Principales lois de probabilités - Lois continues

Tewfik Mahdjoub

Université Abou Bekr Belkaïd, Tlemcen

5 Novembre 2024

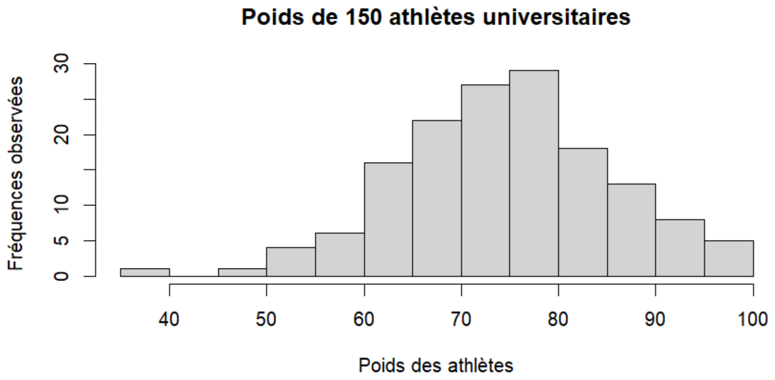
Plan du cours

Sur un échantillon de 150 athlètes universitaires, on a mesuré les variables quantitatives continues suivantes : la glycémie (g/l), le poids (kg) et la taille (cm). Les résultats obtenus sont donnés dans le tableau **Tab** suivant :

N° Etudiant	Glycémie (g/l)	Poids (kg)	Tailles (cm)
1	1.32	87.55	187.97
2	0.85	64.39	152.11
3	1.11	72.82	184.00
4	1.38	80.53	156.91
5	1.13	83.10	174.48
6	1.25	68.44	195.70
7	0.90	96.60	144.61
8	1.19	83.93	173.17
9	0.84	89.20	186.30
10	1.22	92.98	181.58
11	1.08	73.09	149.82
12	1.34	73.63	170.40
13	1.00	75.28	180.68
14	1.28	68.65	189.28

La représentation de la variable "Poids" par un histogramme se fait en tapant la commande :

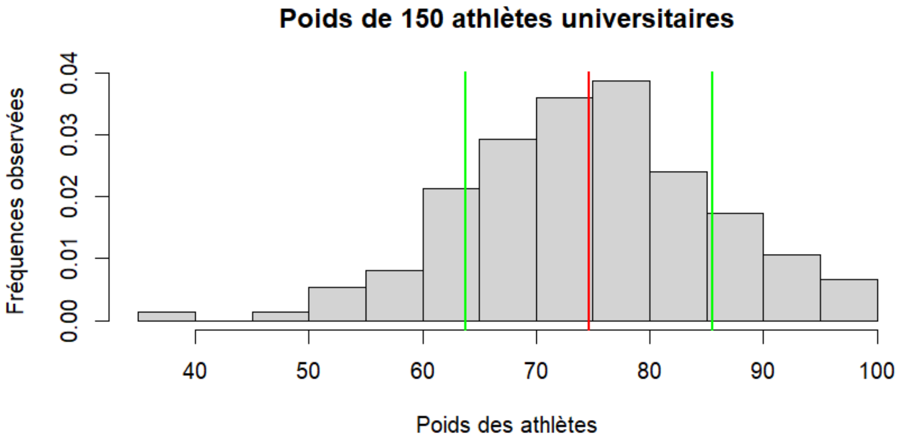
> hist(Tab$Poids, xlab="Poids des athlètes", ylab="Fréquences observées", main="Poids de 150 athlètes universitaires")



L'axe (y'y) peut représenter des pourcentages (fréquences absolues/taille de l'échantillon) par l'instruction **prob=T**

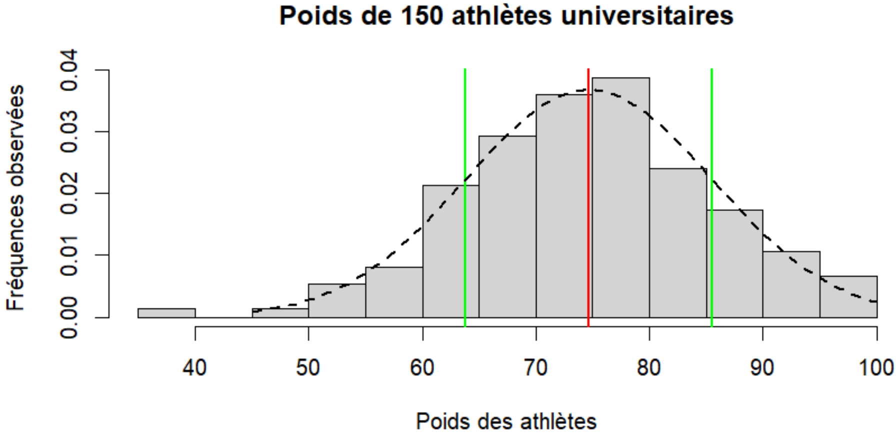
```
> hist(Tab$Poids, xlab="Poids des athlètes", ylab="Fréquences observées", main="Poids de 150 athlètes universitaires", prob=T)
> moy=mean(Tab$Poids)
> var=var(Tab$Poids)*149/150
> et=sqrt(var)
> abline(v=moy, col="red", lw=2)
> abline(v=moy+et, col="green", lw=2)
> abline(v=moy-et, col="green", lw=2)
```

permet d'obtenir :



Cet histogramme peut être approché par la fonction suivante :

> curve(dnorm(x, moy, et), 45, 100 , add=T, lty=2, lw=2)



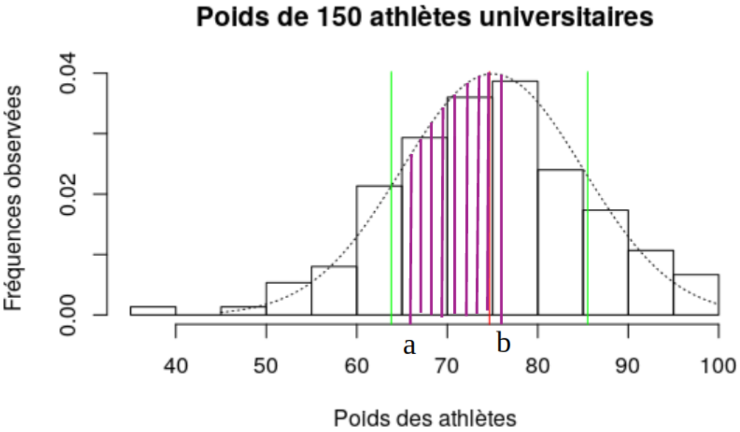
La fonction dnorm(x,μ,σ) est appelée la distribution normale, de moyenne μ et d'écart-type σ

Les valeurs sur (y'y) ne sont pas des pourcentages : $\frac{30athl}{150athl} \neq 0.04$ et donc ne sont pas des probabilités.
Ceux sont les valeurs de la fonction :

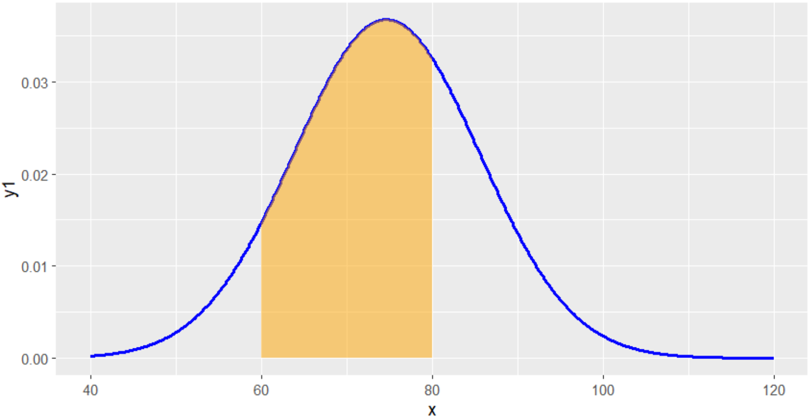
$$dnorm(x, \mu, \sigma) = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \quad (1)$$

La distribution normale est définie pour $x \in \mathbb{R}$: c'est une distribution continue.

La probabilité d'avoir le poids entre a et b est l'aire qui se trouve entre les droites x=a et x=b, la courbe de la loi normale et l'axe (x'x) est :



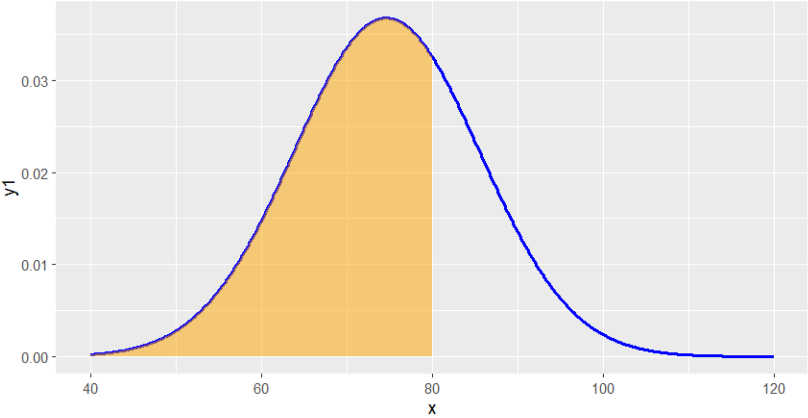
$$P(a \leqslant poids \leqslant b) = \int_a^b dnorm(x, \mu, \sigma)dx \qquad (2)$$



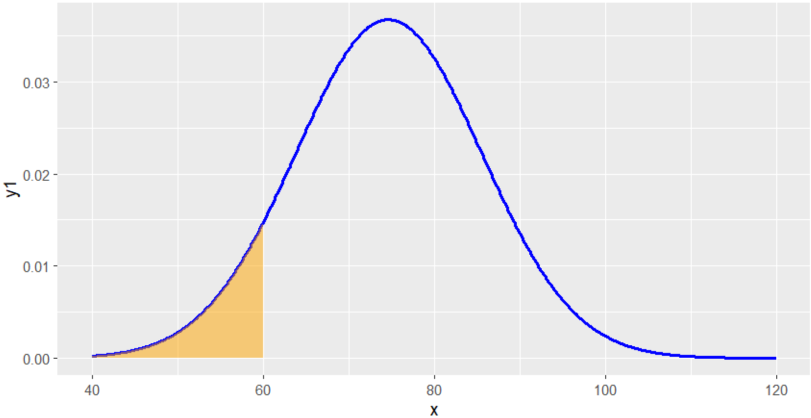
$$P(60 \leqslant poids \leqslant 80) = \int_{60}^{80} dnorm(x, \mu, \sigma) dx \tag{3}$$

Comment calculer cette probabilité ?

On calcule d'abord :



$$P(0 \leqslant poids \leqslant 80) = \int_0^{80} dnorm(x, \mu, \sigma)dx = pnorm(80, \mu, \sigma) \qquad (4)$$



puis :

$$P(0 \leqslant poids \leqslant 60) = \int_0^{60} dnorm(x, \mu, \sigma)dx = pnorm(60, \mu, \sigma) \qquad (5)$$

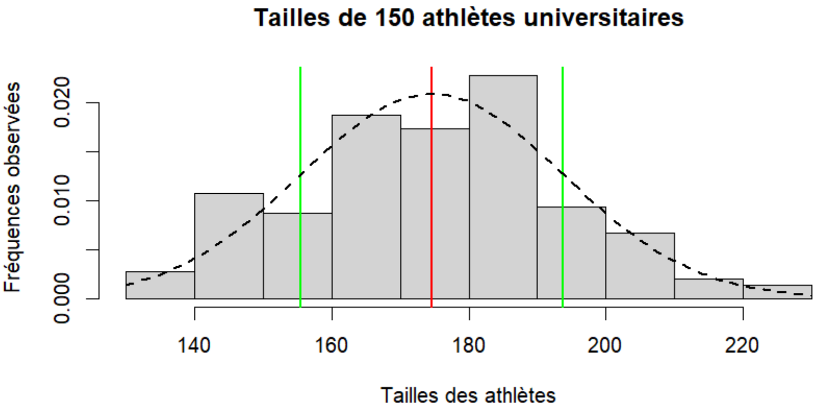
Finalement sous R

$$P(0 \leqslant poids \leqslant 60) = pnorm(80, \mu, \sigma) - pnorm(60, \mu, \sigma) = 0.6003262 \quad (6)$$

La fonction inverse de pnorm() est qnorm()

```
> qnorm(0.25,moy,et)
[1] 67.34255
```

Dans le cas de la variable "taille", on obtient :

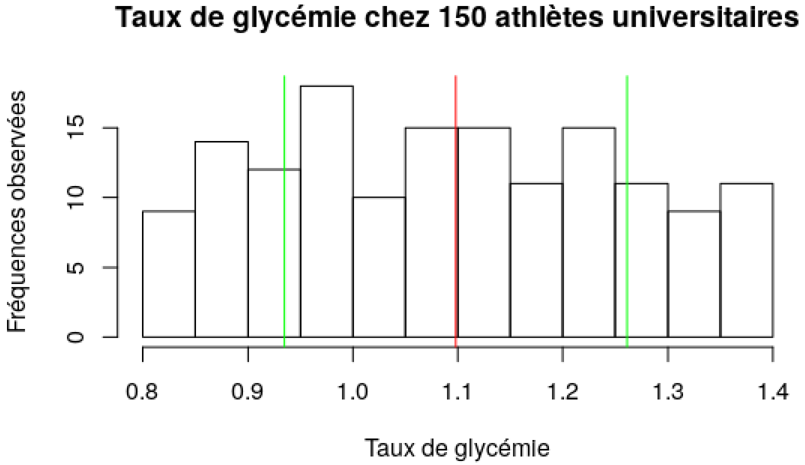


> curve(dnorm(x,moyT,etT), 130 ,230 ,add=T, lty=2, lw=2)

La représentation de la variable "glycémie" par un histogramme, en tapant les commandes :

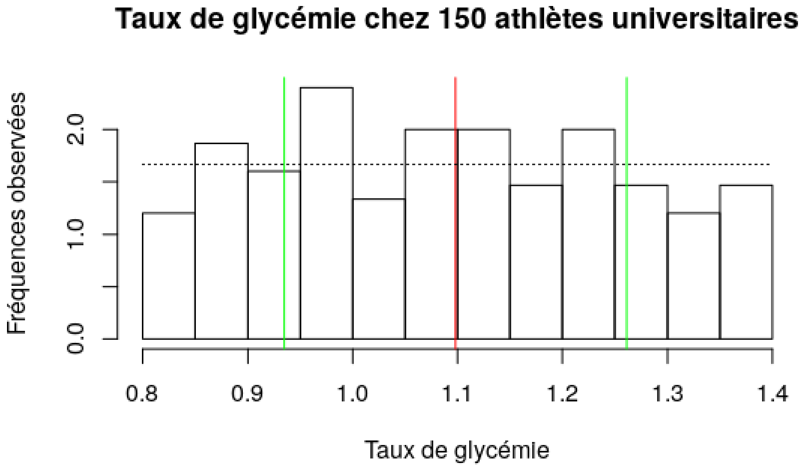
```
> hist(Glycemie,xlab="Taux de glycémie",ylab="Fréquences observées",
main="Taux de glycémie chez 150 athlètes universitaires",prob=T)
> abline(v=mean(Glycemie),col="red")
> abline(v=mean(Glycemie)-sqrt(var(Glycemie)*149/150),col="green")
> abline(v=mean(Glycemie)+sqrt(var(Glycemie)*149/150),col="green")
```

permet d'obtenir :



Cet histogramme peut être approché par la fonction suivante :

```
> curve(dunif(x,0.8,1.4),0.8,1.4,add=T,lty="dotted")
```



La fonction dunif(x,a,b) est appelé la distribution uniforme entre a et b, où b>a.

dunif(x,a,b) est définie de la façon suivante :

$$\begin{array}{rcl} \text{dunif(.,a,b)} : \mathbb{R} & \to & \mathbb{R} \\ x & \mapsto & \text{dunif(x,a,b)} = \begin{cases} \dfrac{1}{b-a} & \text{si } x \in [a,b] \\ 0 & \text{si } x \notin [a,b] \end{cases} \end{array}$$

C'est une distribution continue.

La loi de Student est notée $t(k)$. $k = 1, 2, \dots$: degré de liberté de la loi

> curve(dt(x,1), -3,3)   ici k=1 et $-3 \leqslant x \leqslant 3$

> curve(dt(x,1), xlim=c(-3,3))

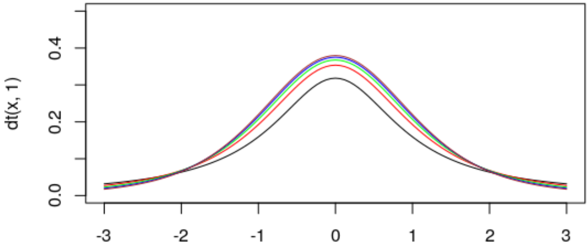
> curve(dt(x,1), xlim=c(-3,3),ylim=c(0,0.5))   ici on a ajouté $0 \leqslant y \leqslant 0.5$

> curve(dt(x,2), xlim=c(-3,3),ylim=c(0,0.5),add=T, col="red")

> curve(dt(x,3), xlim=c(-3,3),ylim=c(0,0.5),add=T, col="green")

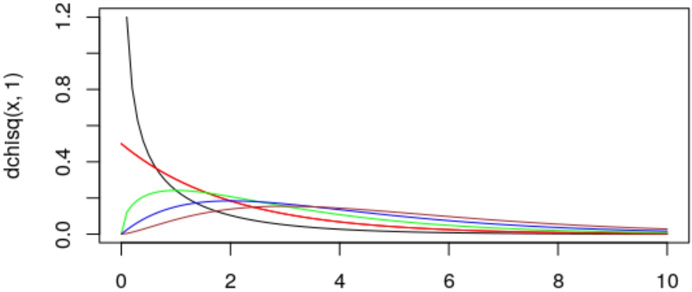
> curve(dt(x,4), xlim=c(-3,3),ylim=c(0,0.5),add=T, col="blue")

> curve(dt(x,5), xlim=c(-3,3),ylim=c(0,0.5),add=T, col="brown")



La loi du Khi-deux est notée $\chi^2(k)$. $k = 1, 2, \ldots$: degré de liberté de la loi $\chi^2(k)$ est définie pour $x \geqslant 0$

```
> curve(dchisq(x,1),0,10)
> curve(dchisq(x,2),add=T,col="red")
> curve(dchisq(x,3),add=T,col="green")
> curve(dchisq(x,4),add=T,col="blue")
> curve(dchisq(x,5),add=T,col="brown")
```



La loi de Fisher est notée $F(\nu_1, \nu_2)$ $\nu_1, \nu_2 = 1, 2, ...$: degrés de liberté de la loi

$F(\nu_1, \nu_2)$ est définie pour $x \geqslant 0$

```
> curve(df(x,1,1),0,10)
> curve(df(x,1,2),0,10,add=T,col="red")
> curve(df(x,2,1),0,10,add=T,col="blue")
```