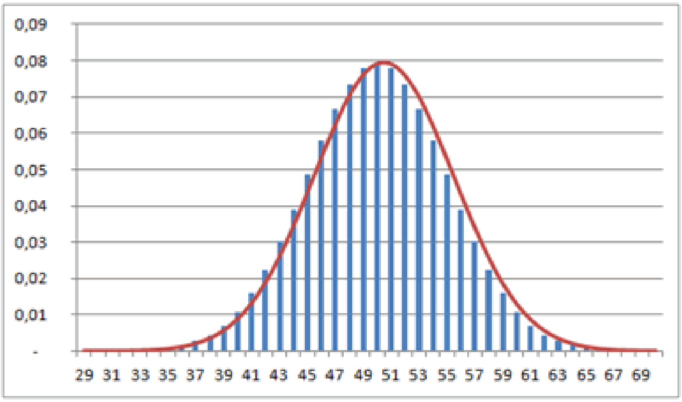
Les lois normales

Du discret au continu :

Considérons une variable aléatoire discrète X suivant la loi binomiale de paramètres $n = 100$ et $p = 0,5$, et dont voici la loi de probabilité en diagramme en bâtons.



Comme X suit la loi $B(n = 100 ; p = 0,5)$, on a alors $E(X) = np = 50$ pour son espérance, la moyenne attendue sur un grand nombre de répétitions de la même expérience, et $V(x) = n\,p\,(1-p) = 25$ pour sa variance, et donc $\sigma = \sqrt{V(X)} = 5$ pour son écart-type.

Lorsqu'on étudie la loi bonomiale sur un grand nombre d'expériences ($n > 50$ par exemple), et à condition que la probabilité de succès sur une expérience ne soit pas trop petite ($p > 0,1$), on peut approximer de manière assez fiable cette loi binomiale par une loi normale dont la représentation est une courbe en cloche ou courbe de Gauss.
On passe ainsi d'une distribution discrète à une distribution continue, beaucoup plus souple.

Sur le graphique précédent, la courbe en rouge est la courbe de Gauss associée à la loi normale de moyenne $\mu = 50$ et d'écart type $\sigma = 5$.

Cette loi normale intervient dans de nombreuses distributions statistiques, lorsqu'un critère d'un indinidu – par exemple la taille d'une femme adulte – dépend d'un grand nombre de facteurs ou paramètres. La répartition de la taille d'une femme adulte dans une population suit une loi normale.

<u>Notation</u> : Lorsqu'on écrira $\int_a^{+\infty} f(t).dt$, il faut comprendre que c'est un réel égal à $\lim\limits_{x \to +\infty} \left(\int_a^x f(t).dt \right)$, ce qui suppose que cette limite existe et est finie.
Même idée pour le réel $\int_{-\infty}^{b} f(t).dt$.

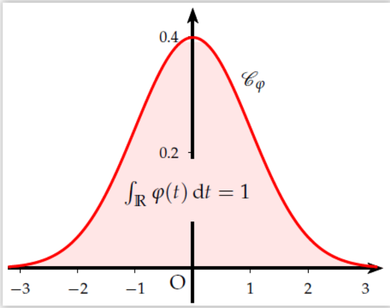
1) Loi normale centrée réduite :

Définissons tout d'abord la fonction densité correspondante.

Définition 1 : On appelle **densité de probabilité de Laplace-Gauss** la fonction ϕ définie sur $\mathbb{R}$ par :

$$\phi(t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}.$$

<u>*Remarques*</u> : Cette fonction correspond bien à une densité de probabilité car :



- ϕ est continue et positive sur $\mathbb{R}$
- ϕ est paire et admet un maximum en zéro : $\phi(0)=\frac{1}{\sqrt{2\pi}}\approx 0,4$
- l'intégrale de ϕ sur $\mathbb{R}$ vaut 1 (propriété admise)

Courbe en cloche ou, **Courbe de Gauss**.

- En revanche, il n'existe pas de primitive explicite de la fonction e^{-t^2}, donc de ϕ.
  Ainsi, pour calculer les probabilités associèes à une loi normale, il faut soit disposer d'un tableau des valeurs de la fonction de répartition, soit utiliser la calculatrice.

- Comme la fonction ϕ est paire, on a :
  $\int_{-a}^{0}\phi(t).dt=\int_{0}^{a}\phi(t).dt$ pour tout réel a, y compris négatif

  et donc naturellement $\int_{-\infty}^{0}\phi(t).dt=\int_{0}^{\infty}\phi(t).dt=\frac{1}{2}$ puisque l'aire totale vaut 1.

Définissons maintenant une variable aléatoire suivant une loi normale centrée réduite.

Définition 2 :

On dit qu'une variable aléatoire **continue** X suit la loi normale centrée réduite, notée $N(0;1)$, si sa densité de probabilité est la fonction ϕ précédente.

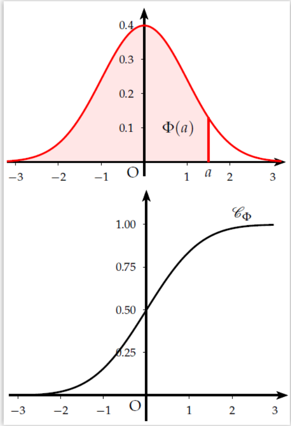
Sa **fonction de répartition** Φ est alors définie par :

$$\Phi(x)=P(X\leqslant x)\text{, ce qui s'écrit aussi : }\Phi(x)=\int_{-\infty}^{x}\phi(t).dt.$$

Remarques :

- Le nombre $\Phi(a)$ est l'aire sous la courbe jusqu'à $x=a$.
- La fonction Φ peut être considérée comme la primitive de la fonction ϕ qui vérifie $\Phi(0)=0,5$.
- Avant l'arrivée des calculatrices, on utilisait des tables de valeurs.

On observe en dessous la courbe de la fonction de répartition Φ.



Propriété 1 : Calcul des probabilités.

Si une variable aléatoire continue X suit la loi normale centrée réduite, alors pour tout réels a et b tels que $a \leqslant b$ on a :

$$P(a \leqslant X \leqslant b) = \Phi(b) - \Phi(a)$$

$$P(X \geqslant a) = 1 - \Phi(a) \quad \text{et} \quad P(X \leqslant -|a|) = 1 - \Phi(|a|)$$

preuve : La première égalité est liée à la relation de Chasles pour l'intégrale, en utilisant des soustraction d'aires sous la courbe.
La deuxième égalité est liée à l'événement contraire : $P(X \geqslant a) = 1 - P(X \leqslant a)$
Enfin la troisième est liée à la symétrie de la représentation de la fonction ϕ :
l'aire sous la courbe de la partie gauche ($X \leqslant -|a|$) est égale à l'aire sous la courbe de la partie droite ($X \geqslant |a|$).

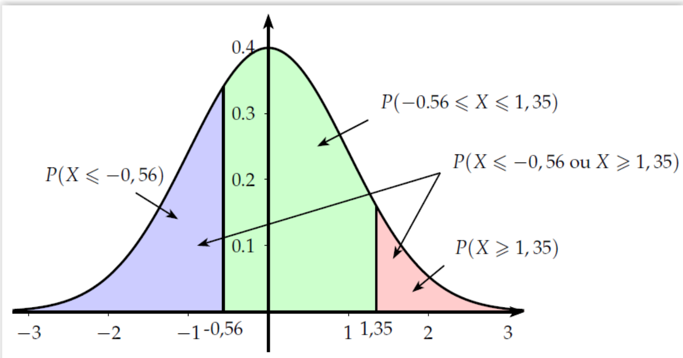
Exemple : Considérons X suivant la loi $N(0;1)$, de fonction de répartition $\Phi(x) = P(X \leqslant x)$.
On lit sur un tableau les valeurs suivantes : $\Phi(1,35) \approx 0,9115$ et $\Phi(0,56) \approx 0,7123$.

Calculons : a) $P(X \geqslant 1,35)$

b) $P(X \leqslant -0,56)$

c) $P(-0,56 \leqslant X \leqslant 1,35)$

d) $P(X \leqslant -0,56 \text{ ou } X \geqslant 1,35)$



---

a) $P(X \geqslant 1,35) = 1 - P(X \leqslant 1,35) = 1 - \Phi(1,35) = 1 - 0,9115 = 0,0885$.
b) Par symétrie de la courbe, on a $P(X \leqslant -0,56) = P(X \geqslant 0,56) = 1 - P(X \leqslant 0,56)$
donc $P(X \leqslant -0,56) = 1 - \Phi(0,56) = 1 - 0,7123 = 0,2877$.
c) $P(-0,56 \leqslant X \leqslant 1,35) = \Phi(1,35) - \Phi(-0,56) = 0,9115 - 0,2877 = 0,6238$.
d) $P(X \leqslant -0,56 \text{ ou } X \geqslant 1,35) = P(X \leqslant -0,56) + P(X \geqslant 1,35) = 0,2877 + 0,0,0885 = 0,3762$.

Remarque : Certaines valeurs interviennent souvent et il est utile de les mémoriser.

$$P(-1 \leqslant X \leqslant 1) = 0,683$$
$$P(-2 \leqslant X \leqslant 2) = 0,954$$
$$P(-3 \leqslant X \leqslant 3) = 0,997$$

Propriété 2 : Si une variable aléatoire continue X suit la loi normale centrée réduite, alors son espérance est nulle et sa variance est égale à 1.

Remarque : C'est pour cela que cette loi est dite centrée, car $E(X) = 0$, et réduite $V(X) = 1$.

On vérifie bien que $E(X) = \int_{-\infty}^{+\infty} t\phi(t)dt = 0$, puisque la fonction $t\phi(t)$ est impaire.

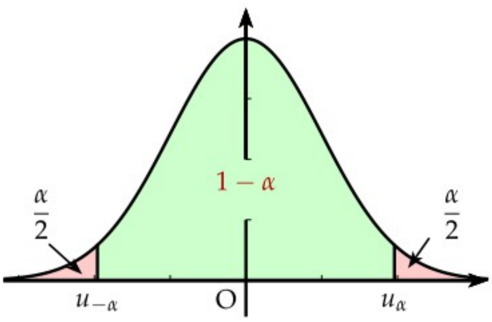
Propriété 3 : Probabilité d'intervalles centrés en zéro

Soit X une variable aléatoire qui suit la loi normale centrée réduite, et $\alpha \in]0;1[$.
Alors il existe un **unique** réel **strictement positif** u_α tel que :

$$P(-u_\alpha \leqslant X \leqslant u_\alpha) = 1 - \alpha$$

preuve : On cherche un réel x strictement positif tel que $P(-x \leqslant X \leqslant x) = 1 - \alpha$.

$$P(-x \leqslant X \leqslant x) = 1 - \alpha$$
$$\Phi(x) - \Phi(-x) = 1 - \alpha$$
$$\Phi(x) - (1 - \Phi(x)) = 1 - \alpha$$
$$\Phi(x) - 1 + \Phi(x) = 1 - \alpha$$
$$2\Phi(x) = 2 - \alpha$$
$$\Phi(x) = 1 - \frac{\alpha}{2}$$



Comme $\Phi(x) = \int_{-\infty}^{x} \phi(t).dt$, que la fonction ϕ est continue et strictement positive, d'après le théorème fondamental de l'intégration on sait que Φ est continue, de dérivée ϕ, donc strictement croissante sur $]0;+\infty[$.

De plus : $\lim\limits_{x \to 0} \Phi(x) = \Phi(0) = \frac{1}{2}$ et $\lim\limits_{x \to +\infty} \Phi(x) = 1$, et $0 < \alpha < 1$ donne $\frac{1}{2} < 1 - \frac{\alpha}{2} < 1$.
Donc, d'après le théorème des valeurs intermédiaires, il existe un unique réel strictement positif $x = u_\alpha$ tel que $\Phi(u_\alpha) = 1 - \frac{\alpha}{2}$.

Exemple : Considérons une variable aléatoire continue X suivant la loi normale centrée réduite.

Déterminons l'intervalle I centré en 0 tel que $P(X \in I) = 0{,}8$.
On donnera les bornes de l'intervalle avec une précision de 10^{-2}.

On est dans la situation où $1-\alpha = 0{,}8$, donc $\alpha = 0{,}2$.
On doit donc avoir $\Phi(u_\alpha) = 1 - \dfrac{\alpha}{2} = 0{,}9$ soit $\Phi(u_\alpha) = P(X \leqslant u_\alpha) = 0{,}9$.
Mathématiquement, inverser la relation $\Phi(u_\alpha) = 0{,}9$ s'écrit : $u_\alpha = \Phi^{-1}(0{,}9)$.
En inversant cette relation à la calculatrice, on obtient $u_\alpha \approx 1{,}28$.

Donc $I = [-1{,}28\,;\,1{,}28]$.

Cette propriété nous permettra par la suite de définir un intervalle de fluctuation de X.

Remarque : Il est intéressant de retenir les valeurs de $u_{0,05}$ et $u_{0,01}$. On obtient alors :

$$P(-1{,}96 \leqslant X \leqslant 1{,}96) = 0{,}95$$
$$P(-2{,}58 \leqslant X \leqslant 2{,}58) = 0{,}99$$

Exercice n°1 : Utilisation de la calculatrice.

Soit X vune ariable aléatoire continue suivant la loi $N(0;1)$.
Donner les probabilités suivantes à 10^{-2} près en utilisant la calculatrice.

$P(X \leqslant 1)$ $\qquad$ $P(X \geqslant 2)$ $\qquad$ $P(-1 \leqslant X \leqslant 1{,}5)$

Exercice n°2 : Utilisation de la calculatrice et inversion de la loi normale.

La calculatrice permet de déterminer le réel u tel que $\Phi(u) = P(X \leqslant u) = 0{,}85$ par inversion de la loi normale centrée réduite. On obtient $u \approx 1{,}036$ dans cet exemple.

A partir des données suivantes, déterminer la valeur de $P(X \leqslant u)$ puis donner une valeur approchée de u à 10^{-3} près en utilisant la calculatrice.

a) $P(X \geqslant u) = 0{,}93$ $\qquad$ b) $P(-u \leqslant X \leqslant u) = 0{,}75$

c) $P(0 \leqslant X \leqslant u) = 0{,}39$ $\qquad$ d) $P(X \geqslant -u) = 0{,}87$

2) Loi normale générale :

Définition 3 :

Une variable aléatoire X suit une loi normale de paramètres μ et σ, notée $N(\mu\,;\sigma^2)$, si et seulement si la variable aléatoire $Z=\dfrac{X-\mu}{\sigma}$ suit la loi normale centrée réduite $N(0\,;1)$.

Par exemple, dire que la variable aléatoire continue X suit la loi normale $N(30;25)$ signifie que la variable aléatoire $Z=\dfrac{X-30}{5}$ suit la loi normale centrée réduite $N(0;1)$.

Propriété 4 :

Si X est une variable aléatoire qui suit une loi normale $N(\mu\,;\sigma^2)$, alors son espérance vaut $E(X)=\mu$ et sa variance vaut $V(X)=\sigma^2$.

preuve :

par linéarité de l'espérance : $E(aX+b)=aE(X)+b$, donc $E(Z)=\dfrac{E(X)-\mu}{\sigma}=0$ et $E(X)=\mu$.

par propriété de la variance : $V(aX+b)=a^2V(X)$, donc $V(Z)=\dfrac{V(X)}{\sigma^2}=1$ et $V(X)=\sigma^2$.

Exercice n°3 :

On considère une variable aléatoire continue X prenant ses valeurs sur un intervalle I, et de densité de probabilité une fonction f continue et positive sur I.

Par définition de l'espérance et de la variance, on a :

$$\mu=E(X)=\int_I x\,f(x).dx \text{ et } V(X)=\int_I (x-\mu)^2 f(x).dx.$$

1) Démontrer que $E(aX+b)=aE(X)+b$ pour tous réels a et b.

2) Démontrer que $V(aX+b)=a^2V(X)$.

3) Démontrer le théorème de König-Huygens : $V(X)=E(X^2)-E(X)^2$.

Exemple :

Les températures de l'eau du mois de juillet, autour du lac Léman, suivent une loi normale d'espérance $18{,}2\,°C$ et d'écart-type $3{,}6\,°C$. Une personne part camper en juillet sur les bords du lac. Que peut-on lui indiquer comme probabilités de température de l'eau des plages dans les cas suivants :

1) température inférieures à $16\,°C$
2) températures comprises entre $20\,°C$ et $24{,}5\,°C$
3) températures supérieures à $21\,°C$

On appelle T la variable aléatoire correspondant aux températures,
et Z la variable aléatoire associée qui vérifie la loi normale $\mathrm{N}(0;1)$.
On a donc $Z = \dfrac{T - 18{,}2}{3{,}6}$.

Il y a deux façons d'obtenir les résultats :

- soit en utilisant directement la variable T et la calculatrice
- soit en utilisant la variable Z, et une table contenant les valeurs de la fonction de répartition de la loi normale centrée réduite $\Phi(a) = \mathrm{P}(\mathrm{Z} \leqslant a)$.

a) Directement avec la calculatrice on cherche $\mathrm{P}(-10^{99} \leqslant \mathrm{T} \leqslant 16)$ avec $\mu = 18{,}2$ et $\sigma = 3{,}6$.
On obtient alors 0,271.

Avec une table, on revient à la variable Z : $T \leqslant 16 \Leftrightarrow Z \leqslant \dfrac{16 - 18{,}2}{3{,}6} \Leftrightarrow Z \leqslant -0{,}611$.
On lit sur la table que $\Phi(0{,}611) = 0{,}729$ car on cherche $\Phi(-0{,}611)$.
Alors la probabilité est : $\Phi(-0{,}611) = 1 - \Phi(0{,}611) = 1 - 0{,}729 = 0{,}271$.

b) Directement avec la calculatrice on cherche $\mathrm{P}(20 \leqslant \mathrm{T} \leqslant 24{,}5)$ avec $\mu = 18{,}2$ et $\sigma = 3{,}6$.
On obtient alors 0,218.

Avec une table, on revient à la variable Z :

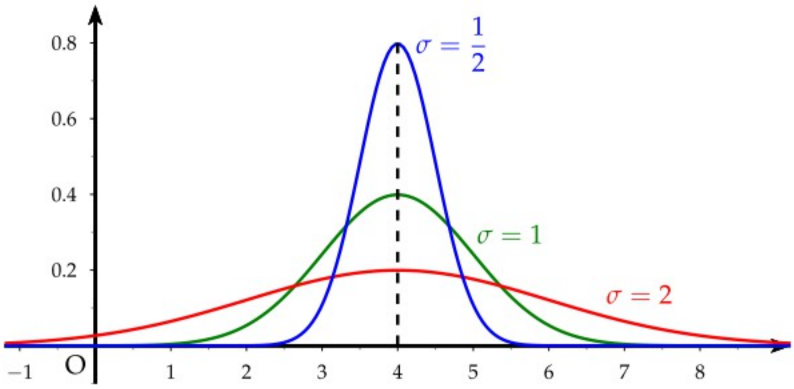
$$20 \leqslant T \leqslant 24{,}5 \Leftrightarrow \frac{20 - 18{,}2}{3{,}6} \leqslant Z \leqslant \frac{24{,}5 - 18{,}2}{3{,}6} \Leftrightarrow 0{,}5 \leqslant Z \leqslant 1{,}75.$$

On lit sur la table que $\Phi(0{,}5) = 0{,}692$ et $\Phi(1{,}75) = 0{,}960$.
Alors la probabilité est : $\Phi(1{,}75) - \Phi(0{,}5) = 0{,}960 - 0{,}692 = 0{,}268$.

c) De même manière on obtient $\mathrm{P}(T \geqslant 21) = \mathrm{P}(Z \geqslant 0{,}78) = 1 - \Phi(0{,}78) = 1 - 0{,}782 = 0{,}218$.

Influence de l'écart type sur la courbe de densité :

Voici les courbes des densités correspondantes à une espérance $\mu = 4$ et aux écarts types respectifs $\sigma = \frac{1}{2}$, $\sigma = 1$ et $\sigma = 2$.

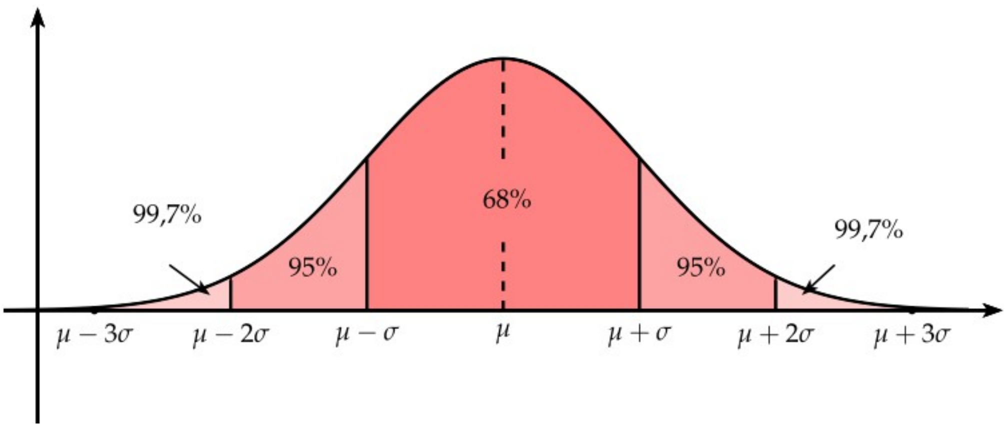


On constate que plus l'écart-type est important, plus la courbe de densité est évasée et plus le maximum est petit. En effet, un écart-type important signifie que la dispersion des données est importante.

Ces différentes courbes peuvent être repérées par 3 intervalles caractéristiques :

$[\mu - \sigma ; \mu + \sigma]$ $\quad$ $[\mu - 2\sigma ; \mu + 2\sigma]$ $\quad$ $[\mu - \sigma ; 3\mu + 3\sigma]$

Les intervalles 1σ, 2σ et 3σ à connaître :



$$P(\mu - \sigma \leqslant X \leqslant \mu + \sigma) = 0{,}68$$

$$P(\mu - 2\sigma \leqslant X \leqslant \mu + 2\sigma) = 0{,}95$$

$$P(\mu - 3\sigma \leqslant X \leqslant \mu + 3\sigma) = 0{,}997$$

3) Approximation normale d'une loi binomiale:

Propriété 5 : Théorème de Moivre-Laplace.

Soit X_n une variable aléatoire qui suit la loi binomiale $B(n, p)$,
et Z_n la variable aléatoire définie de la manière suivante :

$$Z_n = \frac{X_n - E(X_n)}{\sigma(X_n)} = \frac{X_n - np}{\sqrt{np(1-p)}}.$$

Alors pour tous réels $a < b$, on a : $\lim\limits_{n \to +\infty} P(a \leqslant Z_n \leqslant b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} . dt$.

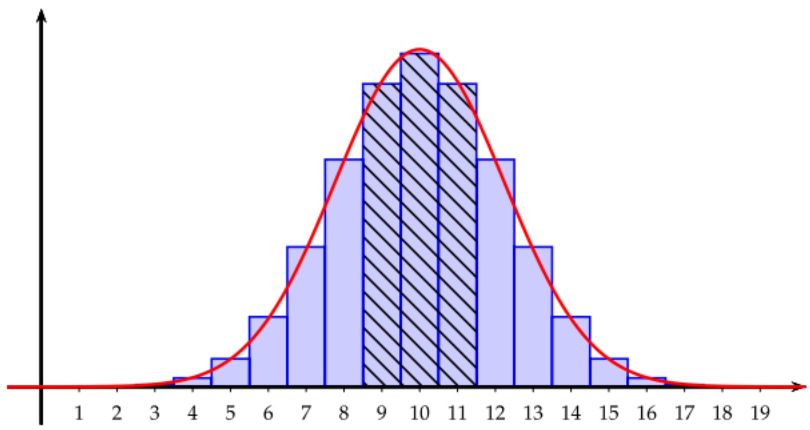
Remarques :

- Cela signifie que la loi binomiale converge en probabilité vers une loi normale.
- Pour les grandes valeurs de n, la loi binomiale est proche de la loi $N(np; np(1-p))$.

En pratique, on pourra faire l'approximation d'une loi binomiale par une loi normale lorsque l'on aura les conditions suivantes : $n \geqslant 30$ $np \geqslant 5$ $n(1-p) \geqslant 5$

Exemple :

On a tracé la loi binomiale $B(n=20; p=0,5)$ ainsi que la densité de la loi normale correspondante, avec $\mu = 20 \times 0,5 = 10$ et $\sigma = \sqrt{20 \times 0,5 \times 0,5} = \sqrt{5}$.



Les deux dernières conditions de l'approximation sont respectées : $np = 10$ et $n(1-p) = 10$.

a) Calculons $P(9 \leqslant X \leqslant 11)$ avec la loi binomiale $B(20; 0,5)$:

Avec la calculatrice, on obtient $P(9 \leqslant X \leqslant 11) \approx 0,1966$.

b) Calculons $P(9 \leqslant X \leqslant 11)$ avec la loi normale $N(10;5)$:

Comme on remplace un digramme en bâtons par un histogramme, il faut intégrer trois rectangles de largeur 1 centrés en 9, 10 et 11. On est donc amené à calculer l'aire sous la courbe densité correspondante à $8{,}5 \leqslant X \leqslant 9{,}5$. C'est ce qu'on appelle la **correction de continuité**.

Avec la calculatrice on détermine alors que $P(8{,}5 \leqslant X \leqslant 9{,}5) \approx 0{,}4977$.

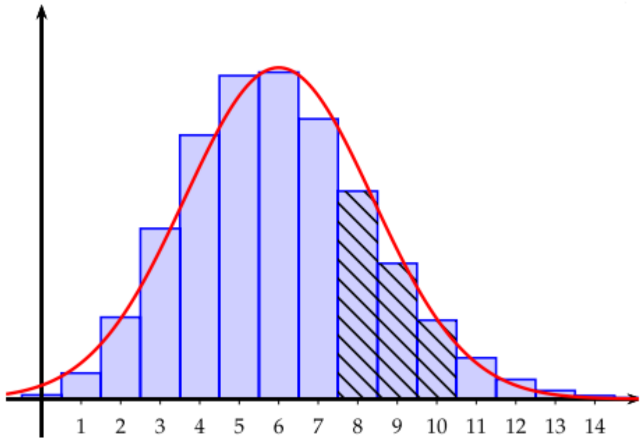
L'erreur est donc de $0{,}1\,\%$.

Il se peut par contre que n soit grand et cependant p trop petit pour qu'on soit dans les conditions de l'approximation normale. Cela se produit par exemple lorsque l'on considère le nombre d'accidents provoqés par un vaccin, le nombre de suicides dans une grande ville, pour une période donnée.
Dans le cas des petites valeurs de p c'est-à-dire pour $p < 0{,}1$, l'approximation de la loi binomiale ne pourra pas se faire avec une loi normale.

Dans l'exemple ci-dessous, on a : $n = 100$ et $p = 0{,}06$ d'où $E(X) = 6$ et $\sigma(X) = \sqrt{5{,}56}$.
Calculons $P(8 \leqslant X \leqslant 10)$.

- Avec la loi binomiale : $P(8 \leqslant X \leqslant 10) \approx 0{,}2140$
- Avec la loi normale : $P(8 \leqslant X \leqslant 10) \approx 0{,}2342$

L'erreur est maintenant de $2\,\%$.



Exercice n°4 :

On lance 180 fois un dé cubique équilibré et on note X la variable aléatoire qui correspond au nombre d'apparition de 6.

1) Déterminer les paramètres de la loi normale associée à X.
2) Vérifier que les conditions de l'approximation normale sont respectées.
3) Calculer les probabilités suivantes en utilisant l'approximation normale :

$P(X \leqslant 20)$ $\qquad$ $P(X > 40)$ $\qquad$ $P(X \leqslant 20 \text{ ou } X > 40)$

4) Retouver les résultats précédents en utilisant la variable $Z = \dfrac{X - \mu}{\sigma}$.

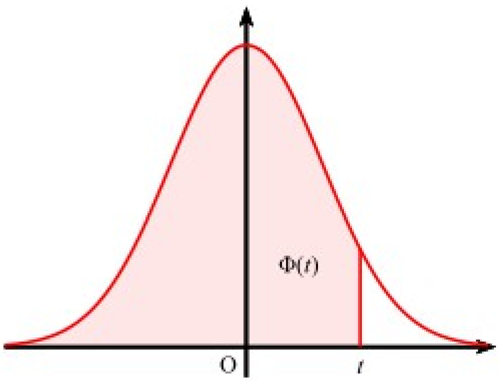
Propriété 6 : Théorème Central-Limit (Hors programme).

Lorsque l'on fait la somme d'un très grand nombre de variables aléatoires de loi quelconque, cette somme suit une loi normale.

Remarque :

Un grand nombre de distributions dans la nature suivent une loi normale, car ces distributions décrivent des phénomènes qui résultent d'un grand nombre de causes de fluctuations indépendantes, comme la taille d'un individu.

Table de la loi normale centrée réduite



$P(-1,96 < T < 1,96) = 0,95$

$P(-2,58 < T < 2,58) = 0,99$

Rappel :

$P(T > t) = 1 - P(T < t) = 1 - \Phi(t)$

$P(T < -t) = P(T > t) = 1 - \Phi(t)$

Exemple :

$P(T < 1,24) = 0,892\,5$

$P(T > 1,24) = 1 - 0,892\,5 = 0,107\,5$

$P(T < -1,24) = P(T > 1,24) = 0,107\,5$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767