# Introduction to R*

Lecture 5: Environments, Running R , Libraries and some probability distributions

Wim R.M. Cardoen

Last updated: 11/17/2022 @ 20:53:25

## Contents

# 1 R Environments

```
under construction
```

# 2 Running R

```
under construction
```

# 3 R Packages

```
under construction
```

## 3.1 Installation of R packages

## 3.2 Using R packages

# 4 Probability distributions

R comes with the most important probability distributions installed.
They can be classified as:

- discrete distributions
- continuous distributions

## 4.1 Discrete distributions

Let $\mathbb{P}(X = k; \{\xi\}))$ be a discrete probability mass function when the random variable $X = k$ and which depends on the parameter set $\{\xi\}$.
Let `distro` be the name of the corresponding distribution. Then,

- **d**`distro`$(k, \ldots)$ : calculates the probability $\mathbb{P}(X = k)$
- **p**`distro`$(k, \ldots)$ : calculates the cumulative probability function (CDF) at $k$:
  $F(X = k; \{\xi\}) := \sum_{j=0}^{k} \mathbb{P}(X = j)$
- **q**`distro`$(p, \ldots)$: calculates the value of $k$ where $p = F(k; \{\xi\})$ or $k = \lceil F^{-1}(p; \{\xi\}) \rceil$
- **r**`distro`$(n, \ldots)$: generates a vector of $n$ random values sampled from the distribution `distro`.

Some common discrete probability distributions

| distro | Name | $\mathbb{P}(X=k;\{\xi\})$ | Parameter set ($\{\xi\}$) |
|---|---|---|---|
| binom | Binomial | $\binom{n}{k}p^k(1-p)^{n-k}$ | $0 \le p \le 1$ |
| nbinom | Negative Binomial | $\binom{k+r-1}{k}(1-p)^k p^r$ | $0 \le p \le 1$ ; $r > 0$ |
| hyper | Hypergeometric | $\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$ | $N \in \{0,1,2,\dots\}$; $K, n \in \{0,1,\dots,N\}$ |
| pois | Poisson | $\frac{\lambda^k e^{-\lambda}}{k!}$ | $0 < \lambda < \infty$ |

### 4.1.1 Examples

- Let's consider the following distribution: `binom`$(p = 0.3, n = 5)$.

```
n <- 5
p <- 0.3
# Own code for the binom distribution
mybinom <- function(n,p){
    v <- vector(mode="double", length=(n+1))
    for(k in 0:n){
      v[k+1] <- choose(n,k)*p^k*(1-p)^(n-k)
    }
    return(v)
}
pvec <- mybinom(n,p)
```

  - Value of the PMF at $k = \{0, 1, \dots, 5\}$:

```
for(k in 0:n){
    cat(sprintf("  P(X=%d):%8.6f and should be %8.6f\n",
                k, dbinom(k, size=n, p), pvec[k+1]))
}
```

```
  P(X=0):0.168070 and should be 0.168070
  P(X=1):0.360150 and should be 0.360150
  P(X=2):0.308700 and should be 0.308700
  P(X=3):0.132300 and should be 0.132300
  P(X=4):0.028350 and should be 0.028350
  P(X=5):0.002430 and should be 0.002430
```

  - Value of the CDF at $k = \{0, 1, \dots, 5\}$:

```
for(k in 0:n){
    cat(sprintf("  F(X=%d):%8.6f and should be %8.6f\n",
    k, pbinom(k,size=n,p), sum(pvec[1:(k+1)]) ))
}
```

```
  F(X=0):0.168070 and should be 0.168070
  F(X=1):0.528220 and should be 0.528220
  F(X=2):0.836920 and should be 0.836920
```

```
F(X=3):0.969220 and should be 0.969220
F(X=4):0.997570 and should be 0.997570
F(X=5):1.000000 and should be 1.000000
```

- The quantile function:

```
pvec <- c(0.0, 0.25, 0.50, 0.75, 1.00)
for(item in pvec){
    cat(sprintf("  P:%4.2f => k=%d\n",
    item, qbinom(item,size=n, prob=p)))
}
```

```
P:0.00 => k=0
P:0.25 => k=1
P:0.50 => k=1
P:0.75 => k=2
P:1.00 => k=5
```

- Sampling random numbers from the distribution:

```
tot <- 15
vec <- rbinom(tot,size=n, prob=p)
print(vec)
```

```
[1] 2 2 0 2 2 1 1 1 1 2 1 1 1 1 0
```

## 4.2 Continuous distributions

Let $f(x;\{\xi\})$ be a continuous probability density function (pdf), which depends on the variable $x$ and the parameter set $\{\xi\}$.
Let `distro` be the name of the corresponding distribution. Then,

- **d**`distro`$(x,\ldots)$ : calculates the value of the pdf at $x$, i.e. $f(x;\{\xi\})$
- **p**`distro`$(x,\ldots)$ : calculates the cumulative probability function (cdf) at $x$: $F(x;\{\xi\}) := \int_{-\infty}^{x} f(t;\{\xi\})\,dt$
- **q**`distro`$(p,\ldots)$: calculates the value of $x$ where $p = F(x;\{\xi\})$ or $x = F^{-1}(p;\{\xi\})$
- **r**`distro`$(n,\ldots)$: generates a vector of $n$ random values sampled from the distribution `distro`.

Some common continuous distributions (see e.g. (Casella & Berger, 2002)) are:

| distro | Name | $f(x;\{\xi\})$ | Dom$(x)$ | Parameter set ($\{\xi\}$) |
|---|---|---|---|---|
| unif | Uniform | $\frac{1}{(b-a)}$ | $a \leq x \leq b$ | $a, b$ |
| norm | Normal | $\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ | $-\infty < x < \infty$ | $-\infty < \mu < \infty$ , $\sigma > 0$ |
| cauchy | Cauchy | $\frac{1}{\pi\sigma} \frac{1}{1+\left(\frac{x-\theta}{\sigma}\right)^2}$ | $-\infty < x < \infty$ | $-\infty < \theta < \infty$ , $\sigma > 0$ |
| t | t-Student | $\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\nu\pi} \frac{1}{\left(1+\frac{x^2}{\nu}\right)^{(\nu+1)/2}}$ | $-\infty < x < \infty$ | $\nu = 1, 2, \ldots$ |
| chisq | Chi-squared | $\frac{1}{\Gamma(\nu/2)2^{(\nu/2)}} x^{(\nu/2)-1} e^{-\frac{x}{2}}$ | $0 \leq x < \infty$ | $\nu = 1, 2, \ldots$ |
| f | F | $\frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \frac{x^{(\nu_1-2)/2}}{\left(1+\left(\frac{\nu_1}{\nu_2}\right)x\right)^{(\nu_1+\nu_2)/2}}$ | $0 \leq x < \infty$ | $\nu_1, \nu_2 = 1, 2, \ldots$ |
| exp | Exponential | $\lambda e^{-\lambda x}$ | $0 \leq x < \infty$ | $\lambda > 0$ |

where $\Gamma(x)$ stands for the gamma function which has the following mathematical form:

$$\Gamma(x) := \int_0^\infty t^{x-1} e^{-x} dt$$

### 4.2.1 Examples

- Let's consider the following distribution: $N(\mu = 5.0, \sigma^2 = 4.0)$.
  Therefore, distro:norm

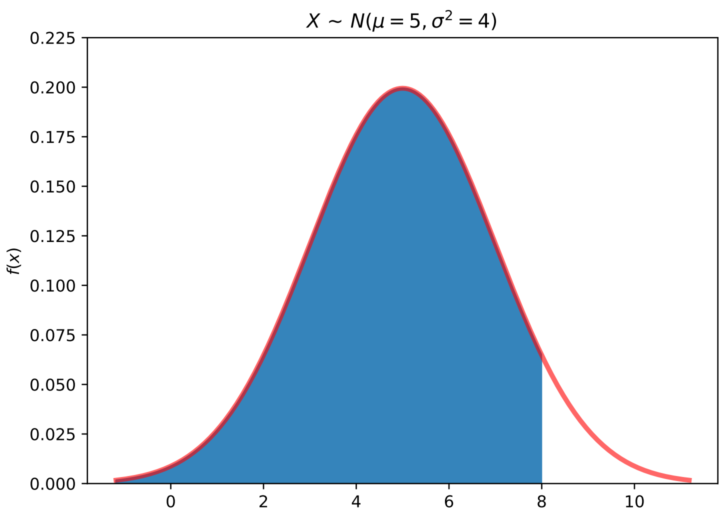


Figure 1: Plot of the normal distribution (red). The area under the curve (blue) represents the cumulative probability at $x = 8.0$.

```r
x <- 8.0
mu <- 5.0
sigma <- 2.0
```

- Value of the PDF at $x$:

  ```r
  cat(sprintf("The density at %f is %12.10f\n", x, dnorm(x,mean=mu, sd=sigma)))
  ```

  ```
  The density at 8.000000 is 0.0647587978
  ```

- Value of the CDF at $x$:

  ```r
  prob <- pnorm(x,mean=mu,sd=sigma)
  cat(sprintf("The Cumulative Probability at %f is %12.10f", x, prob))
  ```

  ```
  The Cumulative Probability at 8.000000 is 0.9331927987
  ```

- The quantile function:

  ```r
  cat(sprintf("The point where the Cumulative Probability is %12.10f: %8.4f",
              prob, qnorm(prob, mean=mu, sd=sigma)))
  ```

  ```
  The point where the Cumulative Probability is 0.9331927987:   8.0000
  ```

- Sampling random numbers from the distribution:

  ```r
  vec <- rnorm(n=10, mean=mu, sd=sigma)
  print(vec)
  ```

  ```
   [1] 5.3621494 6.7463792 2.5564777 7.9355664 3.9496334 3.8057722 5.8507367
   [8] 0.6241145 5.0169393 5.4506935
  ```

# Bibliography

Casella G. & Berger R.L. (2002). Statistical Inference. Duxbury Advanced Series in Statistics and Decision Sciences. Thomson Learning.