# Cours 7 – Estimation dans le cadre de l'échantillon gaussien.

*Eya ZOUGAR* *

***Institut National des Sciences appliqués-INSA***

Génie mathématiques GM3
Thursday 9th March, 2023

INSTITUT NATIONAL DES SCIENCES APPLIQUÉES ROUEN NORMANDIE

*Basé sur le cours de Bruno PORTIER

# 1. Introduction : cadre et objectif.

On dispose de n données réelles $x_1, x_2, \ldots, x_n$.

On suppose que ces données sont les réalisations de $n$ variables aléatoires réelles $X_1, X_2, \ldots, X_n$, indépendantes et de même loi normale $\mathcal{N}(\mu, \sigma^2)$.

On s'intéresse à l'estimation des paramètres d'espérance $\mu$ et de variance $\sigma^2$ supposés inconnus.

Il est facile d'estimer ces paramètres, mais le fait de se placer dans le cadre gaussien, va nous permettre de préciser la loi des estimateurs pour toute taille $n$ de l'échantillon.

Nous allons en particulier pouvoir établir des résultats importants qui seront à la base de tests statistiques utiles.

# 2. Rappels sur quelques lois utiles.
## 2.1. Le vecteur gaussien.

Soit $Z$ un vecteur gaussien de $\mathbb{R}^n$, d'espérance le vecteur $\mu$ de $\mathbb{R}^n$, et de matrice de variance-covariance $\Gamma$, qu'on note $Z \sim \mathcal{N}(\mu, \Gamma)$.

On notera que la matrice $\Gamma$ est définie positive, donc inversible.

La densité de probabilité $f$ de $Z$ est une fonction de $\mathbb{R}^n$ dans $\mathbb{R}_+$. Elle est définie pour tout $z \in \mathbb{R}^n$ par :

$$f(z) = \frac{1}{\sqrt{(2\pi)^n \det(\Gamma)}} \exp\left(-\frac{1}{2}(z-\mu)^T \Gamma^{-1}(z-\mu)\right)$$

# 2.2. La variable gaussienne.

Dans le cas pour $n = 1$ : Si $Z \sim \mathcal{N}\left(\mu, \sigma^2\right)$

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

# 2.2. La variable gaussienne.

Dans le cas pour $n = 1$ : Si $Z \sim \mathcal{N}\left(\mu, \sigma^2\right)$

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

**Lemme**

Si $Z \sim \mathcal{N}\left(\mu, \sigma^2\right)$ alors $\dfrac{Z-\mu}{\sigma} \sim \mathcal{N}(0, 1)$

# 2.2. La variable gaussienne.

Dans le cas pour $n = 1$ : Si $Z \sim \mathcal{N}\left(\mu, \sigma^2\right)$

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

**Lemme**

Si $Z \sim \mathcal{N}\left(\mu, \sigma^2\right)$ alors $\dfrac{Z-\mu}{\sigma} \sim \mathcal{N}(0,1)$

**Moment d'ordre n**

Si $Z \sim \mathcal{N}\left(\mu, \sigma^2\right)$ alors $\mathbb{E}[(Z-\mu)^n] = m_n$ est donné par:

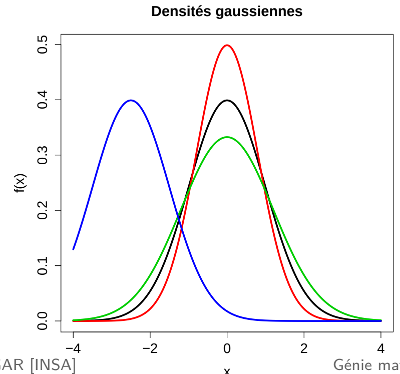
$$\begin{cases} m_{2k+1} = 0 & \textit{si} \text{ n=2k+1} \\ m_{2k} = \dfrac{(2k)!}{2^k k!}\sigma^{2k} & \textit{si} \text{ n=2k} \end{cases}$$

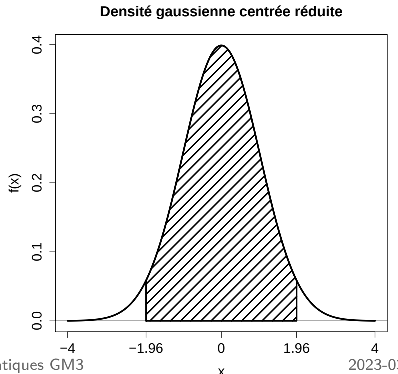
## 2.3. R code

On trouvera ci-dessous le code R permettant de réaliser les figures du slide suivant.

```
# Construction de courbes de densité avec quantiles à 95%
polycurve <- function(x, y, base.y = min(y), ...) {
polygon(x = c(min(x), x, max(x)), y = c(base.y, y, base.y), ...)
}
pdf("Gaussienne_Graphe.pdf")
x   = seq(-4,4,l=200)
d0 = dnorm(x)
d1 = dnorm(x,sd=0.8)
d2 = dnorm(x,sd=1.2)
d3 = dnorm(x,m=-2.5)
matplot(cbind(x,x,x,x),cbind(d0,d1,d2,d3),type="l",lty=1,lwd=3,
    xlab="x",ylab="f(x)",cex.axis=1.4,cex.lab=1.4,
    main="Densités gaussiennes",cex.main=1.5)
xxx=dev.off()
```

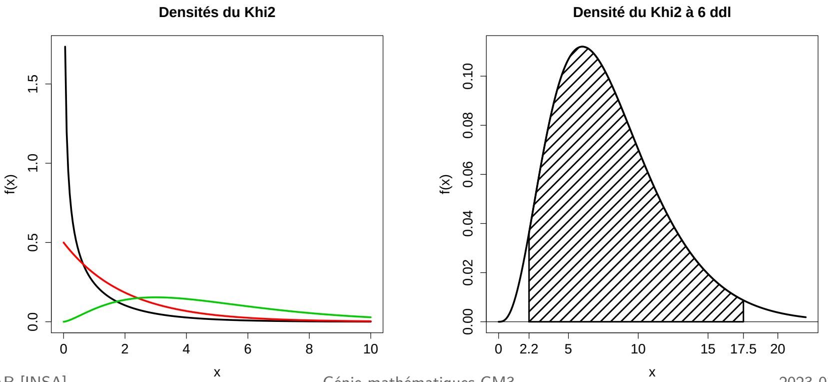
```
pdf("Gaussienne_Quantile.pdf")
plot(x,d0,type="l",xaxt="n",xlab="x",ylab="f(x)",cex.axis=1.4,
cex.lab=1.4,main="Densité gaussienne centrée réduite",cex.main=1.5)
axis(1,at=c(-4,-1.96,0,1.96,4),lab=c("-4","-1.96","0","1.96","4"),
    cex.axis=1.4)
polycurve(xx,yy,0,density=8,angle=45,lwd=2.5)
lines(x,d0,lwd=3)
abline(h=0)
xxx=dev.off()
```



## 2.4 Loi du Khi-deux.

On appelle loi du chi-deux à $p$ degrés de liberté (ddl) la loi de la variable $Z = \sum_{j=1}^{p} X_j^2$ où les variables $X_1, X_2, \ldots, X_p,\ p \geq 1$ sont des variables aléatoires indépendantes et de même loi $\mathcal{N}(0,1)$.
Oa a de plus, $\mathbb{E}[Z] = p$ et $\mathbb{V}ar[Z] = 2p$.

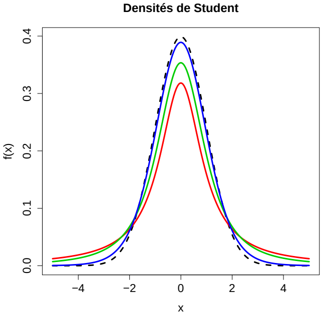


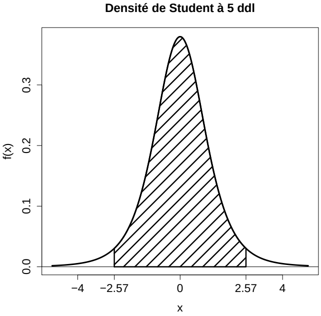
# 2.5. Loi de Student.

On appelle loi de Student à $n$ degrés de liberté, le rapport d'une gaussienne centrée réduite et de la racine carrée d'un khi-deux à $n$ degrés de liberté, la gaussienne et le khi-deux étant indépendants.

**Plus précisément,** si $U \sim \mathcal{N}(0,1)$ et $V \sim \chi^2_n$ avec $U$ et $V$ indépendantes, alors $Z_n = \dfrac{U}{\sqrt{V/n}}$ suit une loi de Student à $n$ ddl

et on note $Z \sim T_n$ et on a $\mathbb{E}(Z_n) = 0$ et $Z_n \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$.



# Lemme de Slutsky.

Soient $(X_n)_{n\geq 1}$ et $(Y_n)_{n\geq 1}$ deux suites de variables aléatoires réelles.
S'il existe une variable aléatoire $X$ et une constante non nulle $a$ telles que

$$X_n \xrightarrow[n\to\infty]{\mathcal{L}} X$$
$$Y_n \xrightarrow[n\to\infty]{P} a$$

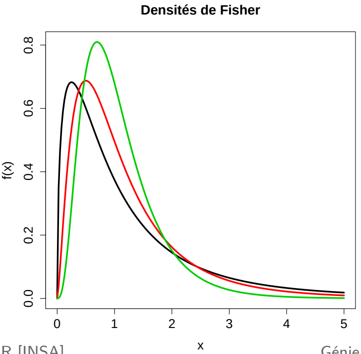
alors

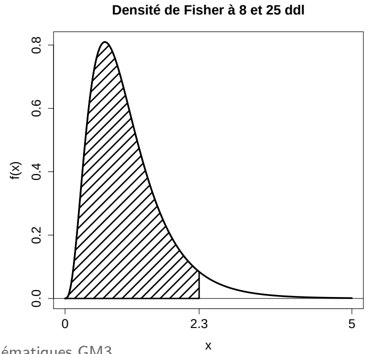
$$Y_n X_n \xrightarrow[n\to\infty]{\mathcal{L}} a X$$

# 2.6. Loi de Fisher.

On appelle loi de Fisher à $p$ et $q$ degrés de liberté le rapport de 2 khi-deux indépendants à $p$ et $q$ degrés de libertés respectivement.

**Plus précisément,** si $U \sim \chi^2_p$ et $V \sim \chi^2_q$ avec $U$ et $V$ indépendantes, alors $Z = \dfrac{U/p}{V/q}$ suit une loi de Fisher à $p$ et $q$ ddl et on note $Z \sim F(p, q)$.



# 3. Théorème de Cochran
## 3.1. Un premier Théorème.

**Rappel.** Si $Z$ est un vecteur gaussien centré et réduit, c'est-à-dire si $Z \sim \mathcal{N}(0, I_n)$, alors $\|Z\|^2 \sim \chi^2(n)$.

En effet, $\|Z\|^2 = \sum_{j=1}^n Z_j^2$ est une somme de $n$ gaussiennes indépendantes, centrées réduites élevées au carré, c'est donc par définition un khi-deux à $n$ degrés de liberté (ddl).

On a par ailleurs le théorème suivant :

**Théorème**

Si

- $Z \sim \mathcal{N}(0, I_n)$
- $P_E$ est un projecteur orthogonal de $\mathbb{R}^n$ sur un sous-espace vectoriel $E$ de dimension $p$

alors $\|P_E(Z)\|^2$ suit une loi $\chi^2(p)$.

# 3.2. Le Théorème de Cochran.

**Théorème de Cochran**

Soient $E_1$ et $E_2$ deux sous-espaces vectoriels orthogonaux de $\mathbb{R}^n$, de dimensions respectives $p_1$ et $p_2$, et soit $Z$ un vecteur gaussien de $\mathbb{R}^n$, centré et de matrice de variance-covariance $I_n$.
Alors,

- $\|P_{E_1}(Z)\|^2$ suit une loi $\chi^2(p_1)$
- $\|P_{E_2}(Z)\|^2$ suit une loi $\chi^2(p_2)$
- les variables aléatoires $\|P_{E_1}(Z)\|^2$ et $\|P_{E_2}(Z)\|^2$ sont indépendantes

**Remarque.** Ce théorème se généralise au cas de $2 \leq m \leq n$ sous-espaces vectoriels orthogonaux $(E_i)_{1 \leq i \leq m}$ de $E$.

# 3.3. Le corollaire de Cochran.

On déduit du théorème précédent, le corollaire suivant, qui nous sera utile dans le cas de l'échantillon gaussien pour étudier la loi des différents estimateurs.

**Corollaire de Cochran.**

Soit $n \geq p$. Soit $E$ un sous-espace vectoriel de $\mathbb{R}^n$ de dimension $p$, et soit $Z$ un vecteur gaussien de $\mathbb{R}^n$, centré et de matrice de variance-covariance $I_n$.
Alors, on a

$$\|Z\|^2 = \|P_E(Z)\|^2 + \|Z - P_E(Z)\|^2$$

et

- $\|P_E(Z)\|^2$ suit une loi $\chi^2(p)$
- $\|Z - P_E(Z)\|^2$ suit une loi $\chi^2(n-p)$
- $\|P_E(Z)\|^2$ et $\|Z - P_E(Z)\|^2$ sont indépendantes

# 4. Estimation des paramètres.
## 2.1. Définition des estimateurs.

Puisque $\mu = \mathbb{E}(X_1)$ et $\sigma^2 = \mathbb{V}\mathrm{ar}(X_1)$, on peut estimer le paramètre de moyenne $\mu$ par:

$$\overline{X}_n = \frac{1}{n} \sum_{j=1}^{n} X_j$$

et le paramètre de variance par:

$$S^2 = \frac{1}{(n-1)} \sum_{j=1}^{n} (X_j - \overline{X}_n)^2$$

On sait que, dans le cas de variables aléatoires $(X_j)$ indépendantes et de même loi, ces estimateurs sont sans biais et convergents.

Cependant, dans le cas de l'échantillon gaussien, on peut préciser leur loi pour toute taille d'échantillon $n$.

# 4.2. Propriétés.

On a les résultats suivants:

**Theorem**

- $\overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.
- $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.
- $\overline{X}_n$ et $S^2$ sont indépendants.

où $\chi^2_q$ désigne la loi du khi-deux à $q$ ddl.

**Corollaire**

$$\frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{S} \sim T_{n-1}$$

où $T_q$ désigne la loi de Student à $q$ ddl.

# 2. Preuves des différents résultats.
## 2.1. Loi de $\overline{X}_n$.

Puisque les variables $X_1, X_2, \ldots, X_n$ sont des variables gaussiennes indépendantes, $\overline{X}_n$ est une combinaison linéaire de variables gaussiennes indépendantes et donc une variable gaussienne.

Par conséquent, on a :

$$\overline{X}_n \sim \mathcal{N}\left(\mathbb{E}(\overline{X}_n), \mathbb{V}\mathrm{ar}(\overline{X}_n)\right).$$

Or, on a déjà montré que

- $\mathbb{E}\left[\overline{X}_n\right] = \mu$ ;
- $\mathbb{V}\mathrm{ar}(\overline{X}_n) = \dfrac{\sigma^2}{n}$.

Le résultat est ainsi établi.

# 2.2. Loi de $S^2$.

## 2.2.1. Notations et Corollaire de Cochran.

Pour tout entier $j = 1, 2, \ldots, n$, on pose $Z_j = \dfrac{X_j - \mu}{\sigma}$.

Les variables aléatoires $Z_1, Z_2, \ldots, Z_n$ sont indépendantes et de même loi normale centrée réduite.

Ainsi, le vecteur $Z = (Z_1, Z_2, \ldots, Z_n)^T$ est un vecteur gaussien centré et de matrice de variance-covariance $I_n$.

On pose $P = \dfrac{1}{n}\mathbf{1_n 1_n^T}$, avec $\mathbf{1_n} = (\mathbf{1}, \mathbf{1}, \ldots, \mathbf{1})^{\mathbf{T}}$.

Il est facile de montrer que $P$ est un projecteur orthogonal sur le sous-espace vectoriel de $\mathbb{R}^n$ engendré par le vecteur $\mathbf{1_n}$. Ce sous-espace vectoriel est de dimension 1.

Ainsi en utilisant le corollaire de Cochran, on déduit que

- $\|PZ\|^2 \sim \chi^2_1$
- $\|(I_n - P)Z\|^2 \sim \chi^2_{n-1}$
- et $\|PZ\|^2$ et $\|(I_n - P)Z\|^2$ sont indépendants.

## 2.2.2. Fin de la preuve.

On pose $\overline{Z}_n = \frac{1}{n}\sum_{j=1}^{n} Z_j$. Il est facile de voir que $\overline{Z}_n = \frac{\overline{X}_n - \mu}{\sigma}$ et que $PZ = \mathbf{1_n}\overline{\mathbf{Z}}_\mathbf{n}$. On déduit alors que

- $\|PZ\|^2 = \left\|\mathbf{1_n}\,\overline{\mathbf{Z}}_\mathbf{n}\right\|^2 = n\overline{Z}_n^2 = \frac{n(\overline{X}_n - \mu)^2}{\sigma^2}$.
- $\|(I_n - P)Z\|^2 = \left\|Z - \mathbf{1_n}\overline{\mathbf{Z}}_\mathbf{n}\right\|^2 = \sum_{j=1}^{n}(Z_j - \overline{Z}_n)^2$.

Or, il est facile de voir que pour tout entier $j$, qu $Z_j - \overline{Z}_n = \frac{X_j - \overline{X}_n}{\sigma}$. En remplaçant alors dans le calcul de $\|(I_n - P)Z\|^2$, on obtient :

$$\|(I_n - P)Z\|^2 = \sum_{j=1}^{n}\frac{(X_j - \overline{X}_n)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2}.$$

Ainsi, en reprenant les résultats fournis par le corollaire de Cochran, on déduit que $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ et $S^2$ est indépendant de $\overline{X}_n$.

# 2.3. Loi de Student.

La preuve du dernier résultat repose sur le résultat de probabilité suivant :

- Soient $U$ et $V$ deux variables aléatoires indépendantes.
  Si $U \sim \mathcal{N}(0\,,\,1)$ et $V \sim \chi^2_n$ alors $\dfrac{U}{\sqrt{V\,/\,n}} \sim T_n$.

On utilise ce résultat avec $U = \dfrac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$ et $V = \dfrac{(n-1)S^2}{\sigma^2}$.

## 2.3. Loi de Student.

La preuve du dernier résultat repose sur le résultat de probabilité suivant :

- Soient $U$ et $V$ deux variables aléatoires indépendantes.

  Si $U \sim \mathcal{N}(0\,,\,1)$ et $V \sim \chi^2_n$ alors $\dfrac{U}{\sqrt{V\,/\,n}} \sim T_n$.

On utilise ce résultat avec $U = \dfrac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$ et $V = \dfrac{(n-1)S^2}{\sigma^2}$.

**En effet:** Puisque $\overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$, alors $\dfrac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \sim \mathcal{N}(0,1)$, et comme $\overline{X}_n$ et $S^2$ sont indépendants, il en est de même pour $U$ et $V$.

ainsi, on obtient par construction le résultat souhaité :

$$\frac{\dfrac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2}\Big/(n-1)}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{S} \sim T_{n-1}$$