Lois de probabilité à densité

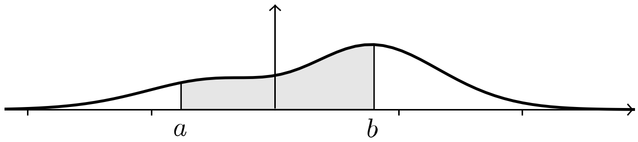
1. Variable aléatoire à densité

Définition. *On dit qu'une variable aléatoire X est une <u>variable aléatoire à densité</u> s'il existe une fonction f à valeurs positives telle que pour tous réels a et b avec $a \leq b$,*

$$P(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x.$$

Cette fonction f est appelé <u>densité</u> de X.

Il résulte de la définition que f est positive et que l'aire sous sa courbe est 1. Une telle fonction est appelée <u>densité</u>.

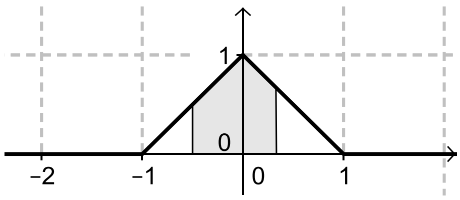


La probabilité $P(a \leq X \leq b)$ correspond à l'aire sous la courbe entre les abscisses a et b.

<u>Exemple A</u>

Soit f la fonction définie sur $\mathbb{R}$ par

$$f(x) = \begin{cases} 1 - x \text{ si } 0 \leq x \leq 1 \\ 1 + x \text{ si} - 1 \leq x \leq 0. \\ \quad 0 \text{ sinon} \end{cases}$$



La courbe de f est constituée de deux segments et deux demi-droites. La fonction f est visiblement continue et positive et l'aire sous la courbe est égale à 1.

Par exemple $P(0 \leq X \leq 1) = \frac{1}{2}$, comme on peut le voir par le calcul suivant :

$$P(0 \leq X \leq 1) = \int_0^1 f(x)\,\mathrm{d}x = \int_0^1 (1 - x)\,\mathrm{d}x = \left[x - \frac{x^2}{2}\right]_0^1 = 1 - \frac{1}{2} = \frac{1}{2}.$$

Calculons $P\left(-\frac{1}{2} \leq X \leq \frac{1}{3}\right)$. Par la relation de Chasles,

$$P\left(-\frac{1}{2} \leq X \leq \frac{1}{3}\right) = \int_{-\frac{1}{2}}^{\frac{1}{3}} f(x)\,\mathrm{d}x = \int_{-\frac{1}{2}}^{0} (1 + x)\,\mathrm{d}x + \int_0^{\frac{1}{3}} (1 - x)\,\mathrm{d}x$$

Chacune de ces intégrales se calcule facilement, on arrive à

$$P\left(-\frac{1}{2} \leq X \leq \frac{1}{3}\right) = \left(\frac{1}{2} - \frac{1}{8}\right) + \left(\frac{1}{3} - \frac{1}{18}\right) = \frac{47}{72}.$$

Remarque. Pour tout réel k et toute variable aléatoire X de densité f, on a $P(X = k) = 0$. En effet $P(X = k) = \int_k^k f(x)\,\mathrm{d}x = 0$. Ainsi pour tous réels a et b

$$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$$

et aussi $P(X < a) = P(X \leq a)$.

Attention, ces égalités sont fausses pour les lois qui ne sont pas à densité, notamment pour la loi binomiale.

<u>Exemple</u>

La fonction f définie sur $\mathbb{R}$ par $f(x) = \begin{cases} \frac{1}{x^2} \text{ si } x \geq 1 \\ 0 \text{ si } x < 1 \end{cases}$ est une densité. En effet l'aire sur $]-\infty; 1]$ est nulle et $\int_1^b f(x)\,\mathrm{d}x = \left[-\frac{1}{x}\right]_1^b = -\frac{1}{b} + 1$, donc $\lim\limits_{b \to +\infty} \int_1^b f(x)\,\mathrm{d}x = 1$.

- **Espérance**

Définition. *Soit X une variable aléatoire de densité f sur I (c'est-à-dire $f(x) = 0$ si $x \notin I$).*
Si I est de la forme $[a; b]$, on appelle <u>espérance</u> de X le réel $E(X) = \int_a^b xf(x)\,\mathrm{d}x$.
Si I est de la forme $[a; +\infty[$, on appelle espérance de X le réel $E(X) = \int_a^{+\infty} xf(x)\,\mathrm{d}x$ défini par $\lim\limits_{b\to+\infty} \int_a^b xf(x)\,\mathrm{d}x$.

<u>Exemple A</u>

La variable aléatoire X est à valeur dans $[-1; 1]$. Ainsi

$$E(X) = \int_{-1}^{1} xf(x)\,\mathrm{d}x = \int_{-1}^{0} x(1+x)\,\mathrm{d}x + \int_{0}^{1} x(1-x)\,\mathrm{d}x$$

La première intégrale de la somme vaut $-\frac{1}{6}$ et la seconde $\frac{1}{6}$, donc $E(X) = 0$.
Le résultat était prévisible car la densité est une fonction paire.

2. Loi uniforme

Définition. *Soit a et b deux réels tels que $a < b$. Dire qu'une variable aléatoire X suit la loi uniforme sur $[a; b]$ signifie que sa densité est définie par $f(x) = \begin{cases} \frac{1}{b-a} & si\ a \leq x \leq b \\ 0\ sinon \end{cases}$.*

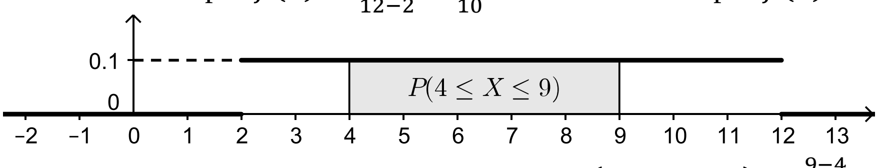
Si $[x; y] \subset [a; b]$, on a $P(x \leq X \leq y) = \frac{y-x}{a-b} = \frac{\text{longueur de } [x;y]}{\text{longueur de } [a;b]}$.

Théorème. L'espérance d'une telle variable aléatoire est $\frac{a+b}{2}$.

Démonstration. Par définition $E(X) = \int_a^b xf(x)\,\mathrm{d}x = \int_a^b \frac{x}{b-a}\,\mathrm{d}x = \frac{1}{b-a}\int_a^b x\,\mathrm{d}x$.
Or une primitive de $x \mapsto x$ est $x \mapsto \frac{x^2}{2}$ donc $E(X) = \frac{1}{b-a}\left(\frac{b^2}{2} - \frac{a^2}{2}\right) = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}$. ■

<u>Exemple</u>

Suite à un problème sur son ordinateur, Jean décide d'appeler le service après-vente du fabricant. Le temps d'attente X, exprimé en minutes, avant d'être en communication avec un technicien suit la loi uniforme sur $[2; 12]$.
La densité de X est définie par $f(x) = \frac{1}{12-2} = \frac{1}{10}$ si $2 \leq x \leq 12$ et par $f(x) = 0$ sinon.



La probabilité que Jean attende entre 4 et 9 minutes est $P(4 \leq X \leq 9) = \frac{9-4}{12-2} = \frac{1}{2}$.
La probabilité que Jean attende moins de 5 minutes est

$$P(X \leq 5) = P(2 \leq X \leq 5) = \frac{5-2}{12-2} = \frac{3}{10}.$$

La probabilité que Jean attende au moins 10 minutes est

$$P(X \geq 10) = P(10 \leq X \leq 12) = \frac{12-10}{12-2} = \frac{1}{5}.$$

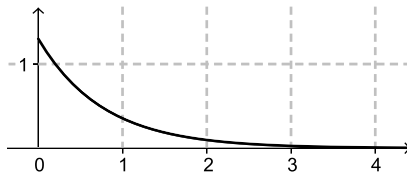
Le temps moyen d'attente est $E(X) = \frac{2+12}{2} = 7$ minutes.

3. Loi exponentielle

Définition. *Soit λ un réel strictement positif. Une variable aléatoire X suit la <u>loi exponentielle de paramètre λ</u> si sa densité est la fonction définie sur $[0;+\infty[$ par $f(x)=\lambda e^{-\lambda x}$.*

Une primitive de f étant $x \mapsto -e^{-\lambda x}$, on a pour tous réels a et b tels que $0 \leq a \leq b$,

- $P(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x = \left[-e^{-\lambda x}\right]_a^b$
  $= e^{-\lambda a} - e^{-\lambda b}$;
- $P(X \leq a) = P(0 \leq X \leq a) = 1 - e^{-\lambda a}$;
- $P(X > a) = 1 - P(X \leq a) = e^{-\lambda a}$.



Densité d'une variable aléatoire d'une loi exponentielle avec $\lambda = 1{,}3$.

La fonction f est clairement positive. Son aire sous la courbe est 1 puisque

$$\lim_{b\to+\infty} \int_0^b f(x)\,\mathrm{d}x = \lim_{b\to+\infty} \left(1 - e^{-\lambda b}\right) = 1.$$

Définition. *Soit X la variable aléatoire correspondant à la durée de vie d'un individu ou d'un objet. On dit que X suit <u>la loi de durée de vie sans vieillissement</u> lorsque la probabilité que l'individu (ou l'objet) soit vivant (ou fonctionne) à l'instant b sachant qu'il est vivant (ou qu'il fonctionne) à l'instant a (avec $a \leq b$) ne dépend pas de son âge a :*

$$P_{X\geq a}(X \geq b) = P(X \geq b - a).$$

Théorème. Une variable aléatoire X suit une loi de durée de vie sans vieillissement si et seulement si elle suit une loi exponentielle.

Démonstration. Si X suit une loi exponentielle, on a

$$P_{X\geq a}(X \geq b) = \frac{P\big((X\geq a)\cap(X\geq b)\big)}{P(X\geq a)} = \frac{P(X\geq b)}{P(X\geq a)} = \frac{e^{-\lambda b}}{e^{-\lambda a}} = e^{-\lambda(b-a)} = P(X \geq b-a).$$

On admet la réciproque. ■

Théorème. L'espérance d'une variable aléatoire X suivant une loi exponentielle de paramètre λ est $\frac{1}{\lambda}$.

Démonstration (exigible). Soit g la fonction définie sur $[0;+\infty[$ par $g(x) = \lambda x e^{-\lambda x}$. Cette fonction est dérivable et

$$g'(x) = \lambda e^{-\lambda x} - \lambda^2 x e^{-\lambda x} = \lambda e^{-\lambda x} - \lambda g(x).$$

On en déduit $g(x) = \frac{1}{\lambda}\left(\lambda e^{-\lambda x} - g'(x)\right)$, d'où

$$\int_0^b g(x)\,\mathrm{d}x = \frac{1}{\lambda}\int_0^b \left(\lambda e^{-\lambda x} - g'(x)\right)\,\mathrm{d}x = \frac{1}{\lambda}\left[-e^{-\lambda x} - g(x)\right]_0^b = \frac{1}{\lambda}\left(-e^{-\lambda b} - \lambda b e^{-\lambda b} + 1\right).$$

Comme $\lim\limits_{b\to+\infty} -\lambda b = -\infty$ et $\lim\limits_{X\to-\infty} Xe^X = 0$, il vient par composition $\lim\limits_{b\to+\infty} -\lambda b e^{-\lambda b} = 0$. On a également $\lim\limits_{b\to+\infty} e^{-\lambda b} = 0$. Finalement

$$E(X) = \lim_{b\to+\infty}\int_0^b x f(x)\,\mathrm{d}x = \lim_{b\to+\infty}\int_0^b g(x)\,\mathrm{d}x = \frac{1}{\lambda}. \blacksquare$$

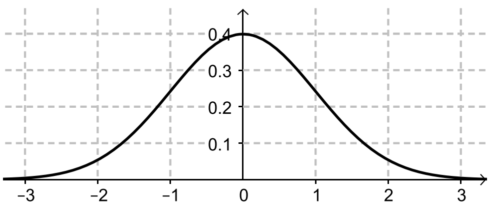
4. Loi normale centrée réduite

Définition. *On dit qu'une variable aléatoire X suit la loi* <u>*normale centrée réduite*</u>*, notée $\mathcal{N}(0;1)$, lorsque sa densité est donnée par la fonction f définie sur $\mathbb{R}$ par* $f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$.

Le fait que la fonction f soit continue est positive est clair. En revanche il n'est pas possible de démontrer au lycée que l'aire sous la courbe est égale à 1, nous l'admettrons. La fonction $x\mapsto e^{-\frac{x^2}{2}}$ n'admet pas de primitives s'exprimant à l'aide de fonctions usuelles.

Théorème. Soit X une variable aléatoire suivant la loi normale $\mathcal{N}(0;1)$. Alors l'espérance de X est 0 et l'écart-type de X est 1.

❖ Étude de la fonction f



La densité f est dérivable sur $\mathbb{R}$ de dérivée

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

donc $f'(x)$ est du signe de $-x$. Il en résulte que f est croissante sur $]-\infty;0]$ et décroissante sur $[0;+;\infty[$.
De plus pour tout x, on a $f(-x)=f(x)$, donc la courbe de f est symétrique par rapport à l'axe des ordonnées. Cette courbe s'appelle courbe de Gauss.

❖ Usage de la calculatrice

Soit X une variable aléatoire suivant une loi $\mathcal{N}(0;1)$.

Puisque la densité de la loi normale tend très rapidement vers 0, on a

$$P(X\leq a)\approx P(-10^{99}\leq X\leq a) \text{ et } P(X\geq a)\approx P(a\leq X\leq 10^{99}).$$

Le menu Distrib, obtenue en faisant 2nde var, permet d'accéder à la commande `normalFRép`.

	Syntaxe TI
Calcul de $P(a\leq X\leq b)$	`normalFrép(a,b)`
Calcul de $P(X\geq b)$	`normalFrép(b,10^99)`
Calcul de $P(X\leq a)$	`normalFrép(-10^99,a)`

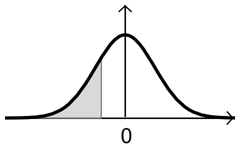
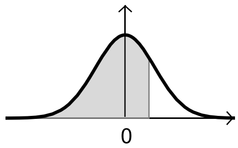
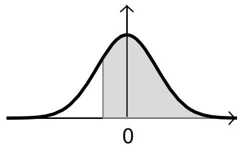
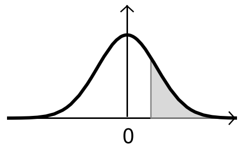
En Casio, la commande est `normCD(a,b)` qui s'obtient par « OPTN », « STAT », puis « DIST ».

<u>Exemple</u>

Soit X une variable aléatoire suivante une loi $\mathcal{N}(0;1)$. On a

- $P(X=2)=0$ car X est une variable aléatoire à densité.
- $P(-1\leq X<3)$ se calcule par `NormalFRép(-1,3)`.
- $P(X<-0{,}3)$ se calcule par `NormalFRép(-10^99,-0.3)`.
- $P(X>-1)$ se calcule par `NormalFRép(-1,10^99)`.

Il peut être utile de savoir que les probabilités du type $P(X < a)$ ou $P(X > a)$ peuvent se ramener à des calculs du type $P(a < X < b)$ grâce aux propriétés du tableau ci-dessous.

Probabilité	$P(X < a)$ où $a < 0$	$P(X < a)$ où $a > 0$	$P(X > a)$ où $a < 0$	$P(X > a)$ où $a > 0$
Graphique				
Calcul	$0{,}5 - P(a < X < 0)$	$0{,}5 + P(0 < X < a)$	$P(a < X < 0) + 0{,}5$	$0{,}5 - P(0 < X < a)$

On peut déterminer le réel k tel que $P(X \leq k) = c$ où c est un réel donné (compris entre 0 et 1 évidemment !). Il faut pour cela utiliser la fonction `FracNormale` dont la syntaxe est FracNormale(c).

Exemple

Soit X une variable aléatoire suivante une loi $\mathcal{N}(0;1)$. Déterminer à 10^{-3} les réels suivants :

a. Le réel b tel que $P(X < b) = 0{,}53$
b. Le réel c tel que $P(X \geq c) = 0{,}3$

Réponse.

a. Le réel b s'obtient par `FracNormale(0.53)`, donc $b \approx 0{,}075$
b. On a $P(X \geq c) = 1 - P(X \leq c)$, donc c se calcule par `FracNormale(0.7)`, d'où $c \approx 0{,}524$.

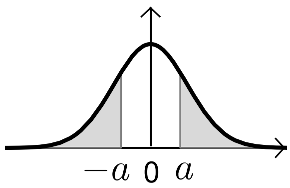
❖ Intervalle centré en 0 de probabilité donnée

Théorème. Soit X suivant la loi $\mathcal{N}(0;1)$. Pour tout réel $a > 0$ on a les relations suivantes

1. $P(X \leq -a) = P(X \geq a)$;
2. $P(-a \leq X \leq a) = 2P(X \leq a) - 1$.

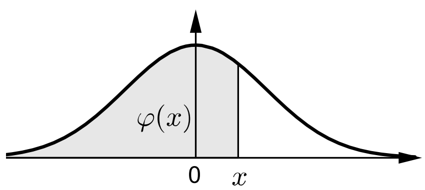
Démonstration.

1. Cela résulte de la symétrie de la courbe de f par rapport à l'axe des ordonnées.
2. Comme $P(X \leq -a) = P(X \geq a) = 1 - P(X < a)$, on a,
$$P(-a \leq X \leq a) = P(X \leq a) - P(X \leq -a)$$
$$= P(X \leq a) - \big(1 - P(X \leq a)\big) = 2P(X \leq a) - 1. \blacksquare$$

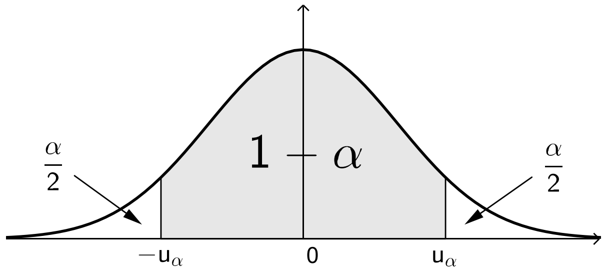


Théorème. Si X est une variable aléatoire suivant la loi normale $\mathcal{N}(0;1)$, alors pour tout réel $\alpha \in]0;1[$, il existe un unique réel strictement positif u_α tel que $P(-u_\alpha \leq X \leq u_\alpha) = 1 - \alpha$.

Démonstration (exigible). Posons $\varphi(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$, fonction qui représente l'aire hachurée ci-contre. Elle est dérivable sur $\mathbb{R}$ et sa dérivée est $\varphi'(x) = f(x)$. Comme $f(x) > 0$ pour tout réel x, il en résulte que φ est strictement croissante sur $\mathbb{R}$. On a de plus $\varphi(0) = \frac{1}{2}$ et $\lim\limits_{x \to +\infty} \varphi(x) = 1$.

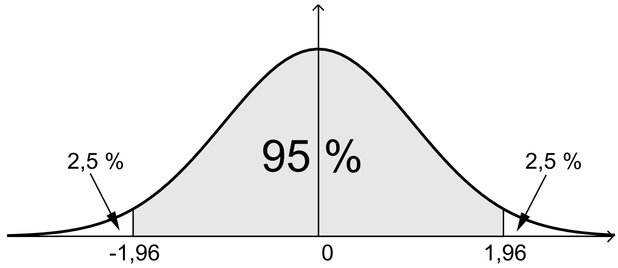


On sait que $P(-a \leq X \leq a) = 2\varphi(a) - 1$, donc l'équation du théorème s'écrit $\varphi(a) = 1 - \frac{\alpha}{2}$. Puisque $\alpha \in]0;1[$, on en déduit $1-\frac{\alpha}{2} \in \left]\frac{1}{2};1\right[$. Le théorème des valeurs intermédiaires appliqué à la fonction φ continue et strictement croissante sur $]0;+\infty[$ montre alors qu'il existe un unique réel $u_\alpha \in]0;+\infty[$ tel que $\varphi(u_\alpha) = 1-\frac{\alpha}{2}$. ■



En pratique u_α se calcule en faisant `FracNormale(1-α/2)`.

Théorème. On a notamment
$u_{0,05} \approx 1{,}96$ et $u_{0,01} \approx 2{,}58$.

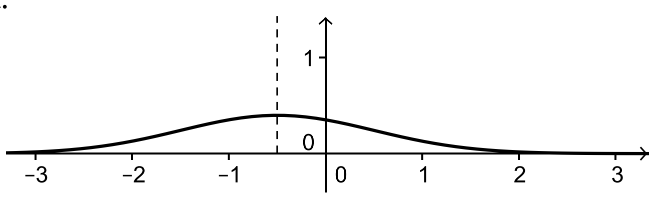


`FracNormale(0.975)` donne la valeur de $u_{0,05}$ et `FracNormale(0.995)` celle de $u_{0,01}$.

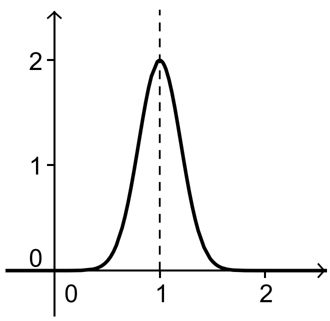
5. Loi Normale

Définition. *Soit μ un réel et σ un réel strictement positif. On dit que la variable aléatoire X suit la <u>loi normale</u> $\mathcal{N}(\mu;\sigma^2)$ si et seulement si la variable aléatoire $\frac{X-\mu}{\sigma}$ suit une loi normale centrée réduite. On parle de la loi normale d'espérance μ et d'écart-type σ (voir théorème ci-dessous).*

Soit f la densité de la loi $\mathcal{N}(\mu;\sigma^2)$. La courbe représentative de f est une courbe en cloche, ayant la droite d'équation $x=\mu$ comme axe de symétrie et d'autant plus « resserrée » autour de cet axe de symétrie que σ est petit.



$\mu = -0{,}5$ et $\sigma = 1$



$\mu = 1$ et $\sigma = 0{,}2$

Théorème. Soit X une variable aléatoire suivant la loi normale $\mathcal{N}(\mu;\sigma^2)$. Alors l'espérance de X est μ et l'écart-type de X est σ.

❖ **Usage de calculatrice**

Les commandes à utiliser sont les mêmes que pour la loi $\mathcal{N}(0;1)$ en rajoutant μ et σ dans les arguments. Si X désigne une variable aléatoire suivant une loi $\mathcal{N}(\mu;\sigma^2)$:

- `NormalFRép(a,b,μ,σ)` permet de calculer $P(a \le X \le b)$;
- `FracNormale(c,μ,σ)` permet de déterminer k tel que $P(X \le k) = c$.

Comme pour la loi normale centrée réduite, on utilisera les approximations

$$P(X \le a) \approx P(-10^{99} \le X \le a) \text{ et } P(X \ge a) \approx P(a \le X \le 10^{99}).$$

En Casio, la commande est `normCD(a,b,σ,μ)`.

Exemple

La distance parcourue par jour ouvré par un technicien en kilomètres suit la loi normale d'espérance 100 et d'écart-type 20, autrement dit cette distance suit la loi $\mathcal{N}(100; 400)$. On choisir un jour ouvré au hasard. On note :

- A l'événement « le technicien parcourt entre 80 et 110 km » ;
- B l'événement « le technicien parcourt plus de 80 km » ;
- C l'événement « le technicien parcourt moins de 90 km ».

Calculer $P(A)$, $P(B)$, $P(C)$ et $P_B(C)$.

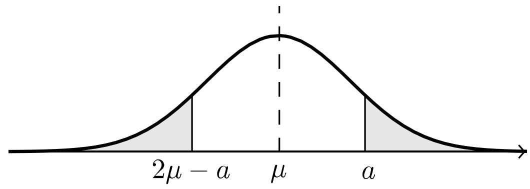
Réponse. Soit X la variable aléatoire qui à un jour associe le nombre de kilomètres parcourus. On a :

- $P(A) = P(80 \leq X \leq 110) \approx 0{,}53$;
- $P(B) = P(X \geq 80) \approx P(80 \leq X \leq 10^{99}) \approx 0{,}84$;
- $P(C) = P(X \leq 90) \approx P(-10^{99} \leq X \leq 90) \approx 0{,}31$;
- $P_B(C) = \frac{P(B \cap C)}{P(B)} = \frac{P(80 \leq X \leq 90)}{P(B)} \approx 0{,}18$.

❖ **Propriété de symétrie**

Théorème. Soit X une variable aléatoire suivant une loi $\mathcal{N}(\mu; \sigma^2)$. Alors pour tout réel a on a
$$P(X \geq a) = P(X \leq 2\mu - a).$$

Cette propriété provient de la symétrie de la courbe de la densité de X par rapport à la droite d'équation $x = \mu$.



Démonstration. Posons $Y = \frac{X-\mu}{\sigma}$. On peut écrire
$$P(X \geq a) = P\left(\frac{X-\mu}{\sigma} \geq \frac{a-\mu}{\sigma}\right) = P\left(Y \geq \frac{a-\mu}{\sigma}\right).$$
Mais Y suit la loi $\mathcal{N}(0; 1)$ et on sait que pour une telle loi, on a $P(X \geq k) = P(X \leq -k)$ pour tout réel k. On en déduit donc
$$P\left(Y \geq \frac{a-\mu}{\sigma}\right) = P\left(Y \leq -\frac{a-\mu}{\sigma}\right) = P\left(Y \leq \frac{\mu-a}{\sigma}\right).$$
En revenant à X, il vient
$$P\left(Y \leq \frac{\mu-a}{\sigma}\right) = P\left(\frac{X-\mu}{\sigma} \leq \frac{\mu-a}{\sigma}\right) = P(X - \mu \leq \mu - a) = P(X \leq 2\mu - a). \quad \blacksquare$$

❖ **Intervalles à connaître**

Théorème. Soit X une variable aléatoire suivant une loi $\mathcal{N}(\mu; \sigma^2)$.

- $P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0{,}683$
- $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0{,}954$
- $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0{,}997$

Démonstration. En posant $Y = \frac{X-\mu}{\sigma}$ on peut écrire
$$P(\mu - \sigma \leq X \leq \mu + \sigma) = P(-\sigma \leq X - \mu \leq \sigma) = P\left(-1 \leq \frac{X-\mu}{\sigma} \leq 1\right) = P(-1 \leq Y \leq 1).$$
Comme la variable aléatoire Y suit une loi $\mathcal{N}(0; 1)$,
$$P(-1 \leq Y \leq 1) = 2P(Y \leq 1) - 1 \approx 0{,}683.$$
Le raisonnement est le même pour 2σ et 3σ. $\blacksquare$

Exemple

Soit X une variable aléatoire suivant la loi $\mathcal{N}(10;\sigma^2)$. Déterminons σ de telle sorte que $P(8 \leq X \leq 12) = 0{,}9$. On a

$$P(8 \leq X \leq 12) = 0{,}9 \Leftrightarrow P\left(-2 \leq \frac{X-10}{\sigma} \leq 2\right) = 0{,}9 \Leftrightarrow P\left(-\frac{2}{\sigma} \leq \frac{X-10}{\sigma} \leq \frac{2}{\sigma}\right) = 0{,}9$$

Comme $\frac{X-10}{\sigma}$ suit la loi $\mathcal{N}(0;1)$, on a donc $\frac{2}{\sigma} = u_{0,1}$, d'où $\sigma = \frac{2}{u_{0,1}} \approx \frac{2}{1{,}645} \approx 1{,}216$.

6. Théorème de Moivre-Laplace

Théorème. Soit $p \in]0;1[$. On suppose que pour tout entier $n \geq 1$, la variable aléatoire X_n suit une loi binomiale de paramètre n et p.
Soit Z_n la variable aléatoire définie par $Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$ et soit Z une variable aléatoire suivant la loi $\mathcal{N}(0;1)$.
Pour tous réels a et b avec $a \leq b$, on a

$$\lim_{n\to+\infty} P(a \leq Z_n \leq b) = P(a \leq Z \leq b).$$

On peut également écrire

$$\lim_{n\to+\infty} P(a \leq Z_n \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,\mathrm{d}x.$$

On dit que Z_n « converge en loi » vers Z. Autrement dit, la loi binomiale $\mathcal{B}(n;p)$ est « très proche » de la loi normale de même espérance et de même variance, $\mathcal{N}\big(np;np(1-p)\big)$.
L'idée est donc que si $X \sim \mathcal{B}(n;p)$ et $Z \sim \mathcal{N}\big(np;np(1-p)\big)$, alors on a, si n est grand,

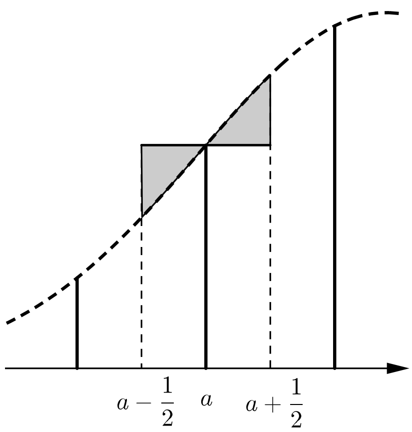
$$P(a \leq X \leq b) \approx P(a \leq Z \leq b).$$

En pratique, on approche les probabilités de la loi binomiale par celle de la loi normale dès que $\boldsymbol{n \geq 30, np \geq 5}$ **et** $\boldsymbol{n(1-p) \geq 5}$.

En fait pour que l'approximation soit la meilleure possible, on procède à la « correction de continuité », car on remplace une loi discrète par une loi continue. Ainsi on écrira

- $P(X = a) \approx P\left(a - \frac{1}{2} \leq X \leq a + \frac{1}{2}\right)$
- $P(a \leq X \leq b) \approx P\left(a - \frac{1}{2} \leq Z \leq b + \frac{1}{2}\right)$
- $P(X \leq a) \approx P\left(Z \leq a + \frac{1}{2}\right)$

La raison se comprend aisément sur le schéma ci-dessous puisque les deux zones hachurées ont des aires presque égales.



Exemple

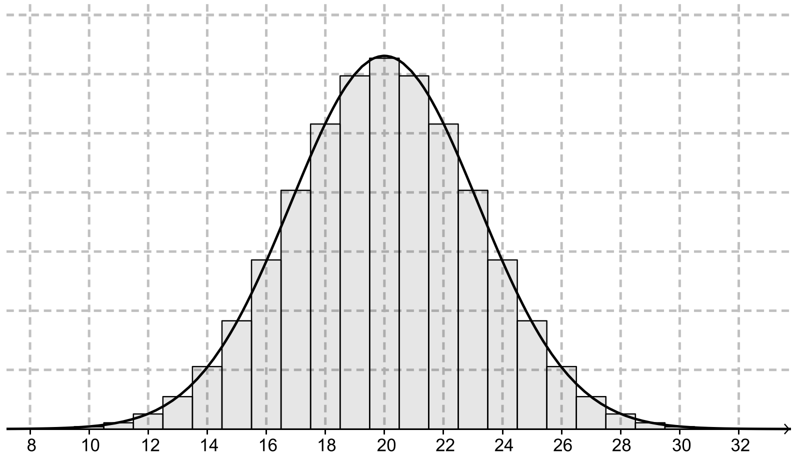
On lance 40 fois un dé cubique équilibré et on note X la variable aléatoire qui indique le nombre d'apparition d'une face paire. X suit la loi binomiale de paramètres $n = 40$ et $p = \frac{1}{2}$. Comme $np = 20$ et $n(1-p) = 20$, les conditions d'approximation de la loi binomiale par la loi normale sont satisfaites.
On a $E(X) = 20$ et $\mathrm{Var}(X) = \sqrt{10}$, donc X suit approximativement la loi normale $\mathcal{N}(20;10)$.
Calculons par exemple $P(18 \leq X \leq 25)$. Soit Z une variable aléatoire suivant une loi

normale $\mathcal{N}(20; 10)$. On a

$$P(18 \leq X \leq 25) = P(17{,}5 \leq X \leq 25{,}5) \approx P(17{,}5 \leq Z \leq 25{,}5).$$

La commande `normalFRép(17.5,25.5,20,√10)` renvoie alors environ $0{,}74441$, à comparer à la valeur exacte qui

$$P(18 \leq X \leq 25) = P(X \leq 25) - P(X \leq 17) \approx 0{,}74486.$$



Exemple

Une école d'ingénieurs organise un QCM de 100 questions pour son examen. Pour chaque question il y a trois réponses possibles dont une et une seule est exacte.
On aimerait trouver un entier k tel que si un candidat a au moins k réponses justes, il y a moins de 5 % de chances que toutes les réponses soient dues au hasard.

Soit X la variable aléatoire donnant le nombre de réponses exactes par un candidat répondant à toutes les questions au hasard. Il s'agit de trouver un entier k tel que

$$P(X \geq k) < 0{,}05 \text{ ou } P(X \leq k-1) \geq 0{,}95.$$

X suit la loi binomiale de paramètre $n = 100$ et $p = \frac{1}{3}$. Les conditions d'approximation par la loi normale sont vérifiées ; on a $E(X) = np = \frac{100}{3}$ et $\mathrm{Var}(X) = np(1-p) = \frac{200}{9}$.
Soit Z' une variable suivant la loi $\mathcal{N}\left(\frac{100}{3}; \frac{200}{9}\right)$. On peut écrire

$$P(X \leq k-1) \approx P\left(Z' \leq k - 1 + \frac{1}{2}\right),$$

ou encore si l'on désigne par Z une variable aléatoire suivant la loi $\mathcal{N}(0; 1)$,

$$P(X \leq k-1) \approx P\left(Z \leq \frac{k-1+\frac{1}{2}-\frac{100}{3}}{\sqrt{\frac{200}{9}}}\right).$$

Comme $P(Z \leq 1{,}64) \approx 0{,}95$, on doit avoir $\frac{k-1+\frac{1}{2}-\frac{100}{3}}{\sqrt{\frac{200}{9}}} \geq 1{,}64$, d'où

$$k \geq 1{,}64\sqrt{\frac{200}{9}} + \frac{100}{3} + 1 - \frac{1}{2} \approx 41{,}56.$$

Par cette méthode, on trouve donc $k \geq 42$.
Un calcul exact confirme cette approximation car

$$P(X \geq 41) \approx 0{,}066 \text{ et } P(X \geq 42) \approx 0{,}043.$$