STATISTIQUES DESCRIPTIVES ET INFÉRENTIELLES

Résumé du cours (Lecture V) + correction des exercices

Mme Marwa Chalgham Abdennadher

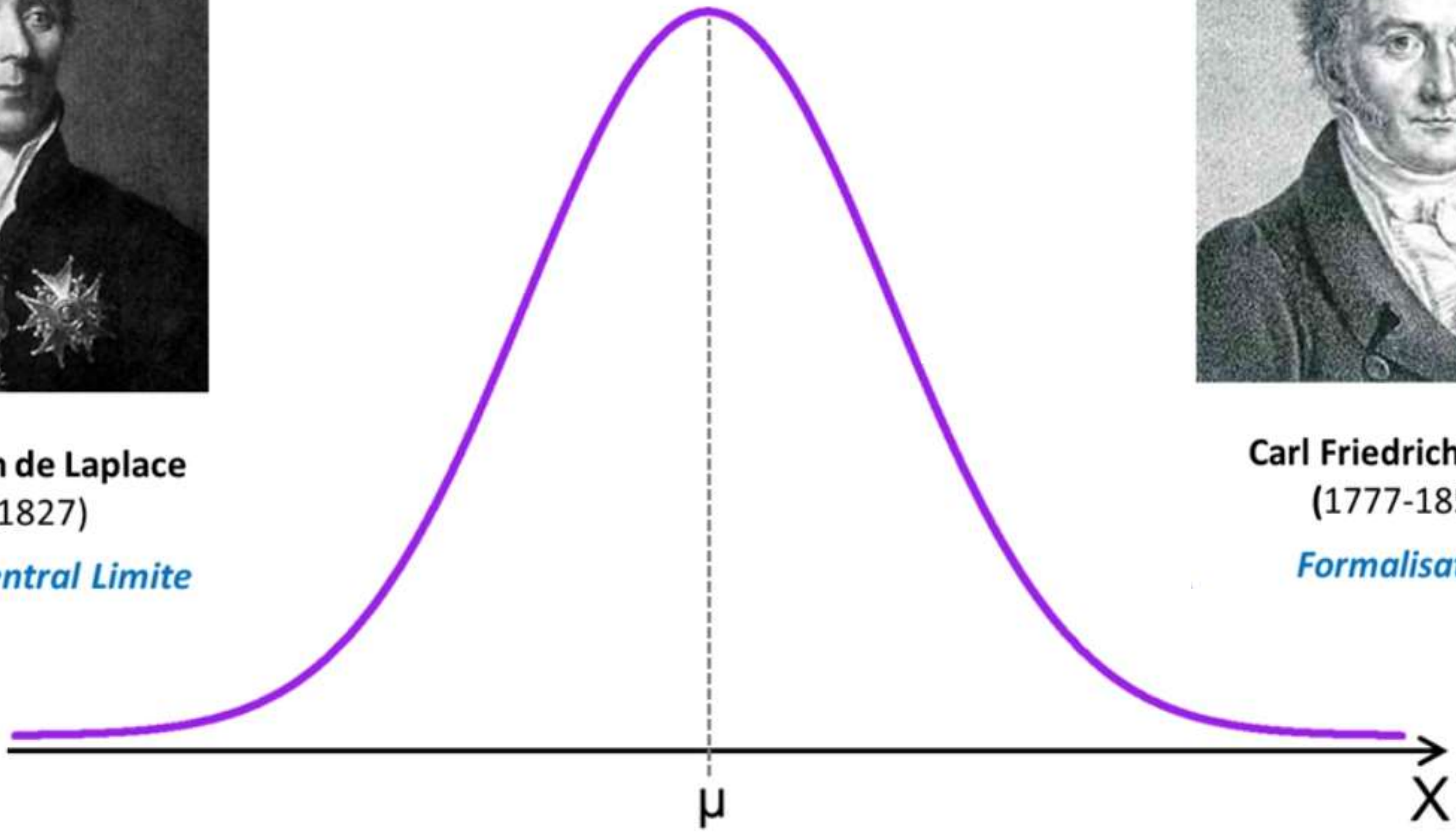
Auditoires: 2ème année Licence en Science de l'Informatique: Analyse des Données et Big Data (D- LSI ADBD)

Normal distribution and sampling distributions

La loi normale a été proposée par Laplace et reprise par Gauss. Aujourd’hui, la loi Normale est appelée aussi Distribution de Gauss puisque les travaux de ce dernier est plus récente.

Loi Normale

Pierre-Simon de Laplace
(1745-1827)
Théorème Central Limite

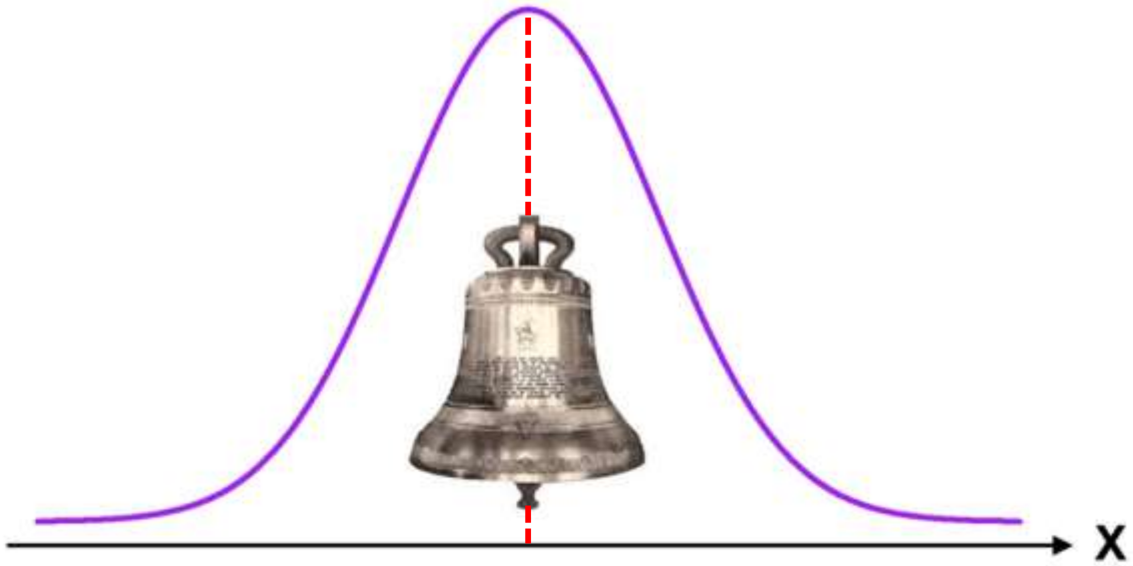


Carl Friedrich Gauss
(1777-1855)
Formalisation

Lorsqu'une ***grandeur (X)*** subie une influence d':

- Un grand nombre de facteurs
- Non tous identifiés
- Ils s'agit des facteurs indépendants
- Ils contribuent très faiblement à faire varier la grandeur étudiée

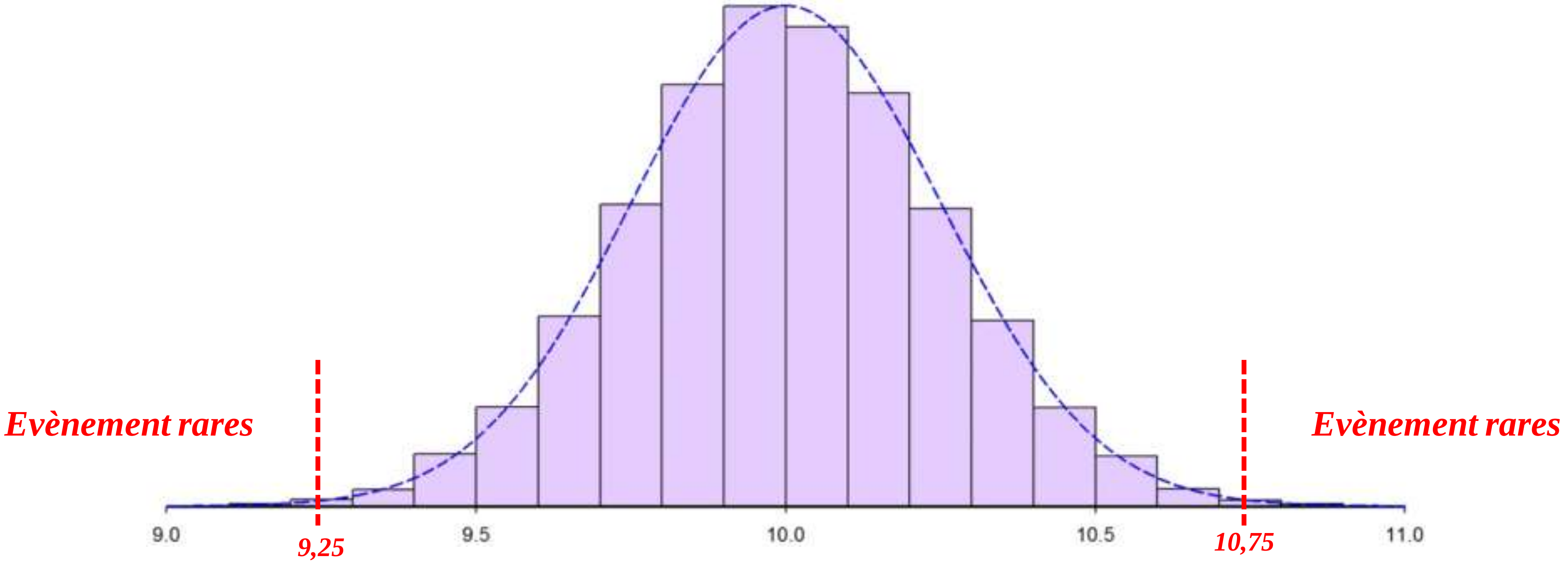
Alors, la valeur prise par la variable aléatoire continue associée se distribue d'une façon continue qui est symétrique avec une tendance centrale. La loi obtenue est une ***loi Normale*** qui a la forme d'une cloche et qui se caractérise par l'espérance et l'écart-type de la variable aléatoire contenue qu'elle décrit.



Loi Normale

Grandeur étudiée

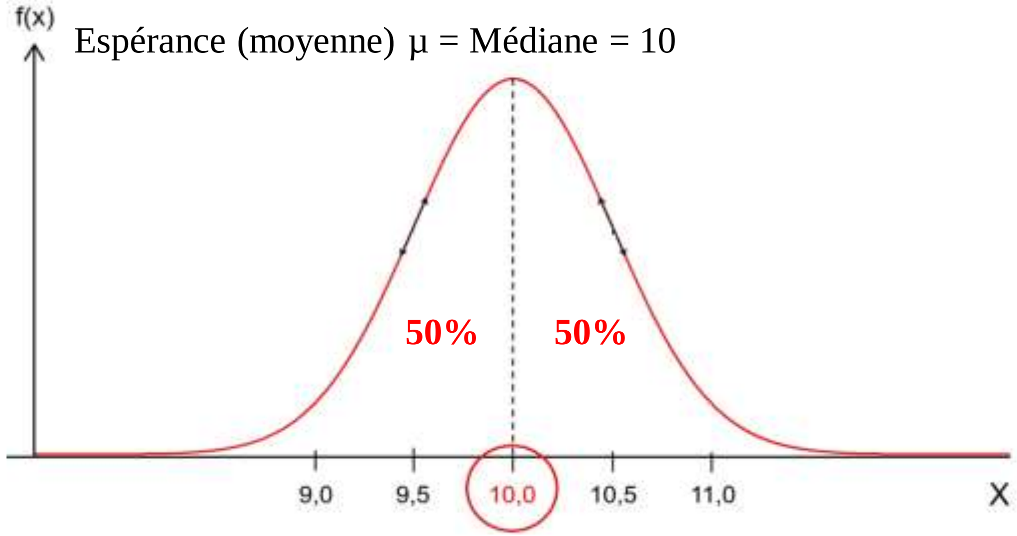
Exemple: Lors de la fabrication d'une pièce en plastique entrant dans la composition d'une appareil électroménager, si on s'intéresse au diamètre de cette pièce, on remarque qu'il dépend d'un grand nombre de facteurs (machine de fabrication utilisée, réglages de la machine, ouvriers intervenants, qualité du plastique, température, humidité, quantité de la matière, etc…). Ces facteurs sont indépendants, contribuant chacun faiblement, et n'interviennent pas en même temps. En conséquent, le diamètre d'une pièce fabriquée durant une année ***suit la loi Normale***. Admettant que le cahier des charges impose que le diamètre d'une pièce doit être ***comprise entre 9,25 et 10,75***. On remarque que la plus part des pièces fabriquées sont à la norme de la qualité exigée puisque il existe un nombre limité de pièces qui sont à l'extérieure des limites indiquées en rouge qui marquent la tolérance de fabrication:



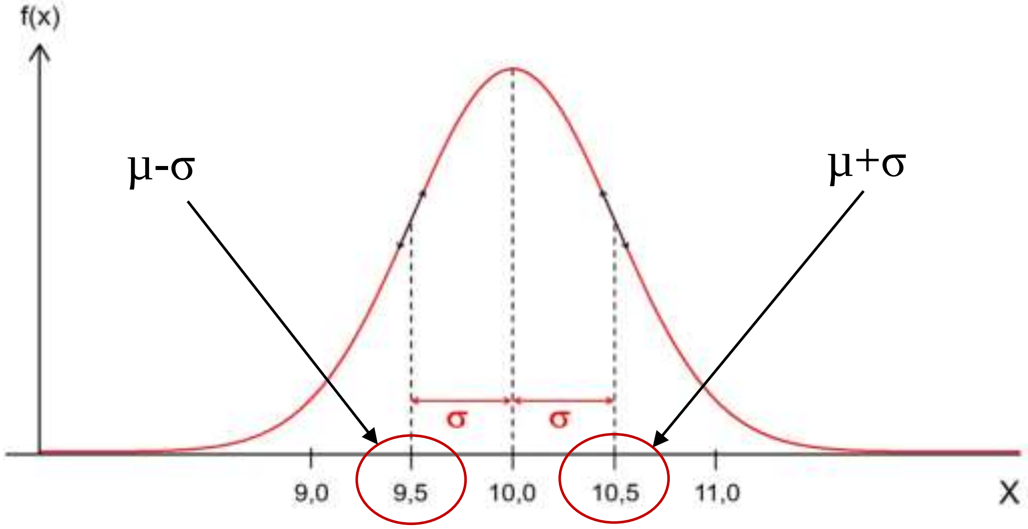
Fabrication en série d'une pièce par une machine suivie jour par jour

Normal distribution and sampling distributions

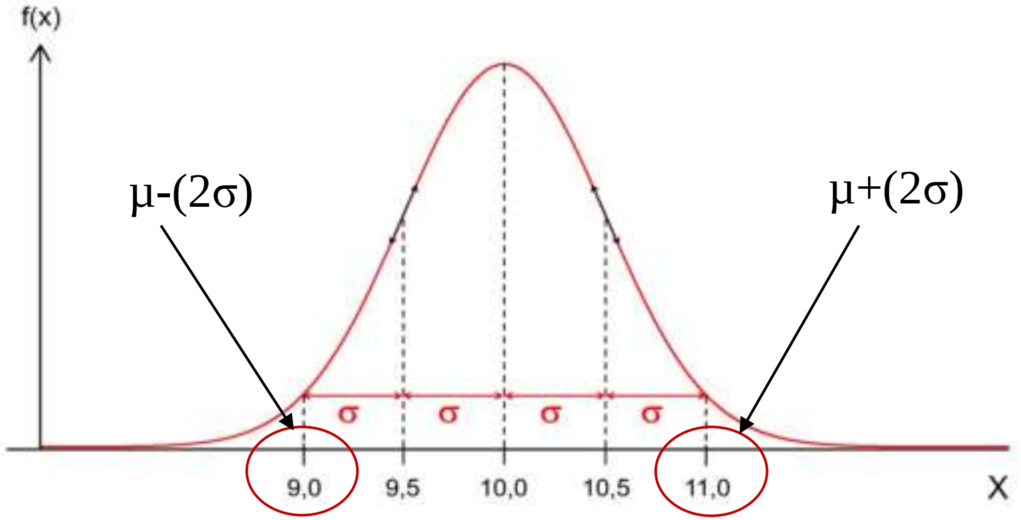
Admettant que le diamètre des pièces fabriquées en une année suivent une loi normale de moyenne μ = 10 mm, et d’écart-type σ = 0,5 mm.



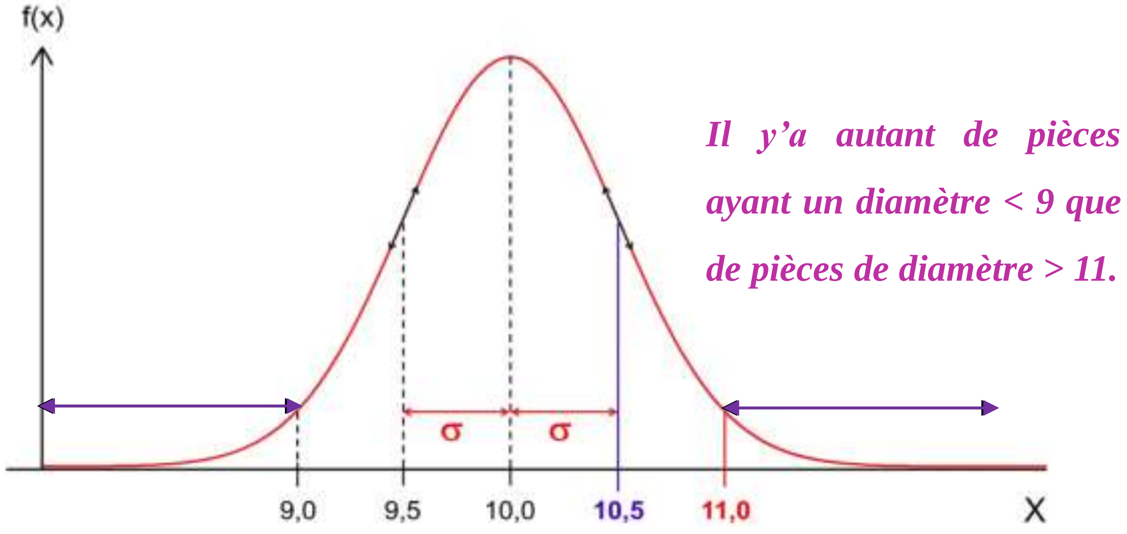
50% des diamètres sont < à 10 mm, et 50% sont > à 10 mm.



10,5 mm, et 9,5 mm sont des valeurs distantes d’un écart-type de la moyenne.

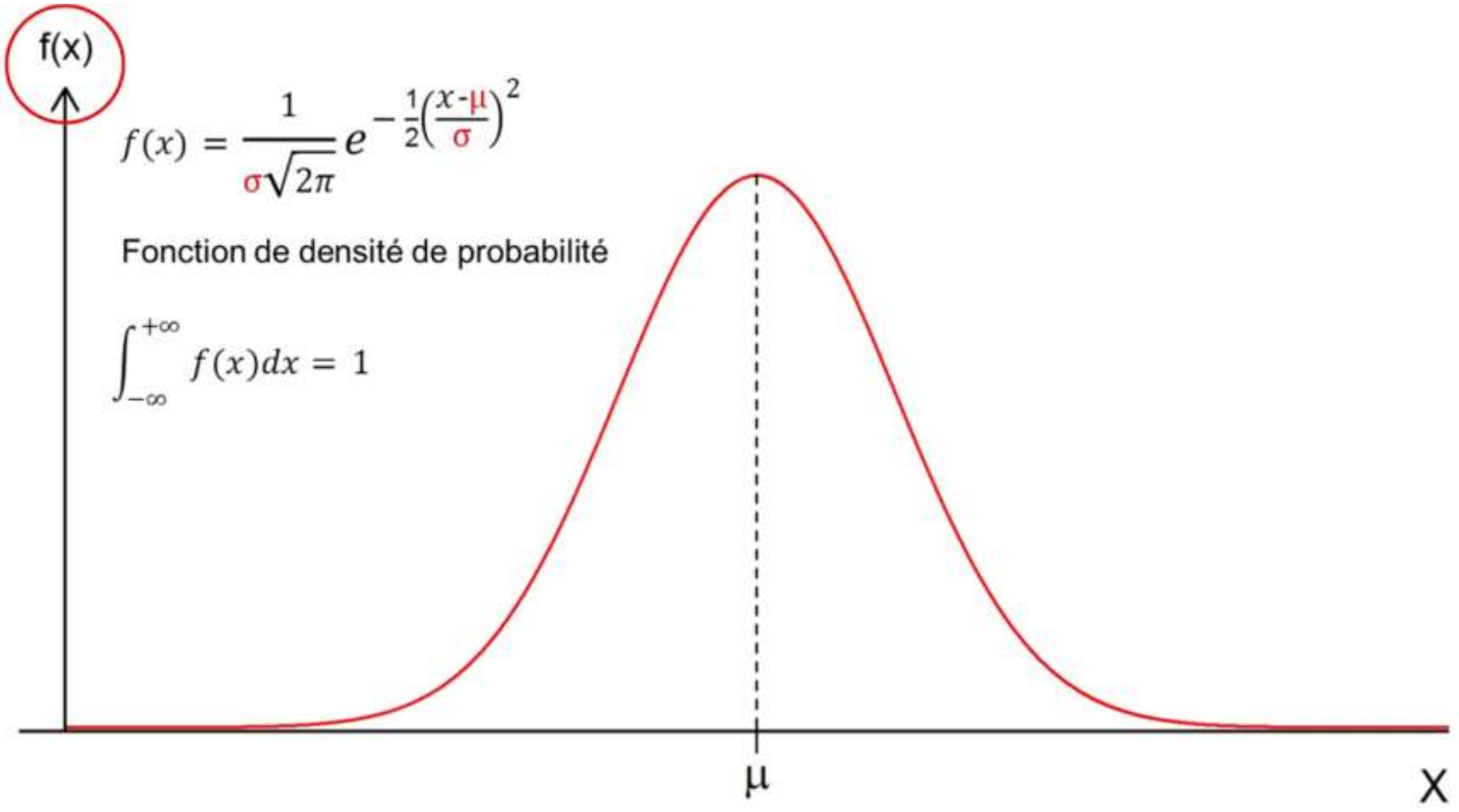


11 mm, et 9 mm sont des valeurs distantes de 2 écart-type de la moyenne.



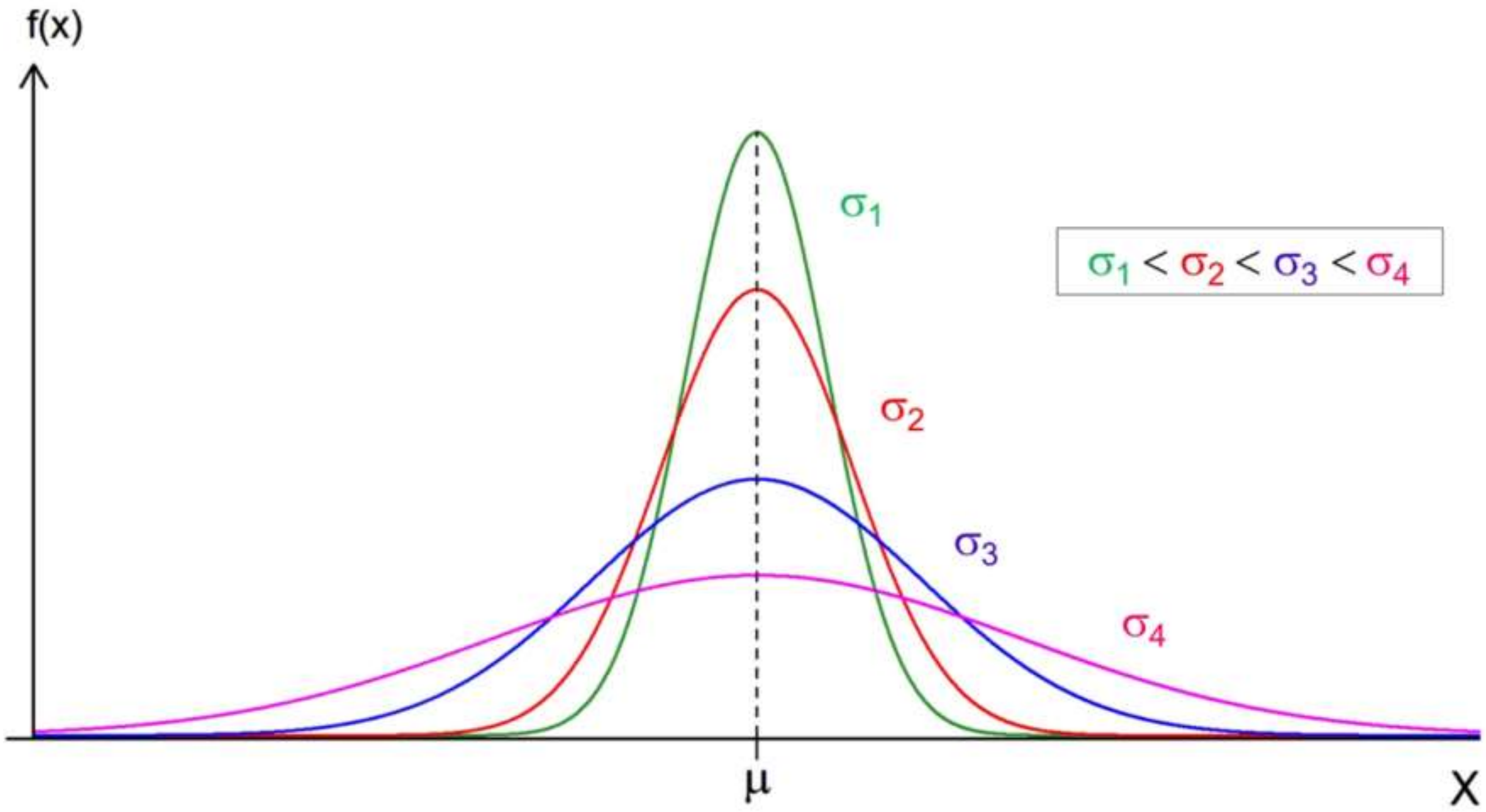
Il y’a autant de pièces ayant un diamètre entre 10,5 et 11 que entre 9 et 9,5mm.

La courbe d’une loi normale est définit par une fonction ***f(x)*** nommée la fonction de densité de probabilité dont les valeurs sont portées sur l’axe des ordonnées.

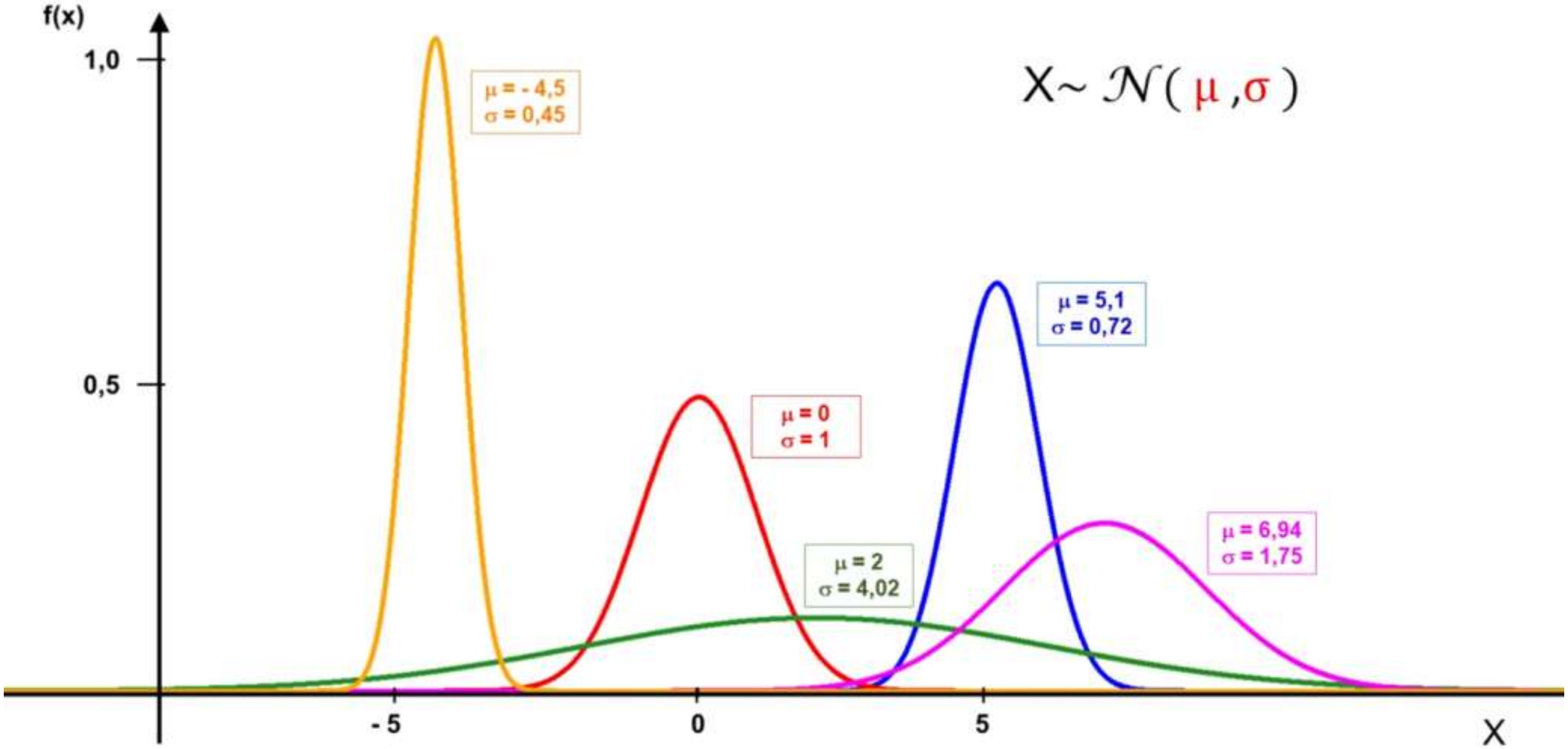


Plus l'écart-type est petit, plus la distribution des valeurs est resserrée par la moyenne c'est-à-dire plus ***la courbe sera haute et étroite***.

À l'inverse, une ***courbe basse et aplatie*** révèle des ***valeurs très dispersés (valeur d'écart-type plus grande)***.

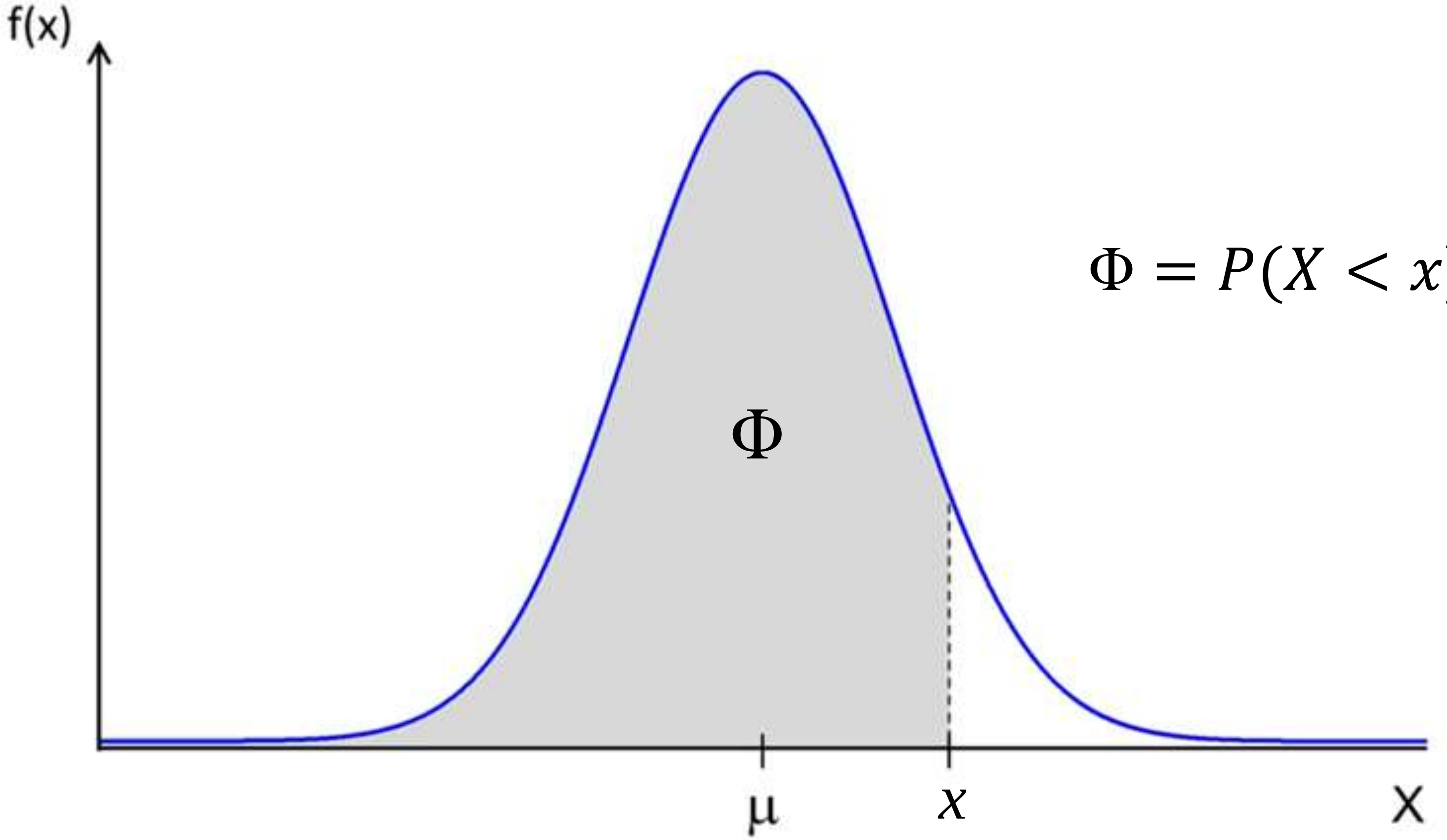


Il existe une infinité de loi Normale chacune correspondant à un couple (μ, σ) différent:



Grace aux propriétés de la loi Normale, on peut calculer aisément des probabilités:

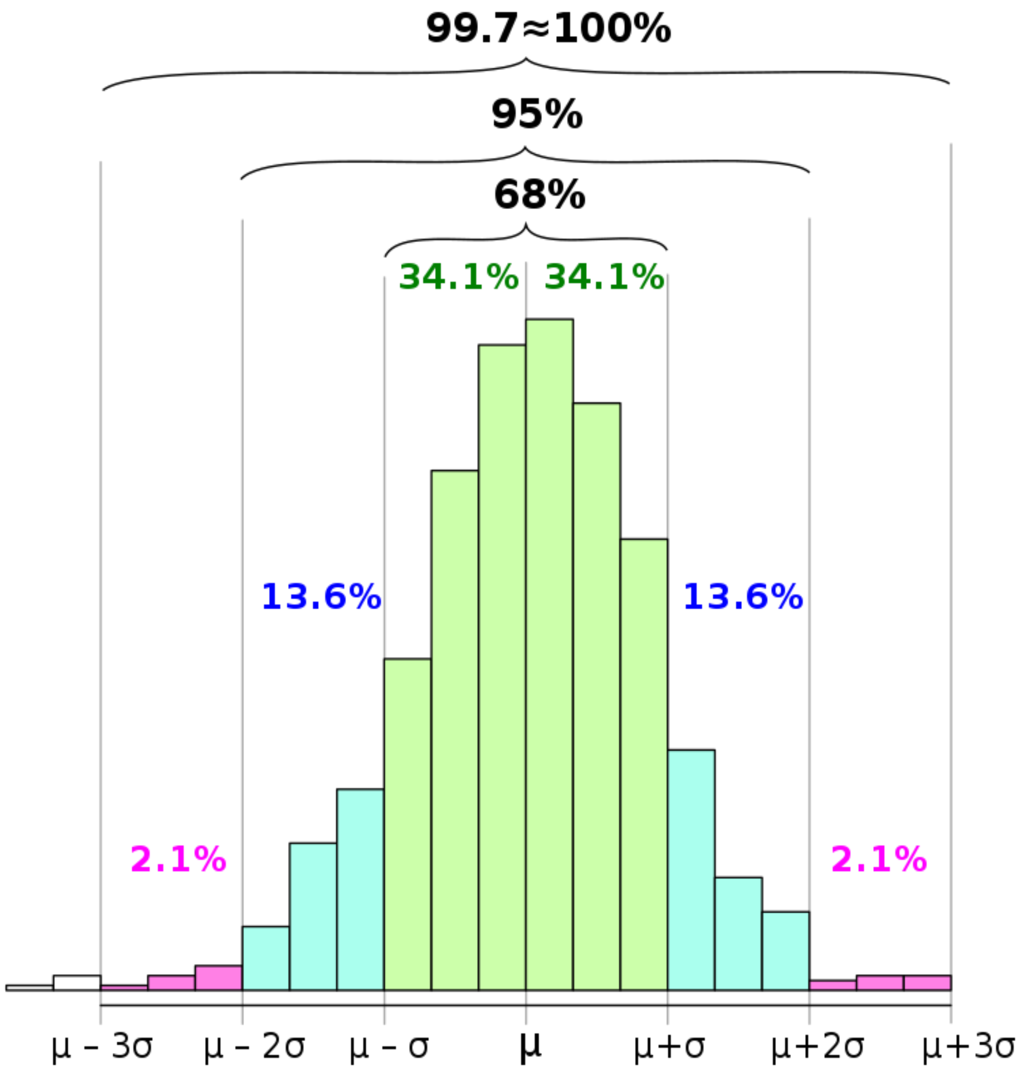
Notez que comme c'est le cas de toute distribution continue, une probabilité est associée à l'aire sous la courbe définissant la loi.



$$\Phi = P(X < x)$$

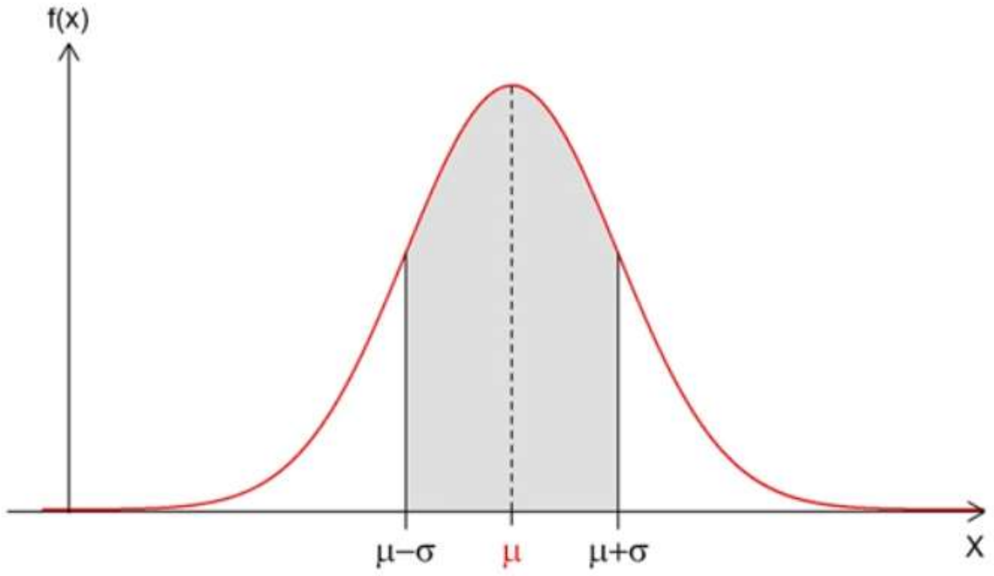
Toute courbe normale (indépendamment de sa moyenne ou de son écart type) est conforme à la règle empirique suivante:

- Environ ***68%*** des valeurs se situent à $\pm$ 1 écart-type de la moyenne.
- Environ ***95%*** des valeurs se situent à $\pm$ 2 écart-types de la moyenne.
- Environ ***99,7%*** des valeurs se situent à $\pm$ 3 écart-types de la moyenne.

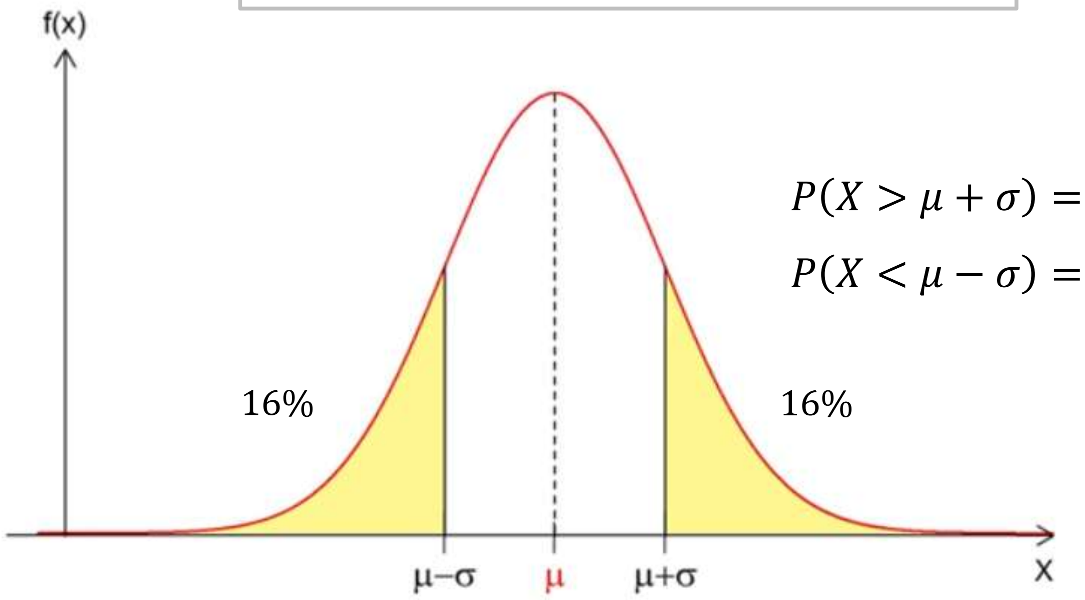


Propriétés et aires remarquables des lois normales:

$$P(\mu - \sigma < X < \mu + \sigma) = 68\%$$



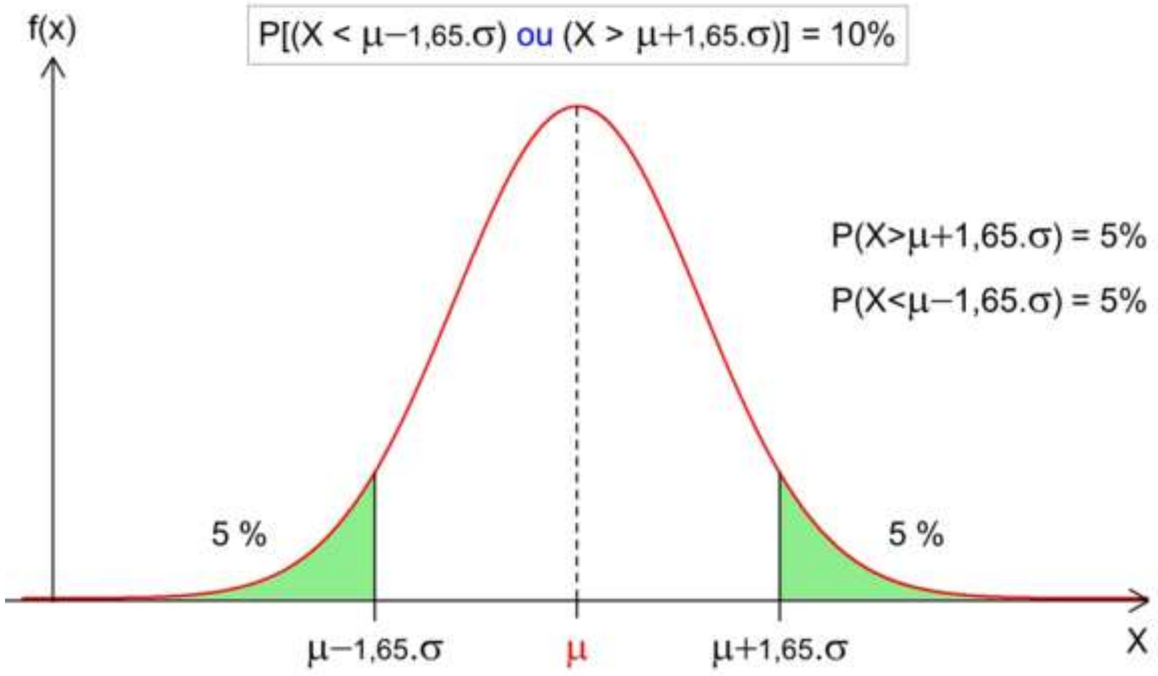
$$P[(X < \mu - \sigma) \ \mathbf{ou} \ (X > \mu + \sigma)] = 33\%$$

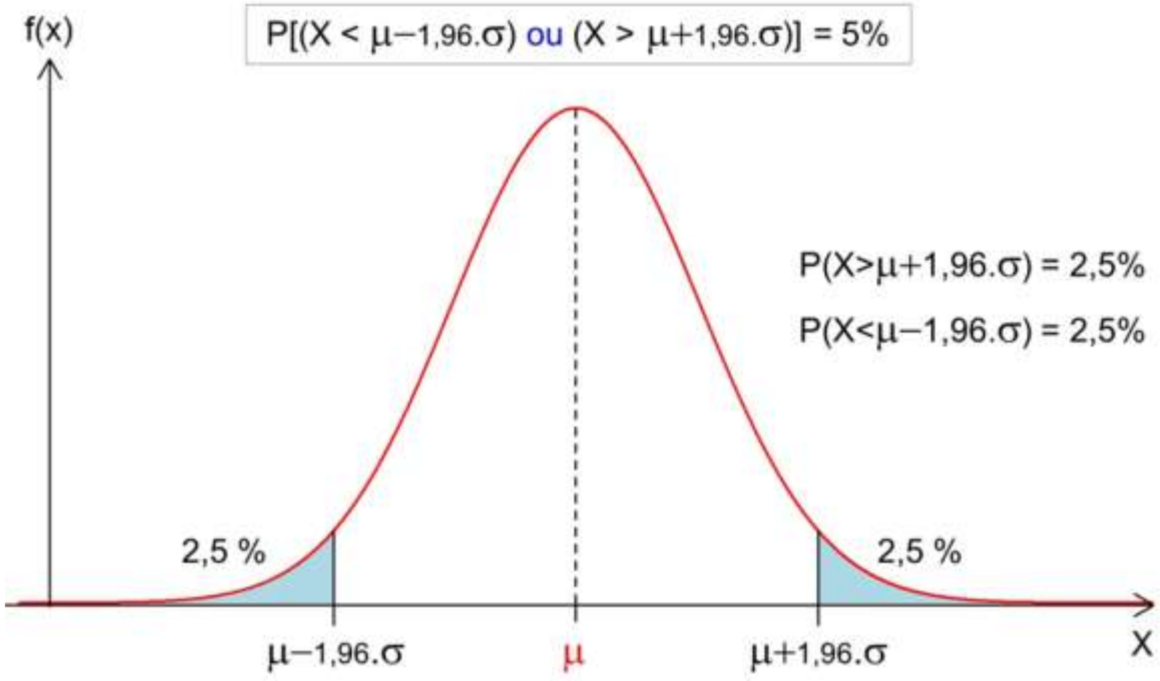


$$P(X > \mu + \sigma) = 16\%$$

$$P(X < \mu - \sigma) = 16\%$$

Propriétés et aires remarquables des lois normales:





La loi Normale centrée réduite:

On rappelle qu'une variable aléatoire est dite centrée si son espérance est nulle, et réduite si son écart-type est égal à 1. Alors, la loi ***Normale centrée réduite***, également connue sous le nom de loi ***Normale standard*** ou loi ***Normale unitaire***, est une distribution de probabilité continue qui a une ***moyenne de zéro et un écart type de un*** notée $\mathcal{N}(0;1)$. Elle est représentée par la fonction de densité de probabilité :

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

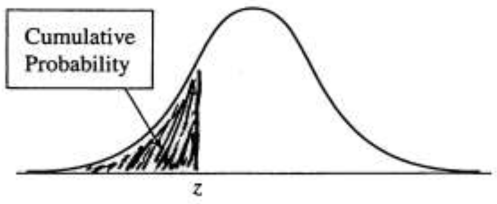
La loi Normale centrée réduite peut être utilisée pour standardiser des données en les convertissant en des valeurs ***Z-scores***, en appliquant la formule suivante:

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

L'idée ici est de convertir les valeurs de l'unité de mesure originale en une nouvelle unité appelé le Z-score qui correspond au nombre d'écarts types séparant un résultat de la moyenne. Autrement dit, les Z-scores représentent le nombre d'écarts-types qu'une observation se situe par rapport à la moyenne.

Tables de loi normale centrée réduite:

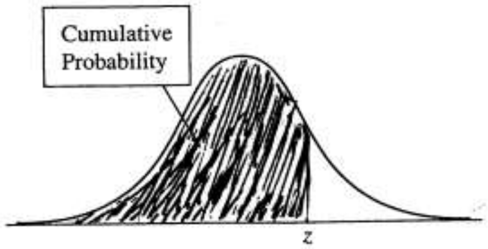
Il est important de savoir qu’il existe plusieurs types de tables de loi normale centrée réduite. La lecture de la probabilité varie d’une table à l’autre. Pour clarifier cela, les tables possèdent généralement un schéma explicatif au sommet de celle-ci. Il faut donc être attentif à la zone grisée qui représente l’aire de la probabilité considérée.



Cumulative probability for z is the area under the standard normal curve to the left of z

TABLE A Standard Normal Cumulative Probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014



Cumulative probability for z is the area under the standard normal curve to the left of z

TABLE A Standard Normal Cumulative Probabilities *(continued)*

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879

Normal distribution and sampling distributions

Exemple de lecture 1 :

2ème chiffre après la virgule de la valeur de Z

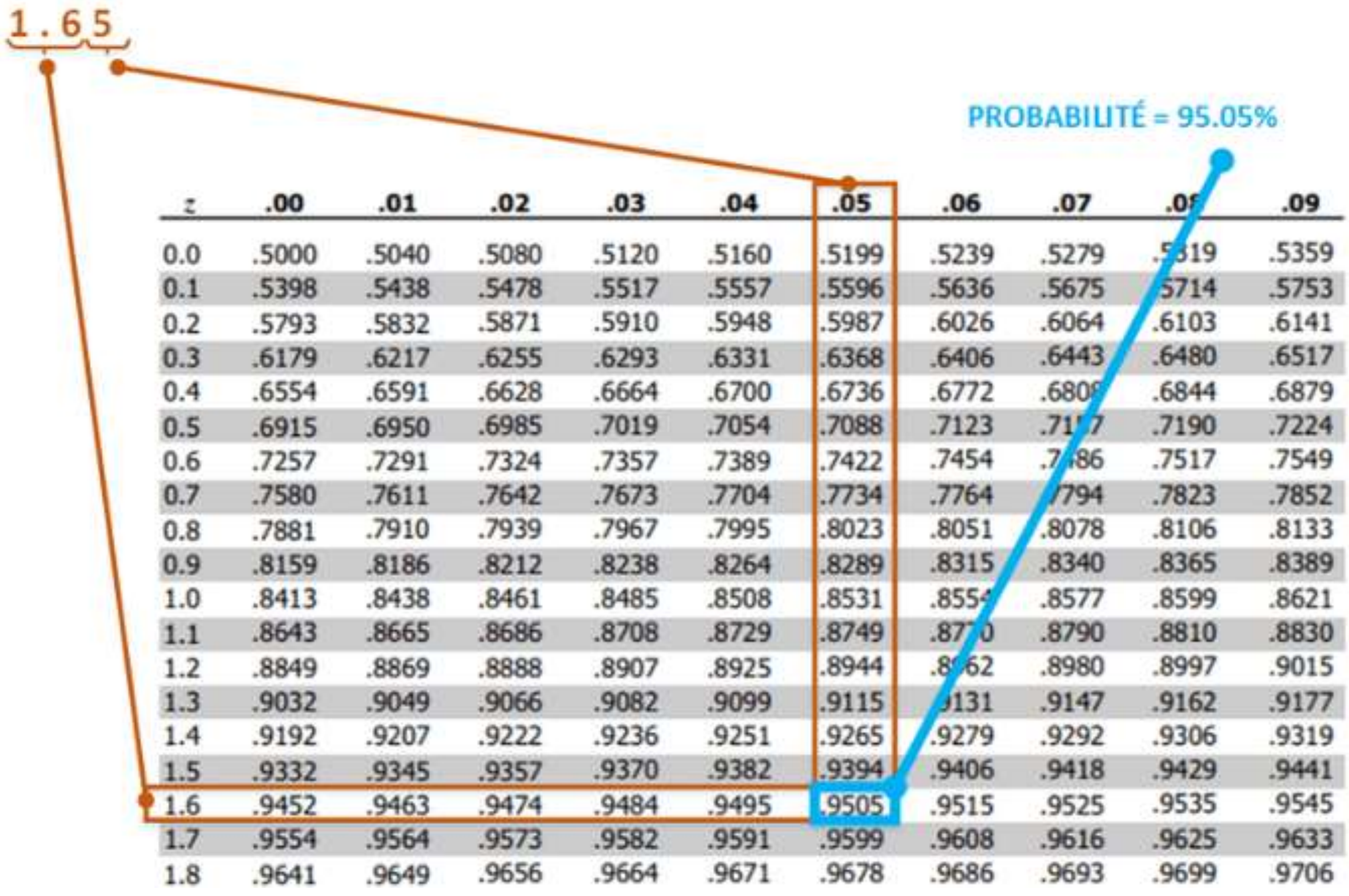
Pour un Z = -2 . 5 1

PROBABILITÉ = 0.6%

Valeur de Z

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	$\boxed{.0060}$	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143

Exemple de lecture 2 :

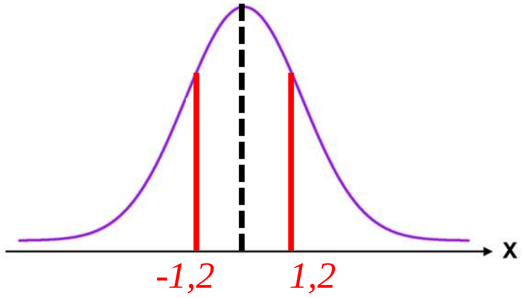


z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706

Application:

Z est une variable aléatoire qui suit la loi normale centrée réduite $\mathcal{N}(0;1)$.

1. On donne $P(Z \leq 1{,}2) \approx 0{,}885$. Déterminer $P(Z \geq -1{,}2)$.
2. Déduire $P(-1{,}2 \leq Z \leq 1{,}2)$.



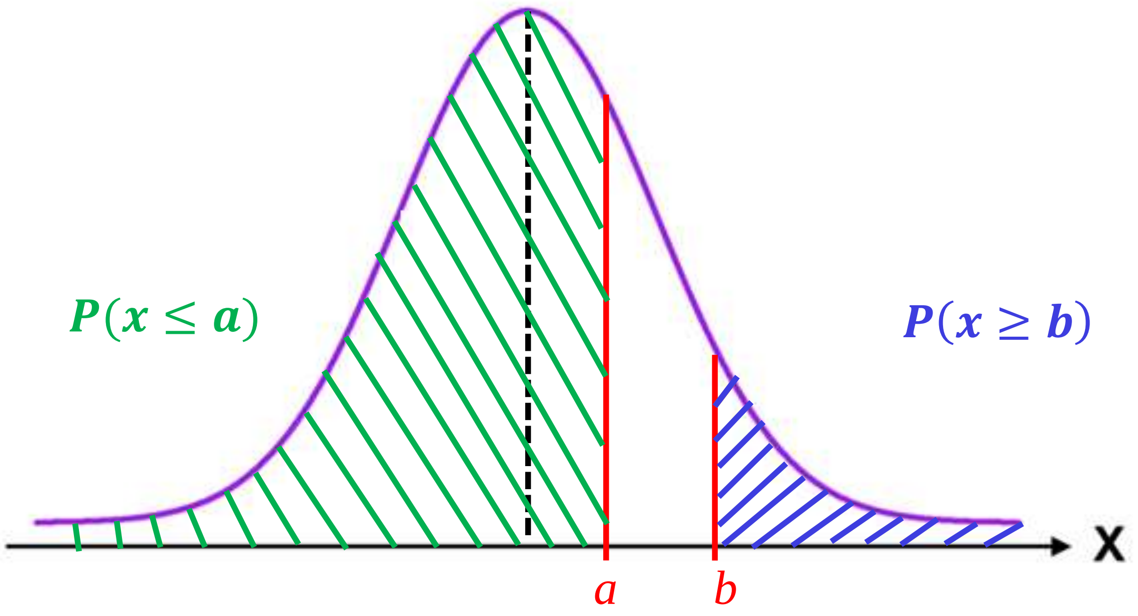
1. On a $P(Z \leq 1{,}2) \approx 0{,}885$.

   Alors par symétrie $P(Z \geq -1{,}2) = P(Z \leq 1{,}2) \approx 0{,}885$

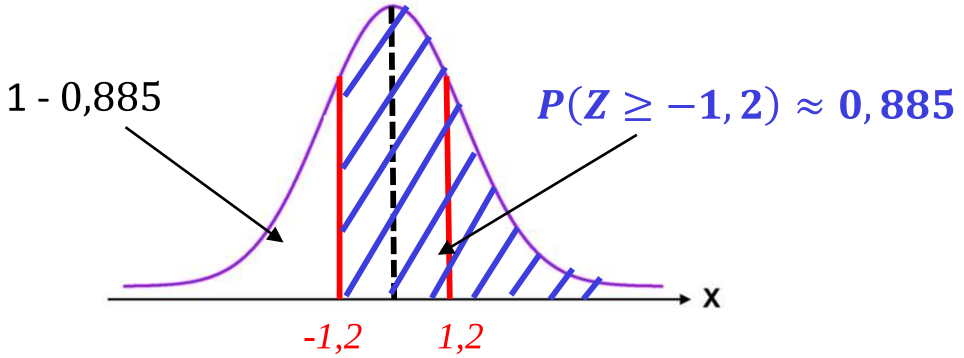
2. $P(-1{,}2 \leq Z \leq 1{,}2) = P(Z \leq 1{,}2) - P(Z \leq -1{,}2)$

$$= 0{,}885 - (1 - 0{,}885)$$

$$\approx 0{,}770$$



$$P(a \leq x \leq b) = P(x \leq b) - P(x \leq a)$$

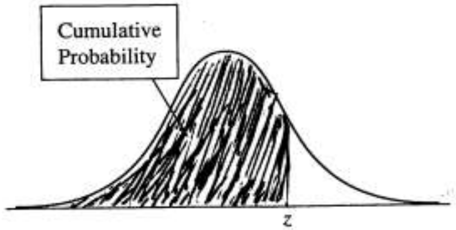


Application:

Reprenons la même application en se basant sur la table de loi normale centrée réduite:

Z est une variable aléatoire qui suit la loi normale centrée réduite $\mathcal{N}(0; 1)$.

1. On donne $P(Z \leq 1{,}2) \approx 0{,}885$. Déterminer $P(Z \geq -1{,}2)$.
2. Déduire $P(-1{,}2 \leq Z \leq 1{,}2)$.

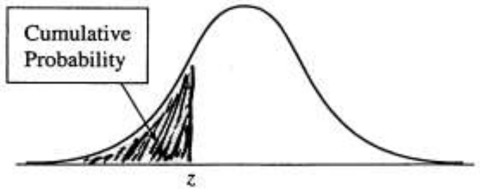


Cumulative probability for z is the area under the standard normal curve to the left of z

TABLE A Standard Normal Cumulative Probabilities (continued)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
...	...	...	...	...	...	...	...	...	...	...
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	$\boxed{.8849}$	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177

On peut vérifier d'après la table qu'on a $P(Z \leq 1{,}2) \approx 0{,}885$.



Cumulative probability for z is the area under the standard normal curve to the left of z

TABLE A Standard Normal Cumulative Probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
...	...	...	...	...	...	...	...	...	...	...
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	$\boxed{.1151}$	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170

D'après la table $P(Z \geq -1{,}2) = 1 - P(Z \leq -1{,}2)$

$$= 1 - 0{,}1151$$

$$= 0{,}8849$$

$$\approx 0{,}885$$

Application:

2. $P(-1{,}2 \leq Z \leq 1{,}2) = P(Z \leq 1{,}2) - P(Z \leq -1{,}2)$

$$= \Phi(1{,}2) - \Phi(-1{,}2)$$

$$= 0{,}8849 - 0{,}1151$$

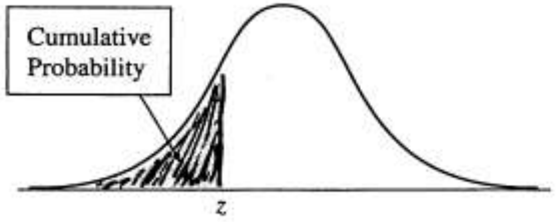
$$\approx 0{,}770$$

3. Déterminer $P(-1{,}2 \leq Z \leq -0{,}74)$.

$$= P(Z \leq -0{,}74) - P(Z \leq -1{,}2)$$

$$= 0{,}2296 - 0{,}1151$$

$$= 0{,}1145$$



Cumulative probability for z is the area under the standard normal curve to the left of z

TABLE A Standard Normal Cumulative Probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
...	...	...	...	...	...	...	...	...	...	...
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451

Souvent, les phénomènes du monde réel suivent une distribution normale (ou quasi-normale). Cela permet aux chercheurs d'utiliser la distribution normale comme modèle pour évaluer les probabilités associées aux phénomènes réels. En général, l'analyse implique deux étapes:

1. ***Transformer les données brutes:***

   Habituellement, les données brutes ne sont pas sous forme de z-scores. Elles doivent être transformées en z-scores, en utilisant l'équation de transformation présentée précédemment : $z = (X - \mu)/\sigma$

2. ***Trouver la probabilité:***

   Une fois les données transformées en z-scores, vous pouvez utiliser des tables de distribution normale standard, ou des calculatrices en ligne.

Exercise 1:

Molly earned a score of 940 on a national achievement test. The mean test score was 850 with a standard deviation of 100. What proportion of students had a higher score than Molly? (Assume that test scores are normally distributed.)

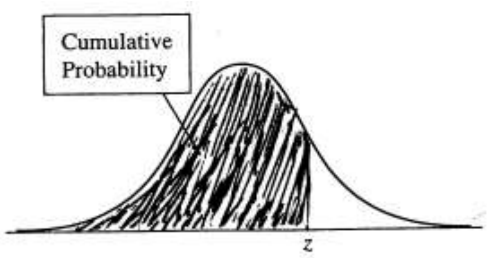
1. ***Transformer les données brutes:***

$$Z = \frac{X-\mu}{\sigma} = \frac{940 - 850}{100} = 0{,}9$$

2. ***Trouver la probabilité:***

$$P(Z \geq 0{,}9) = 1 - P(Z \leq 0{,}9)$$
$$= 1 - \Phi(0{,}9)$$
$$= 1 - 0{,}8159$$
$$= 0{,}1841$$

18,41% of students had a higher score than Molly



Cumulative probability for z is the area under the standard normal curve to the left of z

TABLE A Standard Normal Cumulative Probabilities *(continued)*

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
…	…	…	…	…	…	…	…	…	…	…
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	$\boxed{.8159}$	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621

Normal distribution and sampling distributions

<u>Exercise 2</u>:

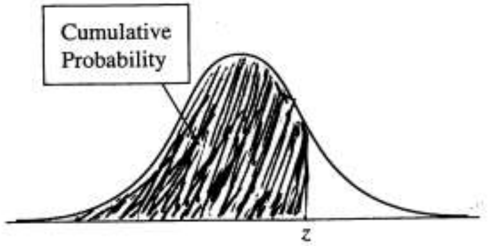
An average light bulb manufactured by the ACME Corporation lasts 300 days with a standard deviation of 50 days. Assuming that bulb life is normally distributed, what is the probability that an ACME light bulb will last at least 365 days?

1. ***Transformer les données brutes:***

$$Z = \frac{X - \mu}{\sigma} = \frac{365 - 300}{50} = 1{,}3$$

2. ***Trouver la probabilité:***

$$P(Z \geq 1{,}3) = 1 - P(Z \leq 1{,}3)$$
$$= 1 - \Phi(1{,}3)$$
$$= 1 - 0{,}9032$$
$$= 0{,}0968$$



Cumulative probability for *z* is the area under the standard normal curve to the left of *z*

TABLE A Standard Normal Cumulative Probabilities *(continued)*

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
...	...	...	...	...	...	...	...	...	...	...
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	$\boxed{.9032}$	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177

An ACME light bulb will last at least 365 days with a probability of 9,68%

Exercise 3:

Suppose scores on an IQ test are normally distributed. If the test has a mean of 100 and a standard deviation of 10, what is the probability that a person who takes the test will score between 90 and 110?

1. ***Transformer les données brutes:***

Pour X = 90: $Z = \frac{X-\mu}{\sigma} = \frac{90 - 100}{10}$ = -1

Pour X = 110: $Z = \frac{X-\mu}{\sigma} = \frac{110 - 100}{10}$ = 1

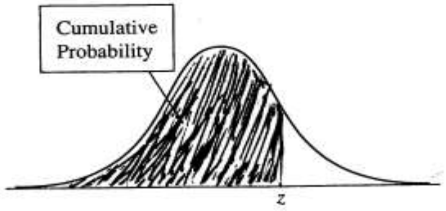
1. ***Trouver la probabilité:***

$$P(-1 \leq Z \leq 1) = P(Z \leq 1) - P(Z \leq -1)$$

$$= \Phi(1) - \Phi(-1)$$

$$= 0{,}8413 - 0{,}1587 = 0{,}6828$$

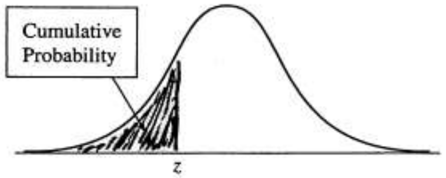
La probabilité qu'une personne qui passe le test obtienne un score compris entre 90 et 110 est de 68,26 %.



Cumulative probability for z is the area under the standard normal curve to the left of z

TABLE A Standard Normal Cumulative Probabilities *(continued)*

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
...	...	...	...	...	...	...	...	...	...	...
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	$\boxed{.8413}$	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621



Cumulative probability for z is the area under the standard normal curve to the left of z

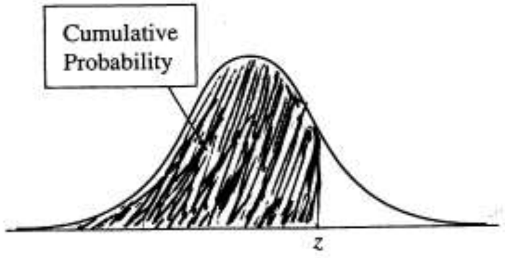
TABLE A Standard Normal Cumulative Probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
...	...	...	...	...	...	...	...	...	...	...
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	$\boxed{.1587}$	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611

Normal distribution and sampling distributions

La ***recherche de percentiles (inverser la distribution normale)*** consiste à trouver la valeur d’une variable aléatoire à un certain pourcentage d’une distribution normale. Pour ce faire, on utilise la fonction de distribution cumulative de la distribution normale standard pour trouver le score Z correspondant au pourcentage donné, puis on utilise la formule $\boldsymbol{x = \mu + z.\sigma}$ pour trouver la valeur de la variable aléatoire.

Par exemple on cherche la plus petite valeur d’un test de QI qui correspond aux 5% des personnes les plus intelligentes. Tout d’abord, on utilise la table de la distribution normale standard pour trouver le score Z correspondant à une probabilité de 95%, qui est de **1,645**. Cela signifie que la personne doit avoir un score de QI qui est au moins de 1,645 écart type au-dessus de la moyenne.



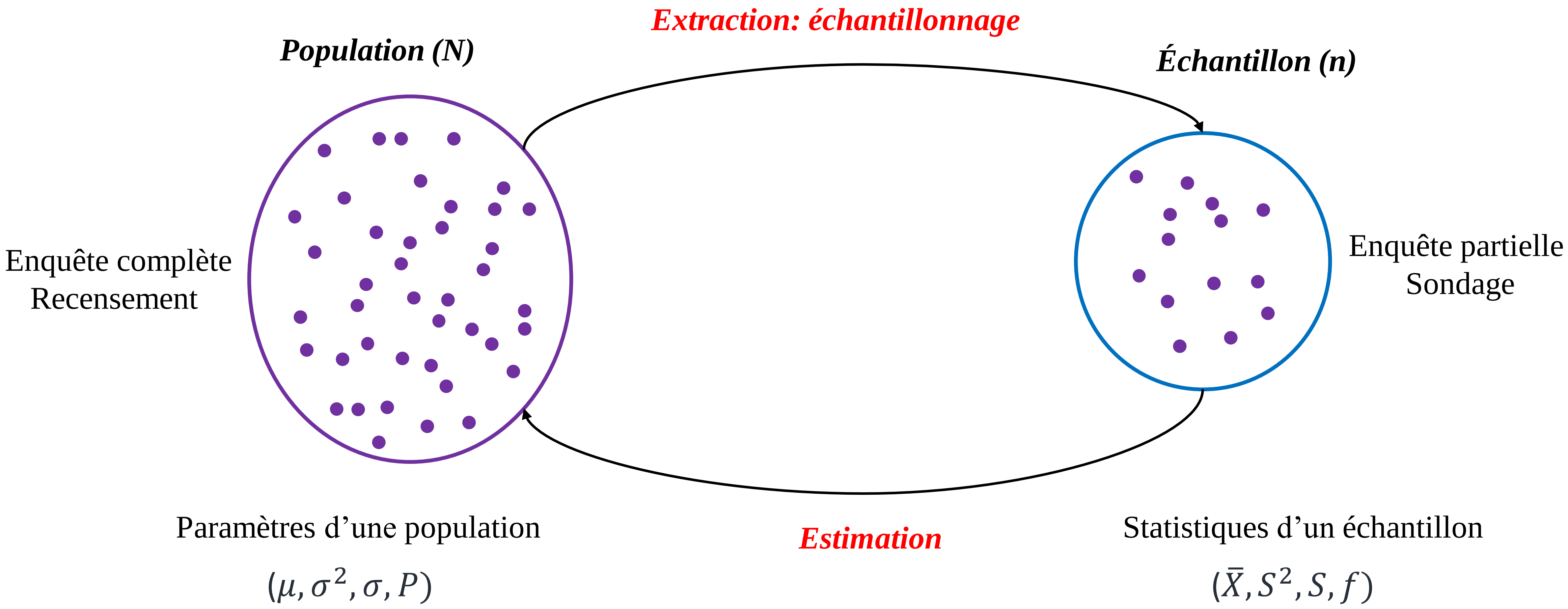
Cumulative probability for z is the area under the standard normal curve to the left of z

TABLE A Standard Normal Cumulative Probabilities *(continued)*

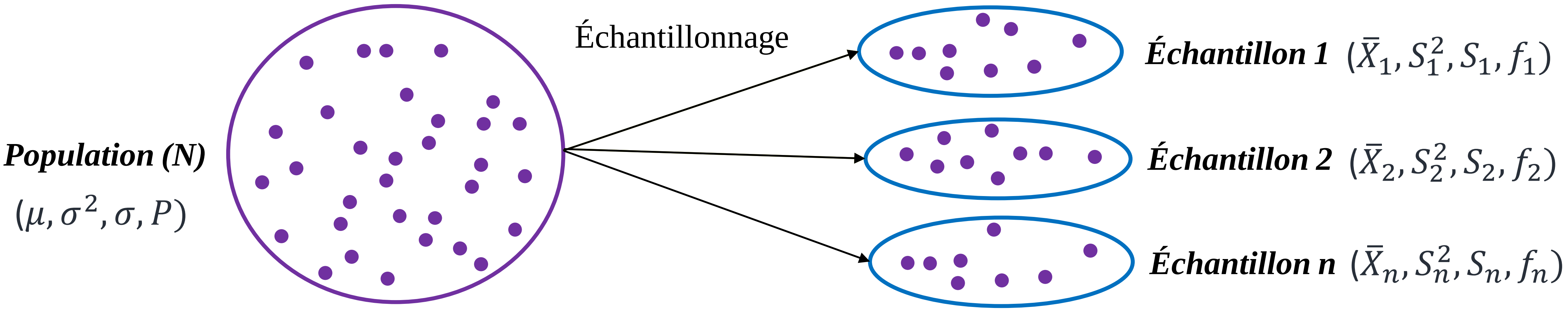
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
...	...	...	...	...	...	...	...	...	...	...
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633

Ensuite, on utilise la formule x = μ + z.σ pour trouver la valeur de la variable aléatoire, qui est le score de QI minimum correspondant à ce percentile. Dans cet exemple, la moyenne du test de QI est de 100 et l’écart type est de 10. Donc, le score de QI minimum correspondant aux 5% les plus intelligents est de : X = 100 + 1.645(10) = 164.5

Sampling distribution :Introduction



Sampling distribution



À retenir !	***Distribution d'échantillonnage de la moyenne***	***On dit que*** $\bar{X} \sim \mathcal{N}\left(\mu\,;\frac{\sigma}{\sqrt{n}}\right)$ ***lorsque l'une des deux conditions est vrai:*** 1. Lorsque la population à étudier suit elle-même la loi normale, 2. Lorsque la taille des échantillons est $n \geq 30$.
	Distribution d'échantillonnage de la proportion	***On dit que*** $F \sim \mathcal{N}\left(P\,;\sqrt{\frac{pq}{n}}\right)$ ***lorsque l'une des deux conditions est vrai:*** 1. Lorsque la population à étudier suit elle-même la loi normale, 2. Lorsque la taille des échantillons est $n \geq 30, ou\ np \geq 15, ou\ encore\ nq \geq 15$.

Exemple : ***(Distribution d'échantillonnage de la moyenne)***

Dans un certain collège, l'âge des étudiants obéit à une loi normale de moyenne 18,7 ans et de variance 13,3 ans. Si on interroge 50 étudiants de ce collège sur leur âge, quelle est la probabilité que l'âge moyen d'un tel échantillon se situe entre 18 et 19,4 ans?

Dans cet exercice on a $\overline{\boldsymbol{X}} \sim \boldsymbol{\mathcal{N}}\left(\boldsymbol{\mu}\,;\frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}\right)$ ***puisque*** la population à étudier suit elle-même la loi normale, et encore la taille des échantillons est $n \geq 30$. Alors $\overline{\boldsymbol{X}} \sim \boldsymbol{\mathcal{N}}\left(18{,}7\,;\frac{\sqrt{13{,}3}}{\sqrt{50}}\right) = \overline{\boldsymbol{X}} \sim \boldsymbol{\mathcal{N}}\left(18{,}7\,;\frac{3{,}646}{\sqrt{50}}\right) = \overline{\boldsymbol{X}} \sim \boldsymbol{\mathcal{N}}(18{,}7\,;0{,}51)$.

On a aussi $Z = \frac{\bar{X}-18{,}7}{0{,}51} \sim \boldsymbol{\mathcal{N}}(0\,;1)$

$$P(18 \leq \bar{X} \leq 19{,}4) = P\left(\frac{18-18{,}7}{0{,}51} \leq \frac{\bar{X}-18{,}7}{0{,}51} \leq \frac{19{,}4-18{,}7}{0{,}51}\right) = P(-1{,}37 \leq Z \leq 1{,}37)$$

Or, $\boldsymbol{P(-a \leq Z \leq a) = \Phi(\mathrm{a}) - \Phi(-\mathrm{a}) = 2\,P(X \leq a) - 1}$

Donc $P(-1{,}37 \leq Z \leq 1{,}37) = 2\,P(Z \leq 1{,}37) - 1 = 2 \times 0{,}9147 - 1 = 0{,}8294 = 82{,}94\%$

Exemple : ***(Distribution d'échantillonnage de la proportion)***

Selon une étude sur le comportement du consommateur, 25% d'entre eux sont influencés par la marque. Si on interroge 100 consommateurs pris au hasard, quelle est la probabilité qu'au moins 35% des consommateurs se déclarent influencés de la marque?

Dans cet exercice on a $\boldsymbol{F} \sim \boldsymbol{\mathcal{N}}\left(\boldsymbol{P}\,;\sqrt{\frac{\boldsymbol{pq}}{\boldsymbol{n}}}\right)$ *puisque* la population à étudier suit elle-même la loi normale, et encore la taille des échantillons est $n \geq 30$, en plus, $np = 100 \times 0{,}25 = 25 \geq 15$, *et encore* $nq = 100 \times (1 - 0{,}25) = 75 \geq 15$.

Alors $\boldsymbol{F} \sim \boldsymbol{\mathcal{N}}\left(\boldsymbol{P}\,;\sqrt{\frac{\boldsymbol{pq}}{\boldsymbol{n}}}\right) = \boldsymbol{F} \sim \boldsymbol{\mathcal{N}}\left(0{,}25\,;\sqrt{\frac{0{,}25 \times 0{,}75}{100}}\right) = \boldsymbol{F} \sim \boldsymbol{\mathcal{N}}\left(0{,}25\,;\sqrt{\frac{0{,}1875}{100}}\right) = \boldsymbol{F} \sim \boldsymbol{\mathcal{N}}(0{,}25\,;0{,}043\,)$.

On a aussi $Z = \frac{F - 0{,}25}{0{,}043} \sim \boldsymbol{\mathcal{N}}(0\,;1)$

$$P(F \geq 0{,}35) = P\left(\frac{F - 0{,}25}{0{,}043} \geq \frac{0{,}35 - 0{,}25}{0{,}043}\right) = P(Z \geq 2{,}31) = 1 - P(Z \geq 2{,}31)$$

$$= 1 - \Phi(2{,}31) = 1 - 0{,}9896 = 0{,}0104$$

$$= 1{,}04\%$$

Sampling distribution

Exercise 4:

The U.S. Census Bureau announced that the median sales price of new houses sold in 2016 was \$316,500, and the mean sales price was \$370,800. Assume the standard deviation of the prices is \$90,000.

1. If you select sample of n = 100, describe the shape of the sampling distribution of X.
2. If you select a random sample of n = 100, what is the probability that the sample mean will be less than \$370,000 ?
3. If you select a random sample of n = 100, what is the probability that the sample mean will be between \$350,000 and \$365,000 ?

---

1. On a Médiane = 316,500 et Moyenne = 370,800 et l'écart-type = 90,000

   Puisque la médiane < Moyenne alors la distribution est étalée à droite ce qui est la même chose pour la distribution de l'échantillon.

2. On a $n \geq 30$, $\bar{X} \sim \mathcal{N}\left(\mu ; \frac{\sigma}{\sqrt{n}}\right)$, on passe au Z score avec la formule suivante: $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0\,;1)$

   $$P(\bar{X} \leq 370{,}000) = P\left(\frac{\bar{X}-370{,}800}{\frac{90{,}000}{\sqrt{100}}} \leq \frac{370{,}000-370{,}800}{\frac{90{,}000}{\sqrt{100}}}\right)$$

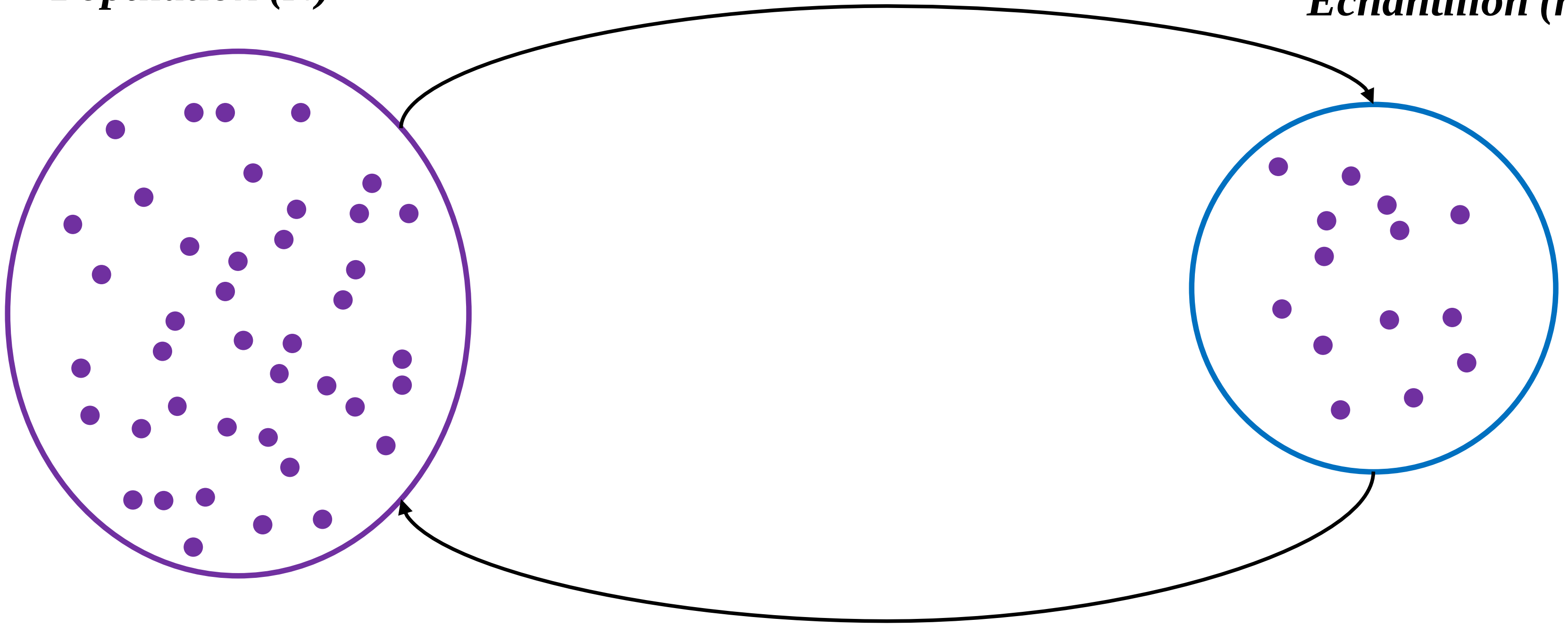
3. $P(350{,}000 \leq \bar{X} \leq 365{,}000) = P\left(\frac{350{,}000-370{,}800}{\frac{90{,}000}{\sqrt{100}}} \leq \frac{\bar{X}-370{,}800}{\frac{90{,}000}{\sqrt{100}}} \leq \frac{365{,}000-370{,}800}{\frac{90{,}000}{\sqrt{100}}}\right)$ = Φ(-0,64) - Φ(2,31) = 1 − Φ(0,64) – (1 − Φ(2,31))

Sampling distribution

Extraction: échantillonnage

Population (N)

Échantillon (n)



Enquête complète
Recensement

Enquête partielle
Sondage

Paramètres d'une population

$(\mu, \sigma^2, \sigma, P)$

Estimation

Statistiques d'un échantillon

$(\bar{X}, S^2, S, f)$

Estimation Ponctuelle

Estimation Par Intervalle de Confiance

1. Méthode des moments **2.** Méthode de Maximum de vraisemblance

3. Méthode des Moindres Carrées Ordinaires

Les propriétés d'un estimateur:

Les propriétés d'un estimateur sont les caractéristiques qui permettent d'évaluer la qualité de l'estimation qu'il fournit. Parmi les propriétés les plus importantes :

1. ***Biais :*** un estimateur est biaisé s'il tend à surestimer ou sous-estimer systématiquement la vraie valeur du paramètre qu'il est censé estimer. ***Un estimateur non biaisé*** est préférable à un estimateur biaisé car il tend à produire des estimations plus précises de la vraie valeur. Pour vérifier si un estimateur est biaisé, on peut calculer sa valeur attendue ou espérance mathématique, qui doit être égale à la valeur du paramètre estimé.
2. ***Efficacité :*** un estimateur est efficace s'il est à la fois non biaisé et a une faible variance ***(c'est-à-dire qu'il est capable de fournir des estimations précises).*** Pour vérifier l'efficacité d'un estimateur, on peut calculer sa variance et le comparer à la variance d'autres estimateurs. Un estimateur est considéré comme efficace s'il a la plus petite variance parmi tous les estimateurs non biaisés.
3. ***Consistance :*** un estimateur est consistant s'il converge vers la vraie valeur du paramètre à mesure que la taille de l'échantillon augmente. Cela signifie que plus de données permettent d'obtenir des estimations plus précises et que l'estimateur ne varie pas de manière erratique ou imprévisible avec la taille de l'échantillon. Pour vérifier si un estimateur est consistant, on peut utiliser le théorème de la limite centrale ou d'autres résultats de la théorie des probabilités pour ***montrer que la distribution de l'estimateur se concentre de plus en plus autour de la vraie valeur du paramètre lorsque la taille de l'échantillon augmente***.

Les propriétés d’un estimateur:

Biais d’un estimateur	*Efficacité d’un estimateur*	*Consistance d’un estimateur*
$E(\hat{\theta}) = \theta$ Exemples: $E(\bar{X}) = \mu$ $E(f) = P$ $E(S^2)$ TAR: $\frac{n-1}{n} \times \sigma^2$ $E(S^2)$ TSR: $\frac{N}{n} \times \frac{n-1}{N-1} \times \sigma^2$ $\Rightarrow$ $E(S^2) \neq \sigma^2$ $Il\ faut\ corriger\ la\ variance$	$V(\hat{\theta}_1) < V(\hat{\theta}_2)$ $\Rightarrow$ $\hat{\theta}_1\ est\ plus\ efficace\ \ que\ \hat{\theta}_2$ $V(\bar{X})$ TAR: $\frac{\sigma^2}{n}$ $V(\bar{X})$ TSR: $\frac{\sigma^2}{n} \times \frac{N-n}{N-1}$ $\Rightarrow$ $V(\bar{X})_{TAR} > V(\bar{X})_{TSR}$ $Le\ TSR\ est\ plus\ efficace\ \ que\ le\ TAR$	$\lim_{n \to +\infty} V(\hat{\theta}) = 0$ $\Rightarrow$ $\hat{\theta}\ est\ un\ estimateur\ convergent\ de\ \theta$ Exemples en cas de TAR: $\lim_{n \to +\infty} V(\bar{X}) = \lim_{n \to +\infty} \left(\frac{\sigma^2}{n}\right) = 0$ $\Rightarrow$ $\bar{X} est\ un\ estimateur\ convergent\ de\ \mu$ $\lim_{n \to +\infty} V(f) = \lim_{n \to +\infty} \left(\frac{pq}{n}\right) = 0$ $\Rightarrow$ $f est\ un\ estimateur\ convergent\ de\ P$

Sampling distribution

Les propriétés d'un estimateur:

Biais d'un estimateur

$$E(\hat{\theta}) = \theta$$

Exemples:

$$E(\bar{X}) = \mu$$

$$E(f) = P$$

$$E(S^2) \xrightarrow{\text{TAR}} \frac{n-1}{n} \times \sigma^2$$

$$E(S^2) \xrightarrow{\text{TSR}} \frac{N}{n} \times \frac{n-1}{N-1} \times \sigma^2$$

$$\Rightarrow \quad E(S^2) \neq \sigma^2$$

Il faut corriger la variance

Proof de $E(\bar{X}) = \mu$

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$ Par définition

$$E(\bar{X}) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right)$$ Propriété de l'Esperance

$$E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$ L'Esperance d'une somme de V.A. est la somme de leurs espérances

$$E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X)$$

$$E(\bar{X}) = \frac{1}{n} n\, E(X)$$ Indépendantes et identiquement distribuées (iid) (de même loi que X)

$$\boldsymbol{E(\bar{X}) = E(X) = \mu}$$ Estimateur sans biais de μ

Exemple qui prouve que $E(\bar{X}) = \mu$

Soit la pop. de 5 étudiants et X la note d'une matière:

Etudiant	A	B	C	D	E
X_i	12	8	10	15	10

$$\mu = \frac{12+8+10+15+10}{5} = \frac{55}{5} = 11$$

Si on divise la pop. en échantillons de taille = 3:

	X_1	X_2	X_3	Notes
ACD	12	10	15	12,33
ABC	12	8	10	10
ABD	12	8	15	11,66
ABE	12	8	10	10
BCD	8	10	15	11
BCE	8	10	10	9,35
BDE	8	15	10	11
CDE	10	15	10	11,66
DEA	15	10	12	12,33
ECA	10	10	12	10,66

$$\bar{X} = \frac{110}{10} = 11$$

$\bar{X} = \mu = 11$

Les propriétés d’un estimateur:

<table>
<tr><th rowspan="2">Biais d’un estimateur</th><th colspan="2">Correction de la variance</th></tr>
<tr><th>Tirage avec remise (TAR)</th><th>Tirage sans remise (TSR)</th></tr>
<tr>
<td>
$$E(\hat{\theta}) = \theta$$
Exemples:
$$E(\bar{X}) = \mu$$
$$E(f) = P$$
$$E(S^2)\begin{cases} \xrightarrow{\text{TAR}} \dfrac{n-1}{n} \times \sigma^2 \\ \xrightarrow{\text{TSR}} \dfrac{N}{n} \times \dfrac{n-1}{N-1} \times \sigma^2 \end{cases}$$
$$\Rightarrow \quad E(S^2) \neq \sigma^2$$
$Il\ faut\ corriger\ la\ variance$
</td>
<td>
$$E(S^2) = \frac{n-1}{n} \times \sigma^2$$
$$E(S^2) = \frac{n-1}{n} \times \sigma^2 \times \frac{n}{n-1}$$
$$E(S'^2) = S^2 \times \frac{n}{n-1}$$
$$or\ S^2 = \frac{1}{n} \sum_{i=1}^{n}(X_i - \bar{X})^2$$
$$Donc\ S'^2 = \frac{1}{n} \sum_{i=1}^{n}(X_i - \bar{X})^2 \times \frac{n}{n-1}$$
$$S'^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$
S'^2 est un estimateur sans biais de σ^2

Donc l’écart-type $S' = S \times \sqrt{\frac{n}{n-1}}$
</td>
<td>
$$E(S^2) = \frac{N}{n} \times \frac{n-1}{N-1} \times \sigma^2$$
$$S^2 = \frac{N}{n} \times \frac{n-1}{N-1} \times \sigma^2 \times \frac{n\,(N-n)}{N\,(n-1)}$$
$$S''^2 = S^2 \times \frac{n\,(N-n)}{N\,(n-1)}$$
$$S''^2 = S'^2 \times \frac{N-1}{N}$$
S''^2 est un estimateur sans biais de σ^2

Donc l’écart-type $S'' = S' \times \sqrt{\frac{N-1}{N}}$
</td>
</tr>
</table>

Sampling distribution

Les propriétés d'un estimateur:

Consistance d'un estimateur	*Proof de $V(\bar{X}) = \frac{\sigma^2}{n}$ pour le cas de tirage avec remise*
$\lim_{n \to \infty} V(\hat{\theta}) = 0$ ➡ $\hat{\theta}$ *est un estimateur convergent de* θ Exemples en cas de TAR: $\lim_{n \to \infty} V(\bar{X}) = \lim_{n \to \infty} \left(\frac{\sigma^2}{n}\right) = 0$ ➡ $\bar{X}$ *est un estimateur convergent de* μ $\lim_{n \to \infty} V(f) = \lim_{n \to \infty} \left(\frac{pq}{n}\right) = 0$ ➡ f *est un estimateur convergent de* P	$V(\bar{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$ Par définition $V(\bar{X}) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right)$ Propriété de la Variance $V(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i)$ La variance d'une somme de V.A. indépendantes est la somme de leurs variances $V(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n} V(X)$ Indépendantes et identiquement distribuées (iid) (de même loi que X) $V(\bar{X}) = \frac{1}{n^2} n\, V(X)$ $\boldsymbol{V(\bar{X}) = \frac{V(X)}{n} = \frac{\sigma^2}{n}}$ $\lim_{n \to \infty} V(\bar{X}) = \lim_{n \to \infty} \left(\frac{\sigma^2}{n}\right) = 0$ ➡ Estimateur convergent

Méthodes d’estimations:

La méthode d’estimation ponctuelle est une approche qui consiste à ***utiliser une seule valeur pour estimer un paramètre d’intérêt***. L’estimation par moments, l’estimation par maximum de vraisemblance et la méthode des moindres carrés ordinaires sont des méthodes spécifiques pour calculer une estimation ponctuelle.

L’estimation par intervalle de confiance est une méthode pour estimer un ***paramètre inconnu*** d’une population à partir d’un échantillon de données. Contrairement à l’estimation ponctuelle qui utilise une seule valeur pour estimer le paramètre, l’estimation par intervalle de confiance fournit une plage de valeurs plausibles pour le paramètre inconnu. L’intervalle de confiance est construit à partir de l’échantillon de données et dépend du niveau de confiance souhaité et de la taille de l’échantillon. Le niveau de confiance indique la probabilité que l’intervalle de confiance contienne réellement la valeur du paramètre inconnu. Par exemple, un intervalle de confiance à 95% signifie qu’il y a 95% de chances que la valeur réelle du paramètre se situe dans l’intervalle de confiance.

La méthode pour calculer un intervalle de confiance dépend du type de paramètre que l’on souhaite estimer. Par exemple, pour estimer la moyenne ou une proportion d’une population, on utilise généralement la distribution ***Normale*** lorsque $n \geq 30$ ou l’σ est connu, et la distribution de ***Student*** lorsque $n < 30$ ou l’σ est inconnu pour calculer l’intervalle de confiance. Alors que pour estimer la variance, on utilise la distribution de ***Khi deux***.

Estimation ponctuelle d'une moyenne:

Soit X la variable aléatoire définie sur une population Ω, on note μ l'espérance de X. μ est le paramètre inconnu que l'on cherche à estimer. On prélève pour cela un échantillon de taille n puis on calcule la moyenne arithmétique $\bar{X}$ des valeurs observées de X.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

➡ $\bar{X}$ est une estimation ponctuelle de la moyenne μ de la population.

Exemple: si on a la valeur de $\bar{X}$ = 11, alors on peut estimer que la moyenne μ de la population est aussi = 11.

Estimation ponctuelle d'une proportion:

Dans une population Ω, on considère la proportion P des éléments de Ω présentant un caractère C et une proportion $q = 1 - p$ des éléments de Ω ne présentant pas ce caractère C. P est le paramètre inconnu que l'on cherche à estimer.

On prélève pour cela un échantillon de taille n puis on calcule la proportion f des éléments de l'échantillon possédant ce caractère C.

➡ f est une estimation ponctuelle de la proportion P de la population.

Exemple: Une population de production des articles de qualité A et de qualité B. On y prélève 100 articles et on trouve 30 articles de qualité A. On s'intéresse à l'estimation de la proportion d'articles de qualité A:

On a $f = \frac{30}{100} = 0{,}3$

On peut alors estimer ponctuellement $P = 0{,}3$.

Estimation ponctuelle d'une variance:

Soit σ^2 la variance de la variable aléatoire X définie sur une population Ω. σ^2 est le paramètre inconnu que l'on cherche à estimer.

On prélève pour cela un échantillon de taille n puis on calcule la variance σ^2 des valeurs observées de X.

On admettra que S^2 n'est pas une bonne estimation ponctuelle de σ^2 . Cela s'explique par le fait que S^2 mesure la dispersion des valeurs de l'échantillon autour de $\bar{X}$ et non pas autour de μ.

➡ Le nombre $S'^2 = \frac{n}{n-1} S^2$ donne une estimation ponctuelle de σ^2.

➡ Le nombre $S' = S \times \sqrt{\frac{n}{n-1}}$ donne une estimation ponctuelle de σ.

Exemple: Une machine met un produit en sachet. Soit X la variable aléatoire qui associe la masse du sachet fabriqué par cette machine représentée dans la table suivante. On considère comme normal un sachet dont la masse est comprise entre 85g et 95g.

On prélève un échantillon de 100 sachets, donner une estimation ponctuelle de :

1. La masse moyenne μ des sachets.
2. La proportion de sachets considérés comme normaux.
3. L'écart-type des masses des sachets dans la production entière.

Masse en g	ni
[75 ; 80[	6
[80 ; 85[	10
[85 ; 90[	19
[90 ; 95[	37
[95 ; 100[	15
[100 ; 105 [	8
[105 ; 110[	5
Total	100

	A	B	C	D	E
1	Masse en g	xi	ni	nixi	nixi²
2	[75 ; 80[	77,5	6	465	36038
3	[80 ; 85[	82,5	10	825	68063
4	[85 ; 90[	87,5	19	1662,5	145469
5	[90 ; 95[	92,5	37	3422,5	316581
6	[95 ; 100[	97,5	15	1462,5	142594
7	[100 ; 105 [	102,5	8	820	84050
8	[105 ; 110[	107,5	5	537,5	57781
9	Total		100	9195	850575
10		Moy	91,95		
11		Variance	50,948		
12		Écart-type	7,1378		

1. La masse moyenne μ des sachets = $\bar{X}$ = 91,95g.

2. $P = f = \frac{19+37}{100} = \frac{56}{100} = 0,56$

3. $\sigma^2 = S^2\sqrt{\frac{n}{n-1}} = 7,14 \times \sqrt{\frac{100}{99}} \approx 7,18g$

EXERCICE 1 : Au Moyen Âge le Maître des Monnaies était chargé par le trésor Anglais pour frapper les pièces de monnaie à partir de l'or. La qualité des pièces se voit évaluée par une épreuve qui s'appelle « The Trial of the Pyx ». Cette épreuve consistait à choisir un échantillon de pièces, de les peser, et d'infliger une punition au Maître des Monnaies si le poids total des pièces ne dépassait pas le poids prescrit, moins une tolérance. Sous l'hypothèse que le Maître des Monnaies est honnête, et n'a pas détourné une partie du métal précieux, le poids de chaque pièce suit une loi $\mathcal{N}\ (128; 1)$.

1. Si 100 pièces sont échantillonnées lors de ce test, quelle doit être la valeur de la tolérance si le pouvoir Britannique veut limiter à 1% la probabilité qu'un Maître des Monnaies honnête soit puni ?
2. Avec la valeur de la tolérance que vous avez trouvée dans la question 1, quelle est la probabilité de détecter un Maître des Monnaies malhonnête qui empoche 0,1 grammes d'or par pièce ?
3. Combien faudrait-il que le Maître des Monnaies empoche par pièce pour que la probabilité de détecter la fraude soit de 50%, toujours avec le seuil trouvé à la question 1.

Sampling distribution

1. Si 100 pièces sont échantillonnées lors de ce test, quelle doit être la valeur de la tolérance si le pouvoir Britannique veut limiter à 1% la probabilité qu'un Maître des Monnaies honnête soit muni ?

Tout d'abord, l'énoncé indique que le poids de chaque pièce suit une loi normale $\mathcal{N}(128{,}1)$, où 128 est la moyenne et 1 est l'écart-type ayant la même valeur de variance ($\sqrt{1} = 1$). Nous pouvons en déduire que le poids de 100 pièces sélectionnées au hasard suit une loi normale $\mathcal{N}(12800{,}10)$, car la somme des poids suit la loi d'une somme de variables aléatoires indépendantes et identiquement distribuées.

Maintenant, nous voulons déterminer la tolérance à appliquer lors du test pour limiter à 1% la probabilité qu'un maître des monnaies honnête soit condamné. Supposons que la tolérance soit de T. Cela signifie que la somme des poids des 100 pièces doit être supérieure ou égale à 12800 - T pour que le maître des monnaies soit considéré comme honnête. Puisque nous voulons limiter la probabilité qu'un maître des monnaies honnête soit condamné à 1%, ça peut être traduit par l'équation suivante:

$$\mathrm{P}(X \leq 12800 \ - T) = 0{,}01$$

Nous pouvons standardiser la somme des poids en la transformant en une variable aléatoire centrée réduite:

$$\mathrm{Z} = \frac{(X-12800)}{10} \sim \mathcal{N}(0\,;1)$$

Sampling distribution

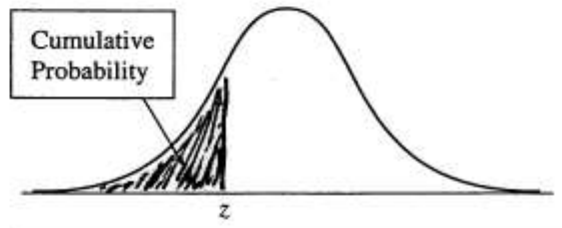
Maintenant, nous pouvons utiliser la formule de transformation pour Z pour obtenir :

$$\mathrm{P}\left(\frac{(X-12800)}{10} \leq -\frac{T}{10}\right) = 0{,}01$$

Nous pouvons utiliser la table de la loi normale standard pour trouver la valeur de - $\frac{T}{10}$ correspondant à cette probabilité. Nous cherchons dans la table la valeur la plus proche de 0,01, qui est 0,0099. On lit la valeur de z correspondante, qui est -2,33 (P(Z < -2,33) = 0,0099).

Nous obtenons donc: $\mathrm{P}\left(\frac{(X-12800)}{10} \leq -2{,}33\right) = 0{,}01$

On a alors $-\frac{T}{10} = -2{,}33$, donc T = 23,3



Cumulative probability for *z* is the area under the standard normal curve to the left of *z*

TABLE A Standard Normal Cumulative Probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110

2. Avec la valeur de la tolérance que vous avez trouvée dans la question 1, quelle est la probabilité de détecter un Maître des Monnaies malhonnête qui empoche 0,1 grammes d'or par pièce ?

Dans ce cas, le poids de chaque pièce suit une loi normale $\mathcal{N}(128 – 0,1\ ;1)$, c'est-à-dire $\mathcal{N}(127,9\ ;1)$, Nous pouvons en déduire que le poids de 100 pièces sélectionnées au hasard suit une loi normale $\mathcal{N}(12790,10)$.

Avec un seuil de tolérance de T= 23, on cherche:

$$P(Y \leq 12800 - 23) = P(Y \leq 12777)$$

$$= P\left(\frac{(Y - 12790)}{10} \leq \frac{(12777 - 12790)}{10}\right) = P(Z \leq -1{,}3) = 1 - P(Z \leq 1{,}3) = 0{,}097$$

<u>Interprétation</u>: Avec une tolérance de T=23, un maître de monnaie malhonnête qui empoche 0,1 grammes d'or par pièce n'a que 9,7% de chance de se faire punir.

Sampling distribution

3. Combien faudrait-il que le Maître des Monnaies empoche par pièce pour que la probabilité de détecter la fraude soit de 50%, toujours avec le seuil trouvé à la question 1.

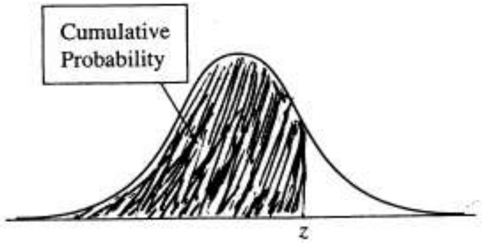
On cherche $E(X)$ avec $X = \sum_{i=1}^{100} X_i$ telque : $P\left(\frac{(X-E(X))}{10} \leq \frac{(12777-E(X))}{10}\right) = 0{,}5$

$$= P\left(Z \leq \frac{(12777-E(X))}{10}\right) = 0{,}5 \text{ avec } Z \sim \mathcal{N}(0,\, 1)$$

D'après la table $\Phi\left(\frac{(12777-E(X))}{10}\right) = \Phi\,(0{,}5) = 0$

$$\rightarrow E(X) = 12777$$

$$\rightarrow X_i \sim \mathcal{N}(12777,\, 1)$$



Cumulative probability for z is the area under the standard normal curve to the left of z

TABLE A Standard Normal Cumulative Probabilities ***(continued)***

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	$\boxed{.5000}$	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753

Interprétation: Le maître de monnaie devrait empocher 0,23 gramme d'or par pièce pour être puni avec une probabilité de 50%.

EXERCICE 2 : Soit S_0le prix d'une action à la date d'aujourd'hui. Supposez que l'évolution du prix de l'action au cours du temps soit $S_t = S_{t-1} + X_t$, où:

$$X_t = \begin{cases} 1 \text{ avec la probabilité de } 0{,}39 \\ 0 \text{ avec la probabilité de } 0{,}20 \\ -1 \text{ avec la probabilité de } 0{,}41 \end{cases}$$

1. Exprimer la variation des prix sur les 700 premières périodes, $\Delta S = S_{700} - S_0$ comme une fonction des X_t.
2. Quelle est la distribution approximative de la variation moyenne du prix de l'action.
3. Quelle est la probabilité pour que l'action augmente de 10 ou plus au cours des 700 premières périodes

Sampling distribution

1. Exprimer la variation des prix sur les 700 premières périodes, $\Delta S = S_{700} - S_0$ comme une fonction des X_t.

$$S_t = S_0 + \sum_{t=1}^{700} X_t \leftrightarrow \Delta S = S_{700} - S_0 = \sum_{t=1}^{700} X_t$$

2. Quelle est la distribution approximative de la variation moyenne du prix de l'action.

La variation moyenne: $\bar{X}_{700} = \frac{1}{700}\sum_{t=1}^{700} X_t$

$$E(\bar{X}_{700}) = E(X_t) = (1 \times 0{,}39) + (0 \times 0{,}20) + (-1 \times 0{,}41) = -0{,}02 = \mu$$

$$Var(X_t) = E(X_t^2) - [E(X_t)]^2$$

Or, $E(X_t^2) = (1^2 \times 0{,}39) + (0^2 \times 0{,}20) + (-1^2 \times 0{,}41) = 0{,}8$

$$\rightarrow Var(X_t) = 0{,}8 - [-0{,}02]^2 = 0{,}7996 = \sigma^2 \rightarrow Var(\bar{X}_{700}) = \frac{\sigma^2}{n} = \frac{0{,}7996}{700}$$

D'après le TCL: $\frac{\bar{X}_{700} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0\,;1) \rightarrow \bar{X}_{700} \sim \mathcal{N}\left(-0{,}02\,;\sqrt{\frac{0{,}7996}{700}}\right)$

3. Quelle est la probabilité pour que l'action augmente de 10 ou plus au cours des 700 premières périodes.

$$P\left(\sum_{t=1}^{700} X_t \geq 10\right) = P\left(\frac{1}{700}\sum X_t \geq \frac{10}{700}\right)$$

$$= P\left(\bar{X}_{700} \geq \frac{1}{70}\right)$$

$$= P\left(\frac{\bar{X}_{700} + 0{,}02}{\sqrt{\frac{0{,}7996}{700}}} \geq \frac{\frac{1}{70} + 0{,}02}{\sqrt{\frac{0{,}7996}{700}}}\right)$$

$$= P(Z \geq 1{,}0144) = 0{,}1552$$

EXERCICE 3 : Un imprimeur de prospectus emballe ces derniers par paquets de 100. On sait qu'en moyenne un prospectus pèse 1 gramme, avec un écart-type de 0,05. Cet imprimeur veut connaître la probabilité pour qu'un de ses paquets de 100 prospectus pèse plus de 100,4 grammes. Quelle est cette probabilité?

Soit X_i le poids d'un prospectus

$X_i \sim$ loi d'espérance $\mu = 1$, et d'écart−type $\sigma = 0{,}05$, alors:

$$P\left(\sum_{i=1}^{100} X_i \geq 100{,}4\right) \rightarrow P\left(\frac{\sum_{i=1}^{100} X_i}{100} \geq \frac{100{,}4}{100}\right) = P\left(\bar{X}_{100} \geq \frac{100{,}4}{100}\right)$$

En se basant sur le théorème central limite, on peut déduire que $\bar{X}_{100} \sim \boldsymbol{\mathcal{N}}\left(\boldsymbol{\mu}\,;\frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}\right)$, donc $\bar{X}_{100} \sim \boldsymbol{\mathcal{N}}\left(\mathbf{1}\,;\frac{\mathbf{0{,}05}}{\mathbf{10}}\right)$

Alors on cherche maintenant:

$$P\left(\bar{X}_{100} \geq \frac{100{,}4}{100}\right) = P(\bar{X}_{100} \geq 1{,}004) = P\left(\frac{\bar{X}_{100}-1}{0{,}005} \geq \frac{0{,}004}{0{,}005}\right) = P(Z \geq 0{,}8) = 1 - \Phi(0{,}8) = 0{,}212$$