UE4 : Biostatistiques

Chapitre 3 : Partie 2

Probabilités

Professeur Philippe CINQUIN

Année universitaire 2011/2012

Plan

- A) Introduction
- B) Statistiques descriptives
- C) Probabilités
- D) Estimation
- E) Intervalles de confiance
- F) Problème récapitulatif
- G) Résumé des objectifs

C.5 Variable Aléatoire Continue

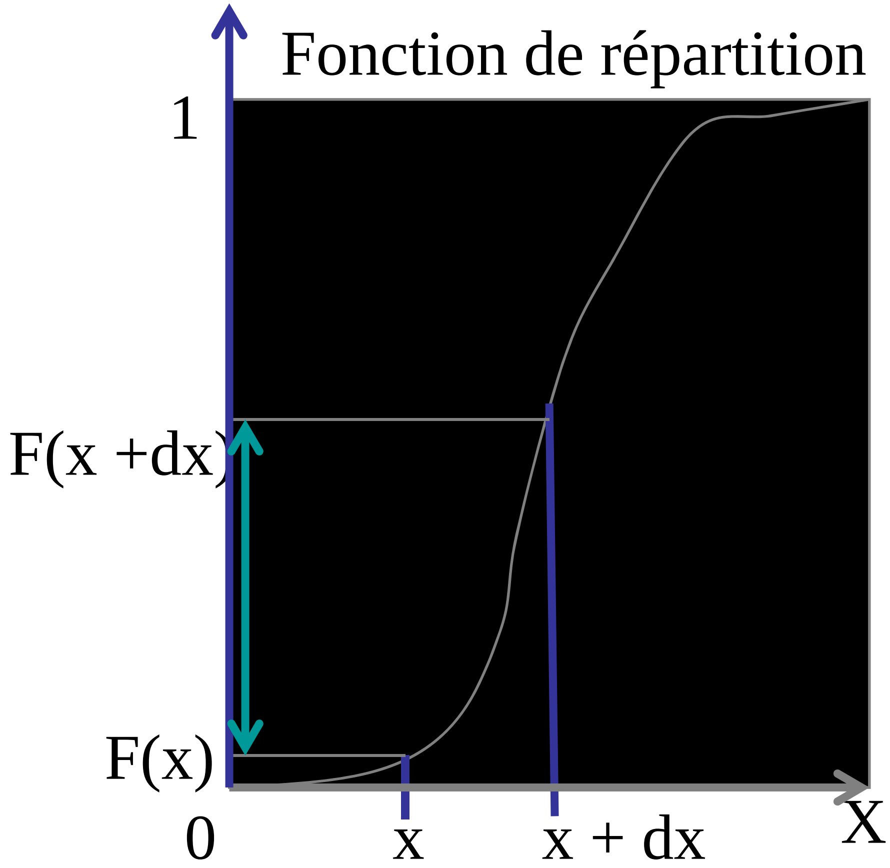
- **1 Définition**
  - Une v. a. X est **continue** si l'ensemble des valeurs qu'elle prend n'est pas **dénombrable** (par exemple, $\Re$ tout entier, ou un intervalle de $\Re$).

2 Fonction de répartition d'une v. a. continue

- **L'expression *P(X=x)* n'a plus de sens !** On pourrait considérer qu'il s'agit de la limite quand $dx \to 0$ de $P(x < X \leq x + dx)$
  - Je lance une bille censée être « ponctuelle » sur un carrelage dont les carreaux sont « grands » : la probabilité que la bille tombe sur un carreau donné a un sens (1/n, si n est le nombre de carreaux)
  - Que se passe-t-il si la taille des carreaux rétrécit sans fin ?
    - n augmente sans fin
    - La probabilité que la bille tome sur un carreau donné se rapproche infiniment de 0, et ne m'apporte donc plus aucune information

Fonction de répartition (suite)

- Soit $F(x) = P(X \leq x)$
- F est la fonction de répartition de X (noter qu'elle a également un sens pour des v. a. discrètes).
- On remarque que
- $P(x < X \leq x + dx) = F(x+dx) - F(x)$



Fonction de répartition (suite)

- Soit f la dérivée de F, définie par

$$f(x) = \lim_{dx \to 0} \frac{[F(x+dx) - F(x)]}{dx}$$

- **f est la densité de probabilité de X**

$$\int_{-\infty}^{x} f(t)dt = F(x) - F(-\infty) = F(x)$$

Paramètres d'une v. a. continue

- Les définitions de l'Espérance, de la Variance et de l'Ecart type se généralisent au cas continu en remplaçant les sommes discrètes par des intégrales.

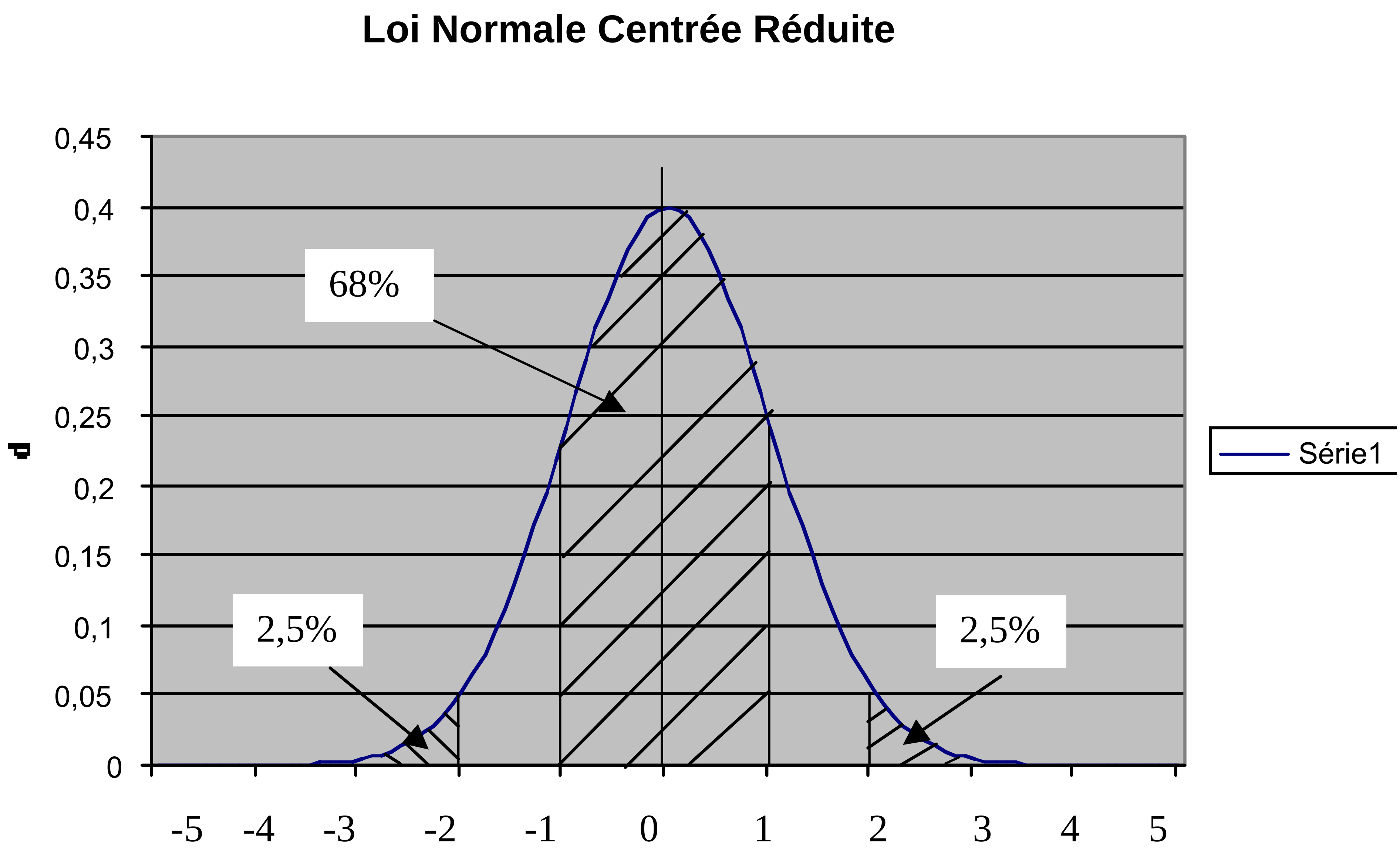
- Toutes les propriétés des paramètres des lois discrètes sont conservées dans le cas continu.

C.6 Loi Normale (ou Loi de Gauss)

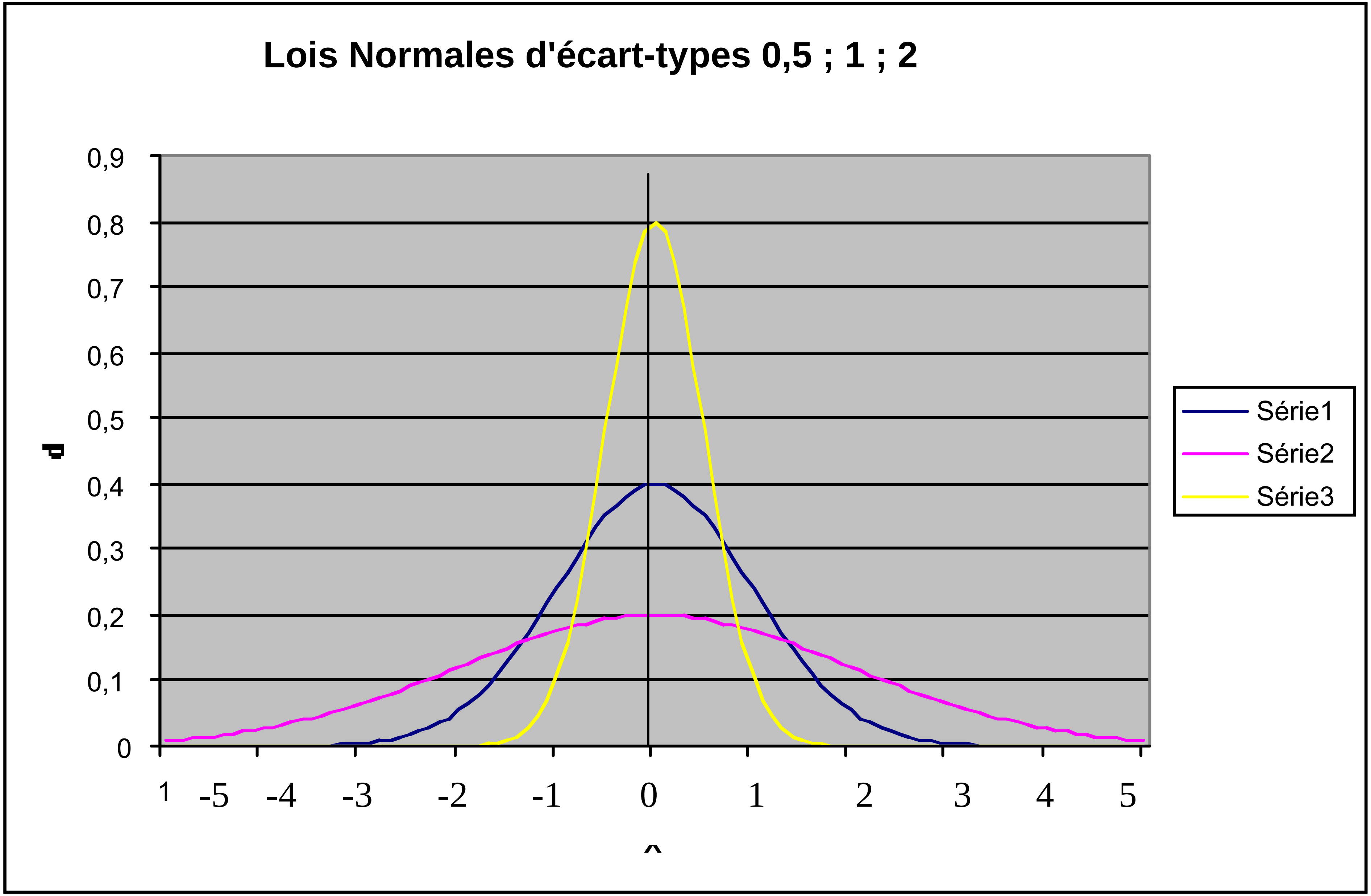
- **1 Loi Normale centrée, réduite**
- Sa densité de probabilité est donnée par

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

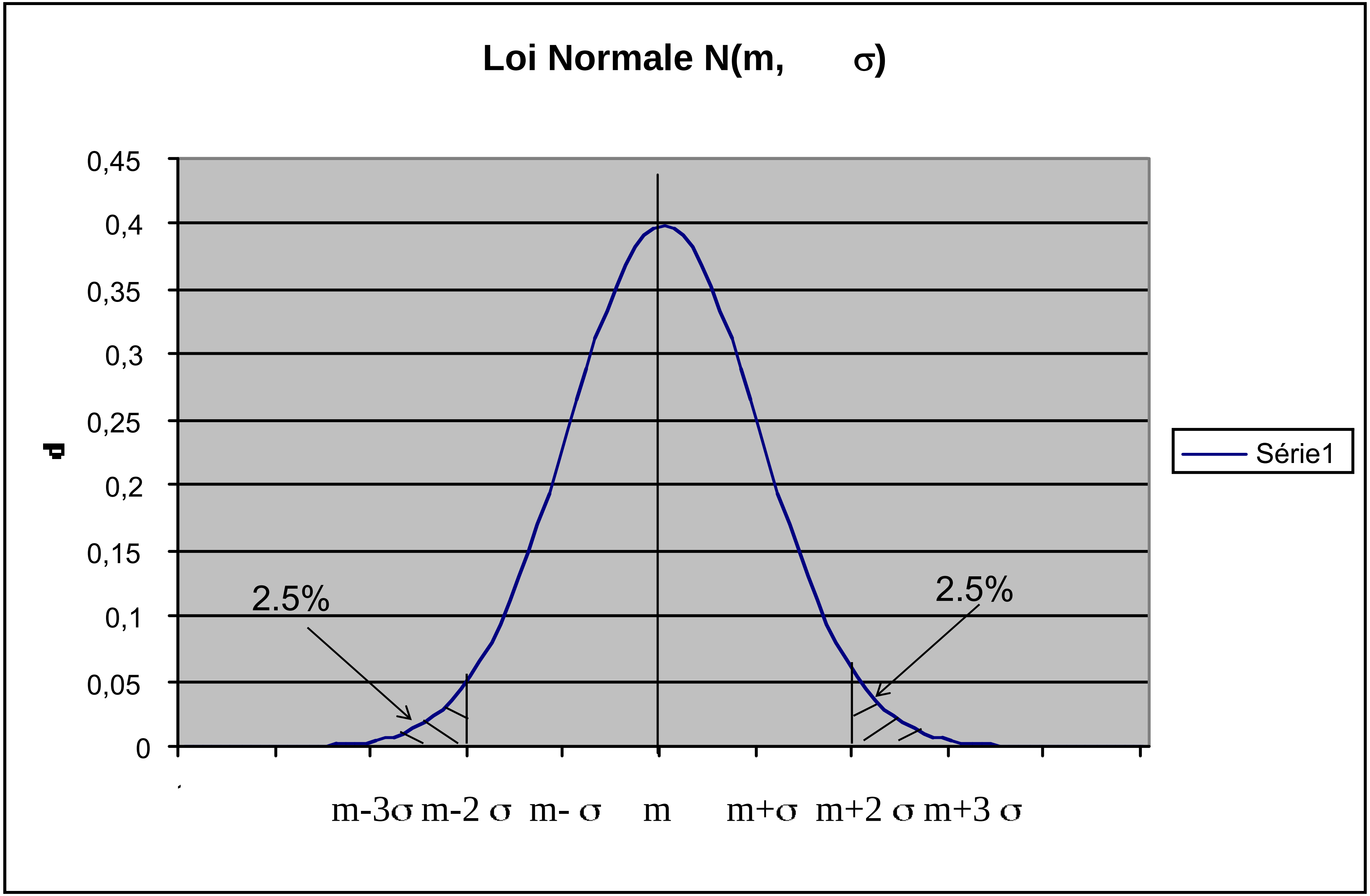
- E(N) = 0
- V(N) = 1
- C’est une courbe “ en cloche ” centrée sur 0.



Loi Normale Centrée Réduite
0,45
0,4
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
68%
2,5%
2,5%
Série1
-5
-4
-3
-2
-1
0
1
2
3
4
5



Lois Normales d'écart-types 0,5 ; 1 ; 2
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
p
1
-5
-4
-3
-2
-1
0
1
2
3
4
5
x
Série1
Série2
Série3



Loi Normale N(m, σ)
0,45
0,4
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
p
2.5%
2.5%
Série1
m-3σ m-2 σ m- σ m m+σ m+2 σ m+3 σ

2 Intérêt de la loi normale

- toute v. a. X, dont la valeur résulte de la sommation de nombreux effets aléatoires indépendants et du même ordre de grandeur, est approximativement une v. a. normale ;

- or, de nombreuses grandeurs biologiques (taille, envergure et poids à la naissance, …) sont la résultante de nombreuses causes aléatoires indépendantes s’exerçant au cours de l’embryogénèse, et sont dont normales.

2 Intérêt de la loi normale (suite)

- les grandeurs mesurant des sensations sont le plus souvent normales, ou log-normales (S = LogE suit une loi normale)

- les grandeurs résultant d'une mesure expérimentale suivent en général des lois normales

- certaines pathologies modifient le type de loi d'une v. a. biologique
  - en conservant son caractère “ normal ”, mais en modifiant m et s,
  - ou en supprimant son caractère “ normal ”

3 Exemple d'utilisation de la loi normale

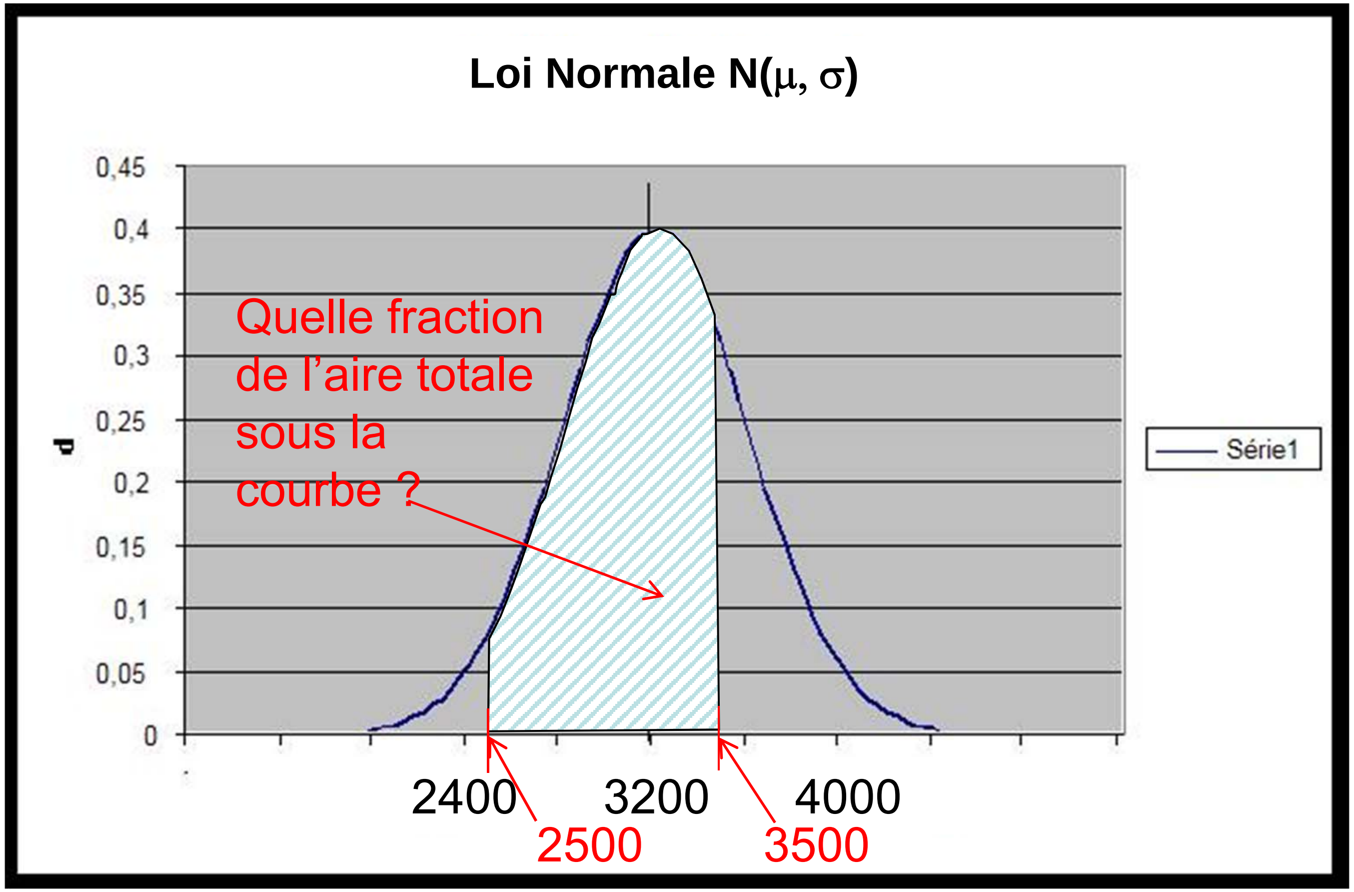
- Soit X la variable aléatoire « poids de naissance ». On suppose que X suit une loi normale d'Espérance μ = 3200 g, et d'écart-type σ = 400 g.
  - Quelle est la probabilité qu'un nouveau né ait un poids supérieur à 4 kg ?
  - Quelle est la probabilité que son poids soit compris entre 2500 g et 3500 g ?

Exemple, suite

- Poids supérieur à 4000 g ?
  - 4000 g - 3200 g = 800 g = 2 x 400 g = 2 σ
  - Donc P (X > 4000 g) = 0.05/2 = 0.025
- Poids compris entre 2500 g et 3500 g ?

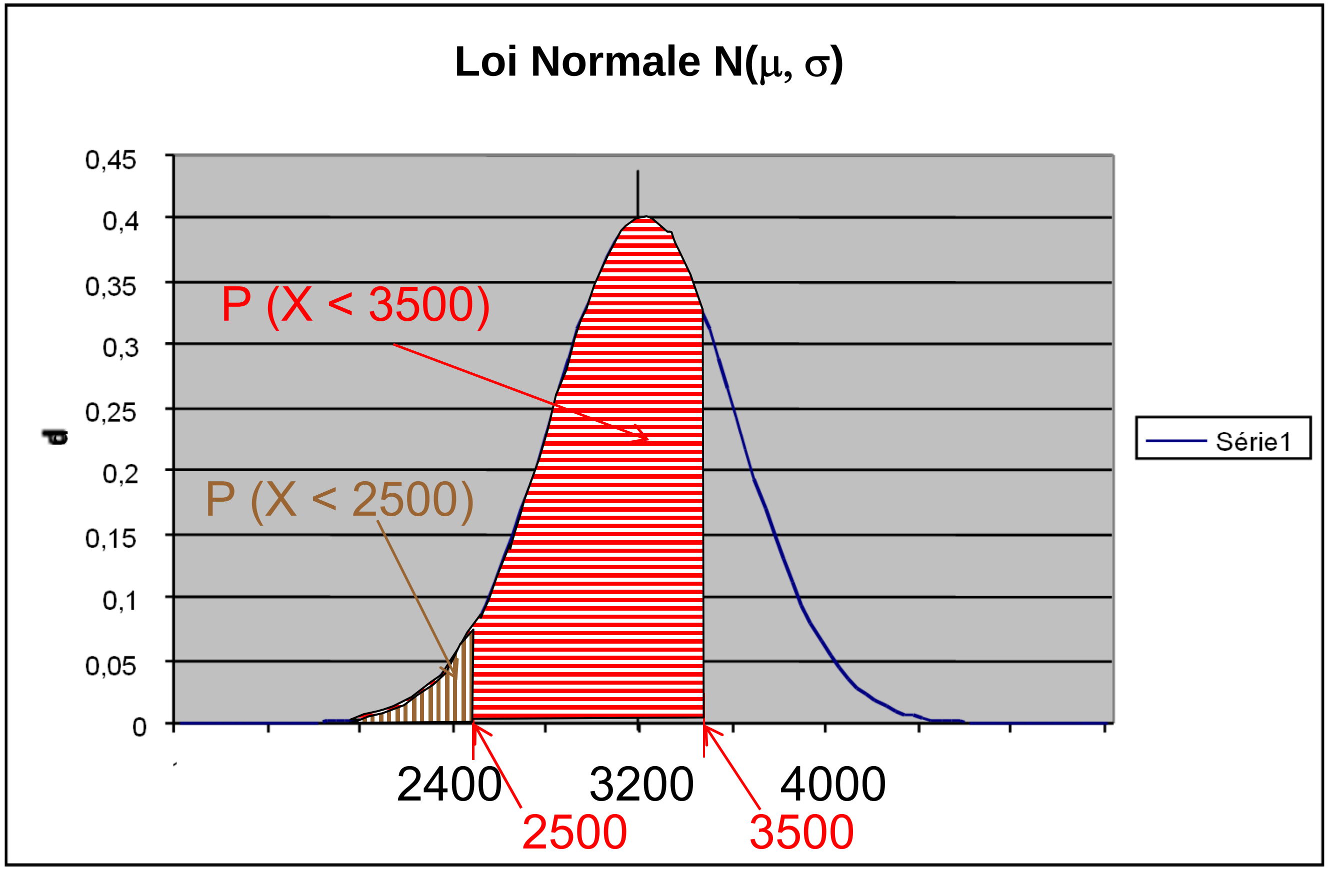
$$P\,(2500 < X < 3500) =$$
$$P\,(2500 - 3200 < X - \mu < 3500\text{-}3200) =$$
$$P\,(-\,700 < X - \mu < 300) =$$
$$P\,(-\,700/400 < (X - \mu)/\sigma < 300/400) =$$
$$P\,(-\,1.75 < (X - \mu)/\sigma < 0.75) =$$
$$P\,[(X - \mu)/\sigma < 0.75] - P\,[(X - \mu)/\sigma < -\,1.75]$$

$N(\mu = 3200, \sigma = 400)$



P (2500 < X < 3500) = « aire sous la courbe » entre $x_1 = 2500$ et $x_2 = 3500$

N(μ = 3200, σ = 400)

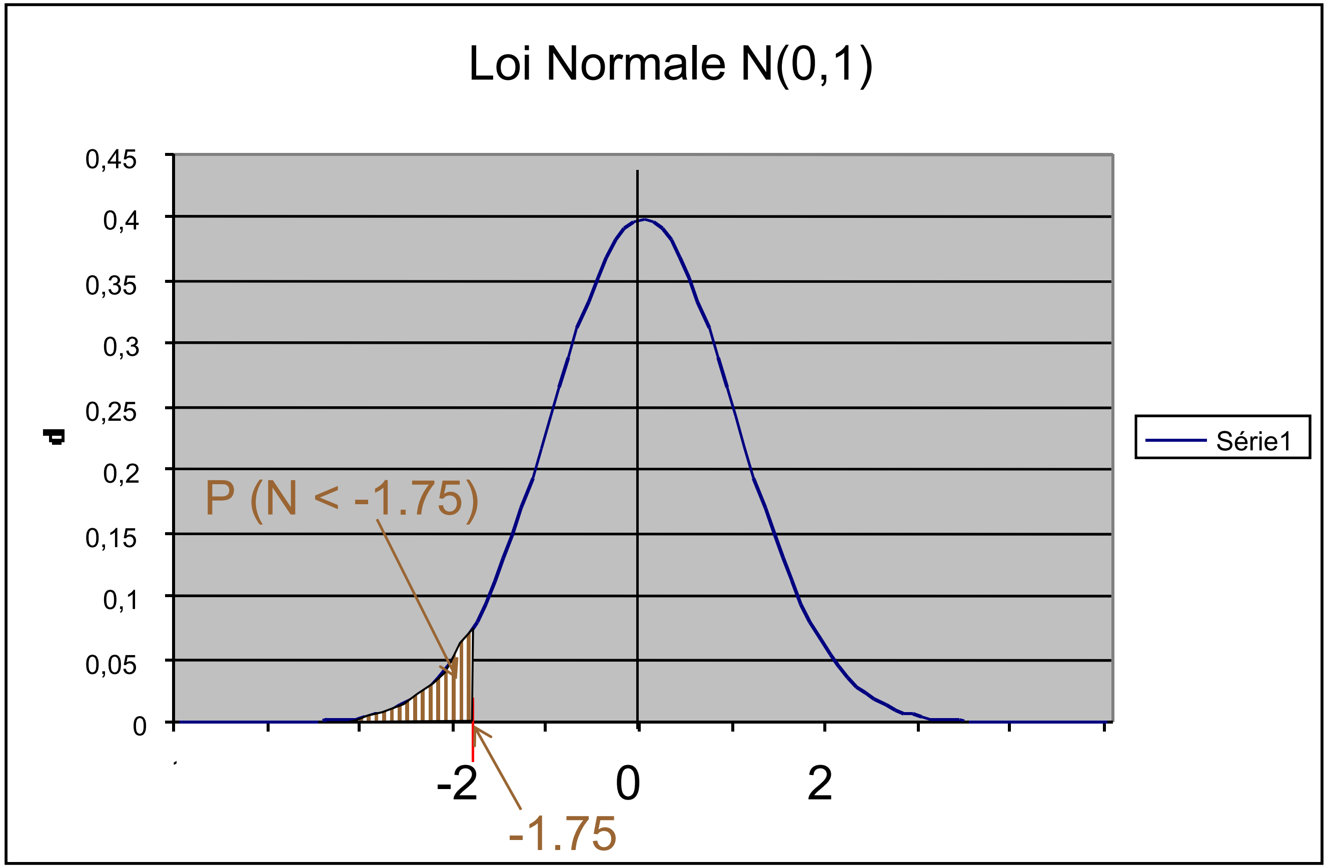


P (2500 < X < 3500) = P (X < 3500) - P (X < 2500)

Exemple, suite

- Or, $N = (X - \mu)/\sigma$ est une loi normale « centrée » (espérance nulle) et réduite (variance 1)
  - En effet, $Y = (X - \mu)$ est normale (soustraction d'une constante à une loi normale), et centrée (on enlève à X sa propre moyenne), et sa variance n'a pas changé (on a enlevé une constante, μ).
  - Et $N = Y/\sigma$ est une loi normale (division d'une loi normale par une constante) et de variance 1 (on a divisé par l'écart type σ de Y, donc la variance de Y se trouve divisée par σ^2)
- Donc $P(2500 < X < 3500) = F(0.75) - F(-1.75)$
- Il faut faire appel à des tables!

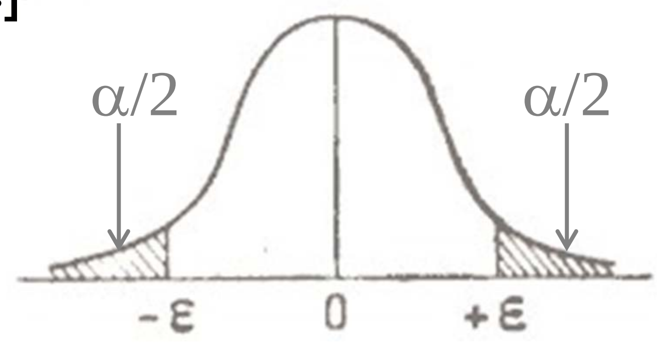
Calcul de P(X < 2500)



- P (X < 2500) = P (X-3200 < 2500 - 3200) = P(X-3200 < -700)
- P (X < 2500) = P [(X-3200)/400 < -700/400] = P(N < -1.75)

Exemple, suite

- Préférer l'utilisation d'un seul type de tables, la table « de l'écart réduit » (particulièrement utile pour les tests d'hypothèses qui suivront)
- Bien observer le « cartouche » de cette table (ne pas hésiter à le redessiner rapidement à la main)
- La table donne la probabilité α pour qu'une loi normale centrée et réduite (Espérance = 0, σ^2 = 1) égale ou dépasse, en valeur absolue, une valeur donnée ε, c'est-à-dire la probabilité extérieure à l'intervalle $[-\varepsilon, \varepsilon]$

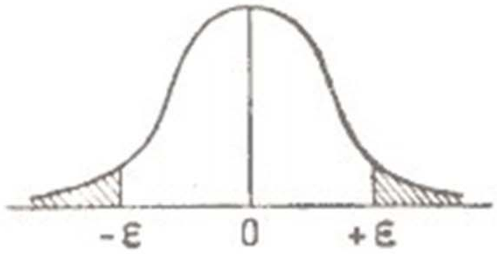


Recherche dans la table « de l'écart réduit »

- Nous cherchons la valeur de α pour les 2 bornes « centrées et réduites », -1.75 et 0.75
- Pour -1.75
  - Compte tenu de la symétrie de la loi normale, les valeurs de ε sont toujours positives, ici donc $\varepsilon = 1.75$

Table de l'écart-réduit (loi normale)

La table donne la probabilité α pour que l'écart-réduit égale ou dépasse, en valeur absolue, une valeur donnée ε, c'est-à-dire la probabilité extérieure à l'intervalle $(-\varepsilon, +\varepsilon)$.



α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	∞	2,576	2,326	2,170	2,054	1,960	1,881	1,812	1,751	1,695
0,10	1,645	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311
0,20	1,282	1,254	1,227	1,200	1,175	1,150	1,126	1,103	1,080	1,058
0,30	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,860
0,40	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,690
0,50	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,60	0,524	0,510	0,496	0,482	0,468	0,454	0,440	0,426	0,412	0,399
0,70	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,80	0,253	0,240	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,90	0,126	0,113	0,100	0,088	0,075	0,063	0,050	0,038	0,025	0,013

La probabilité α s'obtient par addition des nombres inscrits en marge.

Exemple : pour $\varepsilon = 1{,}960$ la probabilité est $\alpha = 0{,}00 + 0{,}05 = 0{,}05$.

Exemple, suite

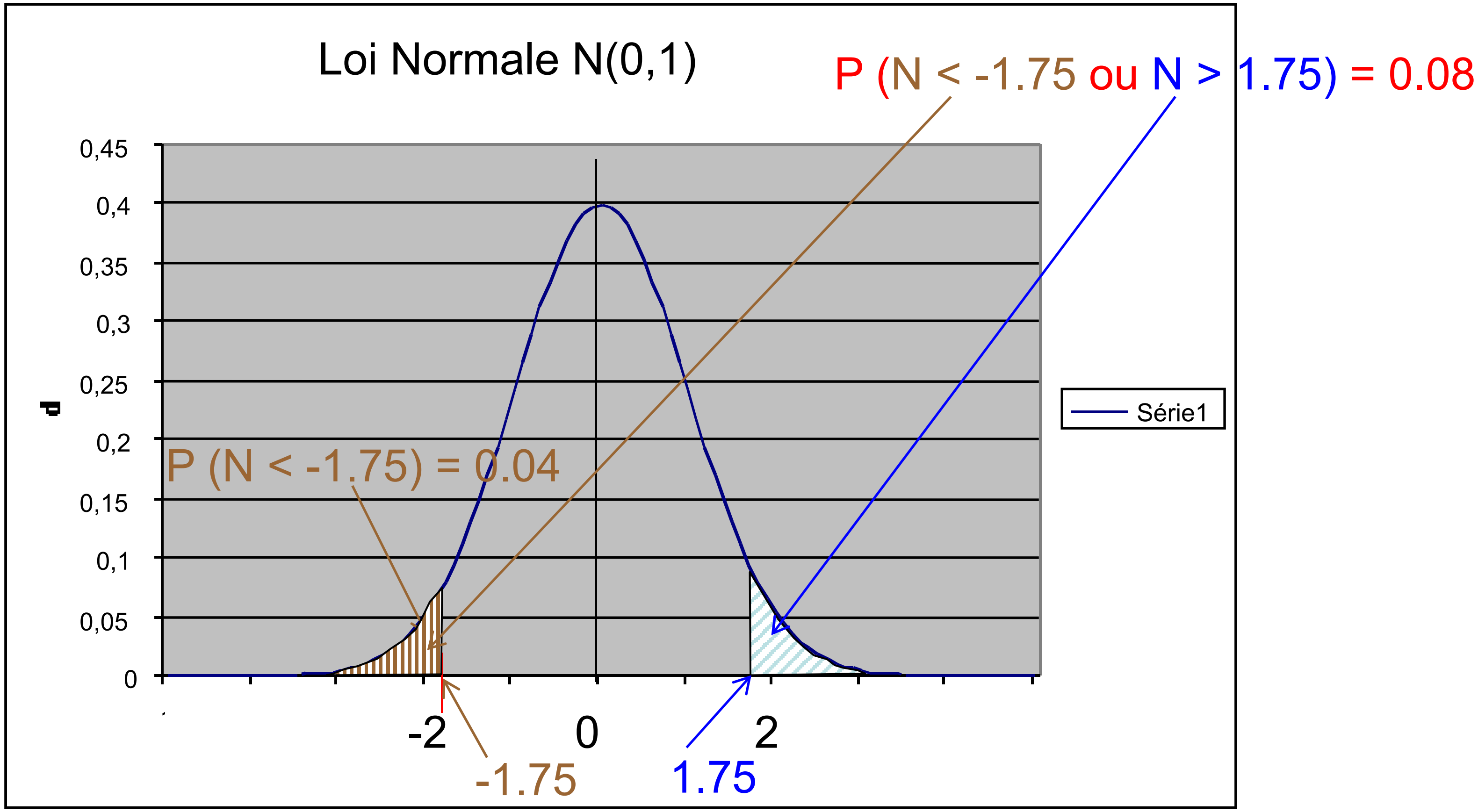
- La table nous dit donc que le α associé à ε = 1.75 vaut 0.08 (= 0.00 + 0.08)

Valeur de la ligne

Valeur de la colonne

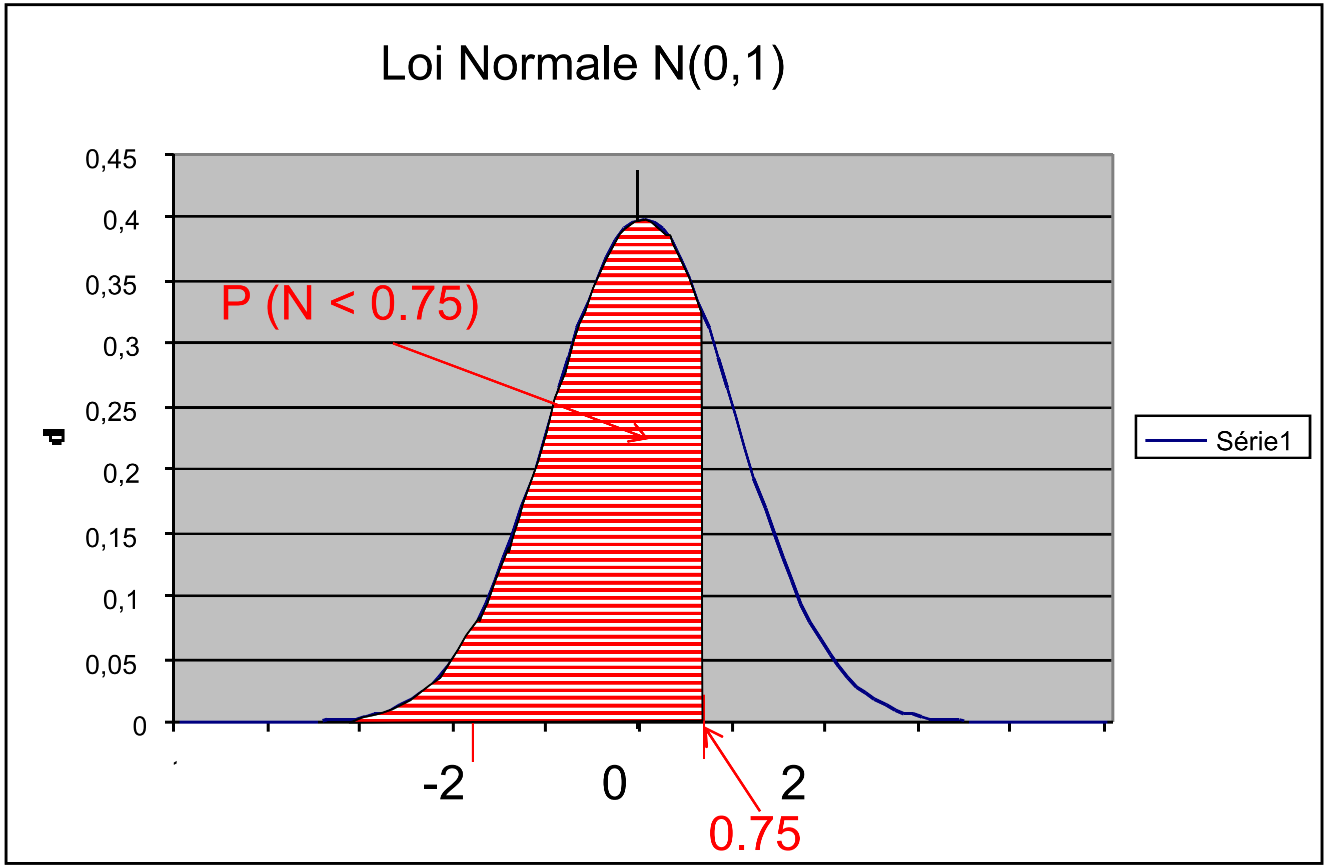
- Autrement dit, si N est une loi N(0,1), P (N < -1.75) + P(1.75 < N) = 0.08
- Compte tenu de la symétrie de la loi Normale, P (N < -1.75) = 0.08/2 = 0.04

Calcul de P(X < 2500) = P(N<-1.75)



- P(N < -1.75) = 0.08/2 = 0.04

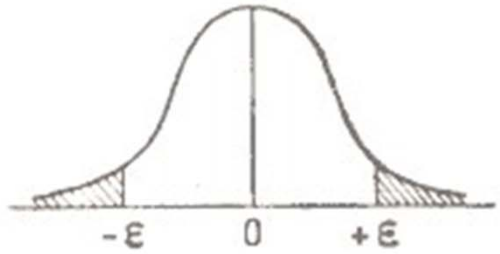
Calcul de P(X < 3500)



- P (X < 3500) = P (X-3200 < 3500 - 3200) = P(X-3200 < 300)
- P (X < 3500) = P [(X-3200)/400 < 300/400] = P(N < 0.75)

Table de l'écart-réduit (loi normale)

La table donne la probabilité α pour que l'écart-réduit égale ou dépasse, en valeur absolue, une valeur donnée ε, c'est-à-dire la probabilité extérieure à l'intervalle $(-\varepsilon, +\varepsilon)$.



α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	∞	2,576	2,326	2,170	2,054	1,960	1,881	1,812	1,751	1,695
0,10	1,645	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311
0,20	1,282	1,254	1,227	1,200	1,175	1,150	1,126	1,103	1,080	1,058
0,30	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,860
0,40	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,690
0,50	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,60	0,524	0,510	0,496	0,482	0,468	0,454	0,440	0,426	0,412	0,399
0,70	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,80	0,253	0,240	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,90	0,126	0,113	0,100	0,088	0,075	0,063	0,050	0,038	0,025	0,013

La probabilité α s'obtient par addition des nombres inscrits en marge.

Exemple : pour $\varepsilon = 1,960$ la probabilité est $\alpha = 0,00 + 0,05 = 0,05$.

Exemple, suite

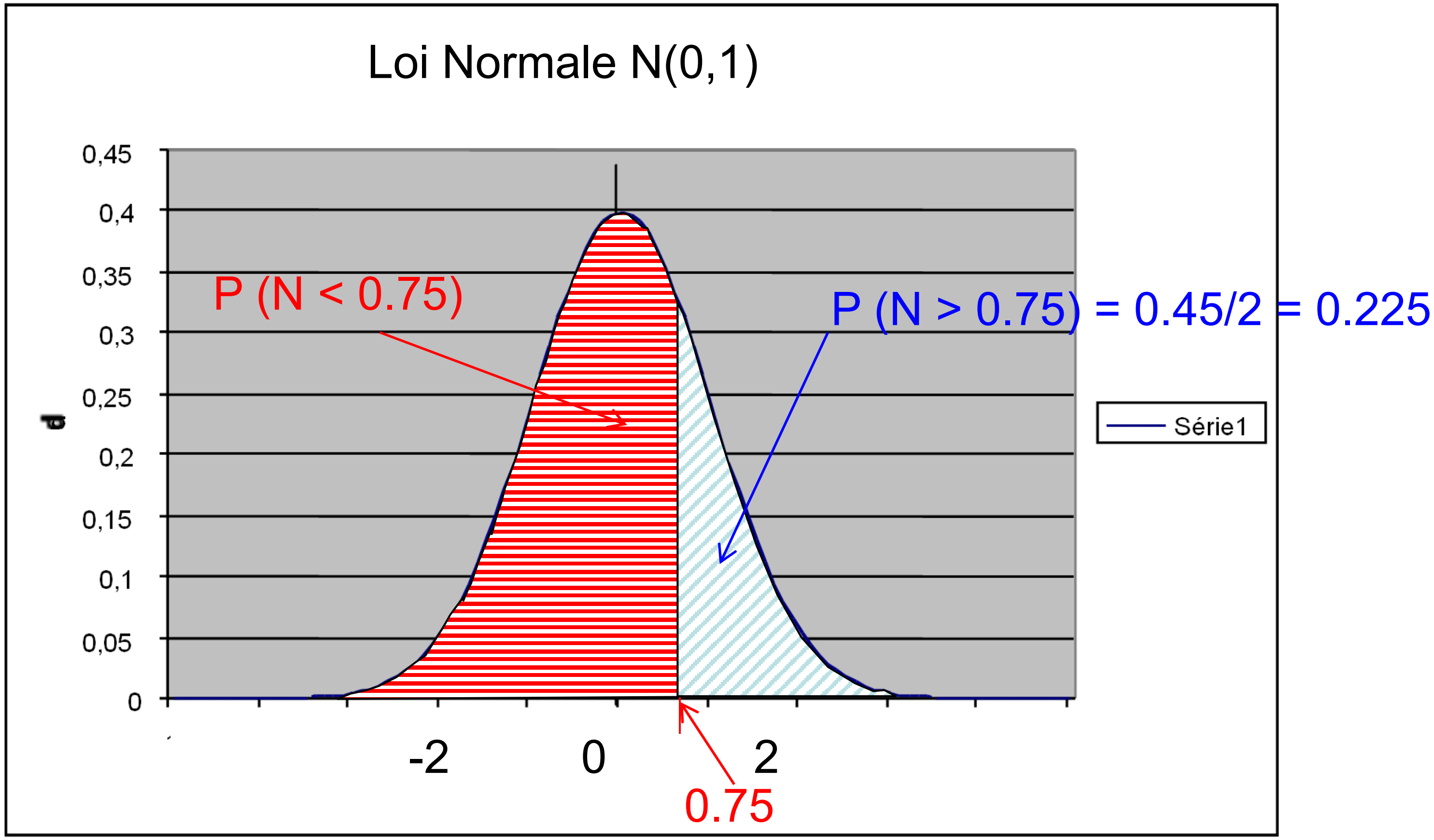
- La table nous dit donc que le α associé à $\varepsilon = 0.75$ vaut 0.45 (= 0.40 + 0.05)

Valeur de la ligne

Valeur de la colonne

- Autrement dit, si N est une loi N(0,1),
  $P(N < -0.75) + P(0.75 < N) = 0.45$
- Compte tenu de la symétrie de la loi Normale,
  $P(0.75 < N) = 0.45 / 2 = 0.225$
- Donc $P(N < 0.75) = 1 - 0.225 = 0.775$

Calcul de P(X < 3500) = P(N<0.75)



- P (X < 3500) = P(N < 0.75) = 1 - 0.225 = 0.775

Exemple, suite

- Donc P (2500 < X < 3500) = F (0.75) - F (-1.75) = P (N < 0.75) - P(N < -1.75) = 0.775 - 0.04 = 0.735
- Pour répondre à ce type de questions, il faut donc :
  - Centrer (soustraire la moyenne)
  - Réduire (diviser par l'écart-type)
  - Interpréter les résultats dans la « table de l'écart-réduit »
- NB cette table est complétée par la ligne suivante, pour les petites valeurs de probabilité :

Table pour les petites valeurs de la probabilité.

α	0,001	0,000 1	0,000 01	0,000 001	0,000 000 1	0,000 000 01	0,000 000 001
ε	3,29053	3,89059	4,41717	4,89164	5,32672	5,73073	6,10941

Mentions légales

Ce document a été réalisé par la Cellule TICE de la Faculté de Médecine et de Pharmacie de Grenoble
(Université Joseph Fourier – Grenoble 1)