Essentials of Data Science With R Software - 1

Probability and Statistical Inference

Probability Theory

:::

Lecture 11

Relative Frequency and Probability

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Relative Frequency and Probability of an Event:

There is a close connection between the relative frequency and the probability of an event.

A random experiment is described by its possible outcomes.

For example, getting a number between 1 and 6 when rolling a die.

Suppose an experiment has *m* possible outcomes (events)

$A_1, A_2, \ldots, A_m$ and the experiment is repeated *n* times.

Relative Frequency and Probability of an Event:

Now we can count how many times each of the possible outcome has occurred.

In other words, we can calculate the absolute frequency $n_i = n(A_i)$ which is equal to the number of times an event A_i , $i = 1, 2, \ldots, m$, occurs.

The relative frequency $f_i = f(A_i)$ of a random event A_i , with n repetitions of the experiment, is calculated as

$$f_i = f(A_i) = \frac{n_i}{n}.$$

Relative Frequency and Probability of an Event:

If we assume that

- **the experiment is repeated a large number of times (mathematically, this would mean that *n* tends to infinity) and**
- **the experimental conditions remain the same (at least approximately) over all the repetitions,**

then the relative frequency f(*A*) converges to a limiting value for *A*.

Relative Frequency and Probability of an Event:

This limiting value is interpreted as the probability of *A* and denoted by

$$P(A) = \lim_{n \to \infty} \frac{n(A)}{n}$$

where *n*(*A*) denotes the number of times an event *A* occurs out of *n* times.

Relative Frequency and Probability of an Event: Example - Coin Toss

Suppose a fair coin is tossed $n = 10$ times.

Number of observed heads $n(A_1) = 3$ times

Number of observed tails $n(A_2) = 7$ times

Meaning of a fair coin: Probabilities of head and tail are equal (i.e., 0.5).

Then, the relative frequencies in the experiment are

$$f(A_1) = 3/10 = 0.3 \text{ and } f(A_2) = 7/10 = 0.7.$$

When the coin is tossed a large number of times and n tends to infinity, then both $f(A_1)$ and $f(A_2)$ will have a limiting value 0.5 which is the probability of getting a head or tail in tossing a fair coin.

Relative Frequency and Probability of an Event: Example - Coin Toss

Suppose the head is denoted by 0 and tail is denoted by 1.

Sample space (Ω) = {0, 1}

Suppose we repeat the experiment 5 times and following outcome is observed:

Head, Head, Tail, Head, Tail

which is expressible as 0, 0, 1, 0, 1.

Relative frequencies of Tail = Probability of Tail

$$\textbf{Probability of Tail} = \frac{\textbf{number of } 1's \; 1+1}{\textbf{Total number of repetitions}} = \frac{2}{5}$$

Relative Frequency and Probability of an Event: Example - Coin Toss

This can be simulated in R by the `sample` command by drawing the observations from 0 and 1 by simple random sampling with replacement.

Suppose we want repeat the experiment 10 times.

This means drawing 10 values and finding its arithmetic mean.

Relative Frequency and Probability of an Event: Example - Coin Toss

For example, the command

```
coin10 = sample(c(0,1), size=10, replace = T)
```

draws a sample of size 10 and stores the values in the data vector `coin10`**. The command**

```
table(coin10)/length(coin10)
```

find the relative frequencies of these 10 values.

So we repeat by increasing the number of repetitions *n* = 10, 100, 1000, 10000, . . .

Relative Frequency and Probability of an Event: Example - Coin Toss

10 repetitions

```
> coin10 = sample(c(0,1), size=10, replace = T)
> table(coin10)/length(coin10)

  0   1
0.7 0.3
```

```
> coin10 = sample(c(0,1), size=10, replace = T)
> table(coin10)/length(coin10)

  0   1
0.3 0.7
```

Relative Frequency and Probability of an Event: Example - Coin Toss

100 repetitions

```
> coin100 = sample(c(0,1), size=100, replace = T)
> table(coin100)/length(coin100)

   0    1
0.55 0.45
```

```
> coin100 = sample(c(0,1), size=100, replace = T)
> table(coin100)/length(coin100)

   0    1
0.45 0.55
```

Relative Frequency and Probability of an Event: Example - Coin Toss

1000 repetitions

```
> coin1000=sample(c(0,1), size=1000, replace = T)
> table(coin1000)/length(coin1000)

    0     1
0.519 0.481
```

```
> coin1000=sample(c(0,1), size=1000, replace = T)
> table(coin1000)/length(coin1000)

    0     1
0.516 0.484
```

Relative Frequency and Probability of an Event: Example - Coin Toss

R Console

```
> coin10 = sample(c(0,1), size=10, replace = T) #10 repetitions
> table(coin10)/length(coin10) #10 repetitions
coin10
  0   1
0.7 0.3
>
> coin10 = sample(c(0,1), size=10, replace = T) #10 repetitions
> table(coin10)/length(coin10) #10 repetitions
coin10
  0   1
0.3 0.7
>
> coin100 = sample(c(0,1), size=100, replace = T) #100 repetitions
> table(coin100)/length(coin100) #100 repetitions
coin100
   0    1
0.55 0.45
>
> coin100 = sample(c(0,1), size=100, replace = T) #100 repetitions
> table(coin100)/length(coin100) #100 repetitions
coin100
   0    1
0.45 0.55
>
> coin1000 = sample(c(0,1), size=1000, replace = T) #1000 repetitions
> table(coin1000)/length(coin1000) #1000 repetitions
coin1000
    0     1
0.519 0.481
> coin1000 = sample(c(0,1), size=1000, replace = T) #1000 repetitions
> table(coin1000)/length(coin1000) #1000 repetitions
coin1000
    0     1
0.516 0.484
```