Essentials of Data Science With R Software - 1

Probability and Statistical Inference

Probability Theory

:::

Lecture 12

Probability and Relative Frequency- An Example

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Relative Frequency and Probability of an Event: Example- Dice Roll

Suppose a fair dice is rolled and its outcome as the number of points on the upper face is recorded as 1, 2, 3, 4, 5, 6.

Sample space (Ω) = {1, 2, 3, 4, 5, 6 }

Suppose we repeat the experiment 100 times and the outcomes are recorded and the relative frequencies are obtained.

Relative Frequency and Probability of an Event: Example- Dice Roll

Suppose we repeat the experiment 100 times and the outcomes are recorded and the relative frequencies are obtained as follows:

Total number of 1's = 15	**f(1) = 15/100**
Total number of 2's = 10	**f(2) = 10/100**
Total number of 3's = 25	**f(3) = 25/100**
Total number of 4's = 14	**f(4) = 14/100**
Total number of 5's = 16	**f(5) = 16/100**
Total number of 6's = 20	**f(6) = 20/100**

Relative Frequency and Probability of an Event: Example- Dice Roll

Meaning of a fair dice: Probabilities of observing 1, 2, 3, 4, 5, 6 are equal, i.e., 1/6.

When the fair dice is rolled a large number of times and *n* tends to infinity, then all f (A_i) , i =1, 2, . . . , 6 will have a limiting value 1/6 which is the probability of getting 1, 2, 3, 4, 5, or 6.

Relative Frequency and Probability of an Event: Example- Dice Roll

This can be simulated in R by the `sample` command by drawing the observations among 1, 2, 3, 4, 5, 6 by simple random sampling with replacement and then finding the relative frequencies using the `table` and `length` commands.

Suppose we want repeat the experiment 100 times. This means drawing 100 values and finding the relative frequencies of 1, 2, 3, 4, 5, and 6.

Relative Frequency and Probability of an Event: Example- Dice Roll

The command

```
dice100 = sample(c(1,2,3,4,5,6), size=100,
replace = T)
```

generates 100 values and stores it in a data vector `dice100`.

Then the following command computes the relative frequencies of the data stored in `dice100`:

```
table(dice100)/length(dice100)
```

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

Relative Frequency and Probability of an Event: Example- Dice Roll

100 repetitions

```
> dice100 = sample(c(1,2,3,4,5,6), size=100,
replace = T)
> table(dice100)/length(dice100)

   1    2    3    4    5    6
0.10 0.17 0.15 0.23 0.14 0.21
```

```
> dice100 = sample(c(1,2,3,4,5,6), size=100,
replace = T)
> table(dice100)/length(dice100)

   1    2    3    4    5    6
0.13 0.18 0.22 0.14 0.16 0.17
```

Relative Frequency and Probability of an Event: Example- Dice Roll

1000 repetitions

```
> dice1000 = sample(c(1,2,3,4,5,6), size=1000, replace = T)
> table(dice1000)/length(dice1000)

    1     2     3     4     5     6
0.147 0.169 0.180 0.181 0.155 0.168
```

```
> table(dice1000)/length(dice1000)

    1     2     3     4     5     6
0.175 0.180 0.174 0.163 0.171 0.137
```

Relative Frequency and Probability of an Event: Example- Dice Roll

1000 repetitions

```
> dice10000 = sample(c(1,2,3,4,5,6), size=10000,
replace = T)
> table(dice10000)/length(dice10000)

     1      2      3      4      5      6
0.1626 0.1680 0.1657 0.1683 0.1718 0.1636

> table(dice10000)/length(dice10000)

     1      2      3      4      5      6
0.1626 0.1680 0.1657 0.1683 0.1718 0.1636
```

Relative Frequency and Probability of an Event: Example- Dice Roll

R Console

```
dice100
   1    2    3    4    5    6
0.10 0.17 0.15 0.23 0.14 0.21
>
> dice100 = sample(c(1,2,3,4,5,6), size=100, replace = T) # 100 repetitions
> table(dice100)/length(dice100) # Relative frequencies
dice100
   1    2    3    4    5    6
0.13 0.18 0.22 0.14 0.16 0.17
>
> dice1000 = sample(c(1,2,3,4,5,6), size=1000, replace = T) # 1000 repetitions
> table(dice1000)/length(dice1000) # Relative frequencies
dice1000
    1     2     3     4     5     6
0.147 0.169 0.180 0.181 0.155 0.168
> table(dice1000)/length(dice1000) # Relative frequencies
dice1000
    1     2     3     4     5     6
0.175 0.180 0.174 0.163 0.171 0.137
>
> dice10000 = sample(c(1,2,3,4,5,6), size=10000, replace = T) # 10000 repetitions
> table(dice10000)/length(dice10000) # Relative frequencies
dice10000
     1      2      3      4      5      6
0.1626 0.1680 0.1657 0.1683 0.1718 0.1636
> table(dice10000)/length(dice10000) # Relative frequencies
dice10000
     1      2      3      4      5      6
0.1626 0.1680 0.1657 0.1683 0.1718 0.1636
>
```