=seq(1.5,12.5,by=1),
  xlab='Value on two dice',ylab='relative frequency',freq=FALSE,
  col='blue3',main='2 Dice Event Probabilities')
```

2 Dice Event Probabilities

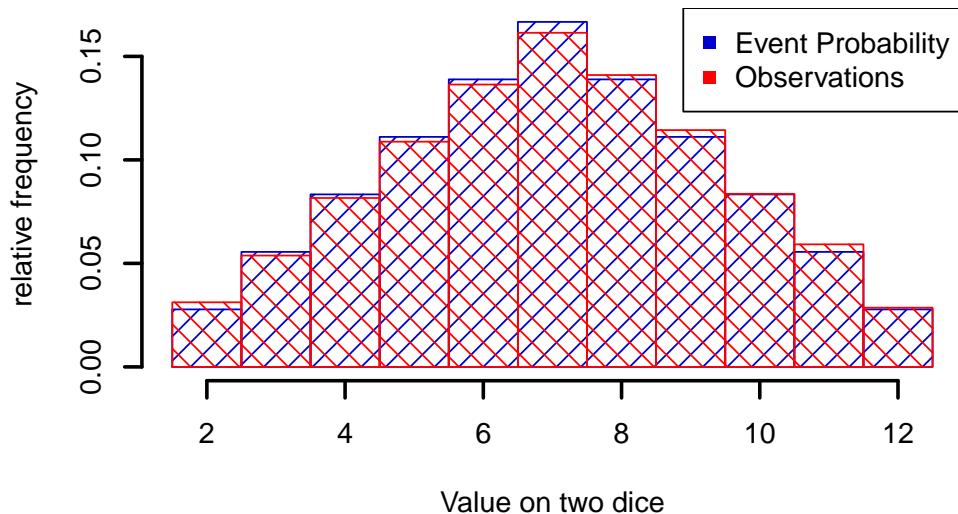


This distribution therefore demonstrates the probability of observing each event (defined as a distinct summed value on the dice).

We can verify this using our simple simulation of dice rolls:

```
#Probability with Dice
hist(rowSums(expand.grid(1:6,1:6)),breaks=seq(1.5,12.5,by=1),
      xlab='Value on two dice',ylab='relative frequency',freq=FALSE,
      col='blue3',density=10,lwd=2,main='2 Dice Event Probabilities')
hist(rowSums(matrix(sample(1:6,size=1E4,replace=TRUE),ncol=2)),
      breaks=seq(1.5,12.5,by=1),lwd=2,
      col='red',density=10,angle=-45,freq=FALSE,add=TRUE)
legend('topright',legend=c("Event Probability","Observations"),
      col=c("blue3","red"),pch=15)
```

2 Dice Event Probabilities



Joint Probability

The **joint probability** of two outcomes is the probability of observing both outcomes at the same time. With our two dice, the joint probability of any two numbers was

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

- Observations are **independent** if and only if

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

Independence

Independence is an important concept in probability, and one that can lead to incorrect inference. Let's look at two examples:

Modelling independence

Consider the case where you are on the jury of a criminal trial. A criminal was identified by matching a sample of DNA to a database of many thousands of people.

The probability of incorrectly matching DNA to a random person is 0.01%, and such incorrect/chance matches are independent. There are 20 000 people in the DNA database. The police find a match, and take the matching person to trial.

The prosecution stands before the jury and says they have *damning* evidence. The probability of the DNA match being wrong is 0.01%, and so there is a 99.99% chance that this person is guilty.

Independence and Trends

You stand at a roulette table in a Casino, where you want to bet on whether the ball will land on red or black. The roulette table has a record of the last 25 rolls:

```
##  
##   Red Black  
##   25     0
```

You know that the probability of rolling “Black” and “Red” is equal:

$$P(B) = P(R)$$

. You know that, in the long term, the relative frequency of black and red should be the same. So you conclude that, therefore, you ought to bet on “Black”.

Independent Probabilities

Let’s look at our criminal case first.

We want to ask the question: what is the probability that we observed *at least 1 match in our database* purely by chance?

$$P(\text{chance match in database}) = 1 - P(\text{no chance match in the database})$$

$$P(\text{no chance match in the database}) = P(\text{no chance match with person 1}) \cap \quad (13)$$

$$P(\text{no chance match with person 2}) \cap \quad (14)$$

$$\vdots \quad (15)$$

$$P(\text{no chance match with person 20,000}) \quad (16)$$

The probability of incorrectly matching DNA to a random person is 0.01%:

$$P(\text{chance match with person } i) = 0.0001$$

So

$$P(\text{no chance match with person } i) = 0.9999$$

Now we compute our joint probability. Chance matches are independent, so we multiply our probabilities together: :

$$P(\text{no chance match in the database}) = \prod_{i \in 20\,000} P(\text{no chance match with person } i) \quad (17)$$

$$= 0.9999^{20000} \quad (18)$$

$$\approx 0.1353 \quad (19)$$

So there is only a 13.5% probability of **not** finding at least 1 match purely by chance! Or rather, an 86.5% probability that the match was random chance.

The Gamblers Paradox

Now let's look at our Roulette table again. The last 25 observations of the table were all Red:

```
##  
##   Red Black  
##   25     0
```

The probability of observing 26 “Red” observations in a row is

$$P(\mathcal{E} = 26 \times R) = 0.5^{26} \approx 1.5 \times 10^{-8}$$

. So surely we should Bet on Black now?!

The problem is that we have considered the wrong probability here. Yes, the probability of observing 26 “Red” spins in a row is very small, *but so is observing 25 “Red” spins and then one “Black” spin!*

$$P(\{R, R, R, \dots, R, R\}) = P(\{R, R, R, \dots, R, B\}) \quad (20)$$

$$P(R)^{25} \times P(R) = P(R)^{25} \times P(B) \quad (21)$$

The logical error here is that we have implicitly assumed that “the universe” knows about the last 25 spins, and now wants to rectify its mistake. In reality, the number that we have calculated above is the probability of seeing 26 “Red” spins *before any spin is made*. But once the spins are made, the probability resets. So the probability of seeing another red is just

$$P(R) = 0.50$$

Let's have a spin and see what we get....

```
##  
##   Red Black  
##   26     0
```

Whoops!