1 Présentation de la loi normale

1.1 Courbe en cloche

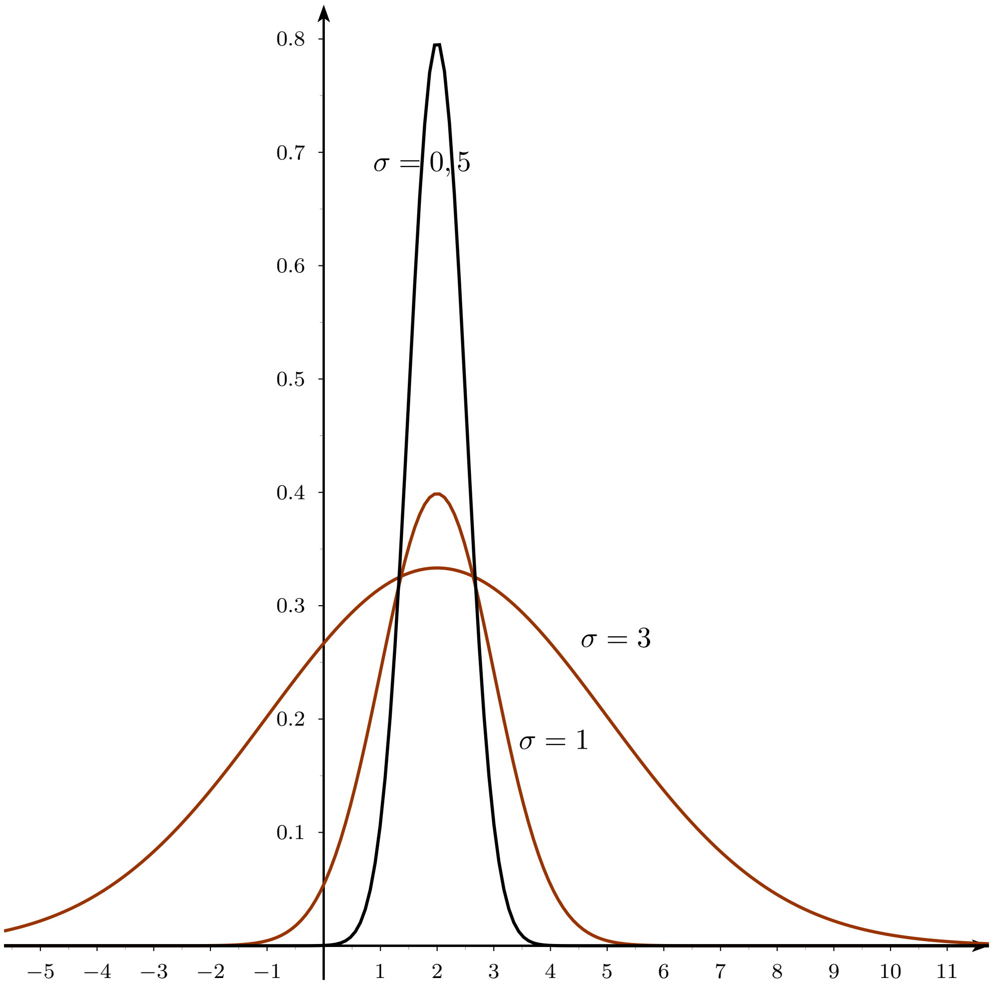
Définition

La courbe "en cloche" de paramètres μ et σ a les propriétés suivantes :

- elle est représentative d'une fonction définie sur $\mathbb{R}$, à valeurs strictement positives ;
- elle est symétrique par rapport à la droite d'équation $x = \mu$.
- l'aire de la surface située entre la courbe et l'axe des abscisses est égale à 1.

Exemple

Les courbes ci-contre ont le même paramètre $\mu = 2$.



Plus σ est petit, plus la courbe est resserrée autour de la moyenne μ.

Plus σ est grand, plus la courbe est "écrasée" et deployée par rapport à l'axe des abscisses.

1.2 Propriétés d'une variable aléatoire qui suit une loi normale

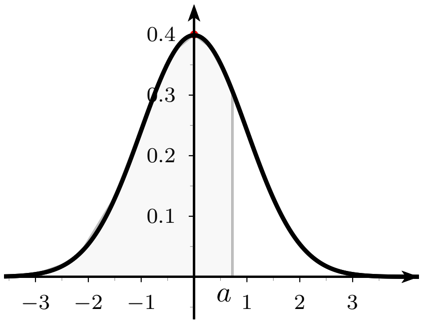
Propriétés

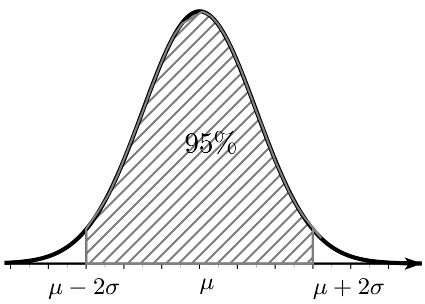
Lorsque X est une variable aléatoire qui suit la loi normale de paramètres μ et σ notée $N(\mu;\sigma)$:

- X peut prendre pour valeur tout nombre réel.

- X est associée à la courbe "en cloche" de paramètres μ et σ (la fonction dont cette courbe est représentative est appelée densité).

- Pour tout réel a, la probabilité $P(X \leqslant a)$ que X prenne des valeurs inférieurs ou égale à a est égale à l'aire de la surface, située entre la courbe de densité et l'axe des abscisses, à gauche de la droite d'équation $x = a$.
- $P(\mu - 2\sigma \leqslant X \leqslant \mu + 2\sigma) \simeq 0,95$; l'intervalle $[\mu - 2\sigma; \mu + 2\sigma]$ est appelé intervalle de fluctuation de la variable alatoire X au seuil approximatif de 95 %.





Bon à savoir :

- Pour tout réel a, $P(X = a) = 0$ donc $P(X < a) = P(X \leqslant a)$. De même,$P(X > a) = P(X \geqslant a)$

- $P(X \leqslant \mu) = P(X \geqslant \mu) = 0,5$

Exercice 1

X suit la loi normale $N(15; 2)$. On veut donner une valeur approchée à la calculatrice de $P(2 \leqslant X \leqslant 16)$.

1. Avec Texas Instruments
   Sur TI 83fr : 2ND VARS NormalFRep(2,16,15,2) NormalFRep (a, b, μ, σ)
   Sur TI 84 : 2ND VARS Normalcdf(2,16,15,2) Normalcdf (a, b, μ, σ)
2. Avec Casio Graph35+
   MENU STAT DIST NormCD NormCD(2,16,2,15)) NormCD(a, b, σ, μ)

On doit trouver : $P(2 \leqslant X \leqslant 16) \simeq 0,6915$.

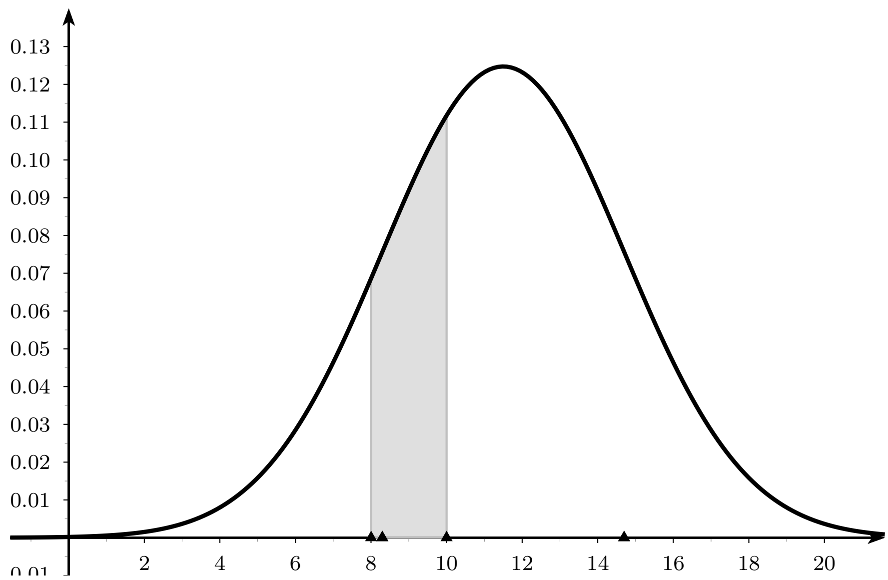
Exercice 2

X suit la loi normale $N(15; 4)$.

1. $P(15 \leqslant X) = \dots\dots\dots$
2. $P(10 \leqslant X \leqslant 20) = \dots\dots\dots$
3. $P(X \leqslant 16) = \dots\dots\dots$

Exercice 3

Un chercheur a étudié l'âge moyen auquel les premiers mots du vocabulaire apparaissent chez les jeunes enfants. L'étude montre que l'âge X d'apparition (en mois) des premiers mots suit une loi normale de moyenne 11,2 et d'écart-type 3,2.
On a tracé la courbe représentant la densité de la loi $N(11,5;3,2)$. Répondre à chaque question en montrant de quelle aire il s'agit.



1. Donner la probabilité qu'un enfant ait prononcé ses premiers mots entre 8 et 10 mois.
2. Donner la probabilité qu'un enfant ait prononcé ses premiers mots avant 7 mois.
3. Donner la probabilité qu'un enfant ait prononcé ses premiers mots après 10 mois.
4. Déterminer l'intervalle de fluctuation de la variable alatoire X au seuil approximatif de 95 %. »

2 Distinction entre échantillonnage et estimation

Propriété

- On considère une population d'individus. Lorsque l'on connaît ou lorsque l'on fait une hypothèse sur la proportion p d'individus ayant une caractéristique donnée dans une population et que l'on effectue un nombre n de tirages avec remise dans cette population, la fréquence observée appartient avec une certaine probabilité à un intervalle appelé intervalle de fluctuation de centre p et de longueur qui diminue lorsque n augmente. On parle alors de situation d'échantillonnage.
- Lorsque l'on ne connaît pas la proportion d'individus ayant une caractéristique donnée, en procédant à un nombre n de tirages avec remise on peut estimer à l'aide de la fréquence f obtenue la proportion p d'individus ayant cette caractéristique.
  Cette estimation se fait à l'aide d'un intervalle de confiance dont l'amplitude diminue lorsque le nombre n de tirages augmente.

3 Échantillonnage

On suppose p connue. On prélève au hasard et avec remise un échantillon de n individus de la population et on note f la fréquence d'apparition du caractère étudié.

Propriété

Pour cette valeur de p, l'intervalle $\left[p - \frac{1}{\sqrt{n}}\ ;\ p + \frac{1}{\sqrt{n}}\right]$ est un intervalle de fluctuation à au moins 95 % de la fréquence f.(dans au moins 95 % des cas, f appartient à cet intervalle)

Remarque :
En pratique, on utilise cette propriété dès que n est assez grand et p pas trop proche de 0 ou de 1.

Test d'hypothèse :
On considère une population dans laquelle on suppose que la proportion d'un caractère est p. On fait l'hypothèse « La proportion dans la population est p »
On observe la fréquence f d'apparition de ce caractère sur un échantillon de taille n et on calcule l'intervalle de fluctuation I à au moins 95 %.

- Si $f \in I$, au risque de 5 % d'erreur (ou au seuil de confiance de 95 %), on accepte l'hypothèse que la proportion dans la population est f.
- Si $f \notin I$, au risque de 5 % d'erreur on rejette l'hypothèse que la proportion dans la population est p.

Exemple : Un fournisseur d'accès à Internet affirme que, sur sa hotline, seuls 20 % des clients attendent plus de 5 minutes pour obtenir un interlocuteur.
Une association de consommateurs interroge au hasard 200 personnes ayant eu à s'adresser à cette hotline. Parmi ces personnes, 53 ont dû attendre plus de 5 minutes. Peut-on mettre en doute l'affirmation du fournisseur d'accès ?

On a $p : 0,2$; $n = 200$.
Les conditions sont vérifiées. L'intervalle de fluctuation est $I = \left[0,2 - \frac{1}{\sqrt{200}}\ ;\ 0,2 + \frac{1}{\sqrt{200}}\right] \approx [0,129\ ;\ 0,271]$.
L'hypothèse à tester est « 20 % des clients attendent plus de 5 minutes ». La fréquence observée est $f = \frac{53}{200} \approx 0,265$.
$f \in [0;129;0;271]$..
Donc au risque de 5 %, on ne peut pas mettre en doute l'affirmation du fournisseur d'accès.

4 Estimation

Propriété

Soit p la proportion inconnue d'apparition d'un caractère. On appelle f la fréquence d'apparition du caractère sur un échantillon de taille n. Alors, l'intervalle $\left[f - \frac{1}{\sqrt{n}}\ ;\ f + \frac{1}{\sqrt{n}}\right]$ contient pour n assez grand la proportion p avec une probabilité supérieure ou égale à 0,95. L'intervalle $\left[f - \frac{1}{\sqrt{n}}\ ;\ f + \frac{1}{\sqrt{n}}\right]$ est appelé intervalle de confiance de p au niveau de confiance 95 %.

Remarque : Un intervalle de confiance au niveau de 95 % a une amplitude de $\frac{2}{\sqrt{n}}$.
L'amplitude diminue lorsque la taille n de l'échantillon augmente.

Exemple : Lors d'un scrutin électoral, on souhaite connaître la proportion p de français votant pour le candidat « A ».
L'institut SOFOS mène une enquête auprès de 1000 personnes tirées au hasard et avec remise (c'est-à-dire qu'une même personne peut éventuellement être choisie plusieurs fois). Le résultat indique qu'une proportion $f = 49\,\%$ d'entre elles voteront pour le candidat « A ».

1. Quelle est la loi suivie par le nombre de personnes votant pour « A » dans une enquête avec remise effectuée auprès de 1 000 personnes ?
2. Étant donné le résultat de l'enquête SOFOS. donner un intervalle de confiance à 95 % pour l'estimation de p.
3. Peut-on affirmer d'après l'enquête que le candidat « A » n'aura pas la majorité des votes, c'est- à-dire que p $<$ 0,5 ?

Solution

1. A chaque étape de l'enquête, il y a une probabilité p de tirer une personne votant pour « A », car on procède à des tirages avec remise. De plus, les 1 000 tirages sont effectués de façon indépendante. L'enquête est donc un schéma de Bernoulli dont chaque succès correspond à tirer une personne votant pour « A »,
Le nombre de personnes votant pour « A » dans l'enquête suit donc une loi binomiale $B(1000\ ;\ p)$ où p est inconnu.
2. Notons F la proportion de personnes votant pour « A » parmi 1 000.
D'après le cours, un intervalle de confiance au niveau 95 % pour p est $\left[F - \frac{1}{\sqrt{1\,000}}\ ;\ F + \frac{1}{\sqrt{1\,000}}\right]$.
L'intervalle de confiance est $\left[0,49 - \frac{1}{\sqrt{1\,000}}\ ;\ 0,49 + \frac{1}{\sqrt{1\,000}}\right] = [0,458\ ;\ 0,522]$.
3. L'intervalle de confiance précédent montre qu'il est tout à fait possible que $p \geqslant 0,5$.
Donc il n'est pas raisonnable d'affirmer suite à l'enquête que le candidat « A » n'obtiendra pas majorité des votes.