Essentials of Data Science With R Software - 1

Probability and Statistical Inference

Probability Theory

:::

Lecture 21
Computation of Probability using R

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Computation of Probability using R:

Various approaches can be employed for the computations of probabilities in terms of approximate relative frequencies.

The logical operator can be used for finding different the relative frequencies.

A function can be written to compute different combinations of these probabilities.

Computation of Probability using R:

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 }

Example:

Suppose we want to compute the probabilities of occurrence of 1's, 2's, 3's, 4's, 5's and 6's.

Note that

$$P(\text{Occurrence of 1's}) = \frac{\text{Total number of 1's}}{\text{Total number of trial}} = \text{Relative frequency of 1's}$$

Computation of Probability using R: *P*(*A*)

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

Example: *P*(*A*)

***A*: Event of occurrence of 2's**

***P*(Occurrence of 2's)**

Total number of 2's = 15

Total number of trials = 100

***P*(Occurrence of 2's) = f(2) = 25/100**

Computation of Probability using R: $P(A)$

This can be demonstrated 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 .

Earlier we used the `table` and `length` commands to compute the relative frequencies.

For illustration, suppose we want repeat the experiment 10 times. This means drawing 10 values and finding the relative frequencies of 1, 2, 3, 4, 5, and 6.

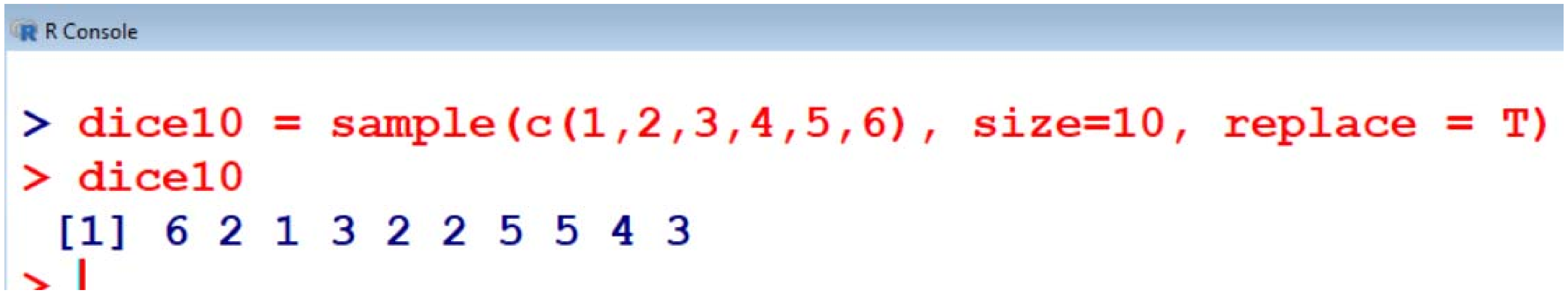
Computation of Probability using R: $P(A)$

The command

```
dice10 = sample(c(1,2,3,4,5,6), size=10, replace
= T)
```

generates 10 values and stores it in a data vector `dice10`.

```
> dice10 = sample(c(1,2,3,4,5,6), size=10,
replace = T)
> dice10
 [1] 6 2 1 3 2 2 5 5 4 3
```



Computation of Probability using R: $P(A)$

Now we can make use of logical operators to compute various type of probabilities.

Example:

Suppose we want to compute the probability of occurrence of 2's.

We need to count the number of 2's in `dice10` and divide by the length of `dice10`.

This is obtained by the following commands:

Computation of Probability using R: *P*(*A*)

```
> dice10
 [1] 6 2 1 3 2 2 5 5 4 3

> length(dice10[(dice10==2)]) # Counts no. Of 2's
[1] 3

> length(dice10)  # Counts total occurrence
[1] 10
```

Probability of occurrence of 2's

```
> p2 = length(dice10[(dice10==2)])/length(dice10)
> p2
[1] 0.3
```

Computation of Probability using R: *P*(*A*)

R Console

```
> dice10
 [1] 6 2 1 3 2 2 5 5 4 3
> length(dice10[(dice10==2)])
[1] 3
> length(dice10)
[1] 10
> p2 = length(dice10[(dice10==2)])/length(dice10)
> p2
[1] 0.3
> 
```

Computation of Probability using R: $P(A \cap B)$

Example: $P(A \cap B)$

Suppose we want to compute the probability of occurrence of 2's or 6's.

A : Event of occurrence of 2's

B : Event of occurrence of 6's

We need to count the number of 2's or 6's in `dice10` and divide by the length of `dice10`.

This is obtained by the following commands:

Computation of Probability using R: $P(A \cap B)$

```
> dice10
 [1] 6 2 1 3 2 2 5 5 4 3

> length(dice10[(dice10==2) | (dice10==6)]) #
Counts no. Of 2's or 6's
[1] 4

> length(dice10)  # Counts total ouotcomes
[1] 10
```

Probability of occurrence of 2's and 6's

```
> p26 = length(dice10[(dice10==2) |
(dice10==6)])/length(dice10)
> p26
[1] 0.4
```

Computation of Probability using R: $P(A \cap B)$

R Console

```
> dice10
 [1] 6 2 1 3 2 2 5 5 4 3
> length(dice10[(dice10==2)])
[1] 3
> length(dice10)
[1] 10
> p2 = length(dice10[(dice10==2)])/length(dice10)
> p2
[1] 0.3
>
> length(dice10[(dice10==2) | (dice10==6)])
[1] 4
> p26 = length(dice10[(dice10==2) | (dice10==6)])/length(dice10)
> p26
[1] 0.4
>
```

Computation of Probability using R: $P(A \mid B)$

In case of computation of conditional probability,

$$P(A \mid B) = P(AB)/ P(B)$$

How to compute $P(AB)$,

Hypothetically

```
length(d10(d10==A) & (d10==B)])
```

We know how to compute all the involved expressions.

Using logical operators, one can also compute them directly also by writing a suitable function.