R HANDBOOK

Synthèse : Mehdi LATIF
mehdi.latif@etu.univ-nantes.fr

Année 2016/2017

Table des matières

I Université FUN
Introduction à la statistique avec R
Bruno Falissard, Christophe Lalanne **3**

II Université de Nantes **5**

III Théorie de la statistique **6**

1 La loi normale **7**
1.1 Le modèle de la loi normale 7
1.2 Calculs pratiques 10
1.3 Compléments mathématiques 16

2 Introduction aux lois de probabilité avec R - Christophe Chesneau **22**
2.1 Syntaxe générale 22
2.2 Densité 23
2.3 Fonction de répartition 26
2.4 Quantiles 29
2.5 Simulation 30
2.5.1 La commande sample 30
2.5.2 Simulation par des lois préprogrammées 32
2.6 Exercices 33
2.6.1 Enoncés 33
2.6.2 Solutions 42

IV Div'R 68

V Représentations graphiques 69

VI Programmation avec R 70

VII Fonctions usuelles et aide mémoire 71

3 Lois 72
3.1 Loi Normale 73
3.2 Loi Normale 74

4 R datasets informations 75

I Université FUN
Introduction à la statistique avec R

Bruno Falissard, Christophe Lalanne

Lien vers le cours de **l'Université FUN** : ici

II Université de Nantes

III Théorie de la statistique

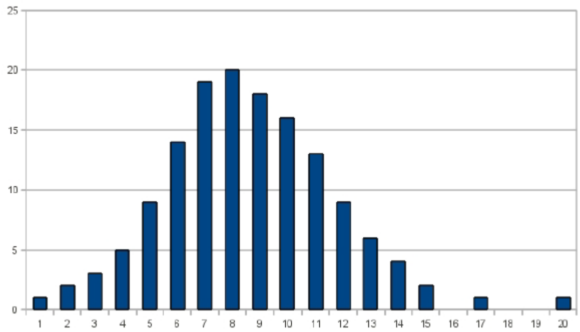
1 La loi normale

1.1 Le modèle de la loi normale

Un exemple : Test de mémoire

Etude de la capacité de mémoire d'adultes atteints d'une maladie neurologique. Chaque individu lit 30 mots et doit ensuite en réciter le plus possible.

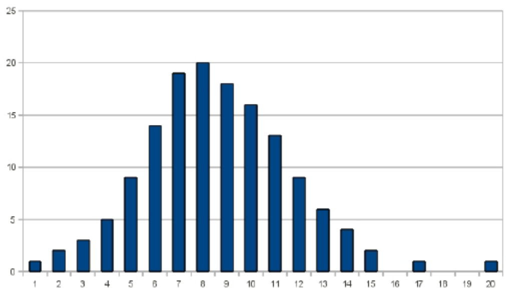
- Population $\mathcal{P}$ = { patients atteints de la maladie }
- Variable quantitative X = *"nombre de mots retenus"*
- 2 paramètres μ ; σ.

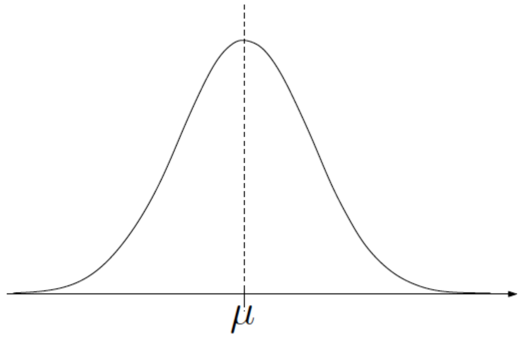


En sciences humaines on observe souvent des distributions :

- plutôt symétriques autour de μ
- avec une forme de cloche

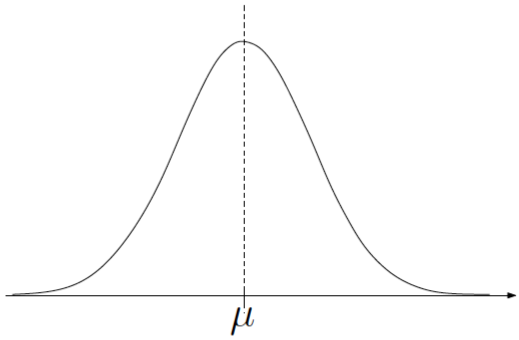
Pour pouvoir faire des calculs, on va parfois supposer que X suit une distribution "modèle", appelée **Loi normale**.



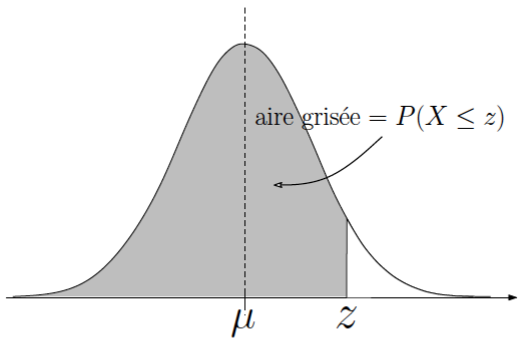


Propriétés de la loi normale

Si X suit cette distribution "modèle", on lui associe une courbe :



- courbe symétrique par rapport à μ
- forme de cloche



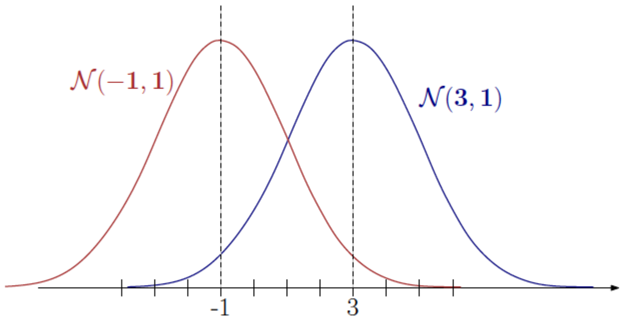
l'aire grisée représente la proportion cumulée.

Paramètres de la loi normale Pour chaque μ ; σ, il existe une **loi normale de moyenne μ et d'écart-type σ.**
Cas particulier : $\mu = 0$ et $\sigma = 1$: *loi normale centrée/réduite*.

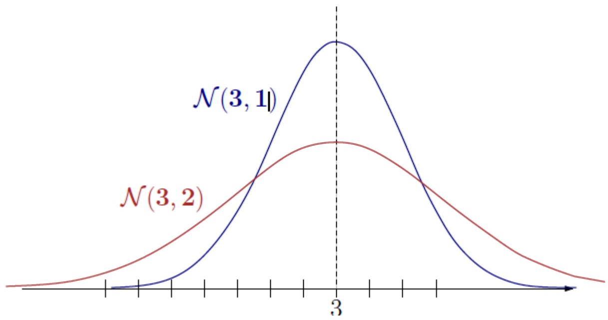
Lorsque l'on suppose qu'une variable X suit le modèle de la loi normale $\mathcal{N}(\mu;\sigma)$, on écrit :

$$X \sim \mathcal{N}(\mu,\sigma)$$

Exemples de lois normales avec moyennes différentes, même écart-type :



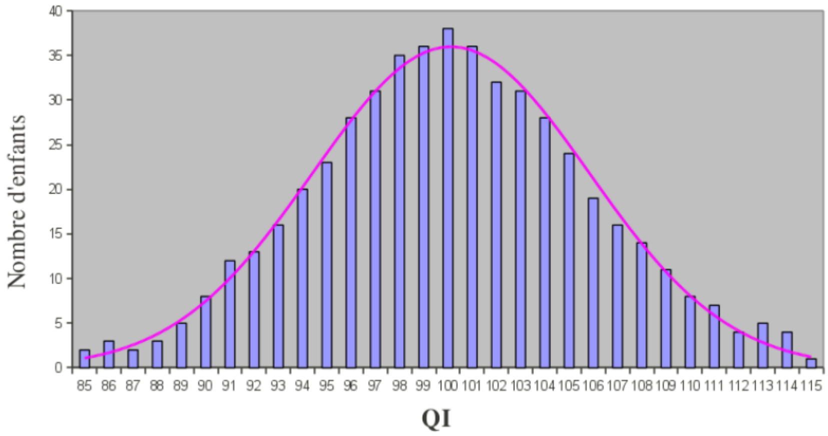
Exemples de lois normales avec même moyenne, écart-types différents :



Pour la tracer :

$$y = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Etude sur le QI de 515 enfants du même âge, $\mu = 100,1$, $\sigma = 5,7$.

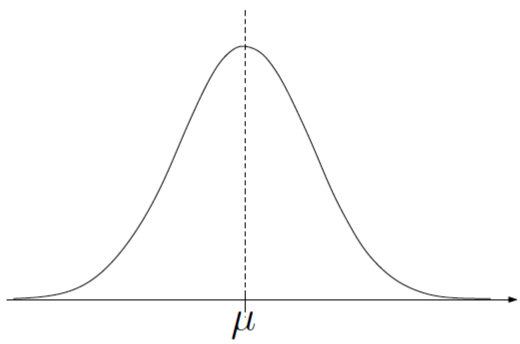


En rose, courbe de la loi normale $\mathcal{N}(\mu = 100,1; \mu = 5,7)$.

A retenir sur :

$$\mathcal{N}(\mu, \sigma)$$

— distribution "modèle" pour des variables **quantitatives continues**
— moyenne μ, écart-type σ
— allure de la courbe :



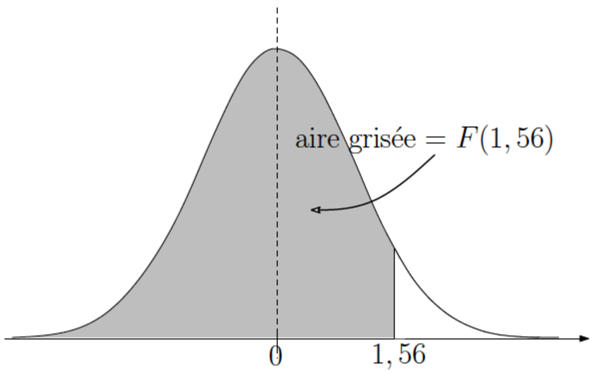
— aires = proportions cumulées

1.2 Calculs pratiques

Loi Normale centrée Réduite $\mathcal{N}(0,1)$

Exemple

1. On suppose qu'une certaine variable $X \sim \mathcal{N}(0.1)$. Pour quelle proportion d'individus est-ce que $X \leq 1.56$?
On cherche $\mathbb{P}(X \leq 1;56)$ *(rappel : on écrit aussi* $F(1.56)$*).*



On cherche $1,56$ dans la table (CF fin du rapport) :

	$\dots$	$0,06$	$\dots$
$\vdots$		$\vdots$	
$1,5$	$\dots$	$0,9406$	
$\vdots$			

Donc $\mathbb{P}(X \leq 1.56) = 0.9406$.
Pour 94.06% des individus, la variable X est inférieure à $1,56$.

2. On suppose qu'une certaine variable $X \sim \mathcal{N}(0.1)$. Pour quelle proportion d'individus est-ce que $X \geq 1.56$?
On cherche $\mathbb{P}(X \geq 1;56)$. *On écrit d'abord*

$$\mathbb{P}(X \geq 1.49) = 1 - \mathbb{P}(X \leq 1.49) = 1 - F(1.49)$$

On cherche $1,49$ dans la table (CF fin du rapport) :

	$\dots$	$0,09$	$\dots$
$\vdots$		$\vdots$	
$1,4$	$\dots$	$0,9319$	
$\vdots$			

Donc

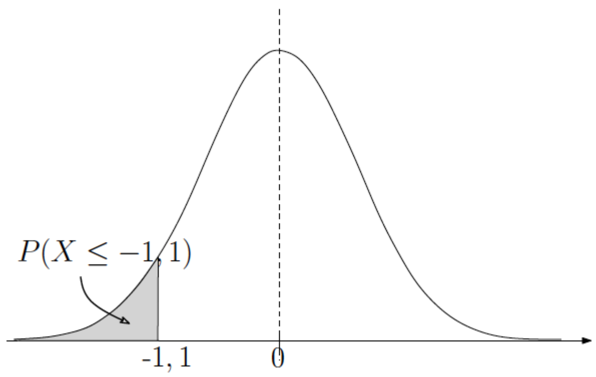
$$\mathbb{P}(X \geq 1.49) = 1 - \mathbb{P}(X \leq 1.49) = 1 - 0,9319$$

Soit

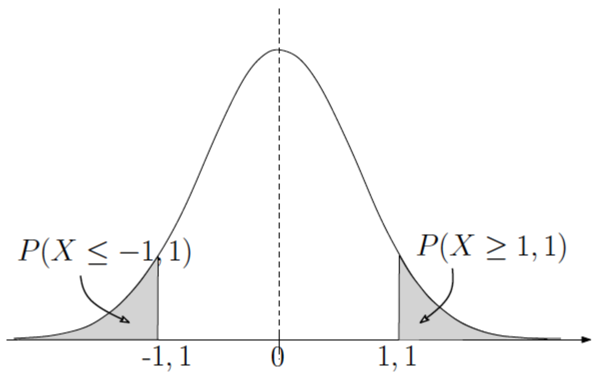
$$\mathbb{P}(X \geq 1.49) = 0,0681$$

Pour 6.81% des individus, la variable X est supérieure à $1,49$.

3. On suppose qu'une certaine variable $X \sim \mathcal{N}(0,1)$. Pour quelle proportion d'individus est-ce que $X \leq -1.1$?
On cherche $\mathbb{P}(X \leq 1,1)$, *c'est à dire* $F(-1,1)$



Mais, on rappelle que la loi normale est symétrique par rapport à μ.



On peut donc ainsi calculer :

$$\mathbb{P}(X \geq 1.1) = 1 - \mathbb{P}(X \leq 1.1) = 1 - F(1.1) = 1 - 0,8643$$

Et ainsi :

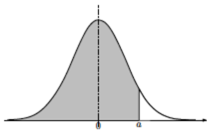
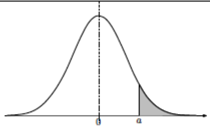
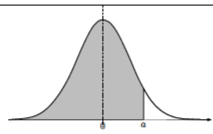
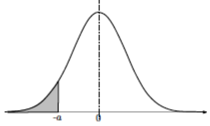
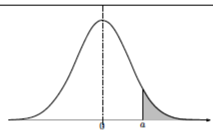
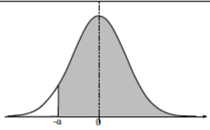
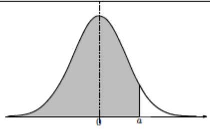
$$\mathbb{P}(X \leq 1,1) = 0,1357$$

A retenir :

$$F(-a) = 1 - F(a)$$

Par exemple : $F(-1,1) = 1 - F(1,1)$

Pour tous type de a :
Pour n'importe quel $a > 0$,

I	$\mathbb{P}(X \leq a)$				$\Rightarrow$ table
II	$\mathbb{P}(X \geq a)$		$= 1 -$		$\Rightarrow$ cas I
III	$\mathbb{P}(X \leq -a)$		$=$		$\Rightarrow$ cas II
IV	$\mathbb{P}(X \geq -a)$		$=$		$\Rightarrow$ cas I

Loi normale quelconque $\mathcal{N}(\mu, \sigma)$

Pour faire des calculs avec une $\mathcal{N}(\mu, \sigma)$, on se ramène à la loi $\mathcal{N}(0, 1)$.
Par le théorème suivant :

$$\text{Si } X \sim \mathcal{N}(\mu, \sigma), \text{ alors } \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

On dit alors que l'on **centre et réduit** X. On utilise la lettre Z pour désigner une loi normale centrée/réduite.

Exemple

1. On suppose qu'une certaine variable $X \sim \mathcal{N}(11, 2)$. Pour quelle proportion d'individus est-ce que $X \leq 14$?
   On cherche $\mathbb{P}(X \leq 14)$.
   (a) On **centre et on réduit** X

   $$\frac{X - 11}{2} \sim \mathcal{N}(0, 1)$$

   (b) On calcule :

   $$\begin{aligned} \mathbb{P}(X \geq 14) = &\ \mathbb{P}\left(\frac{X - 11}{2} \leq \frac{14 - 11}{2}\right) \\ = &\ \mathbb{P}(Z \leq 1,5) \end{aligned}$$

   (c) On cherche $1,5$ dans la table.
   On trouve finalement :

   $$\mathbb{P}(X \leq 14) = 0,9332 \text{ pour } X \sim \mathcal{N}(11, 2)$$

2. On cherche le quantile à $97,5\%$ pour la $\mathcal{N}(0,1)$.
   Cela revient à trouver a tel que

$$P(Z \leq a) = 0,975$$

   On lit alors la table à l'envers.

	$\ldots$	$0,06$	$\ldots$
$\vdots$		$\vdots$	
$1,9$	$\ldots$	$0,9750$	
$\vdots$			

   Donc

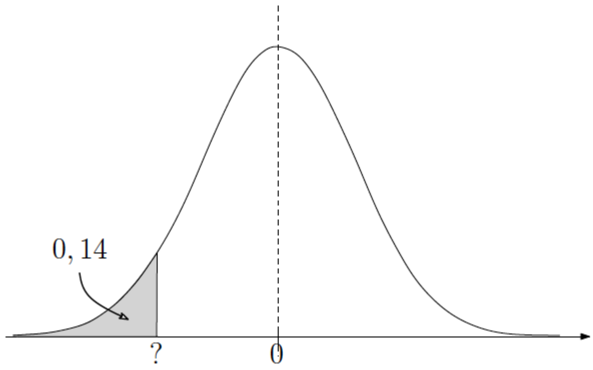
$$\mathbb{P}(X \leq 1,96) = 0,9750$$

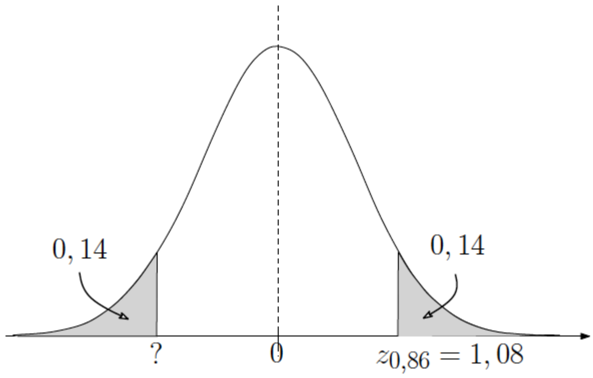
   Le quantile recherché est donc

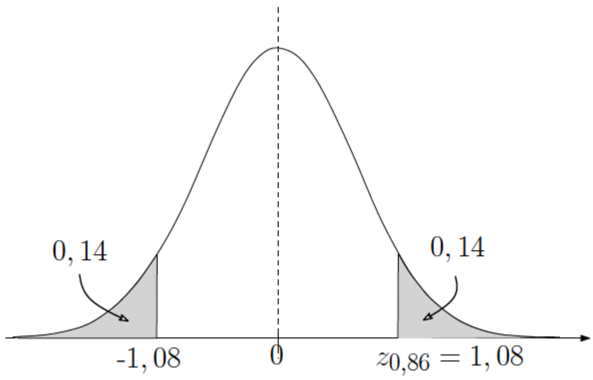
$$a = 1,96$$

   Notation : Le quantile d'ordre α pour la loi normale centrée/réduite est noté $z\alpha$.
   Par exemple, $z_{0,975} = 1,96$ pour $X \sim \mathcal{N}(0,1)$.

3. On cherche le quantile à 14% pour la $\mathcal{N}(0,1)$.
   Cela revient à trouver a tel que $\mathcal{P}(Z \leq a) = 0.14$.
   Mais, Il n'y a pas de nombre inférieur à $0,5$ dans la table de la loi Normale.







Par symétrie de la loi Normale par rapport à μ, on peut en déduire :

$$z_{0,14} = -1,08$$

A retenir :

$$z_\alpha = -z_{1-\alpha}$$

Par exemple : $z_{0,14} = -z_{0,86}$

Quantile d'une loi normale quelconque

Notons Q_α le quantile d'ordre α d'une loi normale quelconque $N(\mu;\alpha)$, On a :

$$Q_\alpha = \mu + \sigma \times z_\alpha$$

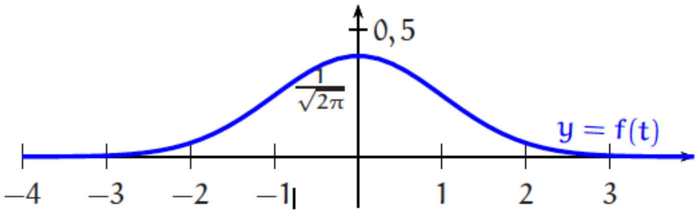
On "déréduit" et on "décentré" le quantile de la loi normale centrée/réduite.

1.3 Compléments mathématiques

La loi normale centrée réduite

La loi normale centrée réduite $\mathcal{N}(0,1)$ est la loi de probabilité dont la densité est la fonction f définie par :

$$\forall t \in \mathbb{R}, f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$



Remarque : Au cours des études post-bac, on sait démontrer que l'intégrale de Gauss [1] :

$$\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} \, \mathrm{d}t = \sqrt{2\pi}$$

En divisant par $\sqrt{2\pi}$, on normalise :

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} \, \mathrm{d}t = 1$$

La fonction de répartition de la loi normale centrée réduite est la fonction F définie par :

$$\forall x \in \mathbb{R}, F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \, \mathrm{d}t$$

1. On appelle intégrales de Gauss les intégrales de la forme :

$$\int_{-\infty}^{+\infty} e^{-at^2} \, \mathrm{d}t$$

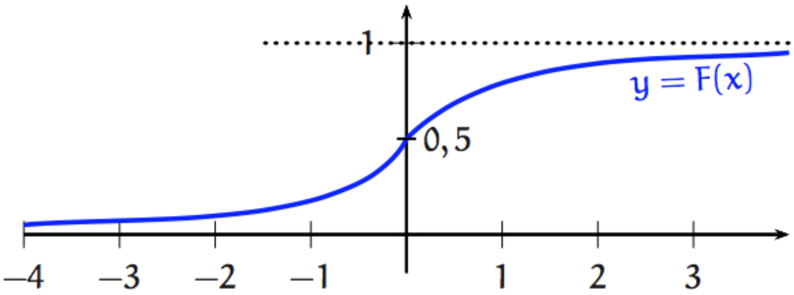
où a est un nombre réel strictement positif. La valeur de l'intégrale de Gauss est lié au nombre π par la relation :

$$\int_{-\infty}^{+\infty} e^{-at^2} \, \mathrm{d}t = \sqrt{\frac{\pi}{a}}$$

Cas où a = 1

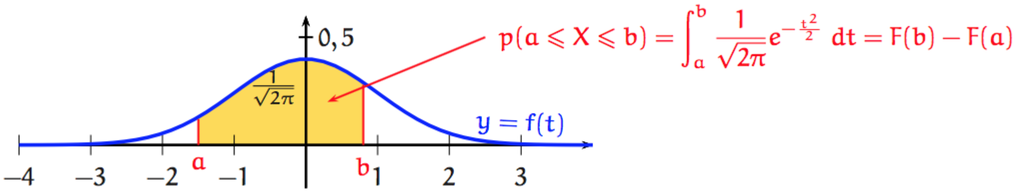
$$\int_{-\infty}^{+\infty} e^{-t^2} \, \mathrm{d}t = \sqrt{\pi}$$

La particularité de l'intégrale de Gauss c'est que la fonction à intégrer n'admet aucune primitive s'exprimant à l'aide des fonctions usuelles.



Ainsi $\{\forall (a,b) \in \mathbb{R}^2 | a < b\}$

$$\mathbb{P}(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt = F(b) - F(a)$$



On ne sait pas exprimer les intégrales précédentes à l'aide des fonctions usuelles à disposition en terminale. Pour obtenir des valeurs de la fonction de répartition de la loi normale centrée réduite, on dispose soit de la calculatrice, soit de tables numériques fournies à la fin de certains livres.
L'espérance de la loi normale centrée réduite est 0 (la loi est centrée) et l'écart-type de la loi normale centrée réduite est 1 (la loi est réduite) ($\mathcal{N}(0,1)$)

Théorème de Moivre-Laplace

La loi normale approche la loi binomiale. Plus précisément :
Théorème : Pour tout entier naturel non nul n, on considère une variable aléatoire X_n qui suit une loi binomiale $\mathcal{B}(n,p)$ puis on considère Zn tel que :

$$Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$$

la variable centrée réduite associée.
Alors, pour tous réels a et b tels que $a < b$

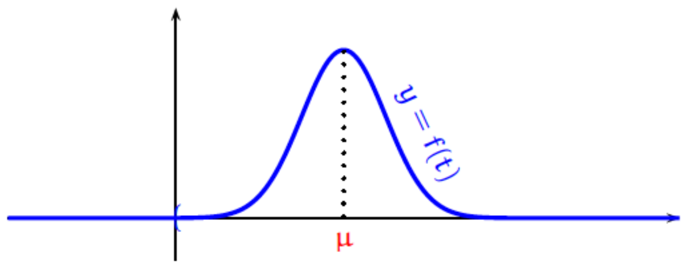
$$\lim_{n \to +\infty} \mathbb{P}\left(np + a\sqrt{np(1-p)} \leq X_n \leq np + b\sqrt{np(1-p)}\right) = \lim_{n \to +\infty} \mathbb{P}(a \leq Z_n \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

Loi normale : cas général

Une variable aléatoire X suit la loi $\mathcal{N}(\mu, \sigma^2)$ si et seulement si la variable aléatoire

$$\frac{X-\mu}{\sigma}$$

suit la loi normale centrée réduite $\mathcal{N}(0,1)$.
μ est l'espérance de $\mathcal{N}(\mu, \sigma^2)$, σ^2 est sa variance et σ est son écart-type.

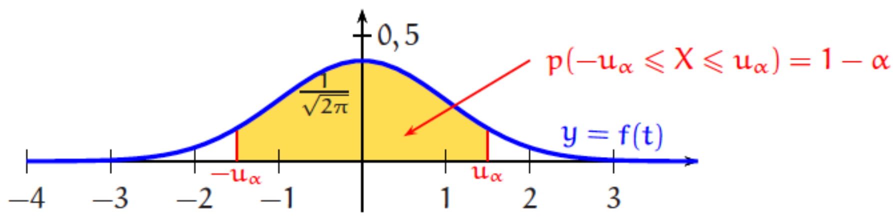


Intervalle associé à une probabilité

Théorème : Soit X une variable aléatoire régie par la loi normale centrée réduite $\mathcal{N}(0,1)$. Pour tout réel $\alpha \in]0,1[$, il existe un réel strictement positif u_α et un seul tel que

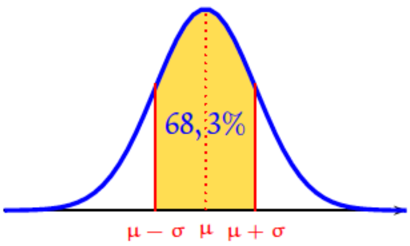
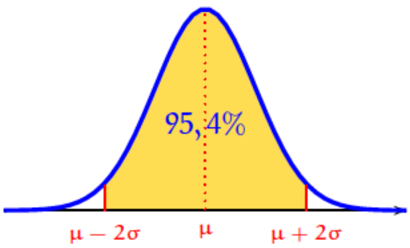
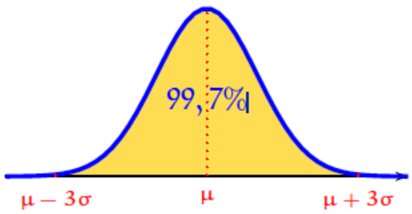
$$\mathbb{P}(-u_\alpha \leq X \leq u_\alpha) = 1 - \alpha$$

.



On doit connaître en particulier $u_{0,05} = 1,96$ (intervalle à 95%) et $u_{0,01} = 2,58$ (intervalle à 99%).
Inversement, pour une loi normale en général, on doit connaître les probabilités associées à des intervalles centrés autour de μ de rayon σ, 2σ ou 3σ :

Théorème : Soit X une variable aléatoire régie par une loi normale $\mathcal{N}(\mu, \sigma^2)$ de paramètres $\mu \in \mathbb{R}$ et $\sigma > 0$.

$p(\mu-\sigma \leqslant X \leqslant \mu+\sigma) = 0,683$	$p(\mu-2\sigma \leqslant X \leqslant \mu+2\sigma) = 0,954$	$p(\mu-3\sigma \leqslant X \leqslant \mu+3\sigma) = 0,997$
		

Conséquence : Si X_n est une variable aléatoire régie par une loi binomiale $\mathcal{B}(n,p)$ et si

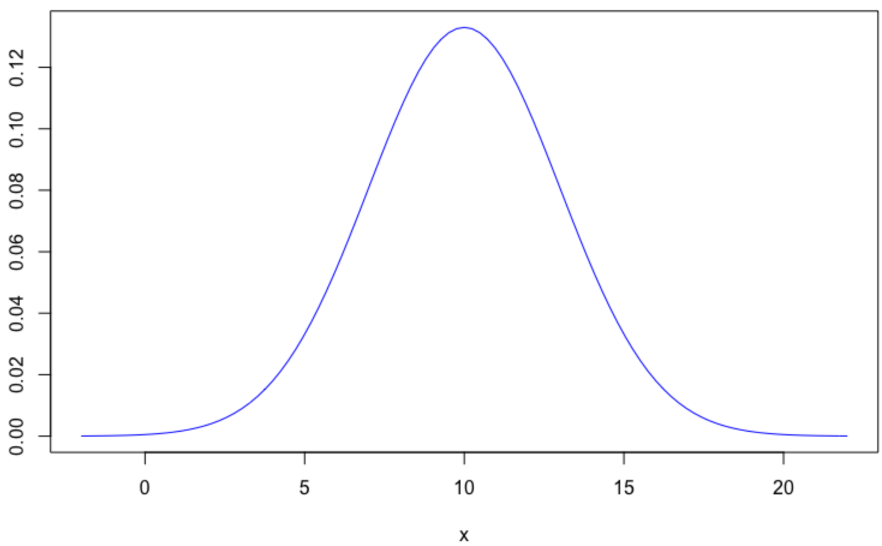
$$Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$$

Alors :

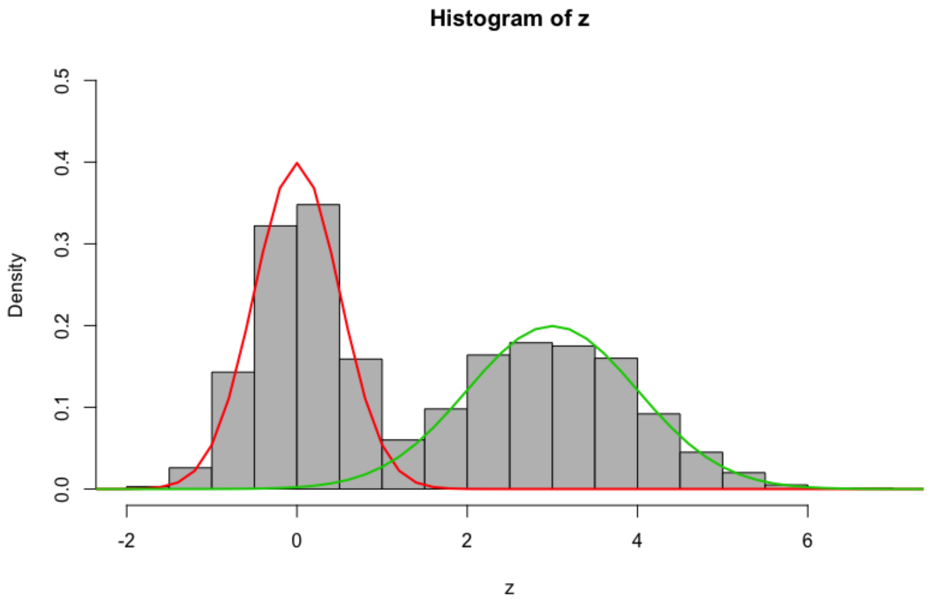
$$\lim_{n\to+\infty} \mathbb{P}\left(np - u_\alpha\sqrt{np(1-p)} \leq X_n \leq np + u_\alpha\sqrt{np(1-p)}\right) = \lim_{n\to+\infty} \mathbb{P}(-u_\alpha \leq Z_n \leq u_\alpha) = 1-\alpha$$

Graphes

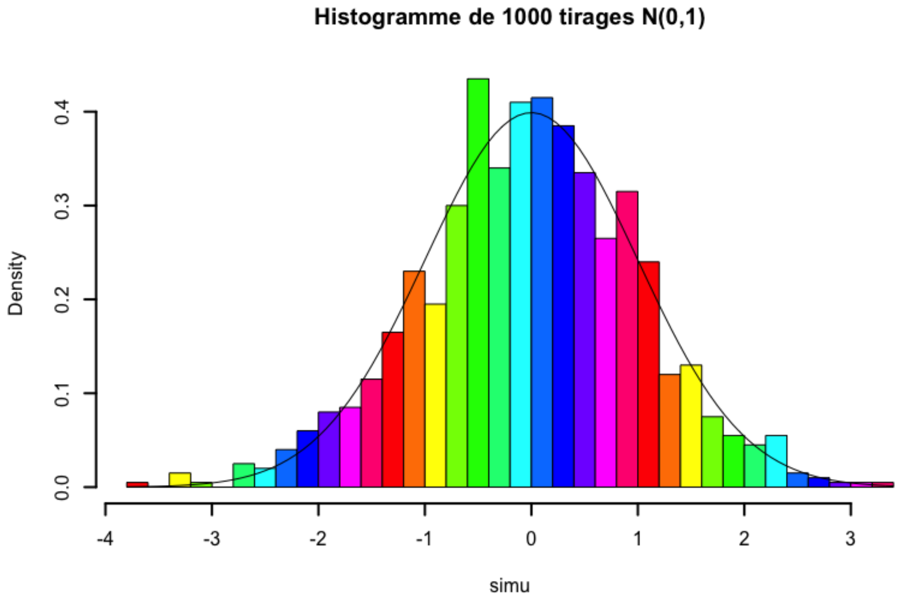
```
> moy <- 10
> std <- 3
> plot(function(x) dnorm(x,moy,std),  moy-4*std,  moy+4*std,xlab="x",ylab="",main = "
  ",col="blue")
```



```
> x <- rnorm(1000,0,0.5)
> y <- rnorm(1000,3,1)
> z <- c(x,y)
> r <- c(-10,10)
> hist(z,freq=FALSE,col="grey",ylim=c(0,0.5),20)
> plot(function(x) 0.5*dnorm(x,0,0.5),xlim=r,col=2,add=TRUE,lwd=2)
> plot(function(x) 0.5*dnorm(x,3,1),xlim=r,col=3,add=TRUE,lwd="2")
```




```
> {hist(simu, prob=T, breaks="FD",
+        main="Histogramme de 1000 tirages N(0,1)",col=rainbow(12), lwd=2)}
> curve(dnorm(x), add=T)
```


Histogramme de 1000 tirages N(0,1)
Density
0.0
0.1
0.2
0.3
0.4
-4
-3
-2
-1
0
1
2
3
simu

2 Introduction aux lois de probabilité avec R - Christophe Chesneau

Auteur : Christophe Chesneau

Note : L'objectif de ce document est de présenter les principales commandes **R** associées aux lois de probabilités et simulations de variables aléatoires réelles (*var*). Quelques éléments théoriques sont consultables ici
Bonne lecture !

2.1 Syntaxe générale

Pour une *var* X suivant une loi notée loi dans **R**, la syntaxe générale est la suivante :

Densité : pour obtenir "la densité" de X, la commande est : dloi ; on ajoute la lettre d devant loi

Fonction de répartition : pour obtenir la fonction de répartition de X, la commande est : ploi ; on ajoute la lettre p devant loi

Quantile : pour obtenir le quantile de X, la commande est : qloi ; on ajoute la lettre q devant loi,

Simulation : pour simuler des réalisations de *var* suivant la même loi que X, la commande est : rloi ; on ajoute la lettre r devant loi.

Ci-dessous, un tableau grossier récapitulatif :

Loi	*Densité*	*Fonction de répartition*	*Quantile*	*Simulation*
Notations	$f(x); \mathbb{P}(X = x)$	$F(x)$	Valeur liée à $F(x)$	$x_1, \dots, x_n$
Commandes	dloi	ploi	qloi	rloi

Les noms loi les plus célèbres sont : *norm* (pour la loi normale), *binom* (pour la loi binomiale) (la loi de *Bernoulli* est une loi binomiale répétée une fois), *unif* (pour la loi uniforme), *geom* (pour la loi géométrique), *pois* (pour la loi de Poisson), *t* (pour la loi de Student), *chisq* (pour la loi du Chi-deux), *exp* (pour la loi exponentielle), *f* (pour la loi de Fisher)...

2.2 Densité

Définition

— Pour une var X discrète, on appelle "la densité" de X en x la probabilité $\mathbb{P}(X = x)$.
— Pour une var X à densité (continue) de densité f_X, on appelle "la densité" de X en x la fonction $f_X(x)$.

Dans les deux cas, la densité ainsi définie caractérise la loi de X.

Commandes

Si la loi de X dépend d'un ou de plusieurs paramètres, disons *par*$_1$ et *par*$_2$, alors la densité de X en x est donnée par les commandes :

dloi(x,par$_1$,par$_2$)

Quelques exemples sont décrits ci-dessous :

Loi	*Binomiale*	*Géométrique*	*Poisson*
Paramètres	$n \in \mathbb{N}*, p \in]0,1[$	$p \in]0,1[$	$\lambda > 0$
$X \sim$	$\mathcal{B}(n,p)$	$\mathcal{G}(p)$	$\mathcal{P}(\lambda)$
$X(\Omega)$	$\{0,\dots;n\}$	$\mathbb{N}$	$\mathbb{N}$
$\mathbb{P}(X=x)$	$\binom{x}{n} p^x(1-p)^{n-x}$	$p(1-p)^x$	$e^{-\lambda}\frac{\lambda^x}{x!}$
Commandes	dbinom(x,n,p)	dgeom(x,p)	dpois(x,lambda)

Loi	*Uniforme*	*Exponentielle*	*Normale*
Paramètres	$(a,b) \in \mathbb{R}^2$	$\lambda > 0$	$\mu \in \mathbb{R}, \sigma > 0$
$X \sim$	$\mathcal{U}([a,b])$	$\varepsilon(\lambda)$	$\mathcal{N}(\mu,\sigma^2)$
$X(\Omega)$	$[a,b]$	$[0,\infty[$	$\mathbb{R}$
$\mathbb{P}(X=x)$	$\frac{1}{b-a}$	$\lambda e^{-\lambda x}$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Commandes	dunif(x,a,b)	dexp(x,$lambda$)	dnorm(x,mu, sigma)

Pour compléter, voir : *help("dgamma"), help("dt"), help("dchisq")* et*help("df")*.

Calculs

On fait :

```
> dbinom(4, 8, 0.3)
[1] 0.1361367
```

Ainsi, on a calculé la densité d'une var $X \sim \mathcal{B}(8;0.3)$ en $x = 4$:

$$\mathbb{P}(X = 4)\binom{8}{4}(0.3)^8(1-0.3)^{8-4}$$

On peut vérifier cela en faisant :

```
> choose(8, 4) * 0.3^4 *(1- 0.3)^(8 - 4)
[1] 0.1361367
```

On fait :

```
> dnorm(1.7, 2, 0.12)
[1] 0.1460692
```

Ainsi, on a calculé la densité d'une var $X \sim \mathcal{N}(2, 0.122)$ en $x = 1,7$:

$$f_X(1.7) = \frac{1}{\sqrt{2\pi(1.12)^2}} e^{-\frac{(1.7-2)^2}{2\times 0.12^2}}$$

On peut vérifier cela en faisant :

```
> (1 / sqrt(2 * pi * 0.12^2)) * exp(- (1.7 - 2)^2 / (2 * 0.12^2))
[1] 0.1460692
```

Pour calculer la densité en plusieurs valeurs, on prend pour x le vecteur ayant pour éléments ces valeurs. On peut faire de même avec un ensemble de paramètres et les arguments correspondants.

On fait :

```
> dbinom(c(4, 6), 8, 0.3)
[1] 0.13613670 0.01000188
```

On a ainsi calculé la densité d'une var $X \sim \mathcal{B}(8, 0.3)$ pour $x \in \{4, 6\}$.

On fait :

```
> dexp(2, c(1, 2, 3))
[1] 0.135335283 0.036631278 0.007436257
```

Ainsi, on a calculé la densité d'une var $X \sim \varepsilon(\lambda)$ en $x = 2$, avec $\lambda = 1$, $\lambda = 2$ et $\lambda = 3$.

On peut aussi mettre ces résultats dans un vecteur pour le réutiliser ultérieurement.
On fait :

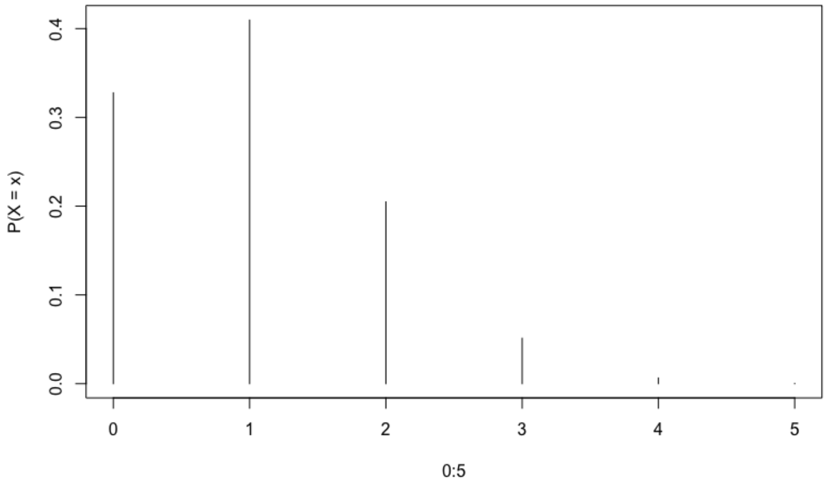
```
> vec = dexp(2, c(1, 2, 3))
> vec
[1] 0.135335283 0.036631278 0.007436257
```

Représentation graphique

On peut représenter le graphe de la densité d'une var X discrète avec la commande plot et l'option *type* $=$ *h*.

```
> plot(0:5, dbinom(0:5, 5, 0.2), type = "h", ylab = "P(X = x)")
```

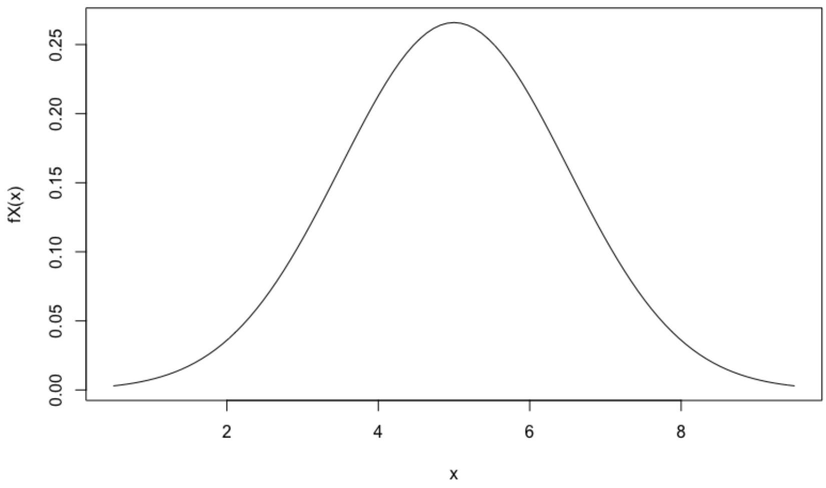
Cela renvoie :



On a ainsi représenté le graphe de la densité d'une var $X \sim \mathcal{B}(5, 0.2)$.

On peut représenter le graphe de la densité d'une var X à densité avec la commande *curve*. On fait :

```
> curve(dnorm(x, 5, 1.5), 0.5, 9.5, ylab = "fX(x)")
```



On a ainsi représenté le graphe de la densité d'une var $X \sim \mathcal{N}(5, 1.5^2)$.

2.3 Fonction de répartition

Définition

On appelle fonction de répartition d'une var X en x la fonction $F_X(x) = P(X \leq x)$.

— Si X est discrète, on a :

$$F_X(x) = \sum_{k \in X(\Omega) \bigcap]-\infty, x]} \mathbb{P}(X = k)$$

— Si X est à densité, on a :

$$F_X(x) = \int_{-\infty}^{x} f_X(t) \mathrm{d}t$$

Commandes

Si la loi de X dépend d'un ou de plusieurs paramètres, disons par_1 et par_2, alors la fonction de répartition de X en x est donnée par les commandes :

ploi(x,par$_1$,par$_2$)

On peut calculer :

$$\mathbb{P}(X > x) = 1 - F_X(x)$$

En faisant :

ploi(x,par$_1$,par$_2$,lower.tail = FALSE)

Calculs

On fait :

```
> pbinom(4, 8, 0.3)
[1] 0.9420324
```

Ainsi, on a calculé la fonction de répartition d'une var $X \sim \mathcal{B}(8, 0.3)$ en $x = 4$:

$$F_X(4) = \mathbb{P}(X \leq 4) = \sum_{k=0}^{4} \mathbb{P}(X = 4)$$

On peut vérifier le calcul précédent en faisant :

```
> sum(dbinom(0:4, 8, 0.3))
[1] 0.9420323
```

On fait :

```
> pnorm(12, 9, 2)
[1] 0.9331928
```

Ainsi, on a calculé la fonction de répartition d'une var $X \sim \mathcal{N}(9, 2^2)$ en $x = 12$:

$$F_X(12) = \mathbb{P}(X \leq 12) = \int_{-\infty}^{12} \frac{1}{\sqrt{2\pi \times 2^2}} e^{-\frac{(t-9)^2}{2\times 2^2}}$$

On fait :

```
> pexp(2, 3, lower.tail = FALSE)
[1] 0.002478752
```

Ainsi, on a calculé : $\mathbb{P}(X > 2), X \sim \varepsilon(3)$

$$\mathbb{P}(X > 2) = \int_{2}^{+\infty} f_X(x)\mathrm{d}x = \int_{2}^{+\infty} 3e^{-3x}\mathrm{d}x = [-e^{-3x}]_2^{\infty} = e^{-6}$$

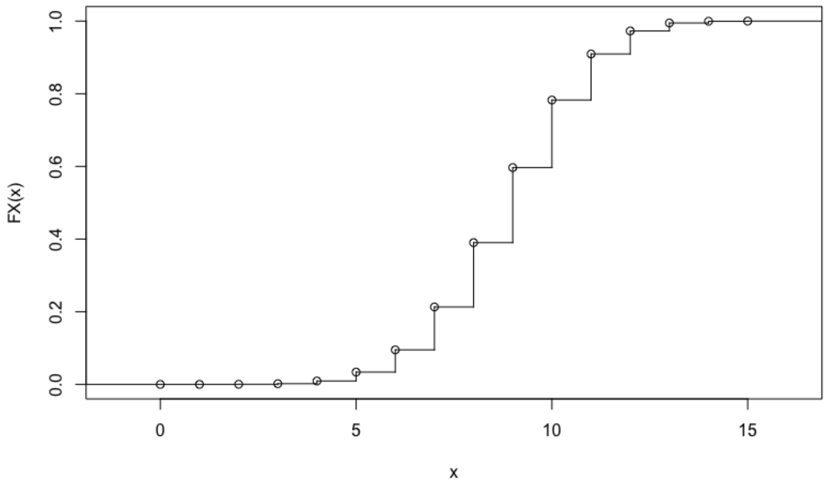
On peut vérifier le calcul précédent en faisant :

```
> exp(-6)
[1] 0.002478752
```

Représentation graphique

On peut représenter le graphe de la fonction de répartition d'une var X discrète avec la commande *stepfun*.
On fait :

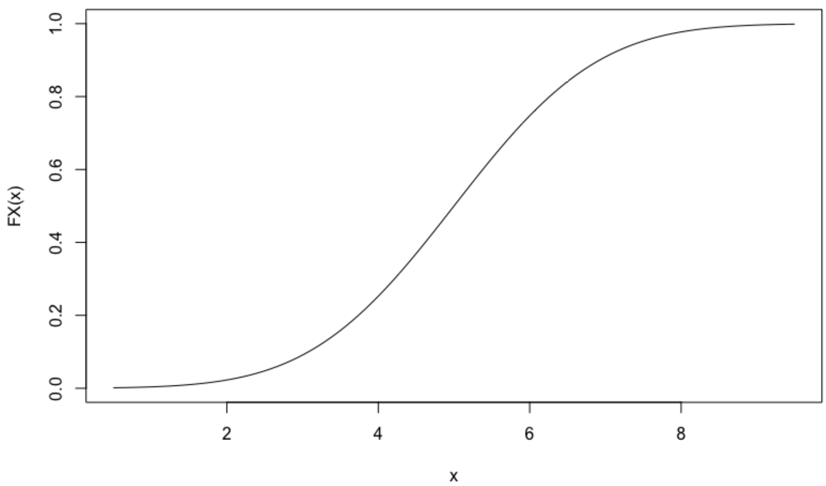
```
> plot(stepfun(0:15, c(0, pbinom(0:15, 15, 0.6))), ylab = "FX(x)", main = "")
```



On a ainsi représenté le graphe de la fonction de répartition d'une var $X \sim \mathbb{B}(15, 0.6)$.

On peut représenter le graphe de la fonction de répartition d'une var X à densité avec la commande *curve*.
On fait :

```
> curve(pnorm(x, 5, 1.5), 0.5, 9.5, ylab = "FX(x)")
```



2.4 Quantiles

Définition

Soit $p \in]0;1[$ et X une var.

— Si X est discrète, on appelle p-ème quantile de X l'entier x_p défini par

$$x_p = \inf\{k \in \mathbb{Z}; F_X(k) \geq p\}$$

— Si X est à densité, on appelle p-ème quantile de X le réel x_p tel que :

$$F_X(x_p) = p$$

Commandes

Si la loi de X dépend d'un ou de plusieurs paramètres, disons *par*$_1$ et *par*$_2$, alors le p-ème quantile de X est donné par les commandes :

qloi(x,par$_1$,par$_2$)

Calculs

On fait :

```
> qbinom(0.25, 5, 0.6)
[1] 2
```

Ainsi, on a calcule le p-ème quantile avec $p = 0,25$ d ?une var $X \sim \mathcal{B}(5, 0.6)$:

$$x_{0.25} = \inf\{k \in \mathbb{Z}; F_X(k) \geq 0.25\}$$

.

On peut vérifier ce résultat en cherchant toutes les valeurs de FX(x) pour $x \in \{0, 1, 2, 3, 4, 5\}$:

```
> pbinom(0:5, 5, 0.6)
[1] 0.01024 0.08704 0.31744 0.66304 0.92224 1.00000
```

On constate alors que $F_X(1) = 0.08704 < 0.25 < 0.31744 = F_X(2)$, donc $x_{0.25} = 2$.

On fait :

```
> qnorm(0.975, 0, 1)
[1] 1.959964
```

Ainsi, on a calcule le p-ème quantile avec $p = 0.975$ d ?une var $X \sim \mathcal{N}(0, 1)$: le réel $x_{0.975}$ tel que $F_X(x_{0.975}) = 0.975$.
On peut vérifier cela en faisant :

```
> pnorm(1.96, 0, 1)
[1] 0.9750021
```

On aurait aussi pu considérer les commandes : *pnorm(1.96)*, sans spécifier $\mu = 0$ et $\sigma = 1$; la commande prend par défaut les paramètres de la loi centrée et réduite.

2.5 Simulation

2.5.1 La commande sample

Utilisation de base

L'utilisation de base est : *sample(x)*, où x est un vecteur numérique, logique ou chaîne de caractères.
Cela renvoie un vecteur dont les éléments correspondent au tirage d'un élément de x **sans remise**.
Cela revient à permuter aléatoirement les éléments de x.
On fait :

```
> x = 1:7
> sample(x)
[1] 7 2 4 6 1 3 5
```

On refait la même commande :

```
> sample(x)
[1] 3 4 1 5 6 2 7
> sample(x)
[1] 4 3 5 6 2 1 7
```

On obtient des résultats différents ; l'ordre de la permutation est due au hasard à chaque fois.
On peut conserver le résultat d'une permutation aléatoire dans un vecteur pour le réutiliser ultérieurement :

```
> y = sample(c("rouge", "vert", "bleu", "blanc", "noir"))
> y
[1] "noir"  "bleu"  "vert"  "blanc" "rouge"
> y = sample(c("rouge", "vert", "bleu", "blanc", "noir"))
> y
[1] "bleu"  "noir"  "blanc" "vert"  "rouge"
```

Options

Il existe des options dans *sample* permettant de modifier le type de tirage. On les active en rajoutant une ou plusieurs commandes dans *sample*.
Par exemple, on fait :

```
> sample(1:3, size = 2, replace = TRUE, prob = c(25 / 100, 20 / 100,55 / 100))
 [1] 1 1
> sample(1:3, size = 2, replace = TRUE, prob = c(25 / 100, 20 / 100,55 / 100))
 [1] 3 3
> sample(1:3, size = 2, replace = TRUE, prob = c(25 / 100, 20 / 100,55 / 100))
 [1] 3 2
```

Quelques options sont présentées ci-dessous.

— *size* : Les commandes : *size* = *"n"*, où n est un entier, précise le nombre de fois n que l'on effectuera le tirage. Par défaut, le nombre de tirage est égal à la longueur de x.
On fait :

```
 > sample(1:10, size = 3)
 [1] 9 5 7
 > sample(1:10, size = 3)
 [1] 3 2 9
```

Dans la syntaxe de sample, l'argument *size* est en deuxième position. On peut donc se contenter de : *sample(1 :10, 3)*.

— *replace* : Les commandes : *replace* = *L*, où L est TRUE ou FALSE, dont que les tirages sont avec remise si L = TRUE et sans remise sinon.
On fait :

```
 > sample(1:5, size = 9, replace = TRUE)
 [1] 3 2 2 3 3 3 2 4 4
 > sample(1:5, size = 9, replace = TRUE)
 [1] 4 2 5 2 1 2 5 4 3
```

— *prob* La commande *prob* permet la simulation de var dont le support est l'ensemble des éléments de x, que l'on suppose au nombre de k. Ainsi, la loi commune de ces var est :

$$\mathbb{P}(X = j\text{-ème élément de } x) = p_j \text{ avec } j \in \{1, \dots, k\}, p_j \in [0, 1] \text{ et } \sum_{j=1}^{k} p_j = 1$$

On ajoute donc *prob* = *vec*, où *vec* est le vecteur des probabilités $(p_1; \dots; p_k)$.
Si *prob* n'est pas indiqué, par défaut $p_j = 1/k$ (tirage équiprobable).
On fait :

```
 > y = c("rouge", "vert", "bleu", "blanc", "noir")
 > sample(y, 2, prob = c(10 / 100, 30 / 100, 10 / 100, 30 / 100, 20 / 100))
 [1] "bleu" "noir"
 > sample(y, 2, prob = c(10 / 100, 30 / 100, 10 / 100, 30 / 100, 20 / 100))
 [1] "blanc" "rouge"
 > sample(y, 2, prob = c(10 / 100, 30 / 100, 10 / 100, 30 / 100, 20 / 100))
 [1] "blanc" "noir"
```

2.5.2 Simulation par des lois préprogrammées

Commandes

Si la loi de X dépend d'un ou de plusieurs paramètres, disons *par*$_1$ et *par*$_2$, alors la simulation de n variables aléatoires indépendantes suivant la même loi que X est donnée par les commandes :

rloi(x,par$_1$,par$_2$)

Calculs

On fait :

```
> rpois(10, 2)
 [1] 2 1 1 1 2 0 1 0 3 1
```

Ainsi, on a simulé des réalisations de 10 var indépendantes suivant chacune la loi de Poisson $\mathcal{P}(2)$.
Encore une pour le fun ?

```
> rpois(10, 2)
 [1] 3 4 2 3 3 2 2 2 2 1
```

Allez, une troisième pour le plaisir

```
> rpois(10, 2)
 [1] 2 1 1 0 2 5 3 2 1 0
```

On fait :

```
> sum(rbinom(80, 1, 0.02))
[1] 3
```

On a alors simulé une réalisation de la var $\sum_{i=1}^{80} X_i$ où $X_1,X_2,\ldots,X_{80}$ sont des var indépendantes suivant chacune la loi binomiale $\mathcal{B}(1,0.02)$, (ainsi, la loi commune est la loi de Bernoulli $\mathcal{B}(0.02)$:

$$\mathbb{P}(X_1 = 0) = 1 - 0.02 = 0.98 \text{ et } P(X_1 = 1) = 0.02.$$

Comme $\sum_{i=1}^{80} X_i \sim \mathcal{B}(80, 0.02)$ cela revient à simulé une réalisation d'une var $Y \sim \mathcal{B}(80, 0.02)$: *rbinom(1, 80, 0.02)*.
On fait :

```
> x = rnorm(15, 22, 2)
> x
 [1] 21.86975 21.27651 21.44174 20.73213 23.46708 23.15454 22.51041 22.90081
 [9] 24.32300 23.98299 23.78338 23.12981 22.93772 21.08650 20.61778
```

Ainsi, on a construit un vecteur dont les éléments sont des réalisations de 15 var indépendantes suivant chacune la loi normale $\mathcal{N}(22, 2^2)$.

2.6 Exercices

2.6.1 Enoncés

Exercice 1

Représenter le graphe de la densité d'une var $X \sim \mathcal{P}(1)$ pour $x \in \{0; \ldots; 8\}$.

Exercice 2

Représenter le graphe de la densité d'une var $X \sim \chi^2(3)$ pour $x \in [0; 10]$.

Exercice 3

Diviser la fenêtre de l'écran en deux fenêtres lignes.

- Représenter, dans la première fenêtre, le graphe de la densité d'une var $X \sim \mathcal{B}(50, 0.08)$ en prenant *ylim = c(0, 0.25)*.
- Représenter dans la deuxième fenêtre, au-dessous de la première, le graphe de la densité d'une var $Y \sim \mathcal{P}(4)$ pour $x \in \{0; \ldots; 50\}$ avec la même option : *ylim = c(0, 0.25)*.

(Ceci illustrera le fait que lorsque $n \geq 31$ et $np \leq 10$ on peut approcher la loi binomiale $\mathcal{B}(n, p)$ par la loi de Poisson $\mathcal{P}(np)$).

Exercice 4

Séparer l'écran graphique en 3 (1 ligne, 3 colonnes).

- Dans la première fenêtre, représenter le graphe de la densité d'une var $X \sim \mathcal{N}(4, 1)$, puis ajouter dans la même fenêtre, avec une autre couleur, celui de la densité d'une var $Y \sim \mathcal{N}(5, 1)$.
- Dans la deuxième fenêtre, représenter le graphe de la densité d'une var $X \sim \mathcal{N}(4, 1)$ puis ajouter dans la même fenêtre, avec une autre couleur, celui de la densité d'une var $Y \sim \mathcal{N}(4, 4)$.
- Dans la troisième fenêtre, représenter le graphe de la densité d'une var $X \sim \mathcal{N}(4, 1)$ puis ajouter dans la même fenêtre, avec une autre couleur, celui de la densité d'une var $Y \sim \mathcal{N}(5, 4)$.

Exercice 5

Représenter le graphe de la densité d'une var $X \sim \mathcal{B}(50, 0.4)$, puis ajouter par dessus ce graphe celui de la densité d'une var $Y \sim \mathcal{N}(20; 12)$ (*cela illustrera le fait que, lorsque $n \geq 31$, $np \geq 5$ et $n(1-p) \geq 5$, on peut approcher la loi binomiale B(n ; p) par la loi normale $\mathcal{N}(np; np(1-p))$*).

Exercice 6

Représenter le graphe de la densité d'une var $X \sim \mathcal{B}(n, p)$ pour $x \in \{0; \ldots; 10\}$ avec $n = 100$ et $p = 0.01$, puis ajouter dans la même fenêtre le graphe de la densité d'une var $Y \sim \mathcal{N}(np, np(1-p))$. Que constatez-vous ?

Exercice 7

Soit $X \sim \mathcal{N}(0,1)$. Calculer les probabilités :

— $\mathbb{P}(X < -0,5)$
— $\mathbb{P}(X > 1,5)$
— $\mathbb{P}(X > -1)$
— $\mathbb{P}(|X| \leq 1,96)$
— $\mathbb{P}(|X| \leq 2,58)$
— $\mathbb{P}(|X| \geq 3)$

Exercice 8

Soit $X \sim \mathcal{N}(15;9)$.

— Calculer les probabilités : $\mathbb{P}(16 \leq X \leq 20)$, $\mathbb{P}(X > 18)$, $\mathbb{P}(X < 6)$ et $\mathbb{P}(|X - 15| > 5,88)$.
— Représenter le graphe de la fonction de répartition de X pour $x \in [6;24]$.

Exercice 9

Représenter, sur une même figure, la fonction de répartition d'une var $X \sim B(50,0.4)$ et celle d'une var $Y \sim \mathcal{N}(20,12)$.

Exercice 10

— Soit $X \sim \mathcal{N}(0,1)$. Calculer les quantiles : $x_{0.00135}$, $x_{0.025}$, $x_{0.95}$, $x_{0.999}$, $x_{0.995}$ et $x_{0.99865}$.
— Soit $Y \sim \mathcal{N}(19,3)$. Calculer les quantiles : $y_{0.975}$ et $y_{0.025}$. Vérifier que $y_{0.975} = 19 + \sqrt{3}x_{0.975}$.

Exercice 11

Construire un vecteur *simul1* composé de 1000 éléments dont les valeurs sont des réalisations d'une var $X \sim \mathcal{N}(15,3)$. Faire l'histogramme des fréquences de *simul1* et superposer la densité de X. Faire le boxplot de *simul1*.

Exercice 12

1. Construire un vecteur *simul2* qui est composé de 1000 éléments dont les valeurs sont des réalisations d'une var $X \sim \mathcal{P}(1)$
2. Construire un vecteur dont le premier élément sera le nombre de réalisations qui vaudront 0, le deuxième élément sera le nombre de réalisations qui vaudront 1...O*On pourra utiliser la fonction compter décrite ci-dessous.*

```
compter = function(a, b) {
  d = numeric()
  for(i in 1:(length(a))) {
    d[i] = sum(b == a[i])
  }
  names(d) = as.character(a)
  d
}
```

3. Faire un barplot et superposer la densité de X.

Exercice 13

Construire dans R une fonction péage qui a 4 arguments : *mu1, mu2, sigma* et *n*. À l'intérieur de cette fonction :

- on définit un vecteur numérique attente de longueur n
- on construit un objet cabines constitué des cabines $C1$ et $C2$
- on ouvre une boucle qui va de 1 à n. Pour chaque i avec $i \in \{1; \ldots; n\}$, à l'intérieur de cette boucle, on tire au hasard une cabine dans l'objet cabines.
  - Si la cabine $C1$ a été choisie, on simule le tirage d'une var d'espérance *mu1* et d'écart-type *sigma* et on affecte le résultat à la i-ème composante de attente
  - Si la cabine $C2$ a été choisie, on simule le tirage d'une var de moyenne *mu2* et d'écart-type *sigma* et on affecte le résultat à la i-ème composante de attente
- après avoir constitué le vecteur *attente*, on construit l'histogramme des fréquences du vecteur *attente*
- on calcule *moy* qui est la moyenne du vecteur attente
- on calcule ecart qui est l'écart-type corrigé des valeurs des éléments du vecteur *attente* en utilisant la fonction **sd**
- on ajoute à l'histogramme la courbe de la densité associée à la loi normale de moyenne moy et d'écart-type *ecart*
- on ajoute à la figure un texte comportant l'indication de la valeur de moy et l'indication de la valeur de *ecart* (garder deux décimales seulement)
- on ajoute à l'histogramme le graphe de la fonction suivante :

$$f(x) = 0.5(f_1(x) + f_2)$$

  où $f_1(x)$ est la densité associée à la loi $\mathcal{N}(\mu_1; \sigma^2)$ et $f_2(x)$ celle associée à la loi $\mathcal{N}(\mu_2; \sigma^2)$.

On exécute ensuite cette fonction sur deux fenêtres :

- pour la première prendre $mu1 = 50$, $mu2 = 51$, $sigma = 2$ et $n = 1000$
- pour la seconde prendre $mu1 = 50$, $mu2 = 60$, $sigma = 2$ et $n = 1000$

L'objectif est de montrer qu'un mélange de lois normales n'est pas obligatoirement une loi normale.

Exercice 14

Diviser l'écran en 4 fenêtres.

- Dans la première fenêtre, représenter le graphe de la densité d'une var $X \sim \mathcal{N}(0, 1)$ pour $x \in [-3; 3]$, puis ajouter celui de la densité d'une var $Y \sim \mathcal{T}(2)$.
- Dans la première fenêtre, représenter le graphe de la densité d'une var $X \sim \mathcal{N}(0, 1)$ pour $x \in [-3; 3]$, puis ajouter celui de la densité d'une var $Y \sim \mathcal{T}(3)$.
- Dans la première fenêtre, représenter le graphe de la densité d'une var $X \sim \mathcal{N}(0, 1)$ pour $x \in [-3; 3]$, puis ajouter celui de la densité d'une var $Y \sim \mathcal{T}(10)$.
- Dans la première fenêtre, représenter le graphe de la densité d'une var $X \sim \mathcal{N}(0, 1)$ pour $x \in [-3; 3]$, puis ajouter celui de la densité d'une var $Y \sim \mathcal{T}(30)$.

Exercice 15

Construire dans R une fonction *stu1* qui a 4 arguments : n, nb, mu et $sigma$.
À l'intérieur de cette fonction, on définit un vecteur *vec* de longueur nb.
On construit les composantes du vecteur par une boucle allant de 1 à nb.
La i-ème composante de vec est calculée ainsi :

- On simule un vecteur *simul* qui est composé de n réalisations d'une var $X \sim \mathcal{N}(mu; sigma^2)$.
- On calcule la moyenne $\bar{x}$ et l'écart-type corrigé s des valeurs des éléments de *simul*.
  La i-ème composante de *vec* est égale à :
  $$\text{vec}[i] = \sqrt{n}\left(\frac{\bar{x} - mu}{s}\right)$$
- Diviser l'écran en 2 fenêtres superposées.
  - Dans la première fenêtre, construire l'histogramme des fréquences de vec. Superposer le graphe de la densité associée à loi $\mathcal{N}(0,1)$ et, avec une autre couleur, le graphe de la densité associée à la loi $\mathcal{T}(n-1)$.
  - Dans la deuxième fenêtre, représenter la fonction de répartition empirique de vec avec les commandes : *plot(ecdf(vec))*. Ajouter en rouge le graphe de la fonction de répartition associée à la loi $\mathcal{T}(n-1)$). Ajouter en jaune le graphe de la fonction de répartition associée à la loi $\mathcal{N}(0,1)$.

Essayer cette fonction avec $n = 3$, $nb = 1000$, $mu = 10$ et $sigma = 2$, puis avec $n = 30$, $nb = 1000$, $mu = 10$ et $sigma = 2$.

Rappel : La fonction de répartition empirique associée à un n-échantillon $(X_1; \dots; X_n)$ d'une var X de fonction de répartition F_X est

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq x\}}$$

Le théorème de Glivenko-Cantelli nous assure que :

$$\max_{-\infty < x + \infty} |\hat{F}(x) - F_X(x)| \to 0 \textit{ converge presque sûrement quand } n \to +\infty$$

Exercice 16

Diviser l'écran en 6 fenêtres.

- Dans la fenêtre 1, représenter le graphe de la densité d'une var $X \sim \chi^2(1)$ pour $x \in [0.01; 10]$.
- Dans la fenêtre 2, représenter le graphe de la densité d'une var $X \sim \chi^2(2)$ pour $x \in [0.01; 10]$.
- Dans la fenêtre 3, représenter le graphe de la densité d'une var $X \sim \chi^2(3)$ pour $x \in [0.01; 10]$.
- Dans la fenêtre 4, représenter le graphe de la densité d'une var $X \sim \chi^2(5)$ pour $x \in [0.01; 10]$.
- Dans la fenêtre 5, représenter le graphe de la densité d'une var $X \sim \chi^2(10)$ pour $x \in [0.01; 20]$.
- Dans la fenêtre 6, représenter le graphe de la densité d'une var $X \sim \chi^2(20)$ pour $x \in [0.01; 60]$.

Exercice 17

Construire dans R une fonction *khi2* qui a 4 arguments : n, nb, mu et $sigma$.
À l'intérieur de cette fonction, on définit un vecteur *vec* de longueur nb.
On construit les composantes de ce vecteur par une boucle allant de 1 à nb. La i-ème composante de vec est calculée ainsi :

- On simule un vecteur *simul* qui est composé de n réalisations d'une var $X \sim \mathcal{N}(mu, sigma^2)$.
- On calcule la variance corrigée s^2 de la série des n composantes de *simul*.

La i-ème composante de *vec* est égale à

$$\text{vec}[i] = \frac{(n-1)s^2}{sigma^2}$$

Une fois ce vecteur construit, on ne garde que les composantes comprises entre 0.01 et 10 et on construit l'histogramme des fréquences.
On superpose le graphe de la densité associée à la loi $\chi^2(n-1)$.
Essayer cette fonction avec $n = 4$, $nb = 10000$, $mu = 10$ et $sigma = 2$.

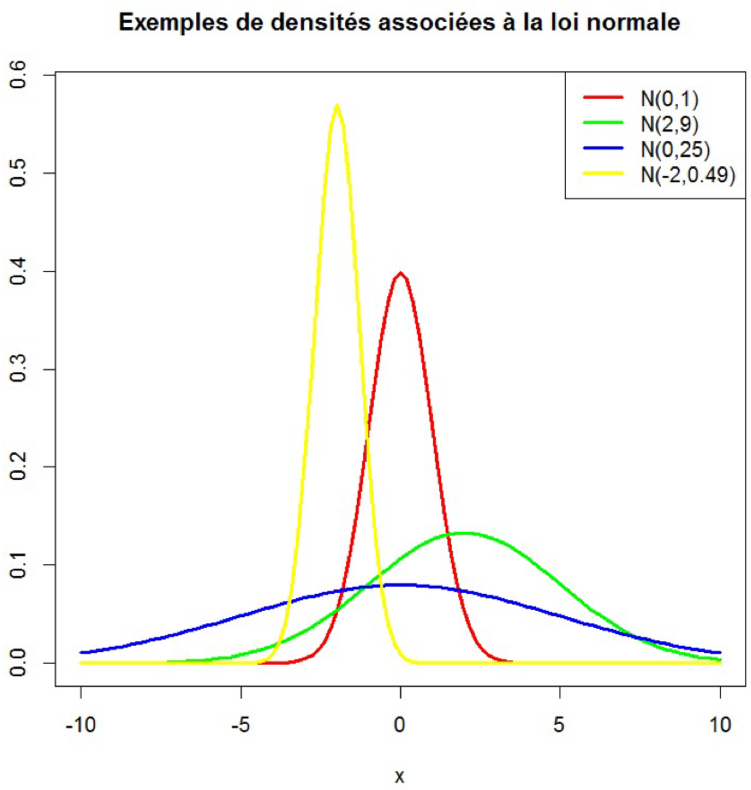
Exercice 18

Diviser l'écran en 4 fenêtres.

- Dans la fenêtre 1, représenter le graphe de la densité d'une var $X \sim \varepsilon(1)$ pour $x \in [0, 3]$.
- Dans la fenêtre 2, représenter le graphe de la densité d'une var $X \sim \varepsilon(2)$ pour $x \in [0, 3]$.
- Dans la fenêtre 3, représenter le graphe de la densité d'une var $X \sim \varepsilon(0.5)$ pour $x \in [0, 20]$.
- Dans la fenêtre 4, représenter le graphe de la densité d'une var $X \sim \varepsilon(0.1)$ pour $x \in [0, 60]$.

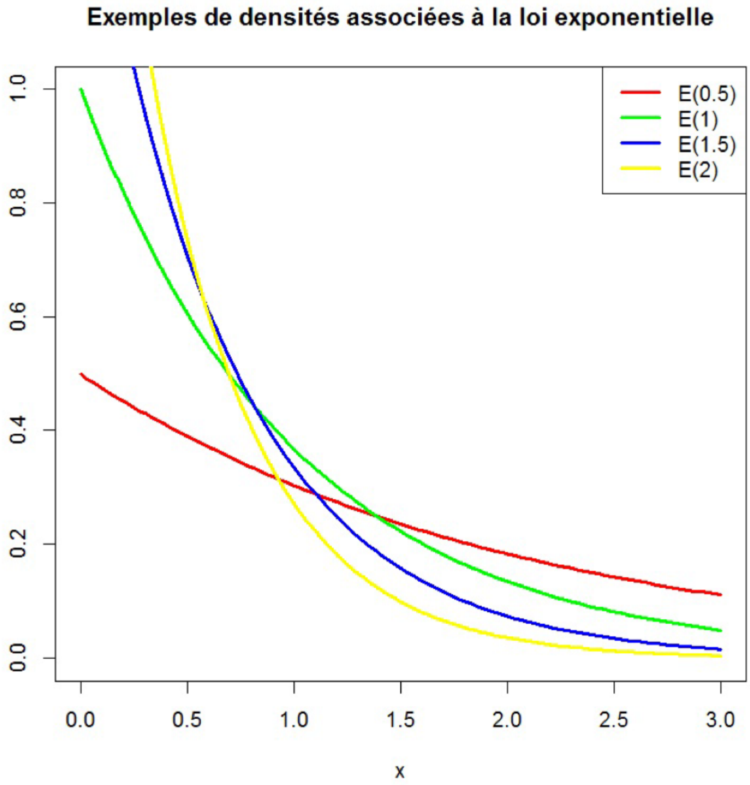
Exercice 19

Proposer des commandes R permettant d'obtenir le graphique suivant :



Exercice 20

Proposer des commandes R permettant d'obtenir le graphique suivant :



Exercice 21

On s'interesse à la durée de vie d'un certain type de voiture. Soit X la var égale à la durée de vie en années d'une de ces voitures. On suppose que X suit la loi exponentielle $\varepsilon(1/10)$, *i.e.* de densité

$$f(x) = \begin{cases} \frac{1}{10} e^{-\frac{x}{10}} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

1. Écrire une nouvelle fonction R équivalente à dexp (on pourra utiliser une structure de contrôle *if . . . else* pour que celle-ci valle 0 quand $x < 0$).
2. Vérifier numériquement que$\int_0^{10} f(x)\mathrm{d}x \approx 1$ avec la commande *integrate*.
3. Séparer l'écran graphique en 2 : dans la fenêtre 1, représenter le graphe de la densité f et, dans la fenêtre 2, celui de la fonction de répartition de X.
4. Calculer la probabilité que la durée de vie d'une voiture dépasse 10.2 ans.

Exercice 22

Soit X une var suivant la loi gamma $\Gamma(5,1)$, *i.e.* de densité :

$$f(x) = \begin{cases} \frac{1}{24}x^4 e^{-x} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

1. Écrire une nouvelle fonction R équivalente à dgamma.
2. Vérifier numériquement que$\int_0^{10} f(x)\mathrm{d}x \approx 1$
3. Évaluer la fonction de répartition de X pour $x \in \{2; \ldots ; 10\}$
4. Déterminer le réel x vérifiant $\mathbb{P}(X \leq x) = 0.92$.

Exercice 23

Soit X une var suivant la loi normale $\mathcal{N}(0,1)$, *i.e.* de densité :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, x \in \mathbb{R}$$

1. Écrire une nouvelle fonction R équivalente à la fonction dnorm
2. Vérifier numériquement que$\int_0^{10} f(x)\mathrm{d}x \approx 1$
3. Séparer l'écran graphique en 2 : dans la fenêtre 1, représenter le graphe de la densité f et, dans la fenêtre 2, celui de la fonction de répartition de X.
4. Calculer $\mathbb{P}(X \leq 2.2)$, $\mathbb{P}(X \geq 1.7)$, $\mathbb{P}(0.2 \leq X < 1.4)$ et $\mathbb{P}(|X| \leq 1.96)$.
5. Déterminer le réel x vérifiant $\mathbb{P}(X \leq x) = 0.98$.

Exercice 24

On suppose que le poids d'un foie gras peut être modélisé par une var X suivant la loi normale $\mathcal{N}(550, 100^2)$, l'unité étant le gramme. Quelle est la probabilité qu'un foie gras pèse :
— moins de 650 grammes ?
— plus de 746 grammes ?
— entre 550 grammes et 600 grammes ?

Exercice 25

Soit X une var suivant la loi normale $\mathcal{N}(0,1)$. La fonction de répartition de X, *i.e.* $F_X(x) = \mathbb{P}(X \leq x)$, avec $x \geq 0$, peut s'écrire comme

$$F_X(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \sum_{n=0}^{\infty} \frac{x^{2n+1}}{\prod_{m=0}^{n}(2m+1)}$$

Utiliser cette expression pour écrire une nouvelle fonction R équivalente à la fonction pnorm (*on tronquera la somme infinie à la valeur 50*).

Exercice 26

Soit X une var dont la loi est donnée par

$$\mathbb{P}(X=0)=0.2,\ \mathbb{P}(X=2)=0.5,\ \mathbb{P}(X=5)=0.3$$

Simuler 1000 réalisations de X et préciser les effectifs associés aux valeurs de X.

Exercice 27

Une urne contient $p+q$ boules, dont p rouges et q noires. Construire dans R une fonction Urne qui a 3 arguments : k, p et q. L'enjeu de celle-ci est de modéliser le résultat de k tirages sans remise d'une boule de l'urne. Par exemple :

```
> Urne(6, 8, 5)
[1] "Rouge" "Noire" "Noire" "Rouge" "Noire" "Rouge"
```

Exercice 28

Lorsqu'on effectue n tirages indépendants d'une même expérience aléatoire, on appelle fréquence du résultat k le rapport entre le nombre de fois où k est tiré, et n.
Par exemple, si on jette 7 fois un dé cubique équilibré, avec pour résultats : $1;1;5;2;6;5;3$, alors la fréquence de 5 est $2/7$, celle de 4 est 0.
Construire dans R une fonction Freq qui a un argument : n. L'enjeu de celle-ci est de renvoyer la fréquence de 5 lors de n tirages indépendants d'un dé cubique équilibré. Comparer les fréquences pour $n \in \{10, 100, 1000\}$, avec la probabilité (théorique) d'obtenir un 5 lorsqu'on lance un dé.

Exercice 29

Une urne contient 15 boules, dont 5 blanches et 10 noires. On considère les deux expériences suivantes :
- *E1* : On tire successivement 10 boules dans l'urne, avec remise
- *E2* : On tire successivement 10 boules dans l'urne, sans remise

1. Simuler une réalisation de chacune des deux expériences avec la fonction sample. On représentera une boule blanche par le chiffre 1 et une boule noire par le chiffre 0.
2. On s'intéresse à la var X égale au nombre de boules blanches tirées lors l'expérience *E1*, et à la var Y , l'analogue mais avec l'expérience *E2*.
   (a) Simuler 500 réalisations de chacune de ces var en utilisant la fonction sample.
   (b) Comparer, suivant le type d'épreuve, le diagramme en barre des fréquences observées avec la distribution des lois binomiales et hypergéométriques correspondantes.

Exercice 30

On considère la marche aléatoire suivante : un mobile est positionné à l'origine d'un axe. À chaque étape, il se déplace d'une distance de longueur 1 vers la droite ou la gauche avec la probabilité 0.5 pour chaque direction. Il effectue n étapes au total. Soit X_i la position du mobile à l'étape i (on pose $X_0 = 0$). Simuler une réalisation du vecteur de var $(X_0; X_1; \dots; X_n)$ (on pourra utiliser la fonction *cumsum* qui donne la somme cumulée d'un vecteur. Par exemple :

```
> cumsum(c(2, 2, 2, 3))
 [1] 2 4 6 9.
```

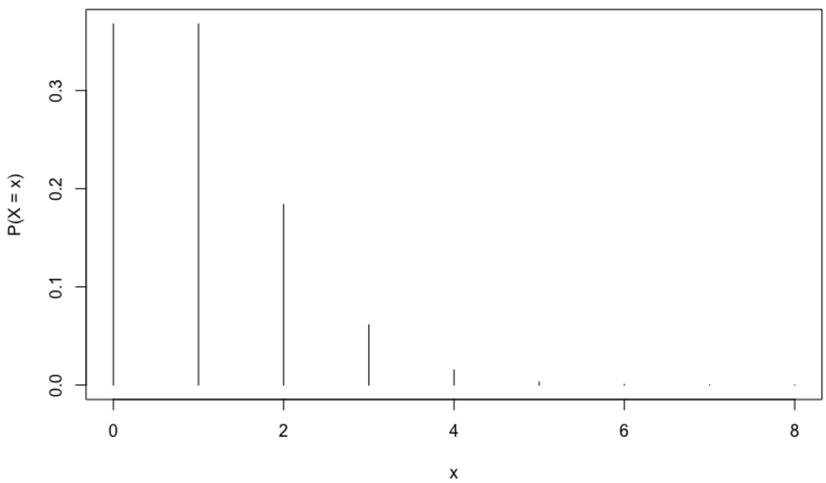
Exercice 31

On lance 5 dés cubiques équilibrés, puis on relance ceux qui n'ont pas fait 6, et ainsi de suite jusqu'à ce que les 5 dés affichent 6.
Simuler une réalisation de cette expérience aléatoire en affichant les chiffres obtenus après chaque lancers et le nombre de lancers total (*on pourra utiliser une boucle* ***while*** *et la commande* ***sum(x != 6)*** *qui calcule le nombre de composantes d'un vecteur x différentes de 6*).

2.6.2 Solutions

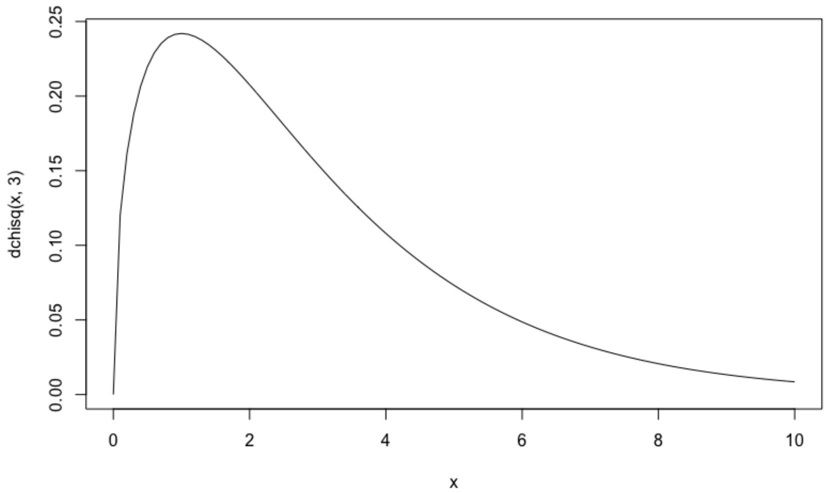
Exercice 1

```
> plot(0:8, dpois(0:8, 1), type = "h", xlab = "x", ylab = "P(X = x)")
```



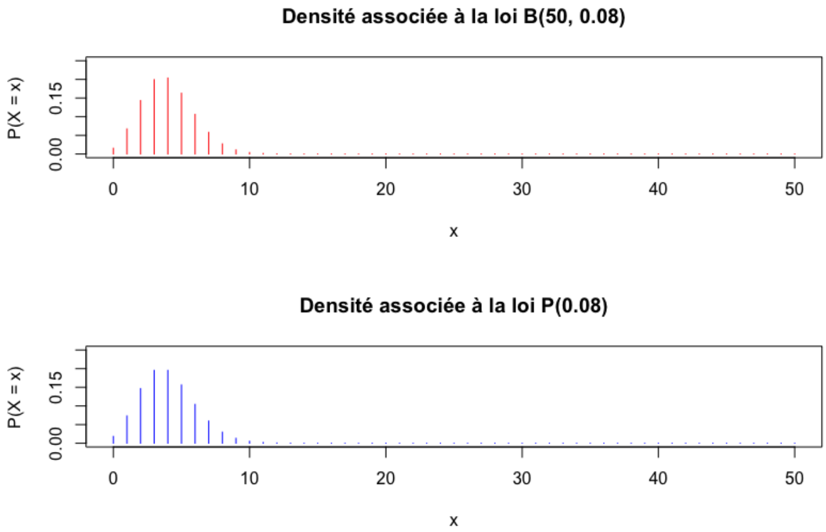
Exercice 2

```
> curve(dchisq(x, 3), 0, 10, xlab = "x")
```



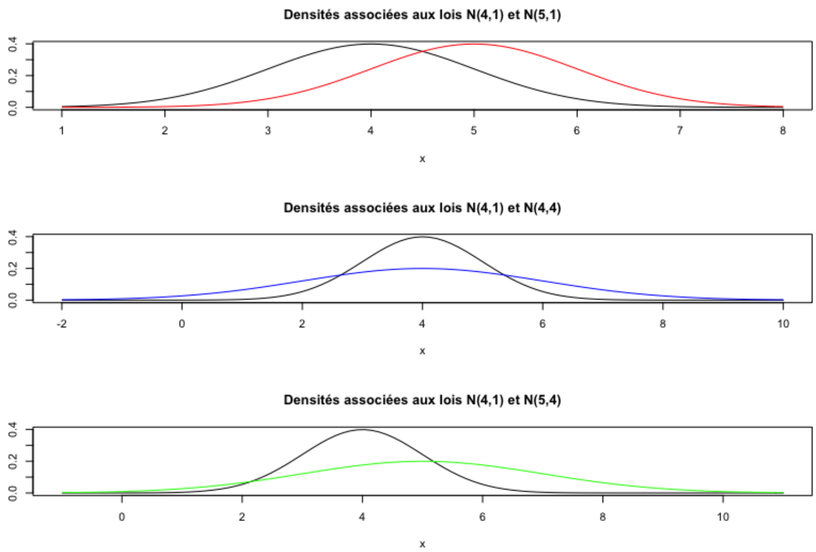
Exercice 3

```
## On split la fenêtre en deux
> par(mfrow = c(2,1))
# On veut les densité donc : dloi
## Var Bino -> B(50,0.08)
> plot(0:50, dbinom(0:50, 50, 0.08), type = "h", xlab = "x", ylab = "P(X = x)",
+      ylim=c(0, 0.25), main = "Densité associée à la loi B(50, 0.08)",col="red")
## Var Poiss -> P(0,4)
> plot(0:50, dpois(0:50, 4), type = "h", xlab = "x", ylab = "P(X = x)",
+      ylim = c(0, 0.25), main = "Densité associée à la loi P(0.08)",col="blue")
## On oubli pas de mettre les options graphique par défaut
> par(mfrow = c(1,1))
```



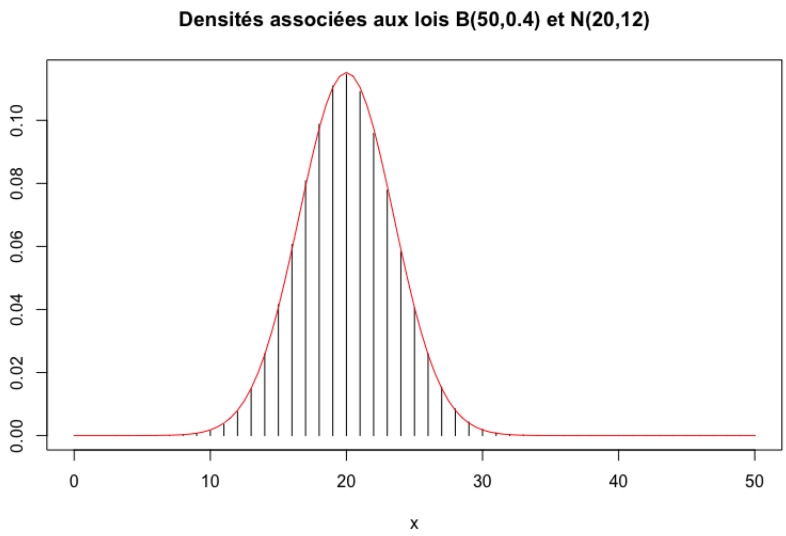
Exercice 4

```
> par(mfrow = c(3, 1))
## ATTENTION : à bien mettre add = TRUE pour pouvoir superposer les graphiques
> curve(dnorm(x, 4, 1), 1, 8, ylab = "",
+        main = "Densités associées aux lois N(4,1) et N(5,1)")
> curve(dnorm(x, 5, 1), 1, 8, col = "red", ylab = "", add = TRUE)
> curve(dnorm(x, 4, 1), -2, 10, ylab = "",
+        main = "Densités associées aux lois N(4,1) et N(4,4)")
> curve(dnorm(x, 4, 2), -2, 10, col = "blue", ylab = "", add = TRUE)
> curve(dnorm(x, 4, 1), -1, 11, ylab = "",
+        main = "Densités associées aux lois N(4,1) et N(5,4)")
> curve(dnorm(x, 5, 2),-1, 11, col = "green", ylab = "", add = TRUE)
## On oubli pas de mettre les options graphique par défaut
> par(mfrow = c(1,1))
```



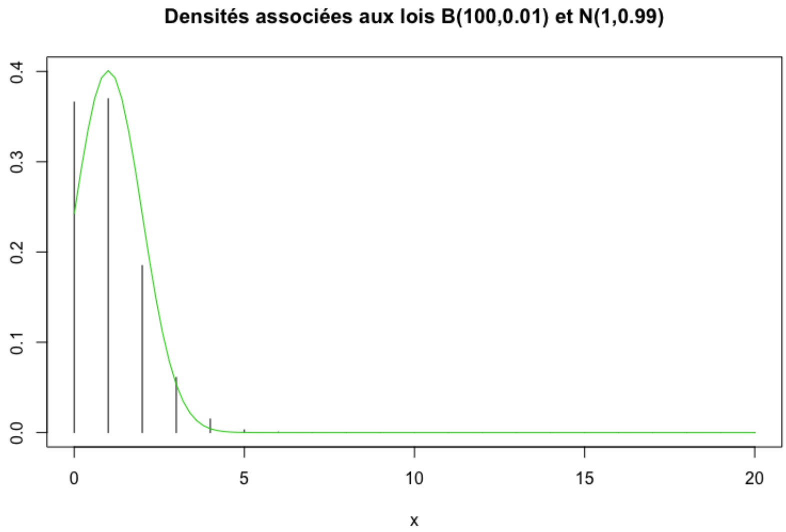
Exercice 5

```
> plot(0:50, dbinom(0:50, 50, 0.4), type = "h", xlab = "x", ylab = "",
+       main = "Densités associées aux lois B(50,0.4) et N(20,12)")
## Ne pas oublier le add = TRUE
> curve(dnorm(x, 20, sqrt(12)), 0, 50, col = "red", ylab = "", add = TRUE)
```



Exercice 6

```
> plot(0:20, dbinom(0:20, 100, 0.01), type = "h", xlab = "x", ylab = "",
+       ylim = c(0, 0.4), main = "Densités associées aux lois B(100,0.01) et N
    (1,0.99)")
> curve(dnorm(x, 1, sqrt(0.99)), 0, 20, col = 3, ylab = "", add = TRUE)
```



On peut établir une approximation de la loi Binomiale($\mathcal{B}(n, p)$) par la loi Normale $(mathcal N(np, np(1-p)))$ pour les valeurs de n et p données comme conditions.

Exercice 7

$X \sim \mathcal{N}(0, 1)$. Cette fois ci, nous voulons calculer les probabilité, nous allons donc utiliser la syntaxe **p**loi

— $\mathbb{P}(X < -0,5)$

```
> pnorm(-0.5, 0, 1)
[1] 0.3085375
```

— $\mathbb{P}(X > 1,5)$ - On passe par le contraire [1]

$$\mathbb{P}(X > 1,5) = 1\mathbb{P}(X < 1,5)$$

```
> 1 - pnorm(1.5, 0, 1)
[1] 0.0668072
```

— $\mathbb{P}(X > -1)$ - Même logique

$$\mathbb{P}(X > -1) = 1 - \mathbb{P}(X < -1)$$

```
> 1 - pnorm(-1, 0, 1)
[1] 0.8413447
```

1. Si tu veux connaître le nombre d'abrutis sur cette terre, commence par compter les malins, tu gagneras du temps - Didier DINET

— $\mathbb{P}(|X| \leq 1,96)$ - Attention, ici il y a des valeurs absolues[2]. Cela signifie que l'on cherche à connaître la quantité comprise entre $-1.96 \leq X \leq 1.96$. On fait donc :

$$\mathbb{P}(|X| \leq 1,96) = \mathbb{P}(X \leq 1,96) - \mathbb{P}(X \leq -1,96)$$

```
> pnorm(1.96, 0, 1) - pnorm(-1.96, 0, 1)
[1] 0.9500042
```

— $\mathbb{P}(|X| \leq 2,58)$ - Même logique

$$\mathbb{P}(|X| \leq 2.58) = \mathbb{P}(X \leq 2.58) - \mathbb{P}(X \leq -2.58)$$

```
> pnorm(2.58, 0, 1) - pnorm(-2.58, 0, 1)
[1] 0.99012
```

— $\mathbb{P}(|X| \geq 3)$ - Là c'est le méga combo (contraire et valeurs absolues)

$$\mathbb{P}(|X| \geq 3) = 1 - \mathbb{P}(|X| \leq 3) = 1 - \mathbb{P}(X \leq 3) - \mathbb{P}(X \leq -3)$$

```
> 1 - (pnorm(3, 0, 1) - pnorm(-3, 0, 1))
[1] 0.002699796
```

On aurait pu utiliser d'autres commandes. Par exemple, pour $\mathbb{P}(X > 1.5)$:

```
> pnorm(1.5, 0, 1, lower.tail = FALSE)
[1] 0.0668072
```

Exercice 8

1. Pour $X \sim \mathcal{N}(15, 9)$, on a :

   — $\mathbb{P}(16 < X < 20)$

   ```
   > pnorm(20, 15, 3) - pnorm(16, 15, 3)
   [1] 0.321651
   ```

   — $\mathbb{P}(X > 18)$

   ```
   > 1 - pnorm(18, 15, 3)
   [1] 0.1586553
   ```

   — $\mathbb{P}(X < 6)$

   ```
   > pnorm(6, 15, 3)
   [1] 0.001349898
   ```

2. Une bonne valeur absolue est une valeur absolue qui n'existe pas - Didier DINET

— $\mathbb{P}(|X - 15| > 5.88)$

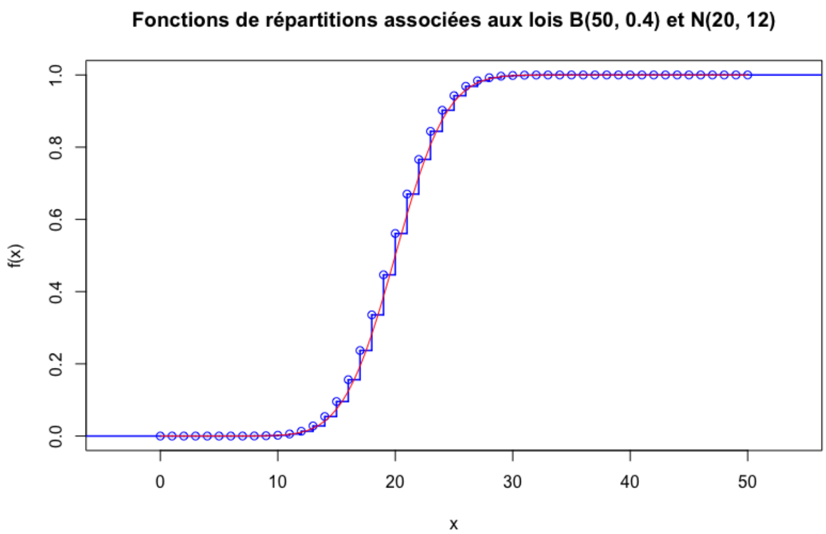
```
 > 1 - (pnorm(15 + 5.88, 15, 3) - pnorm(15 - 5.88, 15, 3))
[1] 0.04999579
```

2. On fait :

```
> curve(pnorm(x, 15, 3), 6, 24)
```

Exercice 9

```
> plot(stepfun(0:50, c(0, pbinom(0:50, 50, 0.4))),
+       main = "Fonctions de répartitions associées aux lois B(50, 0.4) et N(20, 12)"
    ,col="blue",lwd=1.5)
> curve(pnorm(x, 20, sqrt(12)), 0, 50, col = "red", add = TRUE)
```



Exercice 10

1. On fait :

```
 > p = c(0.00135, 0.025, 0.95, 0.999, 0.995, 0.99865)
 ## On définit les valeurs des quantiles
 > x = qnorm(p)
 ## On cherche les quantiles pour les valeurs que l'on a définit
 > cbind(p, x)
 ## On améliore l'affichage
             p          x
  [1,] 0.00135 -2.999977
  [2,] 0.02500 -1.959964
  [3,] 0.95000  1.644854
  [4,] 0.99900  3.090232
  [5,] 0.99500  2.575829
```

```
 [6,] 0.99865  2.999977
```

2. On fait :

```
> p = c(0.975, 0.025)
> y = qnorm(p, 19, sqrt(3))
> cbind(p, y)
         p        y
[1,] 0.975 22.39476
[2,] 0.025 15.60524
```

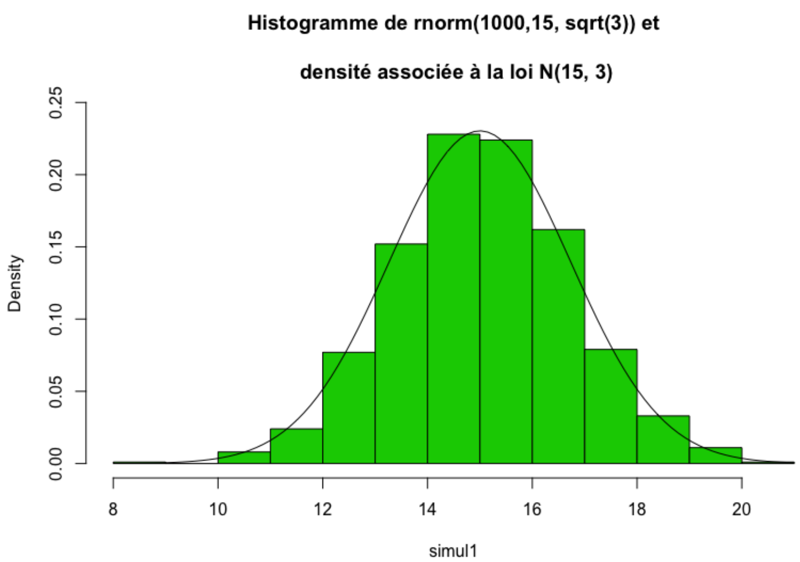
On a

```
> y[1]
[1] 22.39476
> qnorm(0.975)
[1] 1.959964
> y[1] - 19 - sqrt(3) * qnorm(0.975)
[1] 8.881784e-16
```

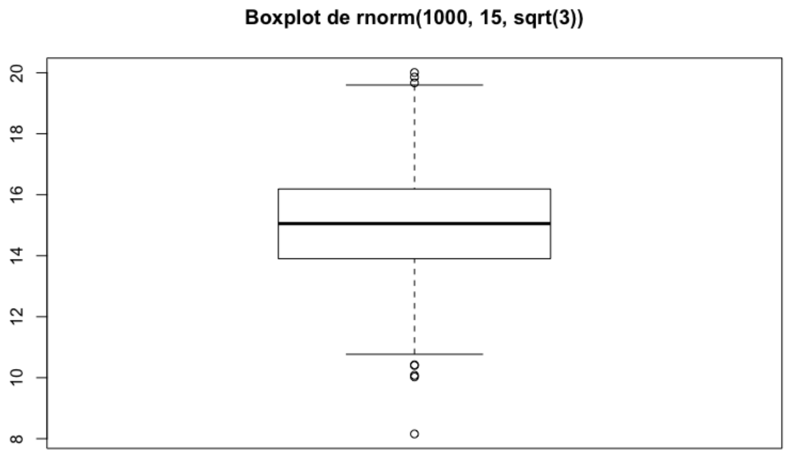
On arrondit cela à 0.

Exercice 11

```
> simul1 = rnorm(1000, 15, sqrt(3))
> head(simul1,5);tail(simul1,5)
[1] 10.40035 13.61364 16.44047 14.85643 14.22192
[1] 12.78735 14.92611 14.51578 15.20232 18.89797
> hist(simul1, col = 3, prob = TRUE, ylim = c(0, 0.25),
+      main = "Histogramme de rnorm(1000,15, sqrt(3)) et
+      \n densité associée à la loi N(15, 3)")
> curve(dnorm(x, 15, sqrt(3)), col = "black", add = TRUE)
```



```
> boxplot(simul1, main = "Boxplot de rnorm(1000, 15, sqrt(3))")
```



Exercice 12

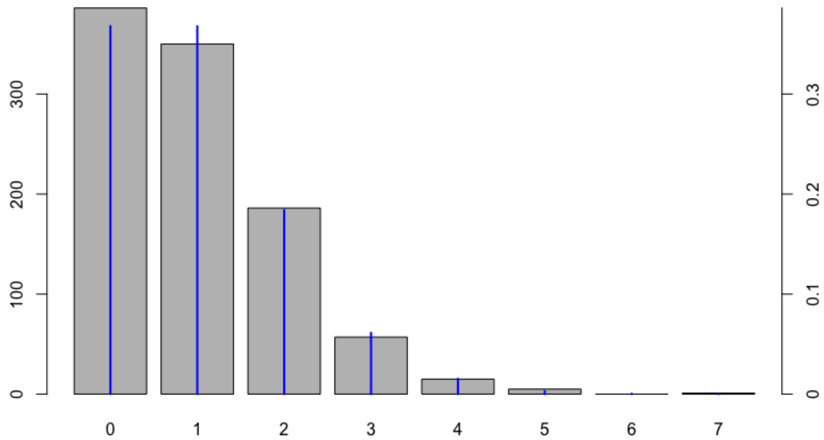
1. On fait :

```
> simul2 = rpois(1000, 1)
> head(simul2,10);tail(simul2,10)
 [1] 0 2 0 1 1 1 0 3 1 1
 [1] 0 0 1 3 1 0 1 0 0 1
> table(simul2)
simul2
  0   1   2   3   4   5   7
386 350 186  57  15   5   1
```

2. En utilisant la fonction compter, on fait :

```
> maxi = max(simul2);maxi
[1] 7
> nbsimul2 = compter(0:maxi, simul2)
  0   1   2   3   4   5   6   7
386 350 186  57  15   5   0   1
```

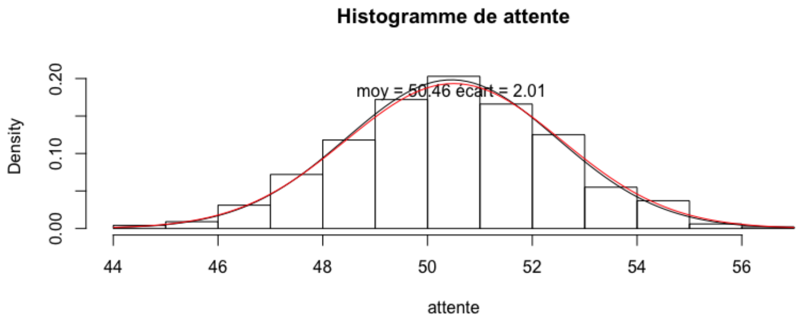
```
> bppois = barplot(nbsimul2)
> points(bppois, dpois(0:maxi,1) * 1000, type = "h",
+          col = "blue", lwd = 2)
> axis(4, at = seq(0, 1000, by = 100), labels = seq(0, 1, by = 0.1))
```

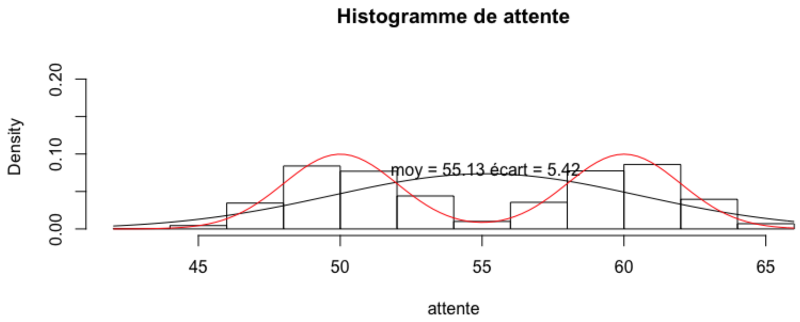


3.

Exercice 13

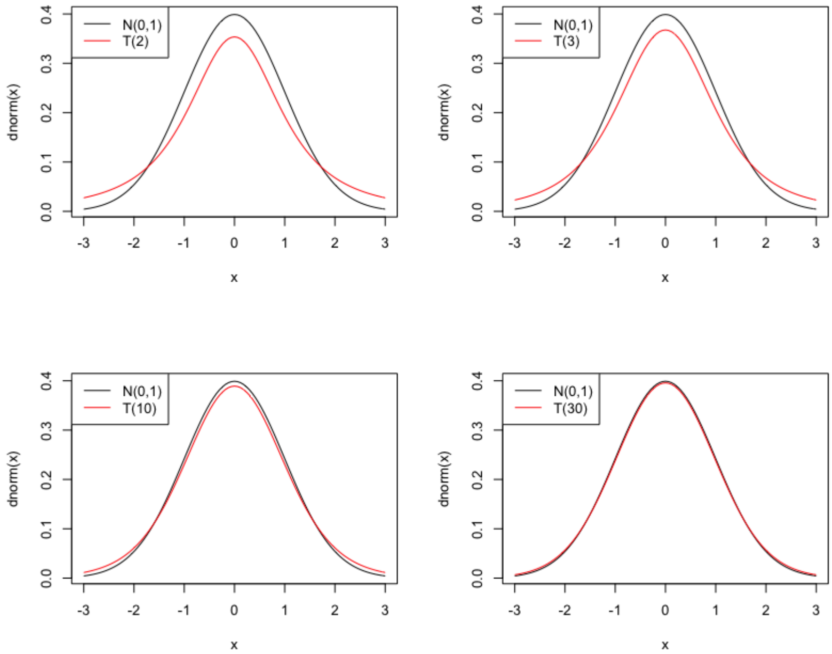
```
> peage = function(mu1, mu2, sigma, n) {
+   attente = numeric(n)
+   cabines = c("C1", "C2")
+   for(i in 1 : n) {
+     tirage = sample(cabines, 1)
+     if(tirage == "C1") {
+       attente[i] = rnorm(1, mu1, sigma)
+     } else {
+       attente[i] = rnorm(1, mu2, sigma)
+     }
+   }
+   a = 0.9 * max(hist(attente, prob = TRUE, ylim = c(0, 0.22), xlab = "attente",
+                      main = "Histogramme de attente")$density)
+   moy = mean(attente)
+   ecart = sd(attente)
+   curve(dnorm(x, moy, ecart), add = TRUE)
+   text(moy, a, paste("moy =", round(moy, 2), "écart =", round(ecart, 2)))
+   curve(0.5 * (dnorm(x, mu1, sigma) + dnorm(x, mu2, sigma)), col = "red", add =
    TRUE)
+ }

> par(mfrow = c(2, 1))
> peage(50, 51, 2, 1000)
> peage(50, 60, 2, 1000)
> par(mfrow = c(1, 1))
```





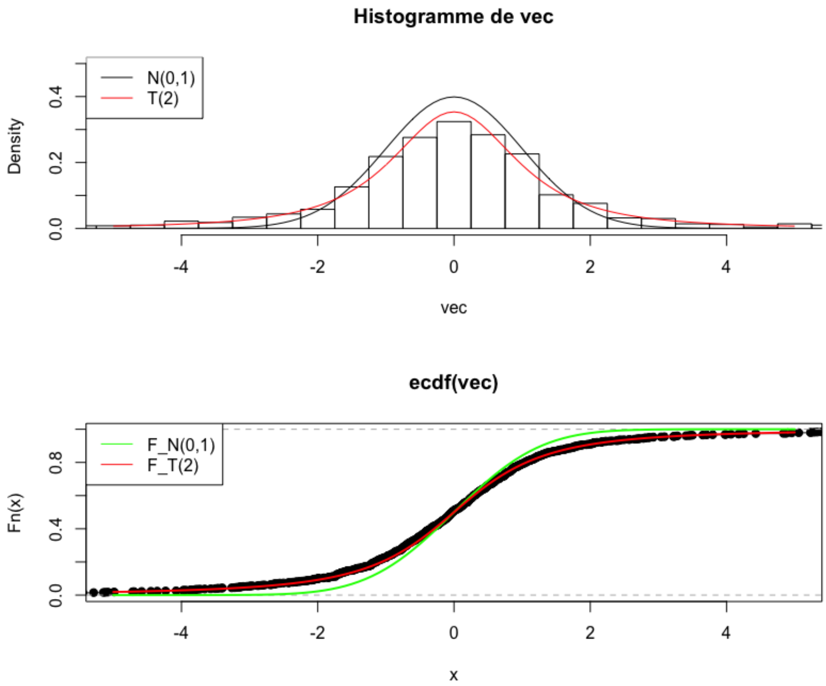
Exercice 14

```
> par(mfrow = c(2, 2))
> curve(dnorm(x), -3, 3)
> curve(dt(x, 2), -3, 3, col = "red", add = TRUE)
> legend("topleft", legend = c("N(0,1)", "T(2)"), col = c("black", "red"), lwd=1)
> curve(dnorm(x), -3, 3)
> curve(dt(x,3), -3, 3, col = "red", add = TRUE)
> legend("topleft", legend = c("N(0,1)", "T(3)"), col = c("black", "red"), lwd = 1)
> curve(dnorm(x), -3, 3)
> curve(dt(x, 10), -3, 3, col = "red", add = TRUE)
> legend("topleft", legend = c("N(0,1)", "T(10)"), col = c("black", "red"), lwd = 1)
> curve(dnorm(x), -3, 3)
> curve(dt(x, 30), -3, 3, col = "red", add = TRUE)
> legend("topleft", legend = c("N(0,1)", "T(30)"), col = c("black", "red"), lwd = 1)
> par(mfrow = c(1, 1))
```

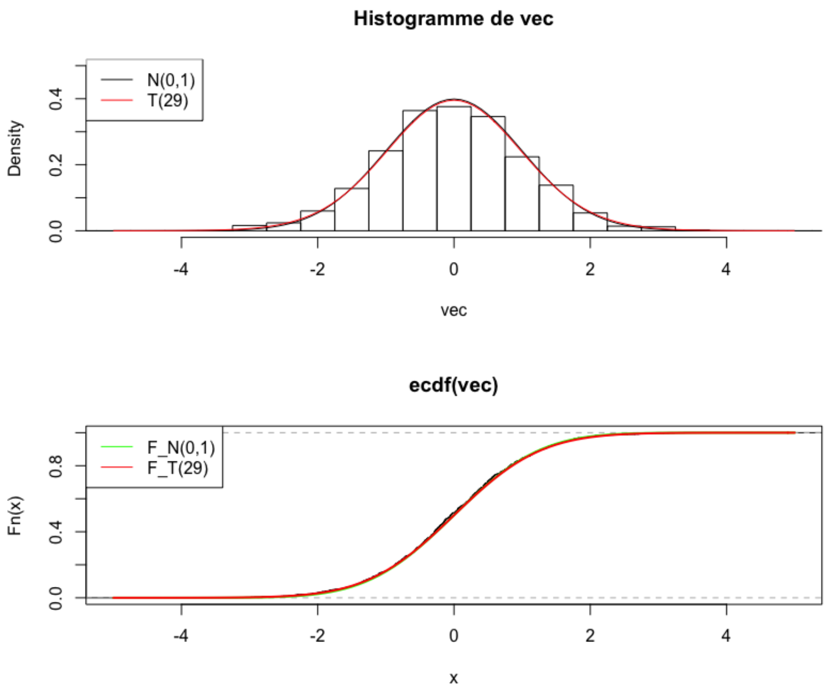


Exercice 15

```
> stu = function(n,nb,mu,sigma) {
+   vec = numeric(nb)
+   for (i in 1:nb) {
+     simul = rnorm(n,mu,sigma)
+     xbar = mean(simul)
+     sech = sd(simul)
+     vec[i] = sqrt(n) * (xbar - mu) / sech
+   }
+   par(mfrow = c(2, 1))
+   hist(vec, prob = TRUE,
+        breaks = seq(-1000.25, 1000.25, by = 0.5), xlim = c(-5, 5), ylim = c(0,
   0.5),
+        main = "Histogramme de vec")
+   curve(dnorm(x), -5, 5, add = TRUE)
+   curve(dt(x, n - 1), -5, 5, col = "red", add = TRUE)
+   legend("topleft",
+          legend = c("N(0,1)", paste("T(", n - 1, ")", sep = "")), col = c("black",
    "red"),
+          lwd = 1)
+   plot(ecdf(vec), xlim = c(-5, 5))
+   curve(pnorm(x), -5, 5, col = "green", lwd = 2, add = TRUE)
+   curve(pt(x, n-1), -5, 5, col = "red", lwd = 2, add = TRUE)
+   legend("topleft", legend = c("F_N(0,1)", paste("F_T(", n - 1, ")", sep = "")),
+          col = c("green", "red"), lwd = 1)
+   par(mfrow = c(1,1))
+ }
> stu(3, 1000, 10, 2)
```

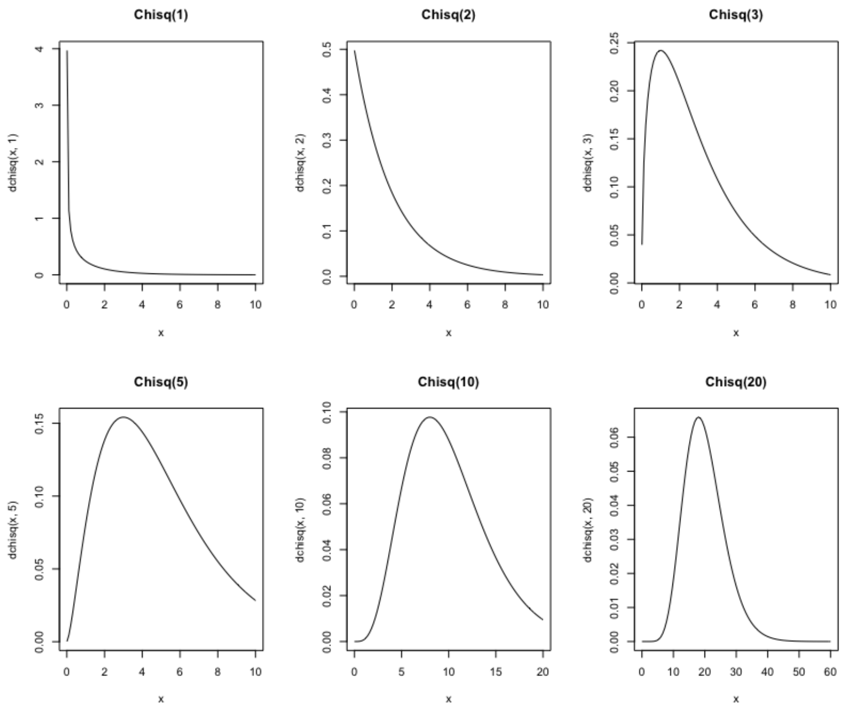



```
> stu(30, 1000, 10, 2)
```



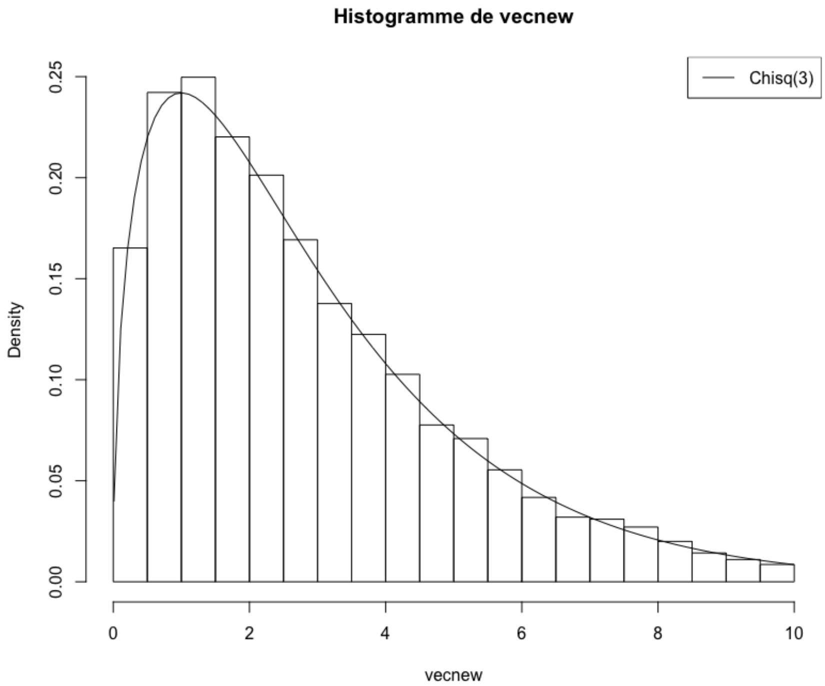
Exercice 16

```
> curve(dchisq(x, 1), 0.01, 10, main = "Chisq(1)")
> curve(dchisq(x, 2), 0.01, 10, main = "Chisq(2)")
> curve(dchisq(x, 3), 0.01, 10, main = "Chisq(3)")
> curve(dchisq(x, 5), 0.01, 10, main = "Chisq(5)")
> curve(dchisq(x, 10), 0.01, 20, main = "Chisq(10)")
> curve(dchisq(x, 20), 0.01, 60, main = "Chisq(20)")
> par(mfrow = c(1, 1))
```



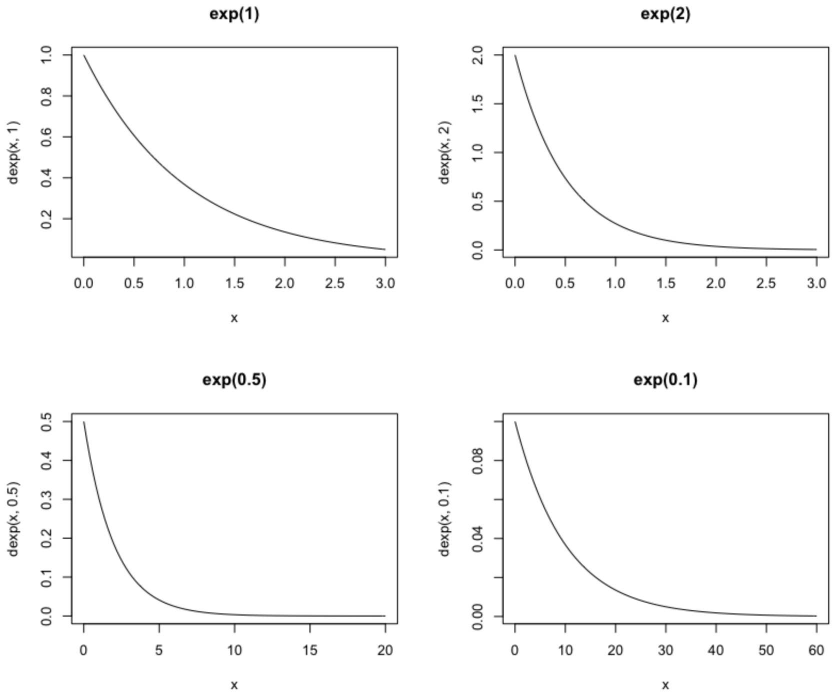
Exercice 17

```
> khi2 = function(n, nb, mu, sigma) {
+   vec = numeric(nb)
+   for (i in 1:nb) {
+     simul = rnorm(n, mu, sigma)
+     sech2 = var(simul)
+     vec[i] = (n-1) * sech2 / sigma^2
+   }
+   vecnew = vec[vec >= 0.01 & vec <= 10]
+   hist(vecnew, prob = TRUE, main = "Histogramme de vecnew")
+   curve(dchisq(x, n-1), 0.01, 10, add = TRUE)
+   legend("topright", legend = paste("Chisq(",n-1,")", sep = ""), col = "black",
+          lwd = 1)
+ }
> par(mfrow = c(1, 1))
> khi2(4, 10000, 10, 2)
```



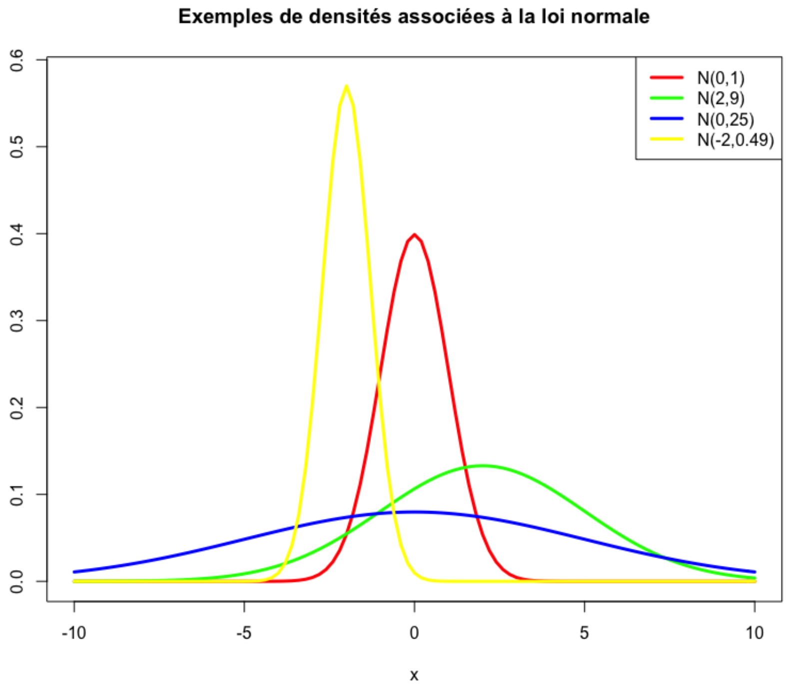
Exercice 18

```
> par(mfrow = c(2, 2))
> curve(dexp(x, 1), 0, 3, main = "exp(1)")
> curve(dexp(x, 2), 0, 3, main = "exp(2)")
> curve(dexp(x, 0.5), 0,20, main = "exp(0.5)")
> curve(dexp(x, 0.1), 0,60, main = "exp(0.1)")
> par(mfrow=c(1, 1))
```



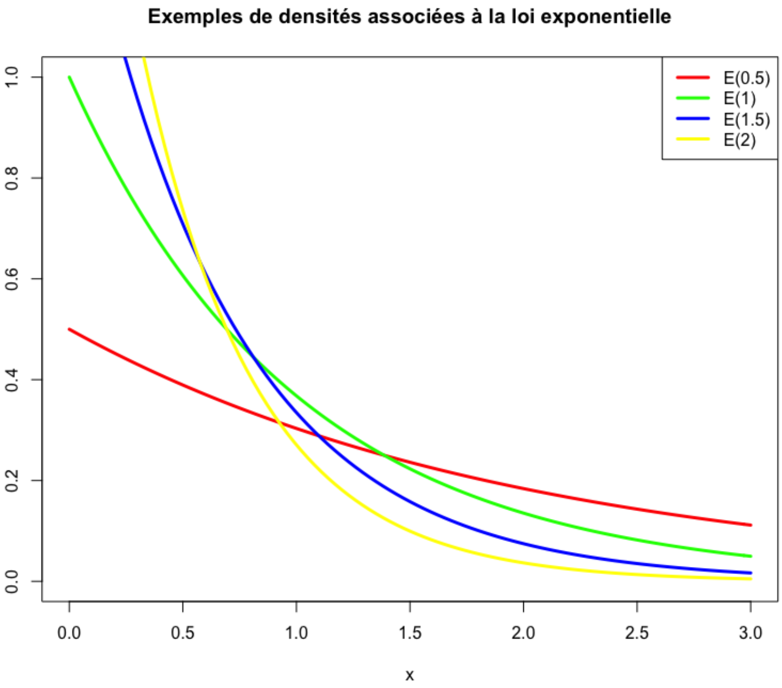
Exercice 19

```
curve(dnorm(x, 0, 1), xlim = c(-10, 10), ylim = c(0, 0.58), col = "red", lwd = 3,
ylab = "", main = "Exemples de densités associées à la loi normale")
curve(dnorm(x, 2, 3), col = "green", lwd = 3, add = TRUE)
curve(dnorm(x, 0, 5), col = "blue", lwd = 3, add = TRUE)
curve(dnorm(x, -2, 0.7), col = "yellow", lwd = 3, add = TRUE)
legend("topright", legend = c("N(0,1)", "N(2,9)", "N(0,25)", "N(-2,0.49)"),
lty = 1, lwd = 3, col = c("red", "green", "blue", "yellow"))
```



Exercice 20

```
> curve(dexp(x, 0.5), xlim = c(0, 3), ylim = c(0, 1), col = "red", lwd = 3,
+       ylab = "", main = "Exemples de densités associées à la loi exponentielle")
> curve(dexp(x, 1), col = "green", lwd = 3, add = TRUE)
> curve(dexp(x, 1.5), col = "blue", lwd = 3, add = TRUE)
> curve(dexp(x, 2), col = "yellow", lwd = 3, add = TRUE)
> legend("topright", legend = c("E(0.5)", "E(1)", "E(1.5)", "E(2)"),
+        lty = 1, lwd = 3, col = c("red", "green", "blue", "yellow"))
```



Exercice 21

1. On fait :

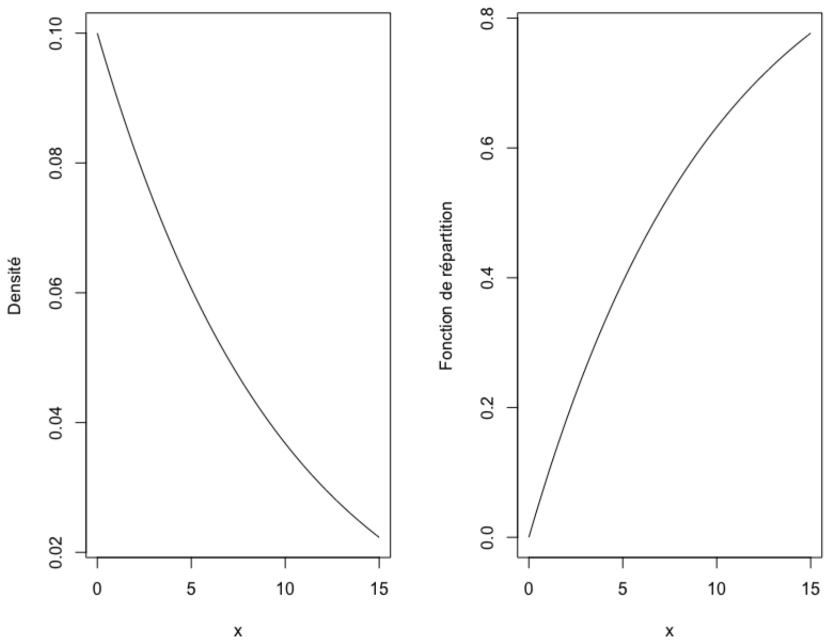
```
 > densexp = function(x, lambda) {
+   ifelse (x >= 0, lambda * exp(-lambda * x), 0)
+ }
```

2. On fait :

```
> integrate(function(x) dexp(x, 1 / 10), 0, 100)
0.9999546 with absolute error < 2.8e-14
```

3. On fait :

```
> par(mfrow = c(1, 2))
> curve(dexp(x, 1 / 10), 0, 15, main = "Densité")
> curve(pexp(x, 1 / 10), 0, 15, main = "Fonction de répartition")
```



4. La probabilité que la durée de vie d'une voiture dépasse 10.2 ans est donnée par :

```
> 1 - pexp(10.2, 1 / 10)
[1] 0.3605949
```

Exercice 22

1. On fait :

```
> densgamma = function(x, m, lambda) {
+   ifelse (x >= 0, (1/factorial(m - 1)) * lambda^m * x^(m - 1) *
+               exp(-lambda * x), 0)
+ }
```

2. On fait :

```
> integrate(function(x) dgamma(x, 5, 1), 0, 100)
1 with absolute error < 2.7e-07
```

3. On fait :

```
> x = 2:10
> pgamma(x, 5, 1)
 [1] 0.05265302 0.18473676 0.37116306 0.55950671 0.71494350 0.82700839
 [7] 0.90036760 0.94503636 0.97074731
```

Ces valeurs correspondent à $F_X(2), \dots, F_X(10)$ respectivement.

4. On fait :

```
> qgamma(0.92, 5, 1)
[1] 8.376739
```

Exercice 23

1. On fait :

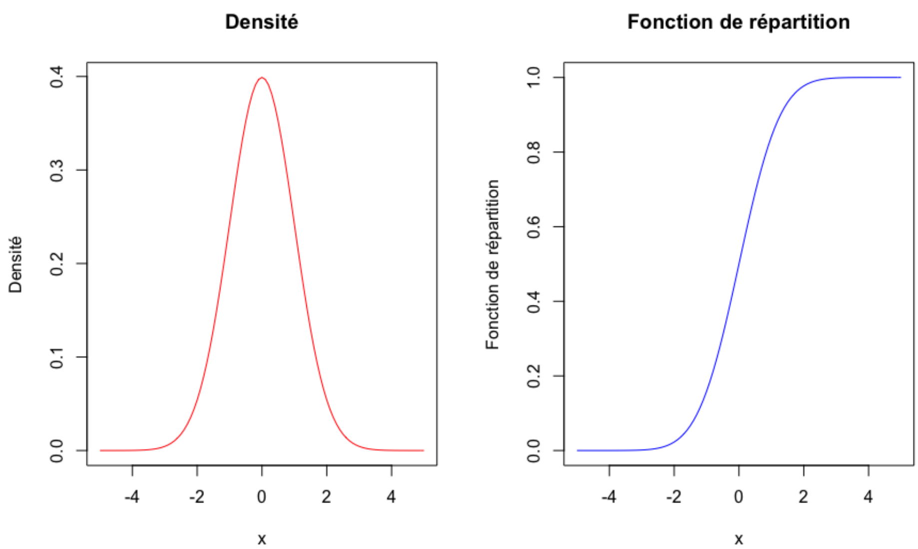
```
densnorm = function(x) {
  (1 / sqrt(2 * pi)) * exp(-x^2 / 2)
}
```

2. On fait :

```
> integrate(function(x) dnorm(x), -100, 100)
1 with absolute error < 3.2e-07
```

3. On considère les commandes :

```
> par(mfrow = c(1, 2))
> curve(dnorm(x), -5, 5, ylab = "Densité",main= "Densité",col="red")
> curve(pnorm(x), -5, 5, ylab = "Fonction de répartition",main ="Fonction
   de répartition",col="blue")
```



4. Les calculs de probabilités :

   (a) $\mathbb{P}(X \leq 2.2)$

```
> pnorm(2.2)
[1] 0.9860966
```

(b) $\mathbb{P}(X \geq 1.7) = 1 - \mathbb{P}(X \leq 1.7)$

```
> 1 - pnorm(1.7)
[1] 0.04456546
```

Ou

```
> pnorm(1.7, lower.tail = FALSE)
[1] 0.04456546
```

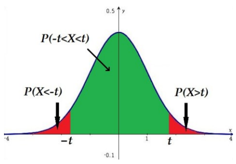
(c) $\mathbb{P}(0.2 \leq X \leq 1.4) = \mathbb{P}(X \geq 1.4) - \mathbb{P}(X \leq 0.2)$

```
> pnorm(1.4) - pnorm(0.2)
[1] 0.3399836
```

(d) $\mathbb{P}(|X| \leq 1.96) = 2\mathbb{P}(X \leq 1.96) - 1$

```
> 2 * pnorm(1.96) - 1
[1] 0.9500042
```

Alors l'explication sur ce dernier point : Nous étudions une variable aléatoire $X \sim \mathcal{N}(0, 1)$, on peut donc jouer avec la symétrie de la loi normale :



On fera de manière générale (et en suivant cette logique) :

Aire verte = 1 - Aire rouge

$$\mathbb{P}(|X| \leq t) = \mathbb{P}(-1 \leq X \leq 1)$$
$$\mathbb{P}(|X| \leq t) = 1 - [\mathbb{P}(X \leq -t) + \mathbb{P}(X \geq t)]$$
$$\mathbb{P}(|X| \leq t) = 1 - \mathbb{P}(X \leq -t) - \mathbb{P}(X \geq t)$$
$$\mathbb{P}(|X| \leq t) = 1 - \mathbb{P}(X \geq t) - \mathbb{P}(X \geq t)$$
$$\mathbb{P}(|X| \leq t) = 1 - 2\mathbb{P}(X \geq t)$$
$$\mathbb{P}(|X| \leq t) = 2\mathbb{P}(X \leq t) - 1$$

5. On fait :

```
> qnorm(0.98)
[1] 2.053749
```

Exercice 24

La probabilité qu'un foie gras pèse

— moins de 650 grammes est donnée par

$$\mathbb{P}(X \leq 650)$$

```
> pnorm(650, 550, 100)
[1] 0.8413447
```

— plus de 746 grammes est donnée par :

$$\mathbb{P}(X \geq 746) = 1 - \mathbb{P}(X \leq 746)$$

```
> 1 - pnorm(746, 550, 100)
[1] 0.0249979
```

Ou

```
> pnorm(746, 550, 100, lower.tail = FALSE)
[1] 0.0249979
```

— entre 550 grammes et 600 grammes est donnée par :

$$\mathbb{P}(550 \leq X \leq 600) = \mathbb{P}(X \leq 600) - \mathbb{P}(X \leq 550)$$

```
> pnorm(600, 550, 100) - pnorm(550, 550, 100)
[1] 0.1914625
```

Exercice 25

On fait :

```
> foncrep = function(x) {
+   n = 1 + 2 * 0:50
+   0.5 + (1 / sqrt(2 * pi)) * exp(-x^2 / 2) * sum(x^n / cumprod(n))
+ }
> foncrep(1.2)
[1] 0.8849303
```

On retrouve bien

```
> pnorm(1.2)
[1] 0.8849303
```

Exercice 26

```
> x = sample(c(0, 2, 5), 1000, replace = T, prob = c(0.2, 0.5, 0.3))
> table(x)
x
  0   2   5
209 486 305
```

Exercice 27

```
> Urne = function(k, p, q) {
+   contenu = rep(c("Rouge", "Noire"), c(p, q))
+   sample(contenu, k)
+ }
> Urne(6,8,5)
[1] "Rouge" "Noire" "Rouge" "Noire" "Rouge" "Noire"
```

Exercice 28

```
> Freq = function(n) {
+   tirage = sample(1:6, n, replace = T)
+   sum(tirage == 5) / n
+ }
> Freq(10)
[1] 0.2
> Freq(100)
[1] 0.15
> Freq(1000)
[1] 0.167
> Freq(10000)
[1] 0.1683
```

On constate que cette dernière probabilité est proche de $1/6$, la probabilité théorique d'obtenir un 5 lorsqu'on lance un dé.

Exercice 29

- On fait
  - Pour $E1$

    ```
    > sample(c(0, 1), 10, replace = T, prob = c(2 / 3, 1 / 3))
    [1] 0 1 1 0 0 0 0 0 1 0
    ```

  - Pour $E2$

    ```
    > sample(c(rep(0, 10), rep(1, 5)), 10)
    [1] 0 1 1 1 0 1 0 0 1 0
    ```

- 1. On fait :

— Pour X

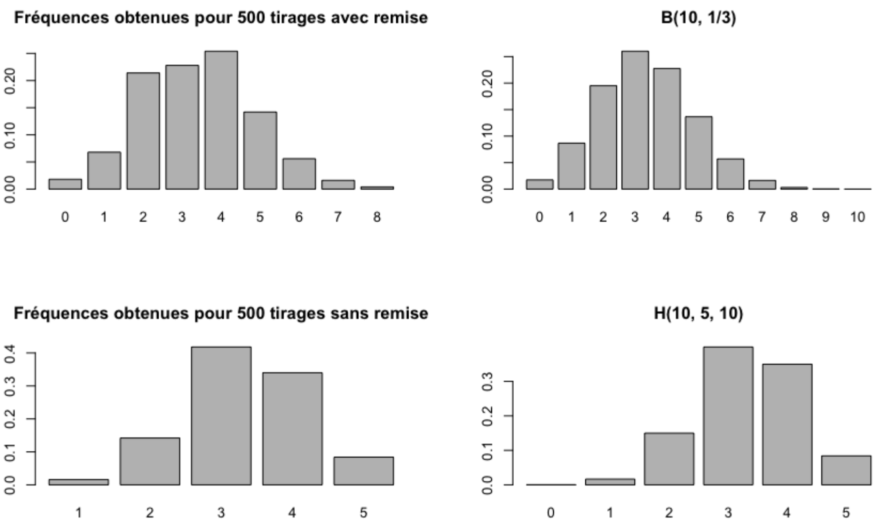
```
> x = numeric()
> for(i in 1:500){
+   x[i] = sum(sample(c(0, 1), 10, replace = T, prob = c(2 / 3,
    1 / 3)))
+ }
> table(x)
x
  0   1   2   3   4   5   6   7   8
  9  34 107 114 127  71  28   8   2
```

— Pour Y

```
> y = numeric()
> for(i in 1:500){
+   y[i] = sum(sample(c(rep(0, 10), rep(1, 5)), 10))
+ }
> table(y)
y
  1   2   3   4   5
  8  71 209 170  42
```

2. On propose les commandes :

```
> titre1 = "Fréquences obtenues pour 500 tirages avec remise"
> titre2 = "B(10, 1/3)"
> titre3 = "Fréquences obtenues pour 500 tirages sans remise"
> titre4 = "H(10, 5, 10)"
> par(mfrow = c(2, 2))
> barplot(table(x) / 500, main = titre1)
> barplot(dbinom(0:10, 10, 1/3), names.arg = 0:10, main = titre2)
> barplot(table(y) / 500, main = titre3)
> barplot(dhyper(0:5, 5, 10, 10), names.arg = 0:5, main = titre4)
```

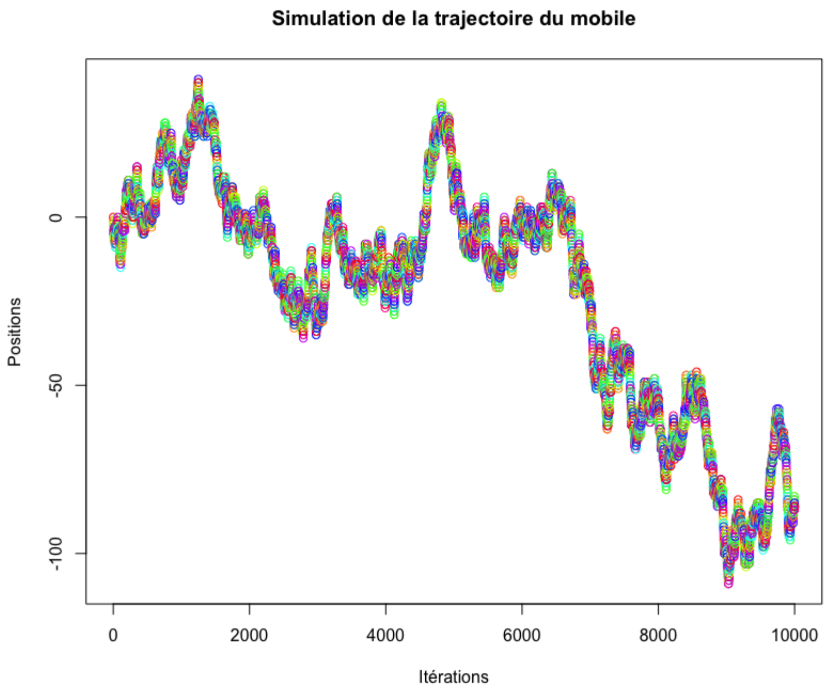


Exercice 30

```
> n = 10000
> u = sample(c(-1,1), n, replace = T)
> x = cumsum(u)
> x = c(0, x)
> head(x,50)
  [1]  0  1  2  1  0 -1  0  1  0 -1 -2 -3 -2 -1 -2 -3 -4 -5 -4 -3 -2 -1  0  1
 [25]  0 -1  0  1  2  3  2  1  0 -1 -2 -3 -2 -1  0 -1 -2 -3 -2 -1 -2 -1  0  1
 [49]  0 -1
```

Nouvelle simulation de u

```
> head(x,50)
  [1]  0 -1 -2 -3 -4 -5 -4 -3 -2 -3 -2 -3 -2 -1 -2 -1 -2 -1  0  1  2  1  2  3
 [25]  4  5  4  3  4  3  2  3  2  3  2  1  2  3  2  3  4  3  2  3  4  3  4  5
 [49]  4  5

> head(x,50)
  [1]  0 -1 -2 -3 -4 -5 -4 -5 -4 -5 -6 -7 -6 -5 -4 -5 -4 -3 -4 -3 -4 -5 -6 -7
 [25] -6 -7 -8 -7 -8 -7 -6 -7 -8 -9 -8 -9 -8 -7 -6 -7 -6 -7 -6 -5 -4 -5 -4 -3
 [49] -2 -3

> plot(x, col=rainbow(10), main="Simulation de la trajectoire du mobile")
```



Exercice 31

```
> n = 0
> x = rep(0, 5)
> while(sum(x != 6) != 0) {
```

```
+     x[x != 6] = sample(1:6, sum(x != 6), rep = T)
+     print(x)
+     n = n + 1
+ }
[1] 4 4 5 6 1
[1] 6 3 3 6 5
[1] 6 1 1 6 6
[1] 6 4 3 6 6
[1] 6 6 2 6 6
[1] 6 6 6 6 6
> print(n)
[1] 6
```

IV Div'R

V Représentations graphiques

VI Programmation avec R

VII Fonctions usuelles et aide mémoire

3 Lois

Loi Normale Centrée Réduite $\mathcal{N}(0;\,1)$.

$$\mathbb{P}(t) = P(X \leq t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}} dx \text{ et } \mathbb{P}(-t) = 1 - \mathbb{P}(t).$$

3.1 Loi Normale

t	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
q 3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

3.2 Loi Normale

4 R datasets informations

```
1 > help(nomDuDataset)
```

— *"ToothGrowth"* : The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC).

1. len numeric Tooth length
2. supp factor Supplement type (VC or OJ).
3. dose numeric Dose in milligrams/day

— *"economics"* : This dataset was produced from US economic time series data available from research.stlouisfed.org. economics is in "wide" format, economics_long is in "long" format.

1. date : Month of data collection
2. psavert : personal savings rate
3. pce : personal consumption expenditures, in billions of dollars
4. unemploy : number of unemployed in thousands
5. uempmed : median duration of unemployment, in weeks
6. pop :total population, in thousands

— *"diamonds"* : A dataset containing the prices and other attributes of almost 54,000 diamonds. The variables are as follows :

1. price : price in US dollars ($326 ?$18,823)
2. carat : weight of the diamond (0.2 ?5.01)
3. cut : quality of the cut (Fair, Good, Very Good, Premium, Ideal)
4. color : diamond colour, from J (worst) to D (best)
5. clarity : a measurement of how clear the diamond is (I1 (worst), SI1, SI2, VS1, VS2, VVS1, VVS2, IF (best))
6. x : length in mm (0 ?10.74)
7. y : width in mm (0 ?58.9)
8. z : depth in mm (0 ?31.8)
9. depth : total depth percentage = z / mean(x, y) = 2 * z / (x + y) (43 ?79)

10. table : width of top of diamond relative to widest point (43 ?95)

— *"world2"* : This is an alternative version of the world database based on latitudes [0, 360), which then has the Pacific Ocean in the centre of the map.
  The data file is merely a character string which specifies the name of an environment variable which contains the base location of the binary files used by the map drawing functions. This environment variable (R_MAP_DATA_DIR_WORLD for the datasets in the maps package) is set at package load time if it does not already exist. Hence setting the environment variable before loading the package can override the default location of the binary datasets.

— *lung* : Survival in patients with advanced lung cancer from the North Central Cancer Treatment Group. Performance scores rate how well the patient can perform usual daily activities.

1. inst : Institution code
2. time : Survival time in days
3. status : censoring status 1=censored, 2=dead
4. age : Age in years
5. sex : Male=1 Female=2
6. ph.ecog : ECOG performance score (0=good 5=dead)
7. ph.karno : Karnofsky performance score (bad=0-good=100) rated by physician
8. pat.karno : Karnofsky performance score as rated by patient
9. meal.cal : Calories consumed at meals
10. wt.loss : Weight loss in last six months

— *mtcars* : The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973 ?74 models).

1. mpg Miles/(US) gallon
2. cyl Number of cylinders
3. disp Displacement (cu.in.)
4. hp Gross horsepower
5. drat Rear axle ratio
6. wt Weight (1000 lbs)
7. qsec 1/4 mile time
8. vs V/S
9. am Transmission (0 = automatic, 1 = manual)
10. gear Number of forward gears
11. carb Number of carburetors

— *mpg :* This dataset contains a subset of the fuel economy data that the EPA makes available on fueleconomy.gov. It contains only models which had a new release every year between 1999 and 2008 - this was used as a proxy for the popularity of the car.

1. manufacturer : manufacturer name
2. model : model name
3. displ : engine displacement, in litres
4. year : year of manufacture
5. cyl : number of cylinders
6. trans : type of transmission
7. drv : f = front-wheel drive, r = rear wheel drive, 4 = 4wd
8. cty : city miles per gallon
9. hwy : highway miles per gallon
10. fl : fuel type
11. class : "type" of car

— *iris* : This famous (Fisher's or Anderson's) iris data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are Iris setosa, versicolor, and virginica. iris is a data frame with 150 cases (rows) and 5 variables (columns) named Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, and Species. iris3 gives the same data arranged as a 3-dimensional array of size 50 by 4 by 3, as represented by S-PLUS. The first dimension gives the case number within the species subsample, the second the measurements with names Sepal L., Sepal W., Petal L., and Petal W., and the third the species.

— *faithful :* Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA.

1. eruptions numeric Eruption time in mins
2. waiting numeric Waiting time to next eruption (in mins)