

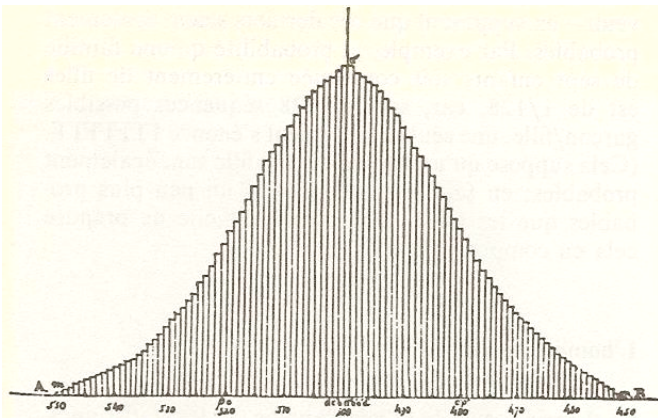
La loi normale

Aux origines de la loi normale ...

« Une grande partie des statistiques se centre autour de ce qu'on appelle la distribution normale. Il s'agit d'une courbe en forme de cloche qui représente de près les proportions d'une population ayant quelques caractéristiques particulières. Par exemple, si 1000 hommes sont tirés au hasard parmi la population de la Mongolie Extérieure, et si l'on trace un graphique représentant combien d'entre eux ont une taille donnée en centimètres, celui-ci ressemblera de près à la courbe en forme de cloche de la distribution normale. (...)

La distribution normale a été originellement appelée loi de l'erreur, et elle est née du travail d'astronomes et de mathématiciens du XVIII^e siècle qui, dans leur effort pour calculer les orbites de corps célestes, se virent obligés de prendre en compte l'effet des erreurs d'observation. La loi de l'erreur décrit comment des valeurs observées s'assemblent autour de leur moyenne, et elle fournit des estimations des risques pour qu'une erreur de taille donnée survienne. Cette loi a été importée dans les sciences sociales par Adolphe Quételet¹, qui a appliqué cette méthode à tout ce qu'il a pu trouver : les mensurations du corps humain, le crime, le mariage, le suicide. Sa Mécanique sociale reçut délibérément ce titre pour faire pendant à la Mécanique céleste de Laplace. Quételet tira rapidement des conclusions générales à partir de la constance supposée des valeurs moyennes de variables sociales, et il proposa la notion séduisante² de « l'homme moyen ». Quételet considérait la condition humaine comme une sorte de dynamique sociale, mais il voulait aussi l'aborder à la manière d'un spécialiste de la régulation. L'accorder, la stabiliser, en atténuer les oscillations. Pour Quételet, « l'homme moyen » n'était pas qu'une abstraction mathématique, c'était un idéal moral. »

Ian STEWART dans « DIEU JOUE-T-IL AUX DÉS ? Les Mathématiques du Chaos » (1989)



Approximation binomiale d'une distribution normale
(Quételet, 1846)



Adolphe Quételet en 1822

Le texte ci-dessus peut certainement susciter d'intéressants débats. Signalons que les idées sociales de Quételet étaient déjà controversées de son temps.

¹ Mathématicien Belge (1796 – 1874). Il est également connu pour être l'auteur, avec Dandelin, des célèbres « théorèmes belges » sur les sections coniques.

² Ce qualificatif est discutable.

1. Les mesures du corps humain

C'est au départ de mesures de tailles que nous allons aborder la distribution normale.

Parmi les statistiques disponibles dans la bibliographie ou sur Internet, une des plus intéressantes date de ... 1947. A l'époque, la direction de « De Bijenkorf », un grand magasin possédant plusieurs succursales aux Pays-Bas, décida de réaliser une vaste enquête sur les mesures corporelles des femmes Néerlandaises. L'objectif était d'optimiser la confection des vêtements pour dames en réalisant des habits mieux adaptés aux tailles des clientes, et en réduisant les dépenses inutiles.

Sur des clientes du magasin, choisies au hasard et ayant marqué leur accord, l'on procéda à quinze mesures : taille, masse, tour de taille, tour de hanches, carrure, etc.

Durant deux semaines, des données concernant 5001 femmes âgées de plus de dix-huit ans furent ainsi récoltées.

Pour la taille, voici les résultats que l'on obtint.



Tailles	Effectifs	Tailles	Effectifs	Tailles	Effectifs	Tailles	Effectifs
139	1	151	71	163	290	175	29
140	1	152	78	164	294	176	49
141	4	153	115	165	291	177	28
142	3	154	149	166	261	178	17
143	2	155	170	167	222	179	5
144	8	156	208	168	184	180	10
145	4	157	208	169	157	181	6
146	17	158	231	170	167	182	3
147	18	159	301	171	109	183	1
148	32	160	302	172	86	184	2
149	51	161	321	173	65	185	0
150	54	162	313	174	62	186	1

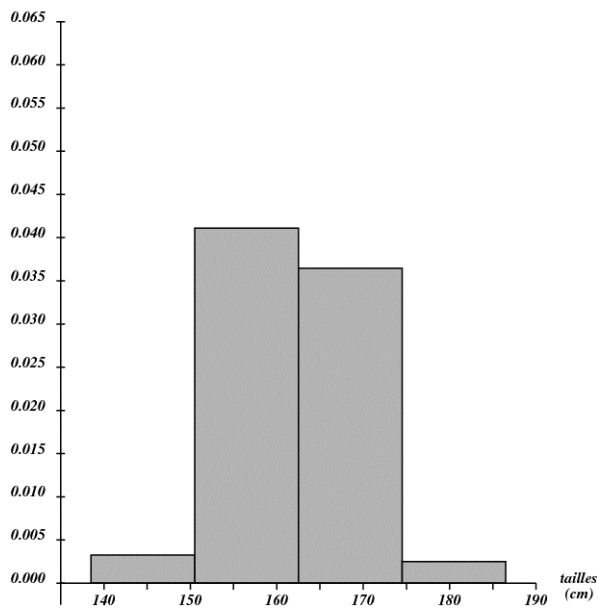
La taille moyenne était égale à 162,05 (cm) tandis que la taille médiane était de 162 (cm).

Nous avons représenté les histogrammes des fréquences en prenant différentes largeurs de classes. Voici comment ils ont été réalisés.

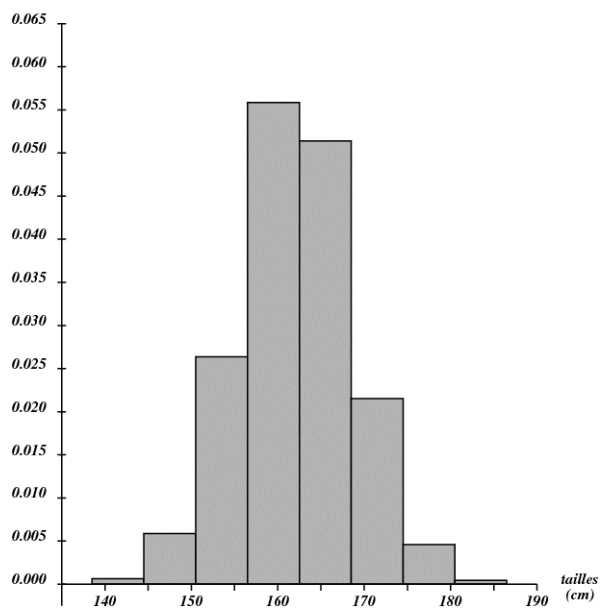
- Pour l'histogramme ❶, nous avons choisi de répartir les tailles en quatre classes de 12 (cm) de largeur chacune :
 $[138.5, 150.5[$, $[150.5, 162.5[$, $[162.5, 174.5[$ et $[174.5, 186.5[$.
 La première classe débute à 138.5 car nous supposons que les femmes à qui l'on a attribué une taille de 139 (cm) sont celles qui mesuraient entre 138.5 (cm) (inclus) et 139.5 (cm) (exclus). Les autres extrémités de classes s'expliquent de la même façon.
- Pour l'histogramme ❷, nous avons choisi de former huit classes de 6 (cm) de largeur chacune : $[138.5, 144.5[$, $[144.5, 150.5[$, ... et $[180.5, 186.5[$.
- Pour les histogrammes ❸ et ❹, nous avons respectivement seize classes de 3 (cm) de largeur chacune, et quarante-huit classes de 1 (cm) de largeur chacune.

Rappelons que dans un histogramme, c'est l'aire d'un rectangle qui correspond à la fréquence de la classe. Par exemple, dans le premier graphique, la fréquence de la classe $[150.5, 162.5[$ est égale à 12 (base du rectangle) fois 0.0411 (hauteur du rectangle), soit environ 0.4932. Cela correspond au fait que 49.32 % des femmes avaient une taille comprise entre 150.5 (cm) et 162.5 (cm).

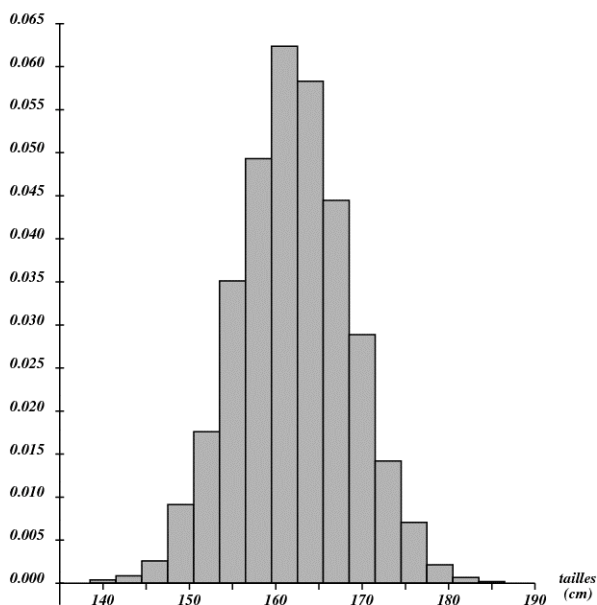
①



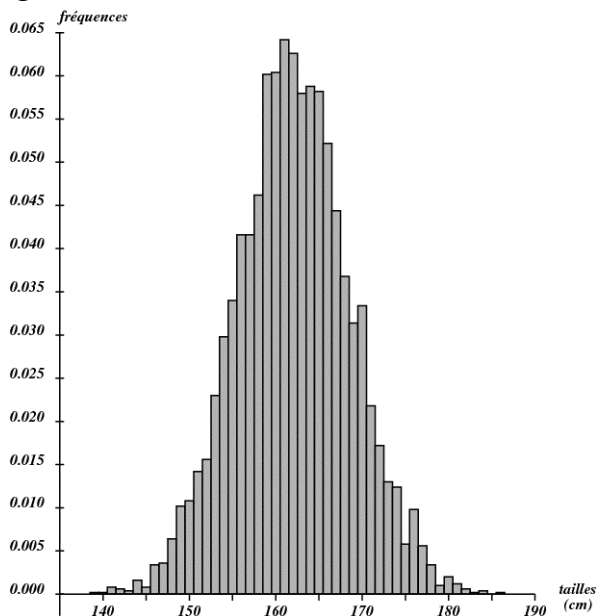
②



③



④



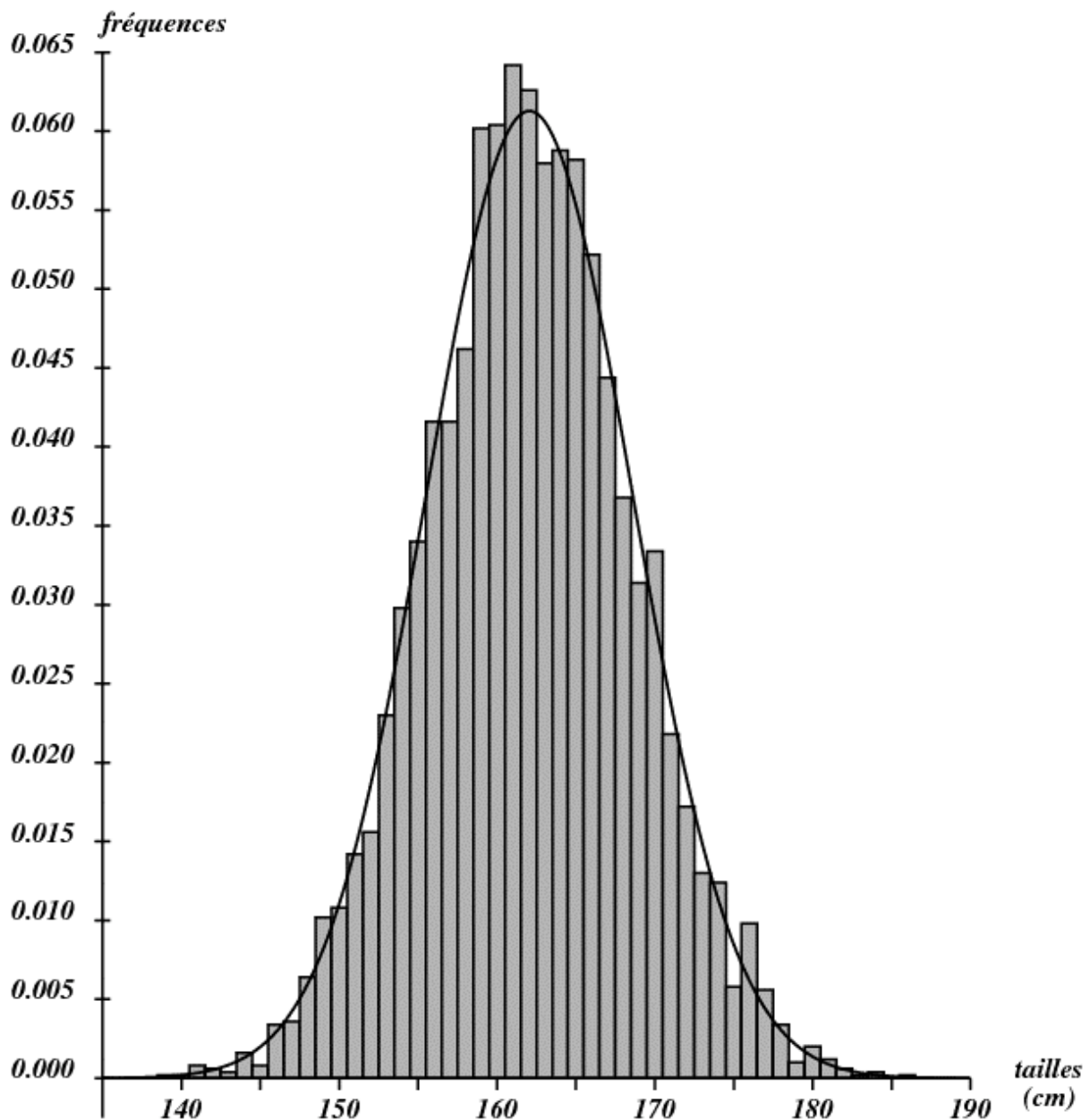
Quelques observations ...

- Les histogrammes sont tous plus ou moins symétriques par rapport à la taille moyenne de la population (162 cm).
- Le dernier graphique est le seul pour lequel la hauteur d'un rectangle est égale à la fréquence de la classe correspondante, puisque la largeur de celle-ci vaut 1.
- Il apparaît clairement que la plus grande partie de la population a une taille assez proche de la taille moyenne ; plus on s'éloigne de la taille moyenne, plus le nombre de femmes diminue : très peu d'entre elles mesurent aux environs de 140 (cm) et de 180 (cm).
- La largeur des classes influence l'allure générale de l'histogramme. Nous pourrions dire que le premier est assez « grossier », le deuxième un peu plus « instructif », le troisième plus riche d'informations et assez bien « équilibré », tandis que le dernier, bien qu'étant le

plus détaillé par rapport aux données récoltées, semble devenir un peu plus « chaotique ». En fait, avoir un « beau » graphique dépend d'un bon accord entre la taille de l'échantillon et la largeur de classe choisie. Il est très probable que le dernier histogramme eût été beaucoup plus « lisse » et symétrique avec un échantillon de 50000 femmes au lieu de 5001 .

De manière générale, on peut comprendre que si l'on a simultanément augmentation de la taille de l'échantillon et diminution de la largeur des classes, les marches de « l'escalier » deviennent de plus en plus étroites et les contours de l'historgramme dessinent une courbe en forme de cloche. Cette courbe est une fonction mathématique qui, dans le cadre de notre exemple, constitue un modèle pour la distribution des tailles des femmes Néerlandaises en 1947. Cette distribution est appelée *distribution normale*.

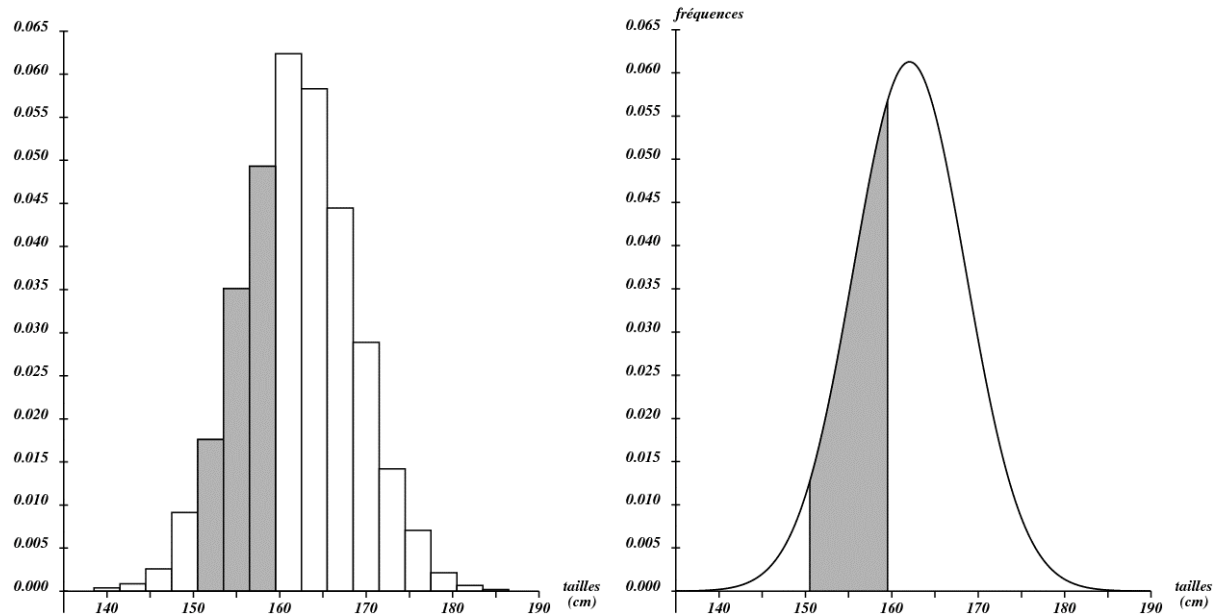
Nous parlerons plus tard de l'expression analytique de cette fonction. Pour le moment, contentons-nous de savoir qu'elle possède un axe de symétrie passant par un sommet dont l'abscisse est la taille moyenne. Superposant cette fonction à l'historgramme ④ , nous voyons que le modèle théorique est assez proche des données réelles.



Une question souvent posée est celle du pourcentage de la population dont la taille est comprise entre deux valeurs données. Par exemple, quel était le pourcentage de femmes mesurant entre 150,5 (cm) et 159,5 (cm) ?

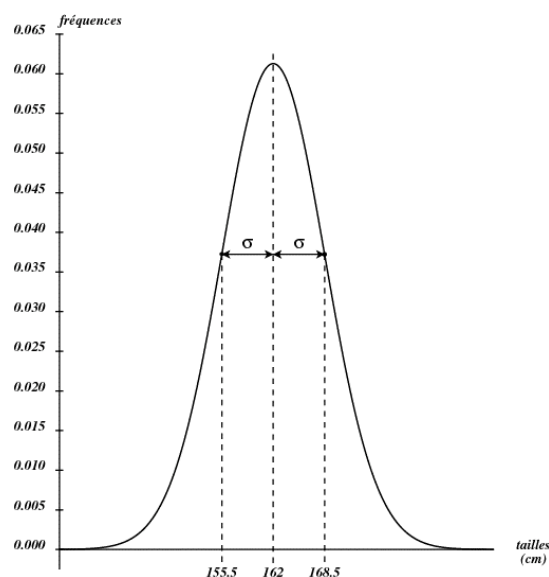
Si l'on dispose d'un histogramme, il faut additionner des aires de rectangles (figure de gauche). Ce calcul donne 30,61 % .

La courbe normale permet également de répondre à la question (avec une certaine approximation) : il faut déterminer l'aire sous la courbe entre les abscisses 150,5 et 159,5 (figure de droite). Il existe une table permettant de trouver cette aire et nous verrons plus loin comment nous en servir. Dans notre cas, elle donne environ 30,96 % .



L'aire grisée sous la courbe représente donc environ 31 % de l'aire totale sous la courbe (celle-ci valant 100 % ou 1).

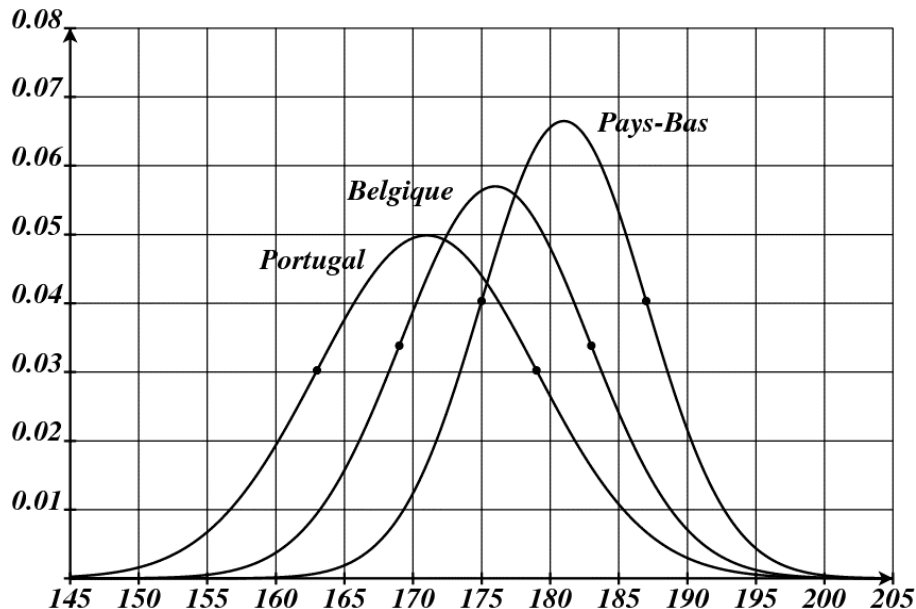
Un autre paramètre important d'une série statistique est l'écart type, noté σ . Rappelons qu'il s'agit d'un paramètre de dispersion : plus il est grand, plus les données de la série sont dispersées autour de leur moyenne. Pour notre exemple, il valait environ 6,5 (cm). Cette valeur se retrouve dans la courbe normale : il s'agit de la distance entre l'axe de symétrie et un des deux points d'inflexion (là où la pente est la plus forte en valeur absolue).



Application

Voici les courbes normales pour les tailles, en centimètres, des hommes Portugais, Belges et Néerlandais en 2005 . Les points d'inflexion sont indiqués.

- Donner la taille moyenne et l'écart type pour chaque nationalité.
- Plus le sommet de la courbe est élevé, plus elle est « étroite ». Pourquoi ?



Lorsque des données statistiques donnent lieu à un histogramme en forme de cloche, symétrique par rapport à la moyenne, nous parlons de *distribution normale*. Si le graphique s'écarte de ce modèle (s'il n'y a pas de symétrie par exemple), alors nous dirons qu'il s'agit d'une *distribution non normale* (ce qui ne veut pas dire « anormale » dans le sens commun du terme !).

Exercices

- Lors de l'enquête de 1947, l'on mesura également la longueur du pied (résultats dans le tableau ci-dessous). Réaliser un histogramme en prenant 1 (cm) comme largeur de classe. La distribution des longueurs semble-t-elle normale ou pas ? Expliquer.

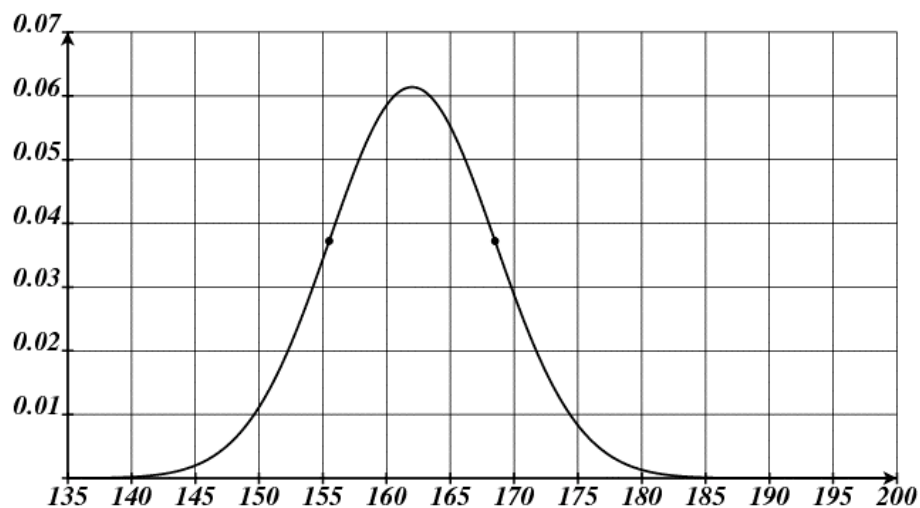
Longueurs	Effectifs	Longueurs	Effectifs
20	4	25	848
20,5	6	25,5	532
21	11	26	303
21,5	33	26,5	179
22	68	27	76
22,5	213	27,5	20
23	419	28	20
23,5	582	28,5	7
24	718	29	1
24,5	968		



2. Même question pour la distribution des âges des 5001 clientes (les classes sont déjà données dans le tableau).

Âges	Effectifs
[18,28[1186
[28,38[947
[38,48[1183
[48,58[1048
[58,68[515
[68,→[122

3. La taille moyenne des Néerlandaises d'aujourd'hui est d'environ 168,5 (cm).
- a) Tracer ci-dessous la courbe normale pour les tailles actuelles, si l'on considère que l'écart type est resté de 6,5 (cm).
- b) Comment la courbe se modifierait-elle avec un écart type double ?



4. Le tableau ci-dessous donne les poids de 507 femmes Belges en 2005. Réaliser un histogramme en utilisant les classes du tableau. La distribution des poids semble-t-elle normale ou pas ? Expliquer.

Poids (kg)	Effectifs
[40, 50[26
[50, 55[45
[55, 60[75
[60, 65[99
[65, 70[76
[70, 75[79
[75, 80[31
[80, 85[26
[85, 90[13
[90, 95[12
[95, 100[9
[100, 110]	12

5. Le tableau ci-dessous donne les poids à la naissance de 124 250 bébés nés en Belgique en 1992 (source : <http://data.un.org/> ; liens : Population → Demographic Statistics → Live births by birth weight). Réaliser un histogramme en utilisant les classes du tableau. La distribution des poids semble-t-elle normale ou pas ? Expliquer.

Poids (g)	Effectifs
[500, 1000[260
[1000, 1500[650
[1500, 2000[1496
[2000, 2500[5404
[2500, 3000[22878
[3000, 3500[49961
[3500, 4000[33925
[4000, 4500[8550
[4500, 5000[1126

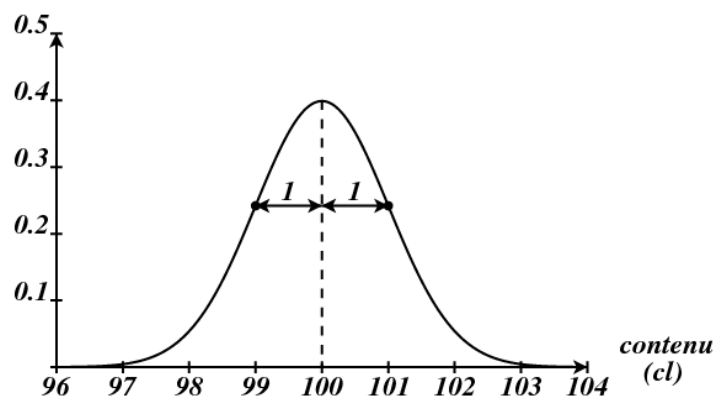
6. Les mesures du corps humain montrent que la taille, la longueur et la largeur du pied, la longueur du dos, la longueur du bras, la hauteur du genou, etc. sont des variables dont la distribution est proche d'une loi normale. Par contre, des caractéristiques comme le poids, le tour de taille ou le tour de hanches ne sont pas distribuées de manière normale, ou alors dans une mesure nettement moindre. Quelle explication peut-on donner à cela ?

2. La loi normale centrée réduite

Exemple

Une bouteille d'un litre de lait ne contient probablement pas *exactement* un litre. Néanmoins, si la machine effectuant le remplissage est correctement réglée, nous pouvons nous attendre à un contenu *moyen* d'environ un litre par bouteille.

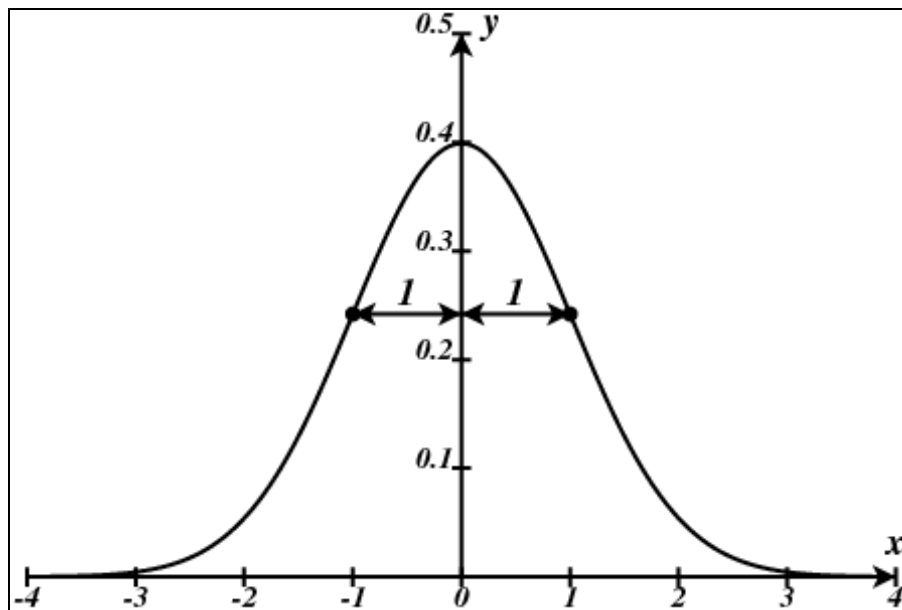
Lors d'un contrôle, les contenus de cent bouteilles sont mesurés avec précision. Supposons que l'inspecteur trouve effectivement un contenu moyen d'un litre. Il constate en outre que les contenus sont distribués de façon normale avec un écart type de un centilitre. Cette situation est illustrée par le graphique suivant.



En réalité, l'inspecteur ne s'intéresse qu'à l'écart entre le contenu réel et un litre. Ainsi, pour chaque bouteille de l'échantillon, il note la différence avec 100 (cl) (positive ou négative selon que la bouteille contient plus ou moins d'un litre).

La moyenne des résultats ainsi notés vaut 0 tandis que l'écart type vaut toujours 1 (expliquer).

Lorsqu'une variable aléatoire est de loi normale de moyenne $\bar{x} = 0$ et d'écart type $\sigma = 1$, on parle de *loi normale centrée réduite*. Elle donne lieu au graphique suivant.



L'équation de la courbe normale centrée réduite est la suivante :

$$y = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$$

Les nombres π et e sont ceux que nous connaissons ($\pi \approx 3,1416$ et $e \approx 2,7183$).

Remarques

- La fonction représentée par cette équation est *paire*. En effet, puisque la variable x est élevée au carré, un réel et son opposé ont la même image. La courbe est donc symétrique par rapport à l'axe des ordonnées.
- Plus la valeur de x augmente (diminue), plus la valeur de y tend vers 0. A gauche de $x = -3,5$ et à droite de $x = 3,5$, l'écart entre la courbe et l'axe des x est déjà si petit qu'ils semblent se toucher. Par exemple, pour $x = 5$ ou $x = -5$, nous avons $y \approx 0,0000015$. L'axe des abscisses est une *asymptote horizontale* à la courbe.

Un problème important : l'aire sous la courbe normale

Nous avons vu que l'aire comprise entre la courbe normale et l'axe des abscisses permet de trouver des fréquences (par exemple, le pourcentage d'individus ayant une taille comprise entre 150,5 et 159,5 centimètres).

Tout le problème est donc de calculer de telles aires. Grâce à l'équation ci-dessus et à des méthodes de quadrature relevant du calcul intégral, des tables ont pu être réalisées.

Celle fournie en annexe donne les aires, avec quatre décimales, entre $-\infty$ et l'abscisse z , pour des valeurs de z comprises entre 0 et 3,99. Nous noterons cette aire $\Phi(z)$.

Exemples

1. Déterminer l'aire sous la courbe entre $-\infty$ et $z = 0,34$.

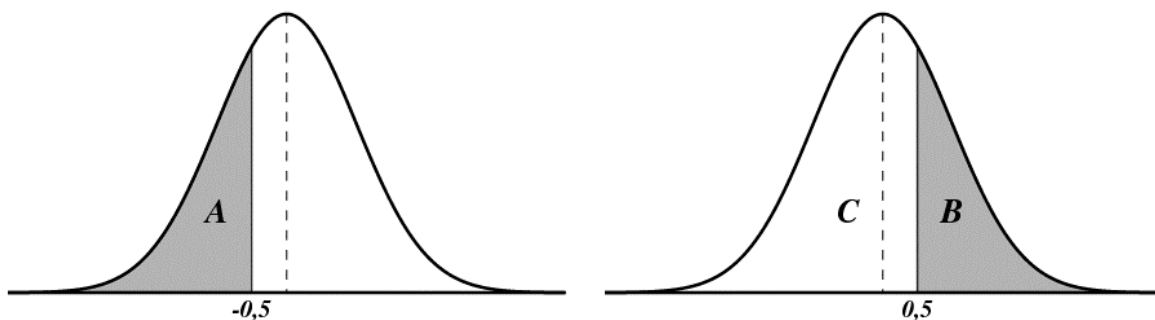
Cette aire se lit directement dans la table (voir ci-dessous).

On trouve ainsi : $\Phi(0,34) \approx 0,6331$.

z	0	1	2	3	4	5
0.0	.5000	.5040	.5080	.5120	.5160	.5199
0.1	.5398	.5438	.5478	.5517	.5557	.5596
0.2	.5793	.5832	.5871	.5910	.5948	.5987
0.3	.6179	.6217	.6255	.6293	.6331	.6368
0.4	.6554	.6591	.6628	.6664	.6700	.6736

2. Déterminer l'aire sous la courbe entre $-\infty$ et $z = -0,5$.

Lorsque la valeur de z est négative, le résultat ne se trouve pas directement dans la table. Pour le trouver, il faut faire le petit raisonnement suivant, basé sur la symétrie de la courbe et sur le fait que l'aire totale sous la courbe est égale à 1.



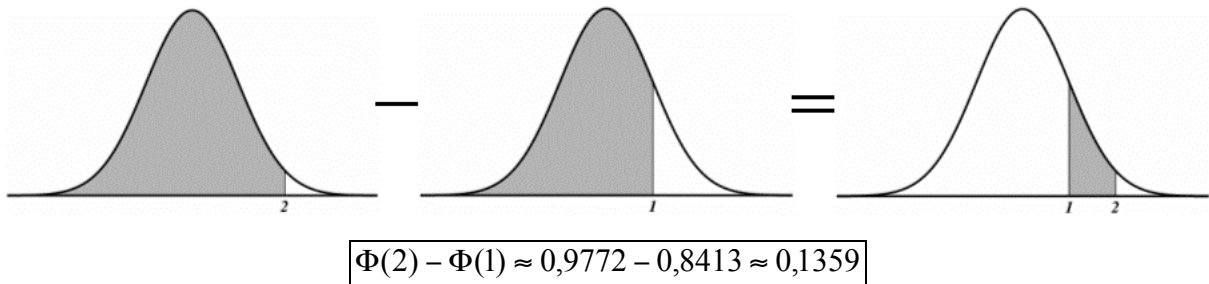
$$\text{aire}(A) = \text{aire}(B) \quad (\text{surfaces symétriques par rapport à } z = 0)$$

$$\text{aire}(B) = 1 - \text{aire}(C) \quad (\text{l'aire } B \text{ est le complément de l'aire } C \text{ par rapport à l'aire totale})$$

$$\text{aire}(C) = \Phi(0,5) \approx 0,6915 \quad (\text{voir table})$$

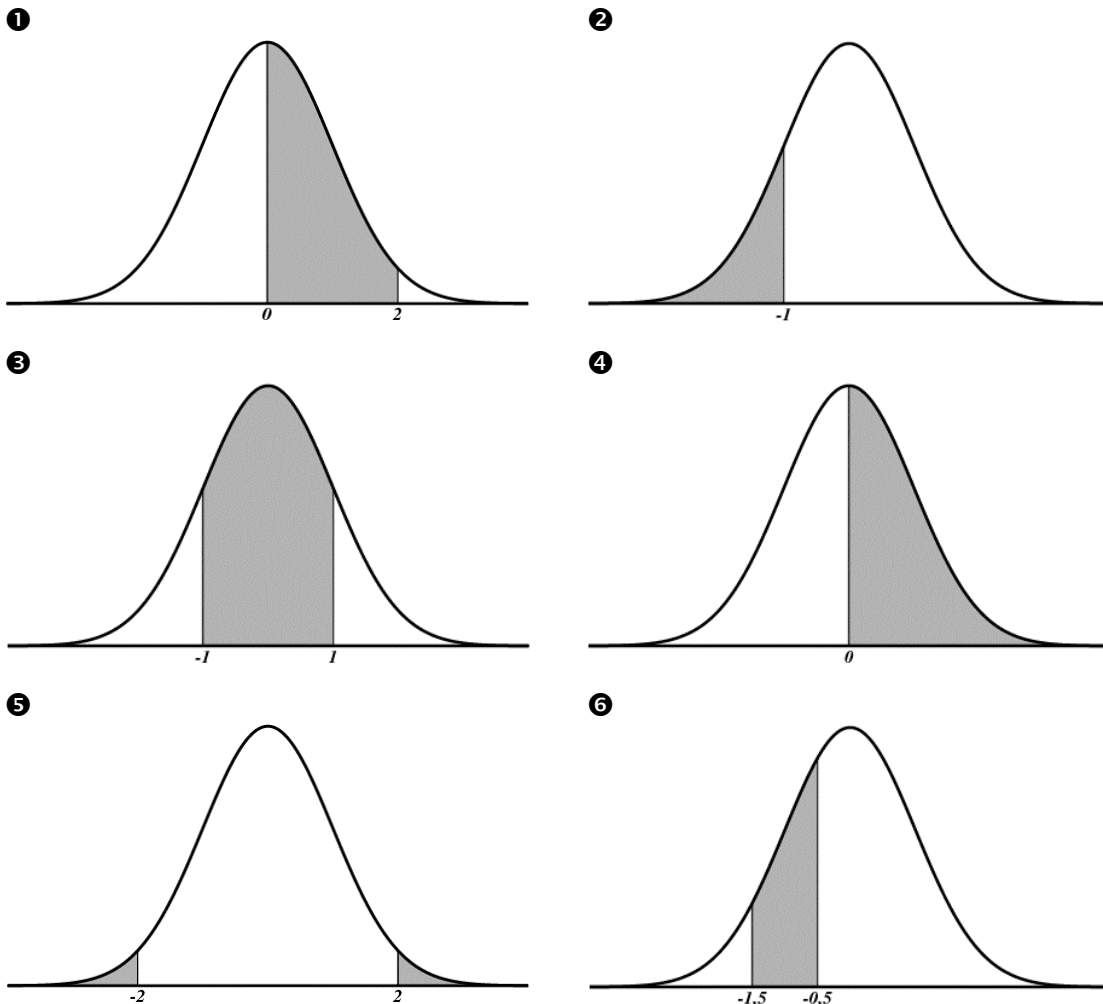
$$\text{aire}(A) \approx 1 - 0,6915 \approx 0,3085$$

3. Déterminer l'aire sous la courbe entre $z = 1$ et $z = 2$.
Il suffit de chercher l'aire entre $-\infty$ et $z = 2$, puis l'aire entre $-\infty$ et $z = 1$, et de faire la différence :



Exercices

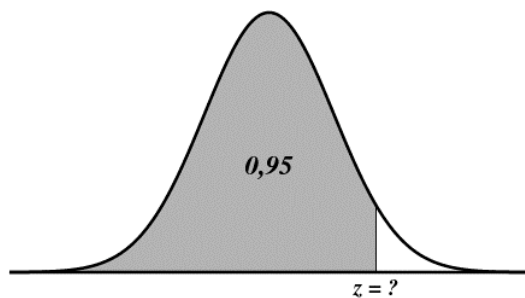
1. À l'aide de la table, déterminer les aires grisées.



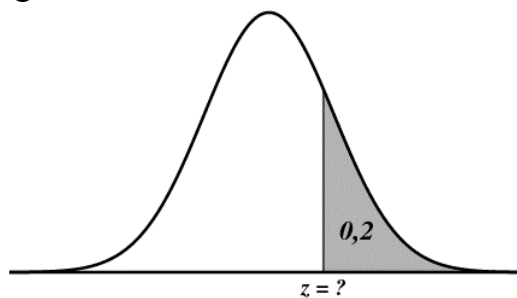
2. Retour au contrôle des bouteilles de lait ...
- Quel est le pourcentage de bouteilles contenant plus de 99(cl) et moins de 101(cl) ?
 - Et quel est le pourcentage de bouteilles dont le contenu se situe entre 98(cl) et 102(cl) ?

3. Les aires grisées sont connues. À l'aide de la table, retrouver les abscisses inconnues.

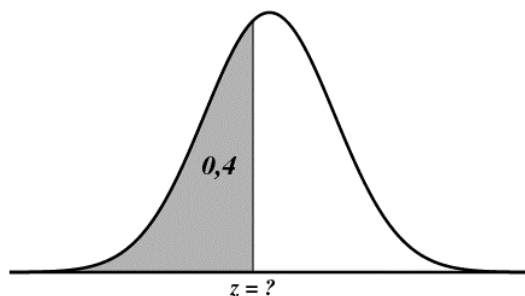
❶



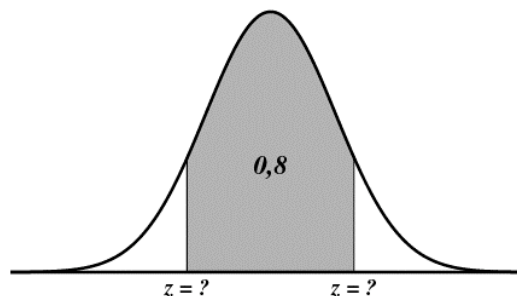
❷



❸



❹



4. Le contrôleur des bouteilles de lait décrète qu'il y a 90 % de bouteilles « normales » : il considère évidemment celles dont le contenu s'écarte le moins de la moyenne. Les bouteilles restantes, représentant 10 % de la production, sont déclarées « anormales », c'est-à-dire avec un contenu insuffisant ou excessif. Préciser pour quelles valeurs du contenu une bouteille est « anormale ».

5. Le producteur de lait, soucieux de sa réputation, tolère au plus 5 % de bouteilles contenant moins de 99 (cl). Il peut atteindre son objectif en réglant la machine de remplissage des bouteilles sur une valeur supérieure à 100 (cl). Laquelle ?

6. Chercher dans la table :

- a) $\Phi(1,96)$ b) $\Phi(2)$ c) $\Phi(-1,96)$ d) $\Phi(-0,04)$

7. À partir de la table, retrouver la valeur de z pour laquelle on a :

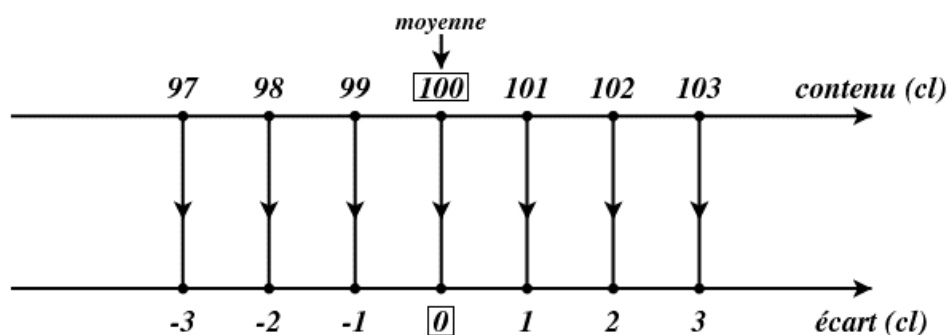
- a) $\Phi(z) \approx 0,96$ b) $\Phi(z) \approx 0,5$ c) $\Phi(z) \approx 0,25$ d) $\Phi(z) \approx 0,04$

8. Un producteur de jus de fruits propose du jus d'orange en boîtes de 50 (cl). Le contenu réel des boîtes est une variable aléatoire normale dont l'écart type vaut 1 (cl). Dans un souci d'économie, le fabricant tolère au plus 2,5 % de boîtes contenant plus de 52 (cl). Sur quelle valeur la machine de remplissage doit-elle être réglée ?

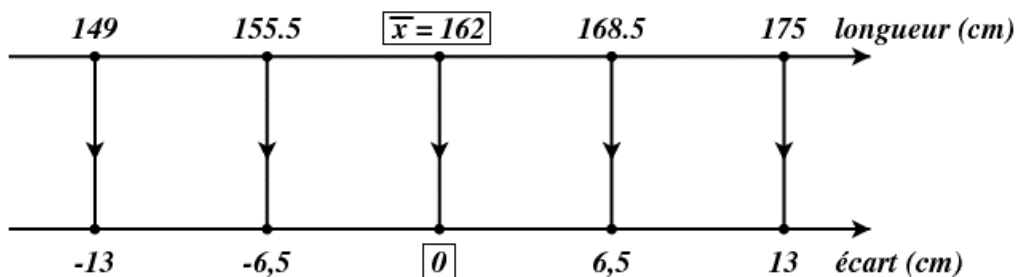
3. D'une loi normale quelconque à la loi normale centrée réduite

Dans la réalité, on ne rencontre pratiquement jamais la loi normale centrée réduite. Pourtant, elle est d'une importance capitale car, comme nous allons le voir, toute loi normale peut se ramener à la normale centrée réduite. Ainsi, avec une seule table, nous allons pouvoir résoudre des problèmes impliquant des lois normales de moyennes et d'écart types quelconques !

Dans le cas des bouteilles de lait, nous avons utilisé un procédé simple pour ramener la moyenne à 0 : au lieu de considérer le contenu réel d'une bouteille, nous avons choisi comme nouvelle variable l'écart entre le contenu et 100 (cl). Cette idée est illustrée par la figure suivante.



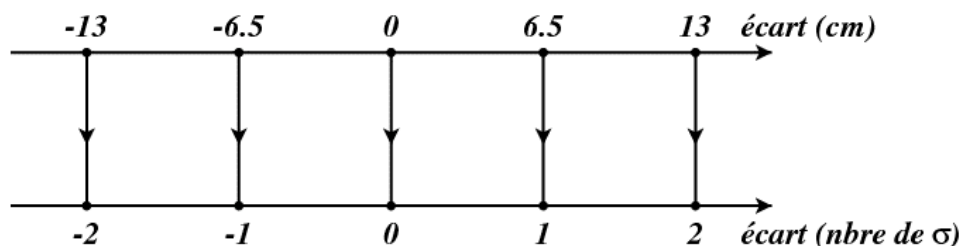
Cette méthode peut aussi s'appliquer aux tailles des Néerlandaises de 1947. Rappelons que la taille moyenne était $\bar{x} = 162$ (cm) et l'écart type $\sigma = 6,5$ (cm).



Dans ce cas, la variable aléatoire représentant l'écart est de loi normale *centrée*, car de moyenne nulle, mais *non réduite*, car l'écart type est différent de 1.

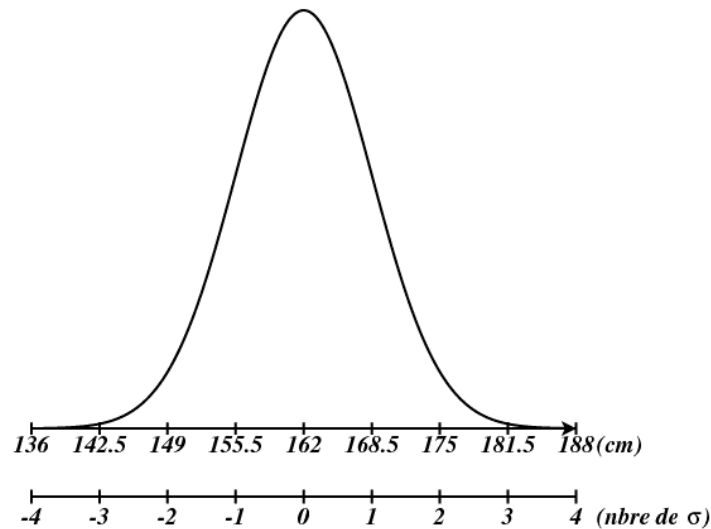
Cet écart type est exprimé en centimètres. Avec une autre unité, nous aurions une autre valeur de σ , par exemple $\sigma = 65$ (mm) ou $\sigma = 0,065$ (m).

Pour exprimer l'écart par rapport à la moyenne, il est astucieux de choisir comme unité ... l'écart type (dans notre cas : 6,5 (cm)) ! L'écart sera ainsi exprimé en multiples de σ .



Par exemple, un écart de 13 (cm) équivaut à un écart de 2 (c'est-à-dire 2σ). Un écart de 2,6 (cm) équivaut à un écart de 0,4 (c'est-à-dire $0,4\sigma$).

Nous avons atteint l'objectif de *réduction* de la loi normale : dans la nouvelle unité, l'écart type de la variable « écart par rapport à la moyenne » vaut 1.



Maintenant, nous pouvons de nouveau nous servir de la table des aires sous la courbe normale centrée réduite !

Exercice résolu

Parmi les 5001 Néerlandaises que nous connaissons bien, 290 mesuraient 163 (cm). En réalité, il faut comprendre que 290 femmes avaient une taille située dans l'intervalle $[162.5, 163.5[$. Elles représentaient environ 5,8 % de l'échantillon.

Vérifier si ce pourcentage est en accord avec le modèle théorique que constitue la loi normale.

Solution

1° Voyons d'abord les écarts entre les extrémités de la classe $[162.5, 163.5[$ et la taille moyenne $\bar{x} = 162$ (cm). Ils valent 0,5 (cm) et 1,5 (cm).

2° Exprimons ces écarts en multiples de σ sachant que $\sigma = 6,5$ (cm).

- $0,5 \text{ (cm)} = \frac{0,5}{6,5} \cdot \sigma \approx 0,08 \cdot \sigma$
- $1,5 \text{ (cm)} = \frac{1,5}{6,5} \cdot \sigma \approx 0,23 \cdot \sigma$

3° La fréquence que nous cherchons, qui est aussi l'aire sous la courbe normale centrée réduite, peut maintenant être trouvée grâce à la table :

$$\Phi(0,23) - \Phi(0,08) \approx 0,5910 - 0,5319 \approx 0,0591 \approx 5,9\%$$

Le modèle normal donne donc un résultat proche de la réalité.

Exercices

1. Utiliser le modèle normal pour répondre aux questions suivantes, concernant les Néerlandaises que nous connaissons bien.
 - a) Quel est le pourcentage de femmes mesurant entre 155,5 (cm) et 168,5 (cm) ?
 - b) Et celui des femmes mesurant entre 149 (cm) et 175 (cm) ?
 - c) Parmi les 5001 clientes, combien peut-on en prévoir avec une taille de 155 (cm) ?
 - d) Comparer chacune des réponses précédentes avec les résultats de l'enquête.

2. Dans un certain pays, la taille moyenne des hommes en âge d'accomplir leur service militaire est de 180 (cm) avec un écart type de 10 (cm). Lors de la visite médicale, 10000 hommes sont contrôlés.
 - a) Estimer le nombre d'hommes mesurant entre 175 (cm) et 185 (cm).
 - b) Si un homme mesure plus de 200 (cm) ou moins de 160 (cm), il est dispensé de service militaire. Prévoir le nombre de dispensés.
 - c) Le ministère de la défense décide de dispenser 10 % des hommes : 5 % car ils sont jugés trop grands, et 5 % car ils sont jugés trop petits. Déterminer la taille au-delà de laquelle un homme sera dispensé, et celle en deçà de laquelle il sera dispensé.

3. Dans un élevage de volailles, les œufs produits ont une masse distribuée selon une loi normale de moyenne 56,3 (g) et d'écart type 7,6 (g).
 - a) Quel est le pourcentage d'œufs de plus de 60 (g) ?
 - b) Les œufs sont répartis en huit catégories, selon leur masse (voir tableau). Déterminer le pourcentage d'œufs dans chaque catégorie.
 - c) Illustrer les résultats précédents en représentant la courbe normale, avec la masse en abscisse. Diviser l'aire sous la courbe en huit parties, chaque partie correspondant à une catégorie. Indiquer le pourcentage d'œufs dans chaque partie.

4. Une entreprise de vente par correspondance livre des colis dont la masse brute (c'est-à-dire emballage compris) est distribuée normalement avec une moyenne de 4,2 (kg) et un écart type de 0,8 (kg). Pour l'année prochaine, les responsables prévoient 20000 livraisons. Compte tenu des tarifs postaux figurant dans le tableau, quel est le montant total des frais de port que l'entreprise devra supporter ?

5. Une poudre est constituée de granulés sphériques dont les diamètres sont distribués normalement avec une moyenne de 170 (μ) et un écart type de 11,6 (μ) (le symbole μ désigne le micron, c'est-à-dire 10^{-6} (m)). Le fabricant propose trois types de poudre (brute, moyenne et fine) et souhaite qu'il y ait le même nombre de granulés dans chaque catégorie.
 - a) Quels devront être les diamètres des trous des tamis qui permettront d'obtenir la répartition souhaitée ?
 - b) Les granulés fins seront-ils distribués normalement ?

6. Une administration communale prévoit de remplacer les lampes de l'éclairage public. La durée de vie des lampes suit une loi normale de moyenne 2500 heures et d'écart type 500 heures.

Pour des raisons économiques, l'administration prévoit le remplacement de *toutes* les lampes lorsque 20 % d'entre elles sont hors service (on ne remplace aucune lampe avant cela).

Dans combien de temps cela arrivera-t-il ?



7. Une sucrerie conditionne du sucre fin en paquets d'un kilo. Des échantillons sont prélevés dans la production. Ils révèlent que le poids net des paquets est distribué normalement avec une masse moyenne qui est effectivement de 1 kilo, et avec un écart type de 15 grammes. Le fabricant veut que 95 % des paquets contiennent au moins 1 kilo de sucre. Sur quel poids net, au gramme près, faut-il régler la machine de remplissage des paquets ?



Conditionnement du sucre et contrôle de qualité à la raffinerie de Tirlemont

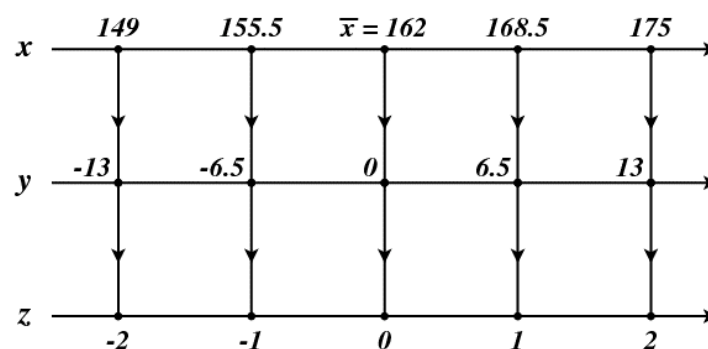
8. La moyenne des capacités respiratoires d'un échantillon de 400 hommes adultes est de 3,7 litres avec un écart type de 0,7 litre. Sachant que les capacités respiratoires sont distribuées suivant une loi normale, combien peut-on s'attendre à trouver de ces personnes ayant une capacité respiratoire comprise entre 3,4 et 4 litres ?

9. Une étude faite par un fabricant de lampes électriques a montré que les durées de vie des lampes suivent une loi normale de moyenne 1200 heures et d'écart type 180 heures. On considère une installation de 10000 lampes neuves.
- Combien peut-on prévoir de lampes hors d'usage au bout de 1000 heures ?
 - Combien peut-on prévoir de lampes qui deviendront hors d'usage entre la 1000^{ème} et la 1500^{ème} heure ?
 - Après combien d'heures y aura-t-il 10 % des lampes hors d'usage ?
-
10. Une entreprise fabrique en grande série des pièces cylindriques en acier. Leur diamètre doit mesurer 6(cm) avec une tolérance de 0,01(cm). À partir d'un échantillon, on a établi que le diamètre moyen des pièces était effectivement de 6(cm) avec un écart type de 0,008(cm).
- Calculer le pourcentage de pièces dont le diamètre est :
 - inférieur à 6,008(cm) ;
 - supérieur 6,008(cm) ;
 - inférieur à 5,995(cm) ;
 - compris entre 5,995(cm) et 6,008(cm) ;
 - compris entre 5,99(cm) et 6,01(cm).
 - Rechercher entre quelles limites se situe 70 % de la production.
-

Une formule de transformation

Pour résoudre les exercices précédents, nous transformons des unités habituelles (*cm, kg, heures, ...*) en une nouvelle unité : l'écart type. Ce changement d'échelle permet de se ramener à la loi normale centrée réduite, quelle que soit la loi normale de départ.

Reprenons l'exemple bien connu.



Nous passons d'abord de la variable x (taille de la personne) à la variable y (écart par rapport à la moyenne) ; nous avons $y = x - 162$.

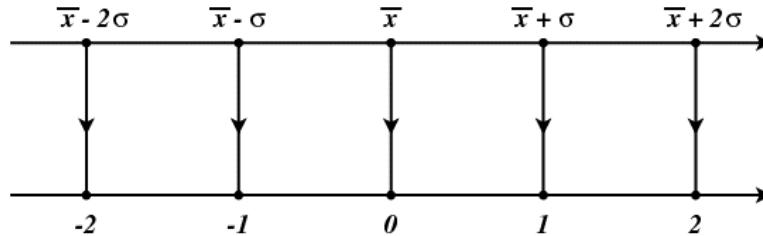
Ensuite, nous passons de la variable y à la variable z (écart par rapport à la moyenne exprimé en multiple de l'écart type) ; nous avons $z = \frac{y}{6,5}$.

La variable z suit une loi normale centrée réduite, et nous avons la relation $z = \frac{x - 162}{6,5}$.

Généralisons ...

Soit x une variable aléatoire normale de moyenne \bar{x} et d'écart type σ . Le passage à la variable aléatoire z de moyenne 0 et d'écart type 1, s'obtient par la formule :

$$z = \frac{x - \bar{x}}{\sigma}$$

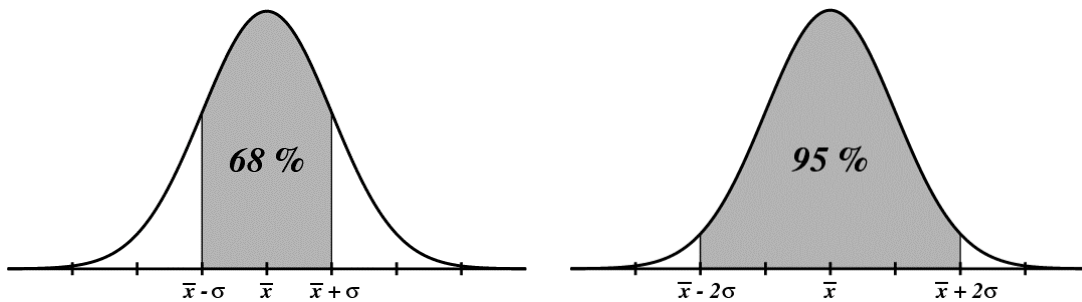


Voici deux règles – approximatives – très connues en statistique :

On considère un échantillon représentatif d'une population. Pour chaque individu de l'échantillon, on mesure une caractéristique donnée (la taille par exemple).

Si cette caractéristique est distribuée selon une loi normale dans la population, on peut s'attendre à trouver dans l'échantillon :

- 68 % d'individus dont la mesure ne s'écarte pas de plus d'un écart type par rapport à la moyenne ;
- 95 % d'individus dont la mesure ne s'écarte pas de plus de deux écarts types par rapport à la moyenne.



Exercice : vérifier ces deux résultats, avec davantage de précision, à l'aide de la table des aires pour la courbe normale centrée réduite.

Exercice

Un brasseur fournit des fûts de bière permettant de servir entre 195 et 205 verres de 25(cl). Il affirme que 95 % des fûts satisfont à cette exigence.

Supposons que les contenus des fûts sont distribués normalement. Sur quel contenu moyen et sur quel écart type (tous deux en litres) le brasseur s'est-il probablement basé ?



4. Approximation normale d'une loi binomiale

4.1. Espérance mathématique et écart type d'une loi binomiale

Rafraîchissons-nous la mémoire : dans un schéma *binomial*, on procède à une expérience aléatoire ne possédant que *deux* résultats possibles (« succès » et « échec »), et cette expérience est répétée *n* fois.

Soit *X* le nombre de succès récoltés. Soit *p* la probabilité de succès et *q* la probabilité d'échec. La probabilité de récolter *k* succès sur les *n* expériences est :

$$P(X = k) = C_n^k \cdot p^k \cdot q^{n-k} \text{ avec } 0 \leq k \leq n$$

Nous allons maintenant calculer l'espérance mathématique et la variance d'une loi binomiale dans un cas particulier.

Une urne contient 7 boules blanches et 3 boules noires. On extrait une boule de l'urne, on note sa couleur, on la replace dans l'urne et on réalise cette expérience 10 fois. Soit *X* la variable aléatoire « nombre de fois qu'une boule blanche est apparue ».

Espérance mathématique de *X*

Quelle est la valeur de *X* la plus probable, celle que l'on s'attend le plus à obtenir après les 10 tirages ?

Etant donné que la proportion de boules blanches dans l'urne est de 70 %, on s'attend à retrouver la même proportion dans notre tirage, à savoir 7 sur 10.

Vérifions en calculant l'espérance mathématique de *X* à l'aide de la formule vue précédemment :

$$E(X) = P(X = 0) \times 0 + P(X = 1) \times 1 + P(X = 2) \times 2 + \dots + P(X = 10) \times 10$$

Dans le tableau suivant, la lettre *k* désigne le nombre de boules blanches obtenues ; il peut prendre des valeurs de 0 à 10 incluses. Dans la deuxième colonne, on trouve la probabilité d'obtenir *k* boules blanches sur 10 tirages. La troisième colonne donne le produit de chaque valeur de *k* par la probabilité correspondante. La somme de tous ces produits est l'espérance mathématique de *X*.

k	$P(X = k)$	$P(X = k) \times k$
0	0,000006	0
1	0,000138	0,000138
2	0,001447	0,002893
3	0,009002	0,027005
4	0,036757	0,147028
5	0,102919	0,514597
6	0,200121	1,200726
7	0,266828	1,867796
8	0,233474	1,867796
9	0,121061	1,089547
10	0,028248	0,282475
		$\Sigma = 7$

Variance et écart type de X

En statistique descriptive, nous avons vu que la variance est un indice permettant de mesurer la dispersion des valeurs d'une variable autour de sa moyenne. Plus précisément, la variance est la moyenne des carrés des écarts entre les valeurs de la variable et la moyenne :

$$V = \sum_{i=1}^{i=n} f_i \cdot (x_i - \bar{x})^2 .$$

La variance d'une variable aléatoire est donnée par une formule semblable, où la fréquence f_i et la moyenne \bar{x} sont remplacées respectivement par la probabilité $P(X = x_i)$ et l'espérance mathématique $E(X)$:

$$V(X) = \sum_{i=1}^{i=n} P(X = x_i) \cdot (x_i - E(x))^2$$

Voici le tableau de calcul de $V(X)$.

$k = x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$P(X = x_i)$	$P(X = x_i) \cdot (x_i - E(x))^2$
0	-7	49	0,000006	0,000289
1	-6	36	0,000138	0,004960
2	-5	25	0,001447	0,036168
3	-4	16	0,009002	0,144027
4	-3	9	0,036757	0,330812
5	-2	4	0,102919	0,411677
6	-1	1	0,200121	0,200121
7	0	0	0,266828	0,000000
8	1	1	0,233474	0,233474
9	2	4	0,121061	0,484243
10	3	9	0,028248	0,282475
				$\Sigma = 2,1$

Exercice à réaliser avec un tableur

Calculer l'espérance mathématique et la variance de X lorsque $n = 20$ et lorsque $n = 100$.

Réponses : - pour $n = 20$: $E(X) = 14$ et $V(X) = 4,2$

- pour $n = 100$: $E(X) = 70$ et $V(X) = 21$

Nous admettrons que l'espérance mathématique et la variance d'une variable aléatoire X de loi binomiale $B(n, p)$ sont données par les formules :

$$E(X) = np \quad \text{et} \quad V(X) = npq$$

L'écart type est la racine carrée positive de la variance : $\sigma(X) = \sqrt{npq}$.

4.2. Approximation normale d'une loi binomiale

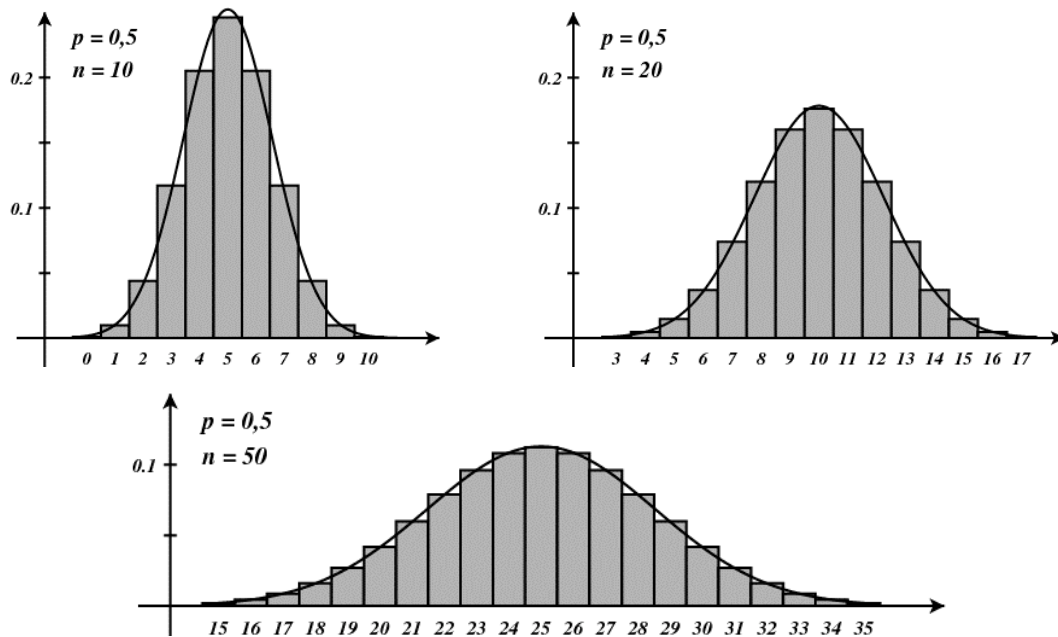
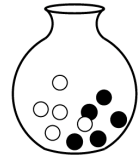
Exemple

Une urne contient des boules blanches et des boules noires. On tire avec remplacement n boules de l'urne et on s'intéresse à la variable aléatoire X « nombre de boules blanches obtenues ». Soit p la probabilité d'obtenir une boule blanche lors d'un tirage. La variable aléatoire X est donc une variable binomiale $B(n, p)$.

Première situation

L'urne contient 5 boules blanches et 5 boules noires ($p = 0,5$).

Tirons n boules de l'urne et réalisons l'histogramme de la loi de probabilité de X .



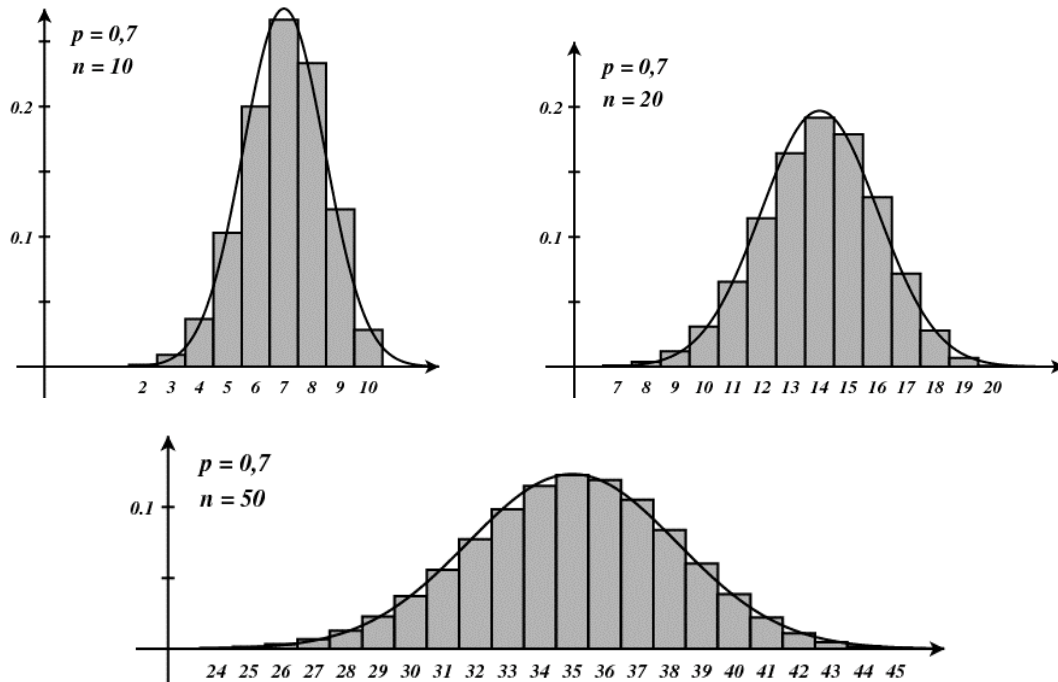
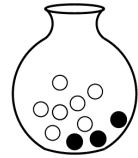
Commentaires

- Les histogrammes sont tous symétriques car une boule blanche a autant de chances d'être tirée qu'une noire ($p = 0,5$).
- Pour $n = 20$, il est très peu probable d'obtenir 0, 1, 2, 18, 19 ou 20 boules blanches. Les rectangles correspondants semblent invisibles car leur hauteur est proche de 0. C'est pour une raison identique que les rectangles visibles sur le troisième histogramme correspondent à un nombre de boules blanches compris entre 15 et 35.
- Les rectangles les plus hauts correspondent à l'espérance mathématique de X . Pour rappel $E(X) = np$, ce qui donne respectivement 5, 10 et 25 pour les trois cas illustrés ci-dessus.
- Lorsque n augmente, la variance augmente aussi, ce qui explique « l'aplatissement » de l'histogramme. Pour rappel, $V(X) = npq$, ce qui donne respectivement 2,5, 5 et 12,5 pour les trois cas illustrés ci-dessus.
- Une courbe normale de même moyenne et de même variance que la binomiale a été superposée à chaque histogramme. Nous constatons que l'aire d'une partie de l'histogramme peut être approchée par une aire sous la courbe normale. Cette approximation est d'autant meilleure que n est grand. Toutefois, avec $p = 0,5$, la symétrie de l'histogramme aidant, l'approximation est déjà satisfaisante pour $n = 10$.

Deuxième situation

L'urne contient 7 boules blanches et 3 boules noires ($p = 0,7$).

Tirons n boules de l'urne et réalisons l'histogramme de la loi de probabilité de X .



Commentaires

- Comme il y a plus de chances de tirer une boule blanche qu'une noire, les histogrammes ne sont plus symétriques.
- Par rapport à la première situation, les histogrammes ont « glissé » vers la droite. En effet, les nombres de boules blanches les plus probables sont maintenant 7, 14 et 35 (au lieu de 5, 10 et 25).
- L'approximation par une courbe normale semble plus douteuse pour $n = 10$. Par contre, pour $n = 20$ et surtout pour $n = 50$, les histogrammes retrouvent une certaine symétrie et cette approximation redevient tout à fait raisonnable.

Troisième situation

L'urne contient 9 boules blanches et 1 boule noire ($p = 0,9$).

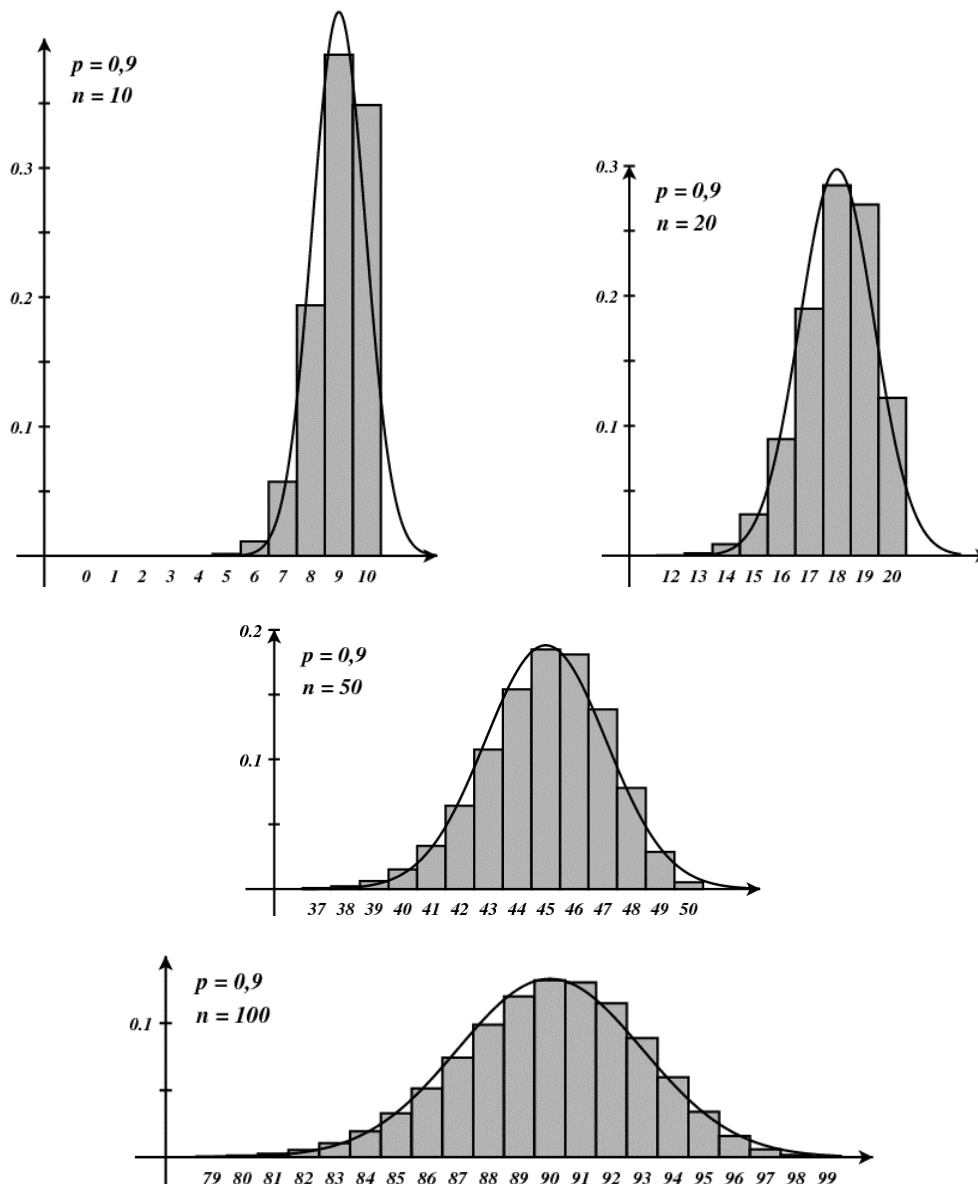
Réalisons de nouveau les histogrammes des lois de probabilités de X , pour $n = 10$, $n = 20$ et $n = 50$.



Anticipons ...

- Maintenant que le tirage d'une boule blanche est nettement plus probable que celui d'une noire, nous pouvons nous attendre à des histogrammes encore plus décalés vers la droite, et beaucoup plus asymétriques pour de petites valeurs de n .
- Il faut attendre des valeurs de n beaucoup plus grandes pour retrouver une certaine symétrie et pour que l'approximation par une courbe normale soit satisfaisante.

Voyons ...



Nous admettrons ce qui suit :

Si n est suffisamment grand, une variable aléatoire X de loi *binomiale* $B(n, p)$ est approximativement de loi *normale* de moyenne np et de variance $npq = np(1 - p)$.
L'approximation est bonne dès que $np > 5$ et $nq > 5$.

Considérons par exemple, une variable aléatoire X de loi binomiale $B(40, 0,3)$.

Nous avons $np = 40 \cdot 0,3 = 12 > 5$ et $nq = 40 \cdot 0,7 = 28 > 5$. Nous pouvons donc procéder à l'approximation normale.

Comme $E(X) = np = 12$ et que $V(X) = npq = 8,4$, notre variable aléatoire pourra être considérée de loi normale $N(12, 8,4)$.

Sous certaines conditions, nous allons maintenant pouvoir calculer des probabilités concernant des variables binomiales grâce à la loi normale.

C'est un grand avantage, car s'il est impossible de disposer de tables binomiales pour des valeurs quelconques de n et de p , nous pouvons en revanche traiter n'importe quelle loi normale à l'aide de la table de la normale centrée réduite !

Exercice résolu

On estime que 20 % des foreuses bon marché produites par une firme d'outillage sont défectueuses. Un supermarché reçoit un lot de 100 foreuses.

Quelle est la probabilité qu'il y ait au plus 25 machines défectueuses ?

Solution

Soit X la variable aléatoire « nombre de foreuses défectueuses ». Chaque machine peut être défectueuse (avec une probabilité de 0,2) ou opérationnelle (avec une probabilité de 0,8). Nous en déduisons que X est une variable aléatoire binomiale $B(100,0.2)$.

La probabilité demandée est $P(X \leq 25)$. Elle se trouve dans les tables binomiales cumulatives :

$$P(X \leq 25) \approx 0,9125$$

Voyons maintenant comment utiliser l'approximation normale.

Nous avons d'abord : $E(X) = 100 \cdot 0,2 = 20$ et $V(X) = 100 \cdot 0,2 \cdot 0,8 = 16$.

Nous utiliserons donc la normale $N(20,16)$ pour approcher la binomiale (notons au passage que les conditions sont remplies : $np = 20 > 5$ et $nq = 80 > 5$).

La variable normale $N(20,16)$ a pour moyenne 20 et pour écart type $\sqrt{16} = 4$.

Pour passer à la normale centrée réduite $N(0,1)$, effectuons le changement de variable bien connu :

$$z = \frac{x - \bar{x}}{\sigma} \quad \text{ou} \quad Z = \frac{X - E(X)}{\sigma(X)} = \frac{X - 20}{4}$$

Dès lors : $P(X \leq 25) = P(X \leq 25,5)$ (correction de continuité : voir page suivante)

$$= P\left(Z \leq \frac{25,5 - 20}{4}\right)$$

$$= P(Z \leq 1,375)$$

$$\approx \boxed{0,91545} \quad (\text{via la table normale centrée réduite})$$

Notons que la table donne les valeurs $P(Z \leq 1,37) \approx 0,9147$ et $P(Z \leq 1,38) \approx 0,9162$.

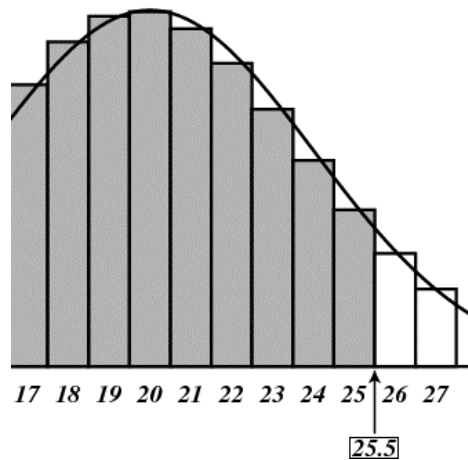
Pour trouver $P(Z \leq 1,375)$, nous avons pris la moyenne arithmétique de ces deux résultats (c'est un cas particulier d'*interpolation linéaire*).

Remarquons que l'approximation normale ne diffère que de 0,00295 du résultat binomial !

Conclusion : il y a environ 91 % de chances d'avoir au plus 25 foreuses défectueuses dans ce lot de 100 foreuses.

Remarque : la correction de continuité

La figure ci-dessous explique pourquoi nous avons remplacé $P(X \leq 25)$ par $P(X \leq 25,5)$.



Pour évaluer l'aire de l'histogramme jusqu'au rectangle 25 inclus, il est plus précis de calculer l'aire sous la courbe normale jusqu'à 25,5 plutôt que jusqu'à 25.

Cette correction est due au fait que nous approchons une variable aléatoire discrète (la binomiale) par une variable aléatoire continue (la normale).

Pour de grandes valeurs de n , l'influence de cette correction est minime et on la néglige parfois. Néanmoins, il est plus prudent de l'appliquer.

Ce procédé permet aussi d'évaluer la probabilité d'une valeur particulière de X .

Par exemple : $P(X = 25) = P(24,5 \leq X \leq 25,5)$.

Exercices

1. Evaluer à l'aide de la table normale centrée réduite :

- a) $P(X \leq 20)$ avec $p = 0,4$ et $n = 50$
- b) $P(X \leq 320)$ avec $p = 0,75$ et $n = 400$
- c) $P(X \leq 20)$ avec $p = 0,5$ et $n = 25$

2. Selon la police d'une certaine ville, un véhicule sur quatre n'est pas en ordre du point de vue du contrôle technique.

Les policiers décident de contrôler 60 véhicules.

Quelle est la probabilité pour qu'il y ait au moins 20 véhicules en infraction ?



3. Parmi les 3000 premiers enfants nés à Graz (Autriche) en 1962, il y avait 1578 garçons. Si l'on suppose qu'il y a autant de chance qu'un nouveau né soit un garçon ou une fille, quelle est la probabilité pour qu'il y ait au moins 1578 garçons sur 3000 naissances ? Commenter la réponse.
-

4. Au cours de l'année 1976, environ 177000 enfants naquirent aux Pays-Bas.
- a) D'après les statistiques, la probabilité que le nouveau-né soit un garçon était de 51,5% . Combien de garçons pouvait-on donc prévoir parmi ces enfants ?
- b) En réalité, il y eut environ 91000 garçons et 87000 filles. Cette répartition est-elle exceptionnelle ? Quelle était la probabilité qu'il n'y ait pas plus de 91000 garçons ?
-

5. Une entreprise achète des fusibles électriques par lots de 8000 pièces. Lors de la réception, un échantillon de 50 pièces est extrait du lot et envoyé à l'inspection. Si les 50 fusibles fonctionnent correctement, le lot est accepté, sinon il est refusé. Supposons qu'un lot comprenant 5 % de pièces défectueuses arrive à l'entreprise. Calculer la probabilité que ce lot de 8000 pièces soit accepté ?³



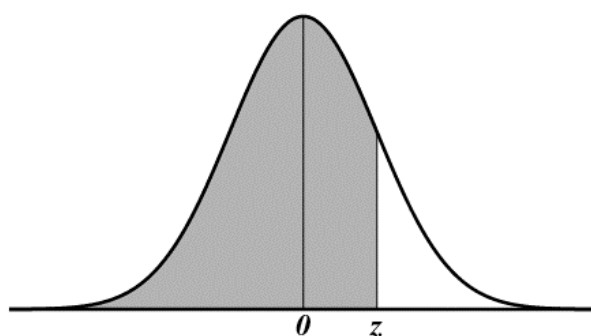
6. Sur deux cent lancers d'un dé, un joueur a obtenu cinquante fois le « 6 » . Ce dé est-il bien équilibré ? Calculer la probabilité d'obtenir au moins cinquante fois le « 6 » . Commenter.
-

7. Un joueur lance cent fois une pièce de monnaie bien équilibrée. Soit X la variable aléatoire binomiale « nombre de faces obtenues ».
- a) Calculer $E(X)$ et $\sigma(X)$.
- b) Calculer la probabilité d'obtenir entre 40 et 60 fois « face ».
- c) Calculer la probabilité d'obtenir entre 45 et 55 fois « face ».
- d) Pouvait-on trouver les deux dernières réponses plus rapidement ? Se rappeler une propriété de la loi normale.
-

8. Un joueur décide de lancer n fois une pièce de monnaie bien équilibrée. Il annonce que la probabilité qu'il obtienne entre 49 % et 51 % de « faces » est égale à 0,95 . Combien vaut n ?

³ Remarque : la constitution de l'échantillon est évidemment un tirage sans remise. Pourtant, si nous calculons comme s'il s'agissait d'un tirage avec remise, l'erreur sera minime. Pourquoi ?

Table des aires soutendues par la courbe normale centrée réduite



z	0	1	2	3	4	5	6	7	8	9
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5754
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	.7258	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7518	.7549
0.7	.7580	.7612	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7996	.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
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.7	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.8	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000