<u>**Statistiques de gestion**</u>

<u>CHAPITRE I</u>

<u>RAPPELS: LOI NORMALE</u>

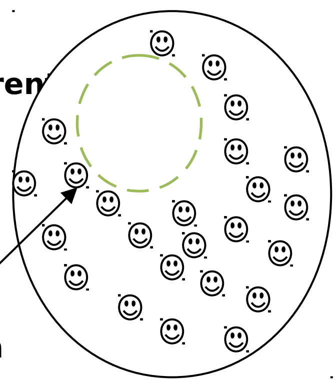
Introduction

Qu’est ce que les statistiques descriptives? On désigne par statistique descriptive l’ensemble des méthodes de collecte et traitement des données. La description des données passe par :

- Une présentation synthétique (tableaux…).
- Une représentation graphique adaptée (histogramme…).
- Un résumé numérique par le calcul de certaines grandeurs typiques (moyenne…).
- Une étude des éventuelles corrélations entre variables. (Déjà vu en)

- **Qu’est-ce que la statistique inféren**

Population Ensemble de référence



Echantillon Sous-ensemble de la population

<u>Exemple</u>

Pour pouvoir effectuer des prévisions il convient d’avoir un échantillon « représentatif ».

Aux élections américaines de 1936, le journal Literary Digestprédit la victoire du républicain Landon en interrogeant par téléphone plus de 2 millions d’électeurs.

Un sondeur, George Gallup, annonce lui la victoire du démocrate Roosevelt en ne sondant que 3000 personnes, mais judicieusement choisies.

C’est ce dernier qui avait raison, l’échantillon du Literary Digest était certes plus important mais « biaisé », car à cette époque seule la bourgeoisie possédait le téléphone…

Pourquoi travaille-t-on sur un échantillon ?

Coût temps impossibilité d’avoir la population entière

La statistique inférentielle est pour sa part l’ensemble des méthodes qui permettent à partir de l’étude d’un échantillon d’induire des informations sur une population.

• Elle nécessite un choix judicieux de l’échantillon, i.e. il doit être représentatif de la population (cf. exemple des élections américaines de 1936 ci-dessus)

• Elle utilise des modèles théoriques de référence, les lois de probabilités.

En effet, on constate en général que la répartition statistique d’une variable sur un échantillon est voisine d’une loi de probabilité.

La statistique inférentielle permet alors :

- D’estimer les paramètres de la loi de probabilité à partir de l’échantillon.

- De mesurer la validité de cette estimation par un intervalle de confiance.

- De mesurer l’adéquation de la loi de probabilité choisie à l’échantillon par des tests statistiques.

Rappels: Variables aléatoires discrètes

Loi Binomiale:

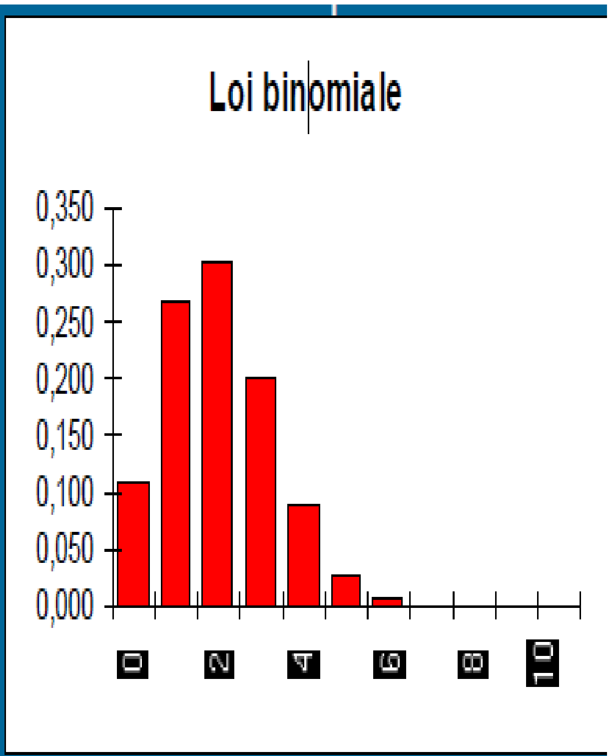
- 1 épreuve élémentaire
- 2 résultats possibles le succès (proba.p) et l’échec (1-p)
- n répétitions de l’épreuve
- Si X est le nombre de succès au cours des n épreuves

$$P(X = k) = C_n^k p^k (1 - p)^{(n-k)}$$

$$\text{avec } C_n^k = \frac{n!}{k!(n-k)!}$$

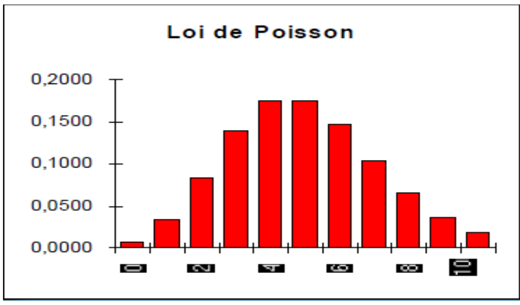
Exemple
p = 0,2
n = 10

x	P(X=x)
0	0,107
1	0,268
2	0,302
3	0,201
4	0,088
5	0,026
6	0,006
7	0,001
8	0,000
9	0,000
10	0,000



La loi de Poisson

La distribution de probabilité de poisson est souvent utilisée pour modéliser les taux d’arrivée dans des situations de file d’attente. (nombre d’arrivées à une station de lavage)



$$P(X = k) = e^{-m}\frac{m^k}{k!}$$

Rappels: Variable aléatoire continue

La probabilité que la v.a.c. X prenne une valeur dans un intervalle entre a et b est donnée par l'aire sous le graphique de la fonction de densité de probabilité f(x) entre a et b.

$$\int_a^b f(x)dx = P(a \leq X \leq b) = F(b) - F(a)$$

La fonction de répartition $F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$

Propriétés de la fonction de densité de probabilité

$$f(x) \geq 0 \text{ pour tout } x \text{ réel}$$

Une probabilité est toujours ≥ à 0.

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

L'aire totale sous la courbe doit être égale à 1.

Distribution exponentielle

La loi exponentielle est utile pour décrire la durée de réalisation d'une tâche

• Exemples:

– dans les files d'attente, la distribution exponentielle est souvent utilisée pour le temps de service

• Liée à la loi de Poisson qui fournit une bonne description du nombre d'occurrences par intervalle

• Loi exponentielle fournit une bonne description de la longueur de l'intervalle entre les occurrences

Distribution exponentielle

Une v.a.c X pouvant prendre toutes les valeurs x dans l'intervalle de 0 à ∞, pour μ>0 dont la fonction de densité est :

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ \dfrac{1}{\mu} e^{-x/\mu} & \text{if } x \geq 0 \end{cases}$$

s'appelle une v.a. exponentielle de paramètre μ.

LA LOI NORMALE

C'est possiblement la loi la plus importante de la théorie statistique. Pourquoi?

– Elle fournit une bonne description de certains phénomènes réels. Les distributions qui sont souvent proches de la normale:

• Résultats de tests d'aptitude pris par plusieurs personnes (résultats d'examens, tests de QI).

• Mesures répétées avec soin de la même quantité avec des équipements de laboratoires.

• Caractéristiques de populations biologiques Bonne approximation d'expériences impliquant un résultat au hasard, comme le lancer d'une pièce un très grand nombre de fois où l'on regarde le nombre de piles.

– Plusieurs procédures statistiques sophistiquées basées sur les lois normales fonctionnent bien lorsque les distributions sont symétriques.

Distribution normale

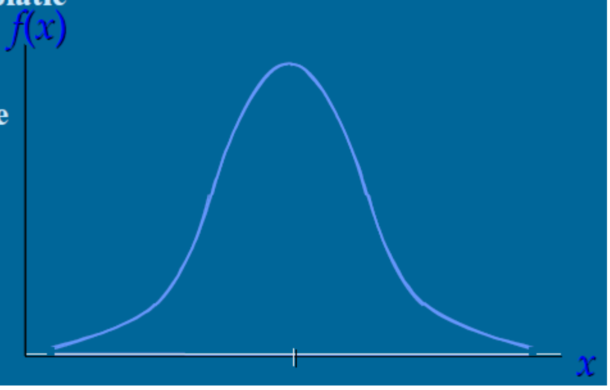
Une v.a.c X pouvant prendre toutes les valeurs réelles x dans l'intervalle de $-\infty$ à $+\infty$, pour $\mu \in \Re$, pour $\sigma \in \Re^+$ dont la fonction de densité est :

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

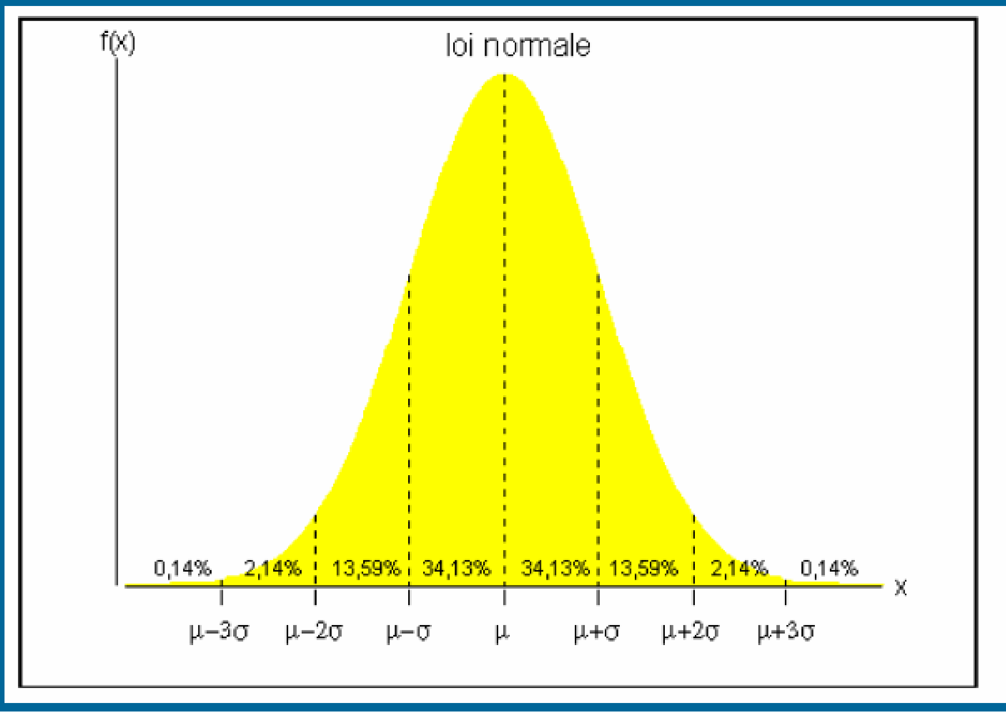
s'appelle une v.a. normale de paramètres μ et σ^2: $X \sim N(\mu, \sigma^2)$

Forme de la distribution normale

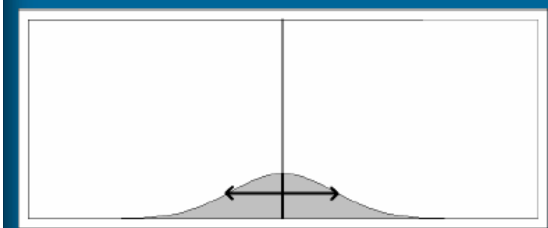
- Il existe une famille entière de lois normales. Elles se différencient par leur moyenne et leur variance
- Courbe en cloche
- Courbe symétrique
- La moyenne, le mode et la médiane correspondent au même point (le point le plus élevé)
- L'écart type détermine la largeur de la courbe, plus il est grand, plus la courbe sera large et aplatie
- L'aire totale sous la courbe est 1
- Aussi appelée loi Gaussienne ou loi de Gauss



Représentation graphique

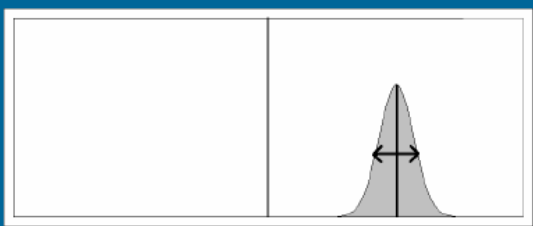


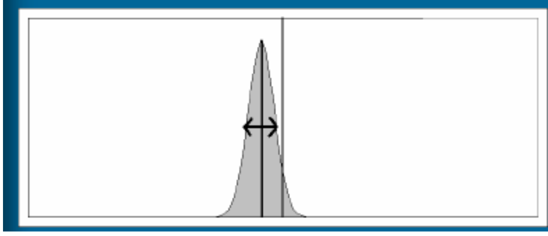
Exemples de lois normales



- Moyenne : 0
- Écart type : 3

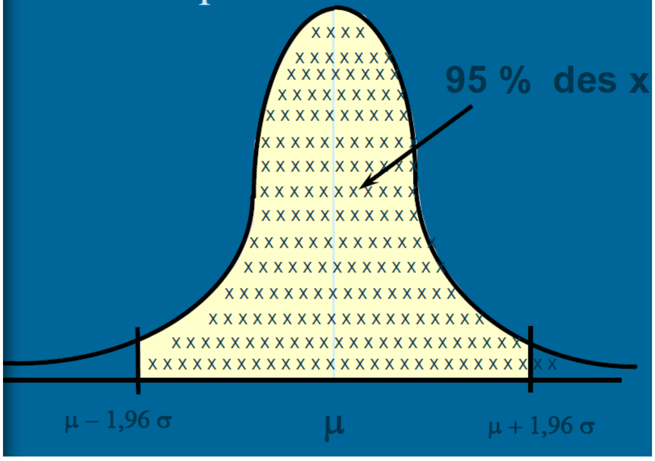
- Moyenne : 4
- Écart type : 1





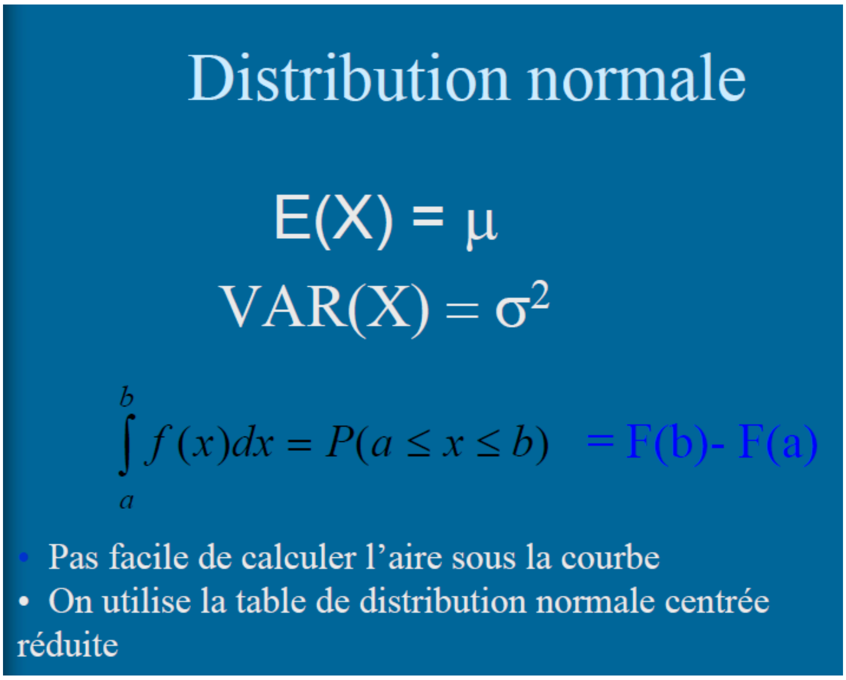
- Moyenne : -1
- Écart type : 0,5

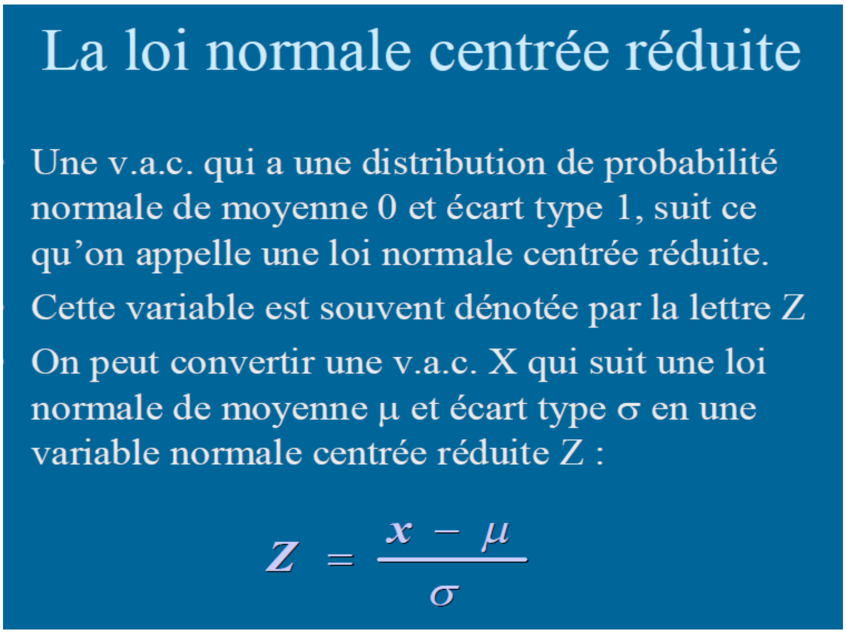
Cas particulier très utile



Distribution normale

- 68,26% des valeurs d'une variable aléatoire normale sont comprises dans l'intervalle [μ−σ;μ+σ]
- 95,44% des valeurs d'une variable aléatoire normale sont comprises dans l'intervalle [μ−2σ;μ+2σ]
- 99,72% des valeurs d'une variable aléatoire normale sont comprises dans l'intervalle

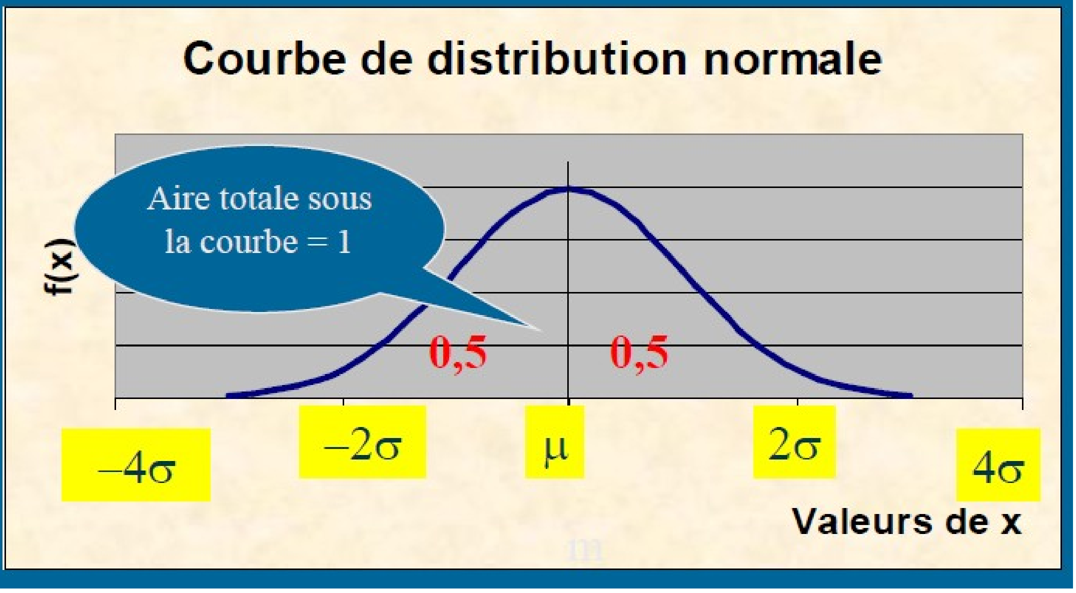




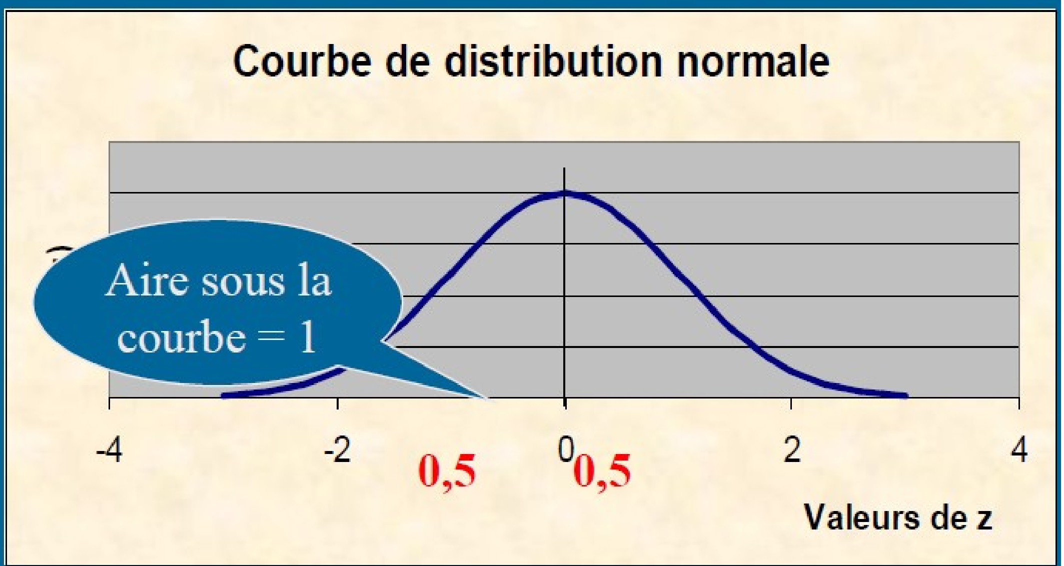
La loi normale centrée réduite

Étant donné une valeur z, nous utilisons la table normale centrée réduite pour trouver la probabilité (l'aire sous la courbe) qui lui est associée.

La loi normale



La loi normale centrée réduite



$$P(X<x)=P(Z<z) \text{ où } z=\frac{x-\mu}{\sigma} \text{ et } Z=\frac{X-\mu}{\sigma}$$

Exemple

Un marchand vend de l’huile à moteur. Lorsque l’inventaire de l’huile à moteur descend à 20 gallons, une commande est effectuée auprès du fournisseur. Le gérant du magasin a remarqué qu’il perdait des ventes lorsqu’il était en attente du nouveau stock. Il avait déterminé que la demande lorsqu’on est en attente du nouveau stock suit une loi normale de moyenne de 15 gallons et un écart type de 6 gallons. Le gérant aimerait savoir quelle est la probabilité d’une rupture de stock?

Exemple:

X est la variable aléatoire qui représente la demande en gallons d’huile à moteur en période d’attente de stock

– On cherche $P(X > 20)$

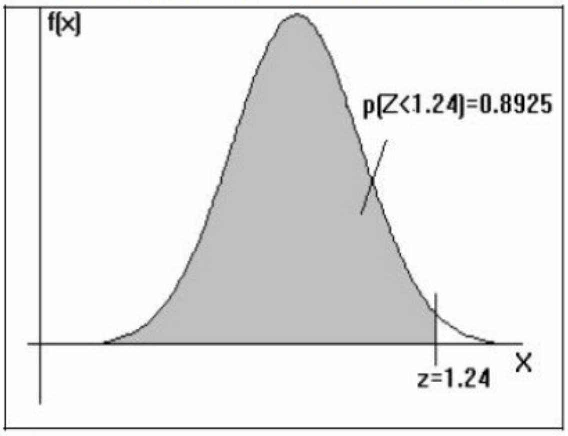
Définissons: $Z = (X - \mu)/\sigma$

– $P(X > 20) = P(Z > (20 - 15)/6) = P(Z > 0,83)$

– On cherche donc: $P(Z > 0,83)$

Utilisation de la table numérique

Lecture de la table: Pour z=1.24 (intersection de la ligne 1.2 et de la colonne 0.04), on a la proportion $P(Z < 1,24) = 0.8925$



$P(Z > 1,96) = 0,025$
$P(Z > 2,58) = 0,005$
$P(Z > 3,29) = 0,0005$

Rappels:

1/ $P(Z > z) = 1 - P(Z < z)$ et 2/ $P(Z < -z) = P(Z > z)$

Exemple: Sachant $P(Z < 1,24) = 0,8925$, on en déduit:

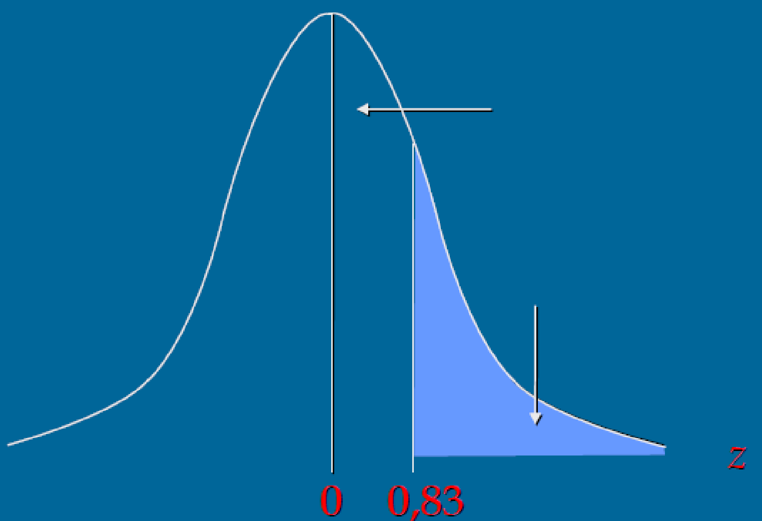
1/ $(P(Z > 1,24) = 1 - P(Z < 1,24) = 1 - 0,8925 = 0,1075$

2/ $P(Z < -1,24) = P(Z > 1,24) = 0,1075$

Un extrait de la table numérique

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	50000	50399	50798	51197	51595	51994	52392	52790	53188	53586
0,1	53983	54380	54776	55172	55567	55962	56356	56749	57142	57535
0,2	57926	58317	58706	59095	59483	59871	60257	60642	61026	61409
0,3	61791	62172	62552	62930	63307	63683	64058	64431	64803	65173
0,4	65542	65910	66276	66640	67003	67364	67724	68082	68439	68793
0,5	69146	69497	69847	70194	70540	70884	71226	71566	71904	72240
0,6	72575	72907	73237	73565	73891	74215	74537	74857	75175	75490
0,7	75804	76115	76424	76730	77035	77337	77637	77935	78230	78524
0,8	78814	79103	79389	79673	79955	80234	80511	80785	81057	81327
0,9	81594	81859	82121	82381	82639	82894	83147	83398	83646	83891
1,0	84134	84375	84614	84849	85083	85314	85543	85769	85993	86214
1,1	86433	86650	86864	87076	87286	87493	87698	87900	88100	88298
1,2	88493	88686	88877	89065	89251	89435	89617	89796	89973	90147
1,3	90320	90490	90658	90824	90988	91149	91309	91466	91621	91774
1,4	91924	92073	92220	92364	92507	92647	92785	92922	93056	93189
1,5	93319	93448	93574	93699	93822	93943	94062	94179	94295	94408
1,6	94520	94630	94738	94845	94950	95053	95154	95254	95352	95449

Exemple



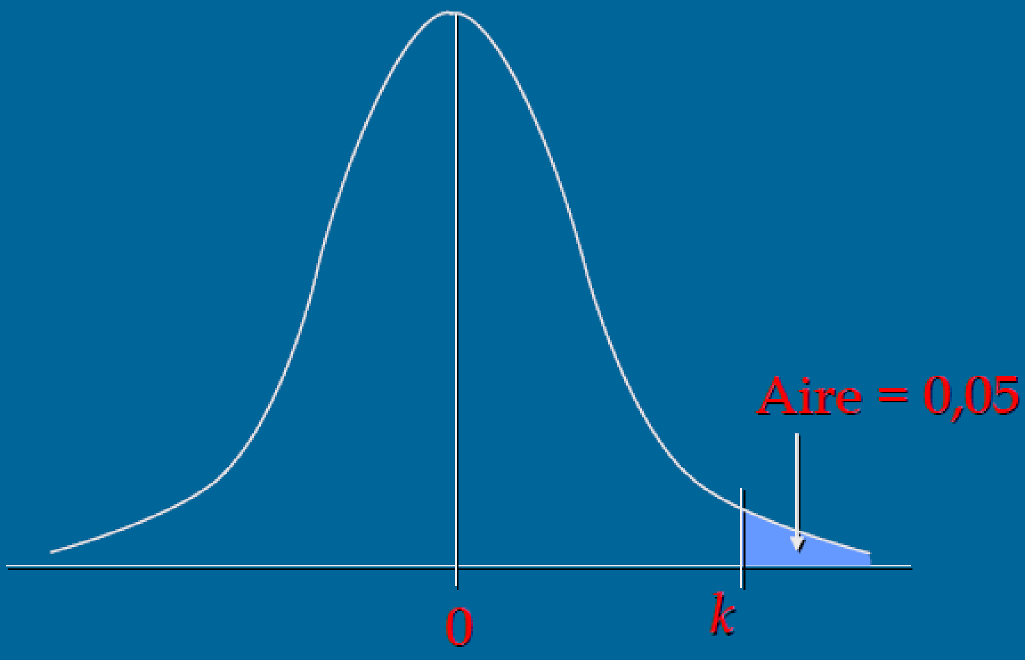
- La table montre une surface de 0,796 pour la région inférieure à $z = 0{,}83$.
- La région de l'extrémité en mauve correspond donc à 1-0.79673 = 0,203.
- La probabilité d'une rupture de stock est 0,203

Exemple

Si le gérant veut que la probabilité de rupture de stock ne dépasse pas 0.05, à quel niveau d'inventaire devrait-il passer une nouvelle commande?

Exemple: Pep Zone

On cherche k tel que $P(Z>k) \leq 0{,}05$



Exemple: Pep Zone

Dans la table, on trouve la valeur de z qui correspond à une probabilité de 0,9500 égal à 1.65.

La valeur correspondante de x est:

$$x = \mu + k\sigma$$
$$= 15 + 1{,}65*6$$
$$= 24{,}9$$

Une commande de 24,9 gallons rendra la probabilité d'une rupture de stock égale à 0,05.

Donc, le gérant devrait passer une nouvelle commande lorsque il lui reste 25 gallons

Exemple: Pep Zone

Dans la table, on trouve la valeur de z qui correspond à une probabilité de 0,9500 égal à 1.65.

La valeur correspondante de x est:

$$x = \mu + k\sigma$$
$$= 15 + 1{,}65*6$$
$$= 24{,}9$$

Une commande de 24,9 gallons rendra la probabilité d'une rupture de stock égale à 0,05.

Donc, le gérant devrait passer une nouvelle commande lorsque il lui reste 25 gallons

Exemple

Une machine fabrique des rondelles de métal dont le diamètre est distribué normalement avec une moyenne de 2,4 cm et un écart type de 0,05 cm. Des rondelles produites par cette machine, trouvons la proportion de celles dont le diamètre:

i) excède 2,5 cm; Rép. 0,0228

ii) n'excède pas 2,32 cm; Rép. 0,0548

iii) est compris entre 2,35 cm et 2,46 cm. Rép. 0,7262

Trouvons la valeur de x telle que 5 % des rondelles présentent un diamètre qui lui est supérieur. Rép. 2,48225

Théorème central-limite

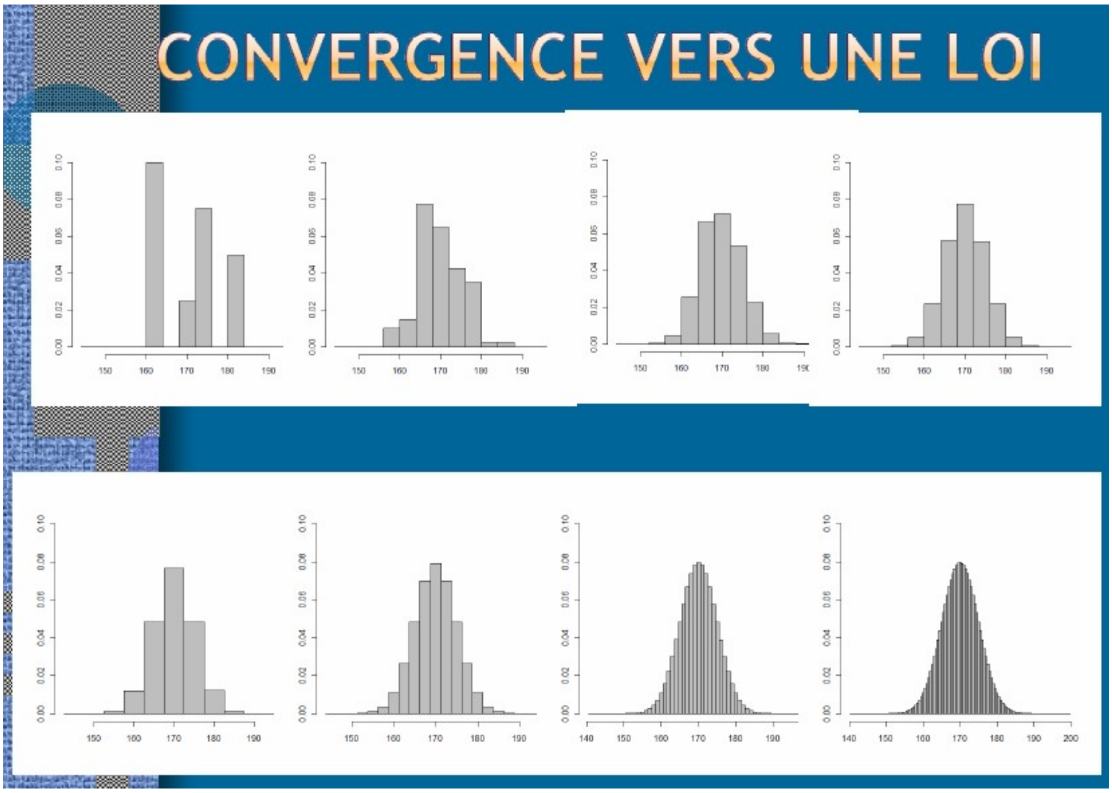
La loi Normale est une « loi limite » pour toutes les lois.

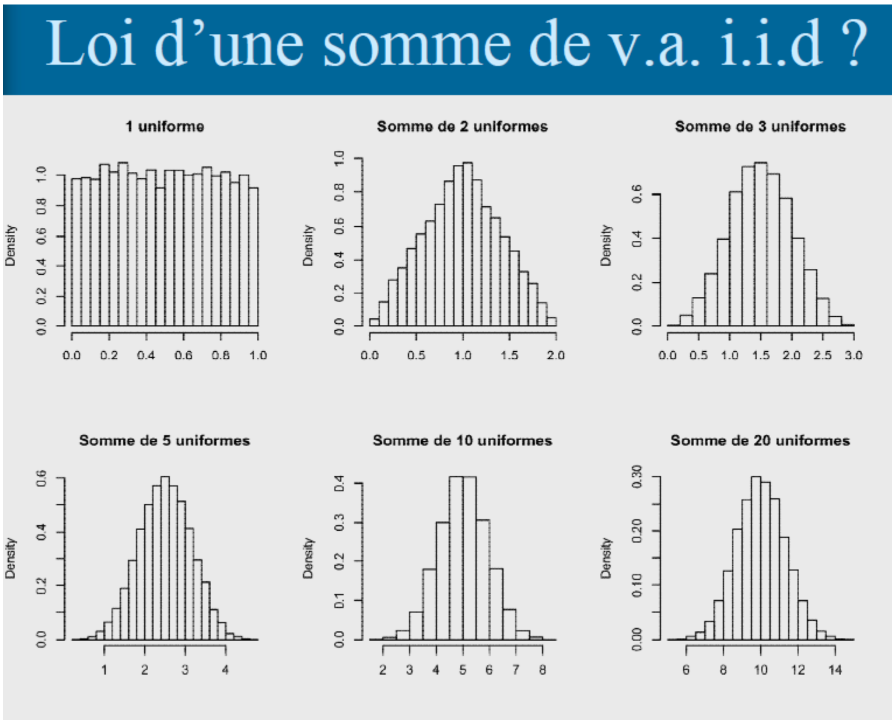
• Soient X1,X2,...,Xn, n variables aléatoires indépendantes et de même type de loi.

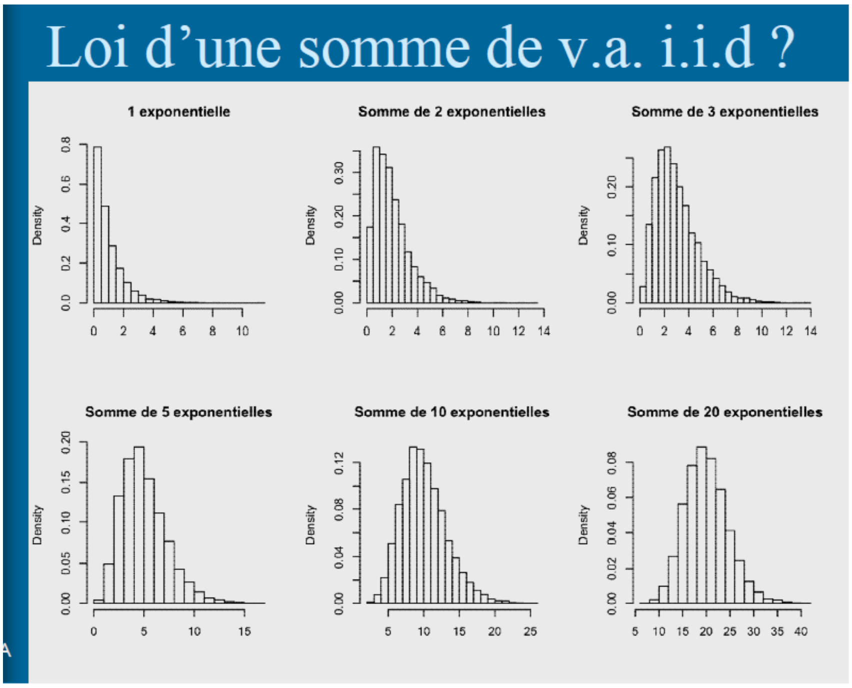
Si n est assez grand, la somme Y = X1+...+Xn suit approximativement une loi normale de paramètres

–moyenne = somme des moyennes des Xi

–variance = somme des variances des Xi







Approximations

- Approximation normale de la loi binomiale

On a la loi binomiale B(n ; p) qui suit à peu prés la loi normale de moyenne np et de variance np(1-p) si n>30, np> 10 et np(1-p)>10.

- Approximation normale de la loi de Poisson

On a $P(m) = N(m ; m)$ si $m>20$