EIDIA

École d'Ingénierie Digitale
et d'Intelligence Artificielle

R Programming

Probabilities with R

UEMF

الجامعة الأورومتوسطية بفاس

EUROMED UNIVERSITY OF FES
UNIVERSITÉ EUROMED DE FÈS

Introduction

In R, each law has an abbreviation (norm for the gaussian for example). To make calculations on a random variable according to this law, a letter is used to indicate the type of calculation that one wishes to do:

- d for density
- p for the distribution function (i.e. calculating probabilities)
- q for quantiles
- r for randomly generated values

Introduction

For the functions p referring to the distribution function and q referring to the calculation of the quantiles, in addition to the parameters specific to each distribution, there is a **lower parameter**.

For example, pnorm (m = average, sd = standard_deviation, lower.tail = **TRUE**).

Its value is **TRUE** by default, indicating that the probabilities concerned are **P (X ≤ x).**

It can sometimes be useful to have an idea of the density function graph of a given law.

The line of code to write is:

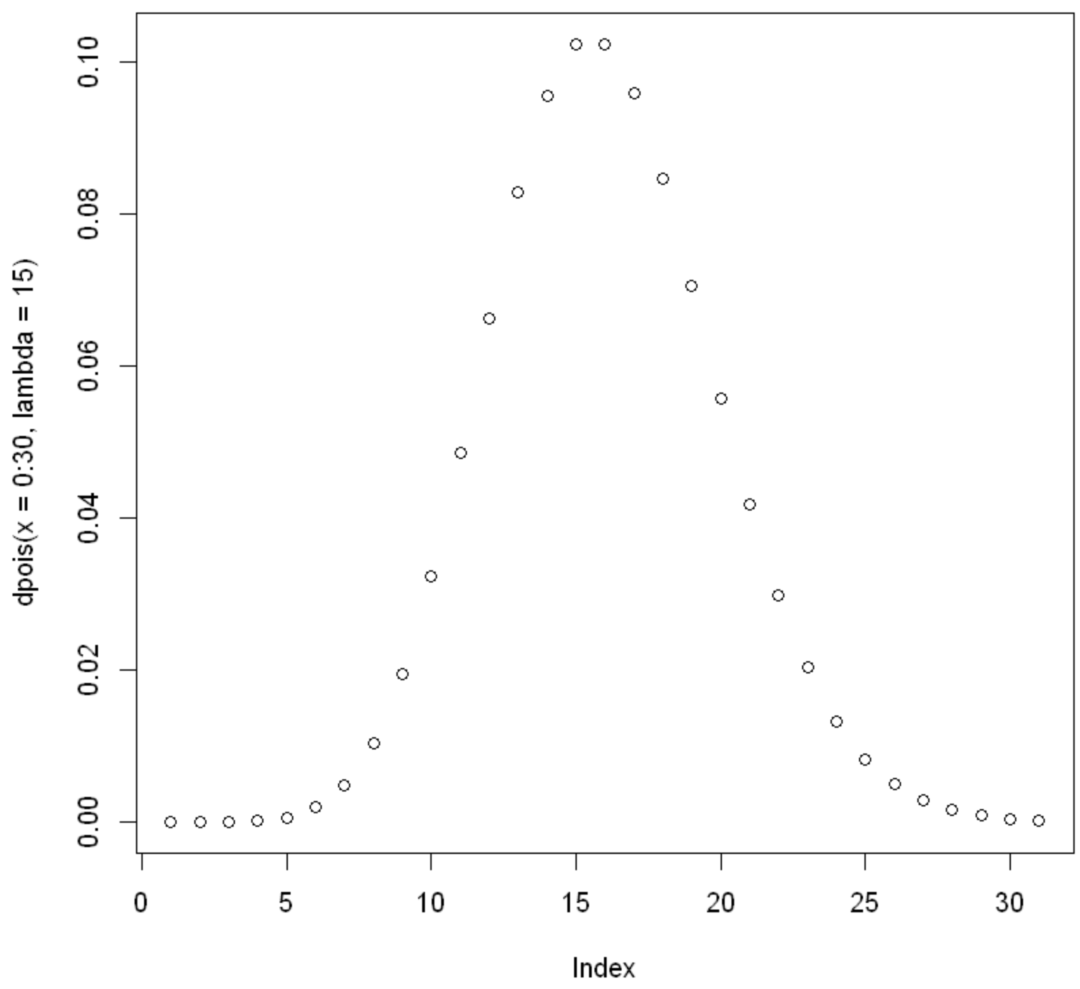
plot (d_functionToBeDefined (x = start: end, parameter of the function))

Introduction

Below an example with the Poisson law described below.

In [15]: **plot** (dpois (x = 0: 30, lambda = 15))

PS: Central Limit Theorem.



Binomial Probability Law

The binomial law admits two parameters:

- the size **n** of the sample or the number of draws, called size in R
- the probability **p** of success.

You can either specify each parameter name and its value (case 1 below), or put the parameters in order without specifying their name (case 2).

For X ~ Binom (n = 20, p = 0.4), calculation of P (X = 5):

In [1]: *# calculation of the probability of having x success among the n draws, Case 1*

```
dbinom(x = 5,size = 20,p = 0.4)
0.074647019528871
```

Binomial Probability Law

In [2]: *# calculation of the probability of having x success among the n draws, Case 2*

```
dbinom(5, 20, 0.4)
0.074647019528871
```

For $X \sim Binom(n = 20, p = 0.4)$, calculation of $P(3 < X < 16)$:

In [4]:
```
sum(dbinom(4:15, 20, 0.4))
0.983721806088823
```

Binomial Probability Law

We can do the calculation in two ways, either by using the sum of the probabilities of each value included in inequality, either using the *cumulative* distribution function, i.e. the notation pbinom of R using the following equality $P(3 < X \leq 16) = P(X \leq 16) - P(X \leq 2)$

In [7]: *# CASE 1 using the sum of the* probability mass function

```
sum(dbinom(3:16, 20, 0.4))
```

0.996341182970179

In [8]: *# CASE 2 using the cumulative function*

```
pbinom(16,20,0.4) - pbinom(2,20,0.4)
```

0.996341182970179

Geometric distribution

The geometric distribution admits only one parameter: the probability (prob) of success of Bernouilli tests.

In R, the definition of the geometric distribution is different.

In R, the geometric law counts the number of tests which precede the first success and the possible values start from 0.

Here, the geometric dist. is defined as the number of tests necessary before obtaining a first success and the set of possible values starts from 1.

Geometric distribution

For $X \sim Geom(p = 0.4)$

calculate P(X = 2):

In [9]: **dgeom**(2,prob = 0.4)

0.144

For $X \sim Geom(p = 0.4)$

calculate $P(X \leq 3)$:

In [10]: **pgeom**(3,prob = 0.4)

0.8704

Poisson distribution

Poisson's law admits a parameter « c », which designates the average number of realizations of the event in the time interval considered or in the space considered.

In R, this parameter is called *lambda*.

For $X \sim Poisson(c = 4)$, calculate P(X = 2):

In [5]: **dpois**(2,lambda = 4)

0.146525111109873

Poisson distribution

For $X \sim Poisson(c = 8)$, calculate $P(X \geq 10)$:

In [6]: 1 - **ppois**(q = 10, lambda = 8)

0.184114207441454

If we look for the smallest value k such as $P(X \leq k) \geq 0,95$

with $X \sim Poisson(c = 15)$

In [8]: **qpois**(0.95,lambda = 15,lower.tail = **TRUE**)

22

$\boxed{\text{FALSE}}$

Poisson distribution

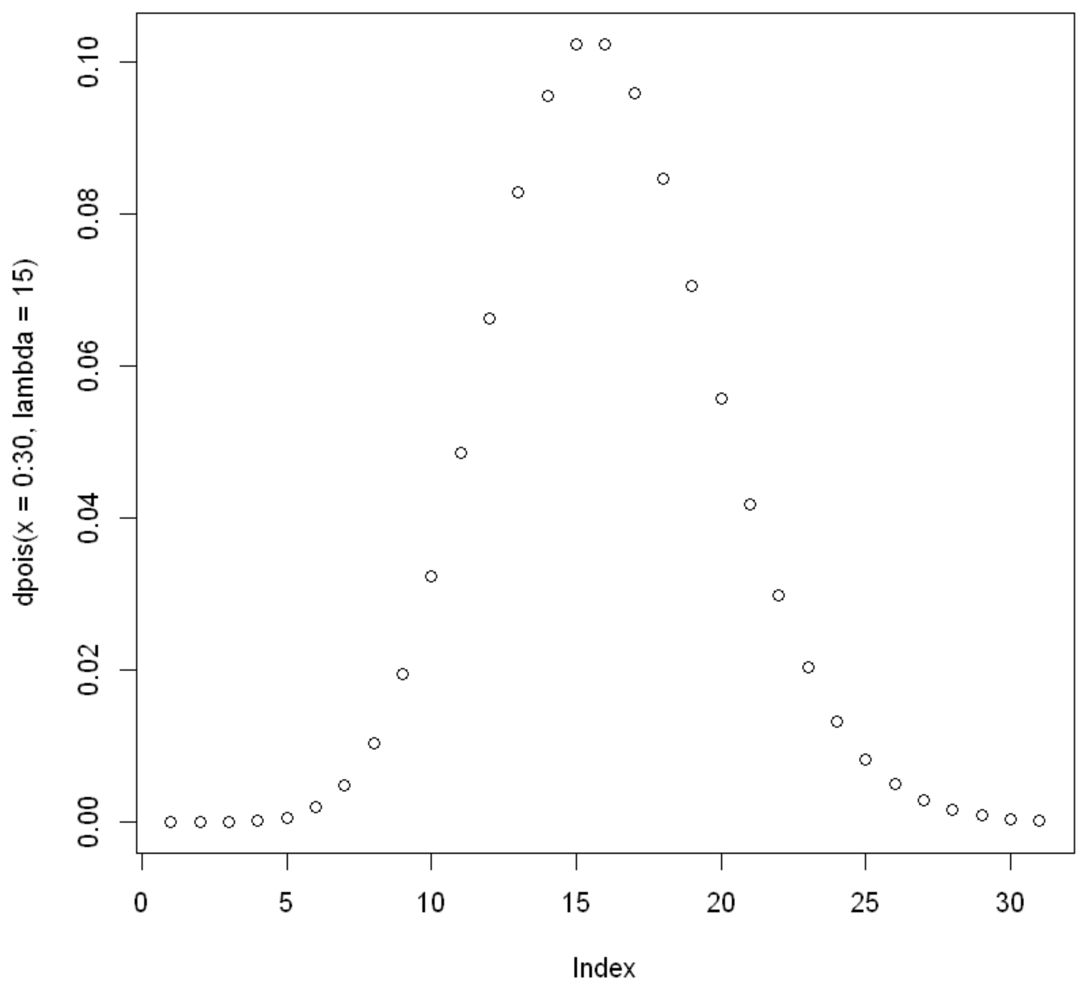
If we look for k such as $P(X > k) \geq 0{,}90$

$With\ X \sim Poisson(c = 15)$

In [9]: **qpois**(0.9,lambda = 15,lower.tail = **FALSE**)

10

In [14]: **plot**(**dpois**(x=0:30, lambda=15))



Normal distribution

The normal distribution admits two parameters: the average noted *m* in R and the standard deviation noted *sd* in R.

- dnorm (x): probability density in x of the **reduced normal centered** distribution
- dnorm (x, mean = m, sd = s): probability density in x of the normal dist. of expectation *m* and standard deviation *s*
- pnorm (x): cumulative function of the reduced normal centered distribution
- pnorm (x, mean = m, sd = s): cumulative function in x for the normal dist. of expectation *m* and standard deviation *s*

Normal distribution

For $X \sim N(10, 4)$, calculate $P(X \leq 8)$:

In [13]: **pnorm**(8,m = 10,sd = 2)

0.158655253931457

For $X \sim N(10, 4)$, we look for k such as $P(X \leq k) = .9$

In [14]: **qnorm**(0.9,m = 10,sd = 2)

12.5631031310892

For $X \sim N(10, 4)$, we look for k such as $P(X > k) = .8$

[15]: **qnorm**(0.8,m = 10,sd = 2)

11.6832424671458

Student Distribution

Student's law admits only one parameter: the degree of freedom denoted by *df* in R.

For $X \sim t_2$, calculate $P(X \leq 3)$:

In [16]: **pt**(3,df = 2)

    0.952267016866645

Calculate the quantile $t_{0.05;10}$:

In [17]: **qt**(1-0.05,10)

    1.81246112281168

Chi-square distribution

Chi-square distribution admits only one parameter: the degree of freedom denoted by *df* in R.

For $X \sim \chi^2_4$, calculate $P(X \leq 3)$:

In [18]: **pchisq**(3,df = 4)

0.442174599628925

Calculate $\chi^2_{0.05,10}$

In [19]: **qchisq**(0.05,10,lower.tail = **FALSE**)

18.3070380532751

Fisher–Snedecor distribution (*F*-distribution)

F-distribution admits 2 parameters:

- *df1* the first degree of freedom
- *df2* the second degree of freedom

For $X \sim F_{3,4}$, calculate $P(X \leq 5)$:

In [20]: **pf**(5,df1 = 3,df2=4)

0.922981284596924

For $X \sim F_{3,4}$, calculate $P(X \geq 4)$:

In [12]: **pf**(4,df1 = 3,df2=4,lower.tail = **FALSE**)

0.106911302347298

Fisher–Snedecor distribution (*F*-distribution)

You can check if $F_{1-\alpha;u,v} = \frac{1}{F_{\alpha;v,u}}$

In [22]: **qf**(0.9,df1 = 4,df2 = 5)

3.52019624553412

In [23]: 1/ **qf**(0.1,df1 = 5, df2 = 4)

3.52019624553412