Variable aléatoires continues

1 Généralités

On a vu dans le chapitre "Variables discrètes" comment décrire le comportement d'une variable aléatoire à valeur dans un ensemble $\{x_1, x_2, \ldots\}$ fini (pour la loi binomiale par exemple) ou infini (pour la loi géométrique). Déterminer une loi de probabilité $\mathbf{P}$ sur $\{x_1, x_2, \ldots\}$ revenait à déterminer les valeurs $\mathbf{P}(X = x_i)$. Il se peut que l'ensemble des valeurs prises par une v.a. X soit un intervalle I de $\mathbb{R}$. Par exemple, on étudie la durée d'une communication téléphonique ou la durée de vie d'un appareil. X peut alors prendre une infinité de valeurs, la probabilité de chacune d'elles serait nulle. Les événements intéressants ne sont plus ≪ obtenir tel réel ≫ mais plutôt ≪ obtenir un réel compris entre a et b ≫ noté $\mathbf{P}(X \in [a, b])$ ou $\mathbf{P}(a \leq X \leq b)$.

Définition 1.1 (Variables aléatoire continue)
*Soit un espace probabilisé d'espace fondamental Ω et de mesure de probabilité $\mathbf{P}$. On dit qu'une variable aléatoire $X : \Omega \to \mathbb{R}$ est **continue** lorsque l'ensemble de ses valeurs est $\mathbb{R}$ ou un intervalle de $\mathbb{R}$.*

Exemple 1.1
Considérons la durée d'une communication téléphonique. Cette durée est aléatoire et on cherche un modèle mathématique qui nous permettrait de répondre par exemple aux questions suivantes :
- Quelle est la probabilité pour que la durée ne dépasse pas x minutes ?
- Quelle est la durée moyenne d'une communication ? Son écart–type ?
- Quelle est la probabilité pour que la communication dure encore x minutes sachant qu'elle a déjà duré y minutes ?

On voudrait proposer comme modèle pour la durée une *fonction* positive qui possède les caractéristiques suivantes
- elle serait d'autant plus élevée sur un intervalle, que la probabilité de tomber dans cet intervalle est forte ;
- la probabilité de tomber dans cet intervalle serait égale à la surface comprise entre la courbe et l'axe des abscisses, sur cet intervalle ;
- par voie de conséquence, la surface totale sous la courbe devrait valoir 1.

Un telle fonction devrait nécessairement être positive ou nulle. On décide ici de se restreindre à des fonctions pour lesquelles la surface est calculable, c'est–à–dire les fonctions que l'on sait intégrer. On justifie ainsi la définition de l'objet introduit ci–après.

1.1 Densité de probabilité

Définition 1.2 (Fonction de densité)
On appelle fonction de densité *une fonction f définie sur $\mathbb{R}$ et à valeurs réelles telle que*
- $(\forall x \in \mathbb{R}), \quad f(x) \geq 0$
- *f est continue sauf éventuellement en un nombre fini de points*
- $\displaystyle\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$

Définition 1.3
Soit f une fonction de densité. On appelle loi de probabilité de densité f, la loi définie pour tout intervalle A de $\mathbb{R}$ par

$$\mathbf{P}(X \in A) = \int_A f(x)\,\mathrm{d}x \ .$$

Remarque 1.1. *Dans cette définition, chacune des bornes de l'intervalle A peut être infinie. De plus, la probabilité de l'événement $\{X = a\}$ est nulle pour tout réel a. En effet, on peut par abus de langage utiliser la définition avec « l'intervalle » réduit au point a. On a alors comme probabilité une intégrale sur un intervalle réduit à un seul point, qui est nulle.*

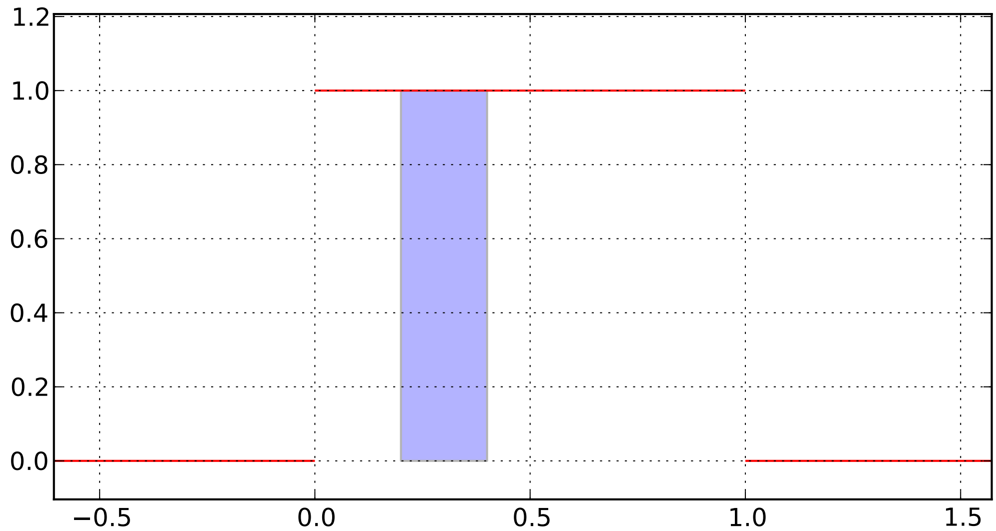
Lecture d'une densité : pour poursuivre notre point de vue de la modélisation, on a maintenant les interprétations suivantes :

- les régions de $\mathbb{R}$ de forte (resp. faible) densité sont les régions de forte (resp. faible) probabilité d'occurrence des réalisations de X. C'est–à–dire que X ≪tombera≫ plus souvent dans des zones de forte densité.
- les régions où f s'annule sont des valeurs interdites pour X.

Exemple 1.2 (Loi uniforme)
Les variables aléatoires continues les plus simples sont celles dont la densité est donnée par la fonction :

$$f : \mathbb{R} \longrightarrow \mathbb{R}^+$$
$$x \longmapsto f(x) = \begin{cases} 1 \text{ si } 0 \leq x \leq 1 \\ 0 \text{ sinon.} \end{cases}$$

La représentation graphique de cette fonction est donnée par la figure ci–dessous :



Des réalisations d'une variable aléatoire d'une telle loi sont données par la touche `random` d'une calculatrice. Il s'agit d'un nombre réel, compris entre 0 et 1. On peut lire sur le graphique qu'aucune zone de l'intervalle $[0; 1]$ n'est privilégiée, d'où le nom de loi *uniforme*. Cette fonction est une densité car elle est

- positive ou nulle (0 ou 1)
- continue par morceaux (intervalles $]-\infty; 0[$, $[0; 1]$, $]1; \infty[$)
- d'intégrale 1 (surface d'un carré de côté 1).

En utilisant les surfaces de rectangles, on notamment que si X est une variable aléatoire de loi uniforme sur $[0; 1]$

$$\mathbf{P}(X \in [0; \frac{1}{2}]) = \mathbf{P}(X \in [\frac{1}{2}; 1]) = \frac{1}{2}, \quad \mathbf{P}(X < 0) = \mathbf{P}(X > 1) = 0 \ .$$

2 Fonction de répartition

On a vu en probabilités discrètes que la fonction de répartition était un objet qui caractérisait complètement une variable aléatoire. Un tel objet existe également pour les variables aléatoires à densité.

Définition 2.1
Soit X une variable aléatoire continue de densité f donnée. On appelle fonction de répartition de X la fonction définie sur $\mathbb{R}$ et à valeur dans $[0;\, 1]$

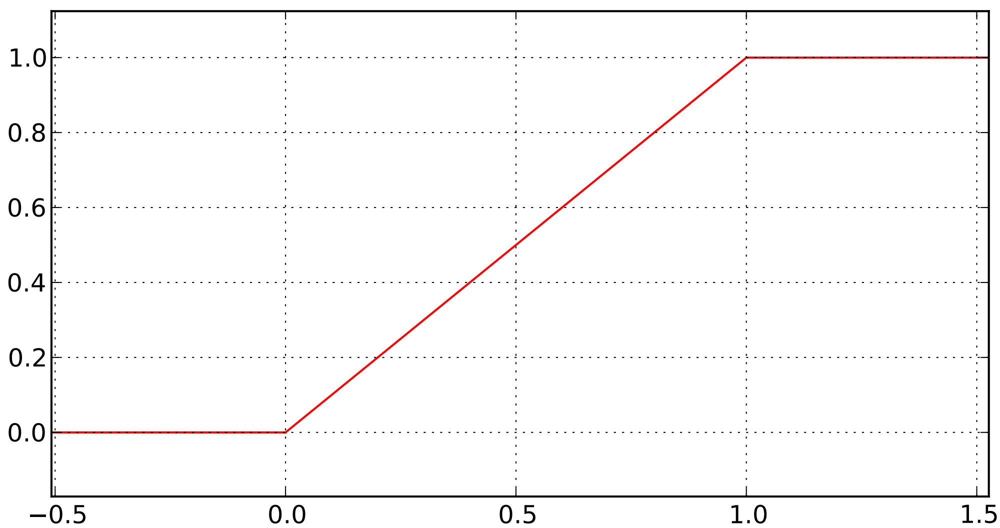
$$\begin{aligned} F : \mathbb{R} &\longrightarrow [0;\, 1] \\ t &\longmapsto F(t) = \mathbf{P}(X \leq t) = \int_{-\infty}^{t} f(x)\,\mathrm{d}x\ . \end{aligned}$$

Proposition 2.1
La fonction de répartition
- *est croissante sur* $\mathbb{R}$*,*
- *est continue partout mais dérivable seulement en tout point où f est continue,*
- *a pour limite 0 en $-\infty$ et 1 en ∞.*

Loi uniforme. Pour l'exemple de la loi uniforme, on peut encore utiliser les surfaces de rectangles pour calculer les intégrales. On obtient la fonction de répartition

$$(\forall\, t \in \mathbb{R}), \quad F(t) = \begin{cases} 0 & \text{si } t < 0 \\ t & \text{si } 0 \leq t \leq 1 \\ 1 & \text{si } t > 1 \end{cases} \quad .$$

Sa représentation graphique est :



Lecture de la fonction de répartition. Le comportement d'une variable aléatoire se lit aussi sur le graphique de sa fonction de répartition : les zones de forte probabilité sont les zones de *forte croissance* de F. Les zones où F est constante sont celles où la «masse» de probabilité ne s'accroît pas. La v.a. X ne tombe pas dans ces zones.

Définition 2.2 (Quantiles)
Soit $p \in]0; 1[$. On appelle quantile d'ordre p le plus petit réel t_p tel que

$$F(t_p) = p \ .$$

On distingue en particulier la médiane qui est le quantile d'ordre $\frac{1}{2}$.
Les trois quartiles Q_1, Q_2 et Q_3 sont les quantiles d'ordre respectif $\frac{1}{4}$, $\frac{1}{2}$ et $\frac{3}{4}$.

Les trois quartiles partagent $\mathbb{R}$ en quatre intervalles de probabilités égales. On peut aussi dire qu'ils partagent le graphique de la densité en quatre parties de surfaces égales.
Le quantile d'ordre 0,95 est la valeur t telle que $\mathbf{P}(X \leq t) = 0.95$.

3 Espérance et variance

Rappelons que pour une variable aléatoire discrète, l'espérance était définie par

$$\mathbf{E}[X] = \sum_i x_i \, \mathbf{P}(X = x_i).$$

Pour avoir une bonne intuition de l'objet à introduire, on considère que les valeurs de x_i sont très nombreuses et on imagine un petit intervalle autour d'un x_i d'amplitude $\mathrm{d}x$. Pour cet x_i et son intervalle, on peut faire l'approximation

$$\mathbf{P}(x_i < X < x_i + dx) \sim f(x_i) \, \mathrm{d}x$$

où f est une densité appropriée. Cela revient à disperser la masse de probabilité placée en x_i sur l'intervalle infinitésimal autour de x_i. Il reste à sommer les contributions de chacun de ces intervalles, ce qui correspond à la notion d'intégration. On justifie donc ainsi la définition suivante :

Définition 3.1
On appelle espérance *de la variable aléatoire X de densité f, le réel* **s'il existe**

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x \, f(x) \, \mathrm{d}x \ .$$

Plus généralement, si ϕ est une application de $\mathbb{R}$ dans $\mathbb{R}$ on définit de même

$$\mathbf{E}[\phi(X)] = \int_{-\infty}^{\infty} \phi(x) \, f(x) \, \mathrm{d}x \ .$$

Il faut remarquer que l'espérance n'existe pas toujours car l'intégrale n'est pas forcément finie.

Les propriétés de linéarité de l'intégrale donnent les résultats intuitifs suivants sur l'espérance :

Proposition 3.1
Dans la mesure où les quantités ci–dessous existent, on a pour $a \in \mathbb{R}$

$$\mathbf{E}[a \, X] = a \, \mathbf{E}[X] \qquad \textit{et} \qquad \mathbf{E}[X + a] = \mathbf{E}[X] + a.$$

Lorsqu'une variable aléatoire X à densité admet une espérance $\mathbf{E}[X]$, on peut se questionner sur la dispersion des réalisations de X autour de cette valeur espérée. Cette dispersion est mesurée, entre autre, par la variance définie ainsi :

Définition 3.2
Soit X une variable aléatoire de densité f, admettant une espérance $\mathbf{E}[X]$. On appelle variance *de X, la valeur positive réelle si elle existe, donnée par*

$$\mathbf{V}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2] = \int_{-\infty}^{\infty} (x - \mathbf{E}[X])^2 \, f(x) \, \mathrm{d}x.$$

*On définit également l'*écart-type *comme la racine carrée de la variance.*

Formule de calcul. La définition dans la forme donnée ci–dessus est utile pour saisir la signification de la variance en terme de dispersion mais n'est pas pratique pour le calcul. On montre, en développant le carré dans l'intégrale, que si la variance existe, alors on a

$$\begin{aligned} \mathbf{V}[X] &= \mathbf{E}[X^2] - (\mathbf{E}[X])^2 \\ &= \int_{-\infty}^{\infty} x^2 \, f(x) \, \mathrm{d}x - \left(\int_{-\infty}^{\infty} x \, f(x) \, \mathrm{d}x \right)^2 \end{aligned}$$

qui s'énonce ≪l'espérance du carré moins le carré de l'espérance≫. On vérifiera à titre d'exercice que l'espérance d'une variable aléatoire de loi uniforme sur $[0,\, 1]$ est égale à $\frac{1}{12}$.

Les résultats suivants sont familiers pour les variances déjà rencontrées en statistique descriptive et en probabilités discrètes. Nous les énonçons donc sans démonstration :

Proposition 3.2
Dans la mesure où les quantités ci–dessous existent, on a pour $a \in \mathbb{R}$

$$\mathbf{V}[a\,X] = a^2\, \mathbf{V}[X] \qquad \textit{et} \qquad \mathbf{V}[X + a] = \mathbf{V}[X] \ .$$

Parmi les lois à densité de variables aléatoires continues, la loi normale que nous étudions dans ce chapitre occupe une place importante. L'une des raisons est qu'elle apparaît de manière naturelle dans un des piliers de la théorie des probabilités : le théorème central limite (que l'on verra dans le chapitre suivant). Elle est proposée pour la première fois comme modèle des erreurs de mesures par le mathématicien allemand Gauss. À cette époque, l'expression de sa densité est déjà connue par les travaux antérieurs de De Moivre, poursuivis par Laplace. Pour cette raison, on rencontrera parfois la loi normale sous le nom de loi de Laplace–Gauss, loi gaussienne, loi de De Moivre–Laplace–Gauss etc ... . La forme caractéristique ≪en cloche≫ de sa densité est observée dans les études statistiques de nombreux phénomènes réels.

4 Lois à densité usuelle

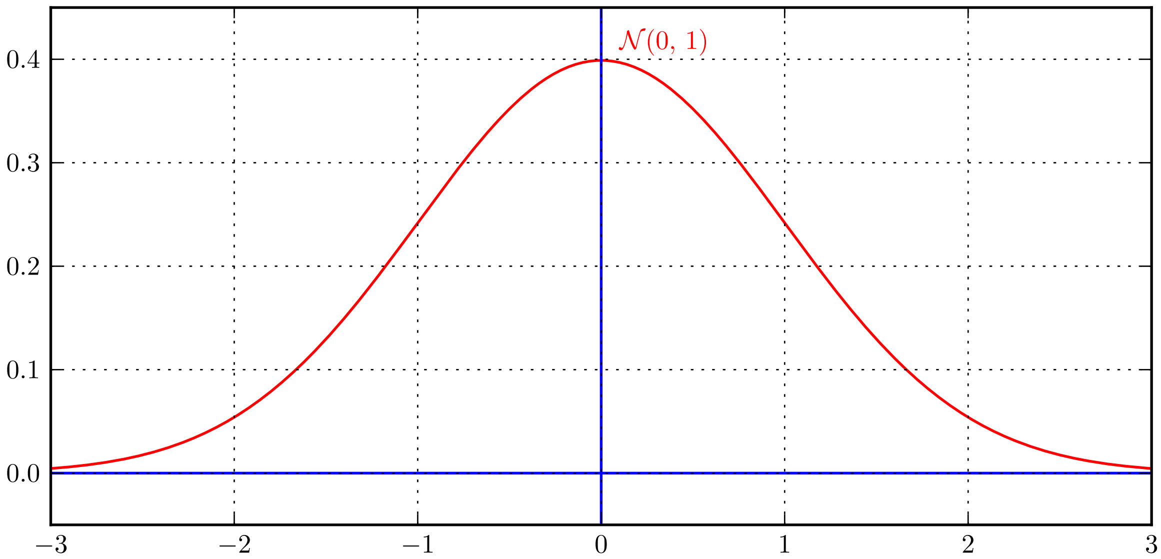
4.1 Densité de la loi normale centrée réduite $\mathcal{N}(0,\, 1)$

Définition 4.1
On suppose donné un espace probabilisé $(\Omega,\, \mathcal{F},\, \mathbf{P})$ qui ne sera pas précisé. Une variable aléatoire continue Z est dite de loi normale centrée réduite si sa fonction de densité est donnée par

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R}^+ \\ x &\longmapsto f(x) = \frac{1}{\sqrt{2\,\pi}}\, \mathrm{e}^{-\frac{x^2}{2}} . \end{aligned}$$

On notera $Z \sim \mathcal{N}(0,\, 1)$.

La représentation graphique de cette densité est donnée ci–dessous.



Cette définition prétend que la fonction définie ci–dessus est bien une densité. Or, s'il est clair qu'elle est continue sur $\mathbb{R}$ (comme composition et produit de fonctions continues) et positive, il reste à prouver que

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x = 1.$$

En tant que fonction continue sur $\mathbb{R}$, la fonction f admet des primitives mais **on n'en connaît pas** d'expression analytique. La démonstration de ce résultat ne s'appuie donc pas sur un simple calcul de primitive mais sur une technique d'intégration qui ne relève pas de ce cours. On admet donc ce résultat important, ainsi que les suivants qui sont du même type :

Proposition 4.1 (Moments d'une normale centrée réduite)
Si $Z \sim \mathcal{N}(0,\ 1)$, alors

$$\mathbf{E}(Z) = 0 \qquad \textit{et} \qquad \mathbf{V}(Z) = 1\ .$$

Il faut noter que la densité f est une fonction paire. Sa représentation graphique admet donc l'axe des ordonnées comme axe de symétrie ce qui suffit pour assurer que l'espérance soit nulle.

L'espérance est aussi la médiane.

4.2 Fonction de répartition, table

Comme toute variable aléatoire à densité, la fonction de répartition de la loi de Z est définie par

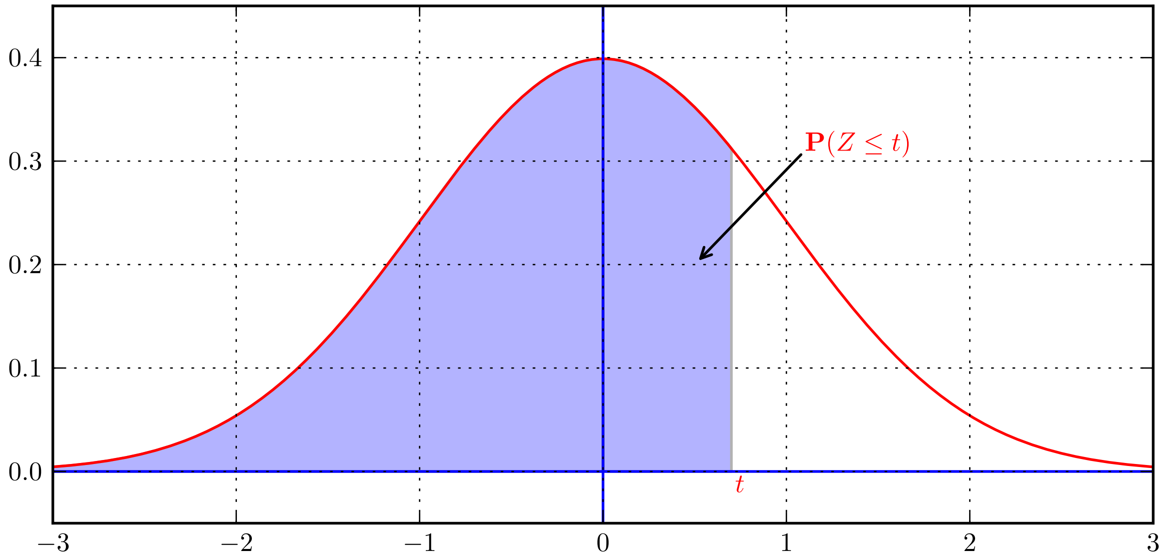
$$\begin{aligned} F\,:\, \mathbb{R} &\longrightarrow [0;\,1] \\ t &\longmapsto F(t) = \mathbf{P}(X \leq t) = \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{t} \mathrm{e}^{-\frac{x^2}{2}} x\,\mathrm{d}x\ . \end{aligned}$$

En pratique, on a recours à des algorithmes d'intégration numériques pour obtenir les valeurs de ces probabilités. Les sorties de ces algorithmes sont rassemblées dans des tables qui donnent la valeur de la fonction de répartition pour suffisamment de t. On distingue deux cas d'utilisation de ces tables.

4.2.1 Lecture directe

On se donne $t \in \mathbb{R}$ et on cherche $F(t) \in [0;\, 1]$. Deux cas se présentent :

$\boxed{\textbf{Si } t \geq 0 \textbf{ :}}$ la lecture est immédiate. Par l'argument de symétrie, la valeur lue sera forcément plus grande que $\frac{1}{2}$.



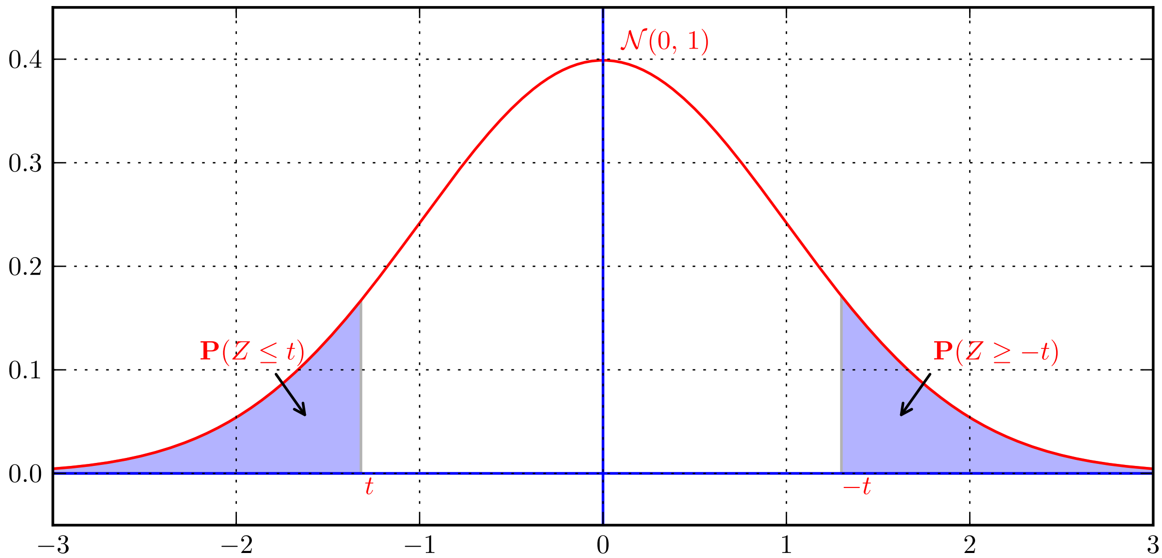
Par exemple, on lit dans la table que :

$$\mathbf{P}(Z \leq 1,56) = 0,9406\ .$$

$\boxed{\textbf{Si } t < 0 \textbf{ :}}$ l'argument de symétrie implique que

$$\mathbf{P}(Z \leq t) = \mathbf{P}(Z \geq -t) = 1 - \mathbf{P}(Z \leq -t)\ .$$

La valeur obtenue sera forcément plus petite que $\frac{1}{2}$.



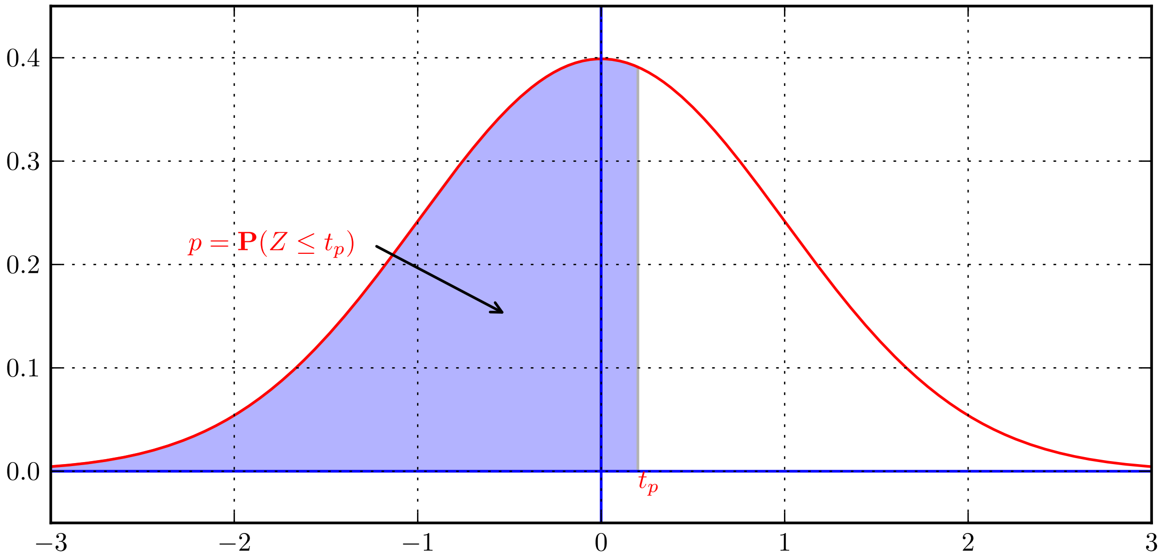
Par exemple, on lit dans la table que :

$$\mathbf{P}(Z \leq -1,78) = 1 - \mathbf{P}(Z \leq 1,78) = 1 - 0,9625 = 0,0375\ .$$

4.3 Lecture inverse (quantile)

On se donne maintenant $p \in [0; 1]$ et on cherche le quantile d'ordre p, c'est–à–dire le réel t_p tel que $\mathbf{P}(Z \leq t_p) = p$. Ici encore, on distingue deux cas.

$\boxed{\textbf{Si } p \geq \frac{1}{2} \textbf{ :}}$ le quantile recherché est nécessairement positif. La lecture se fait en recherchant dans le corps de la table la valeur la plus proche de p. Le quantile se lit alors dans les marges.

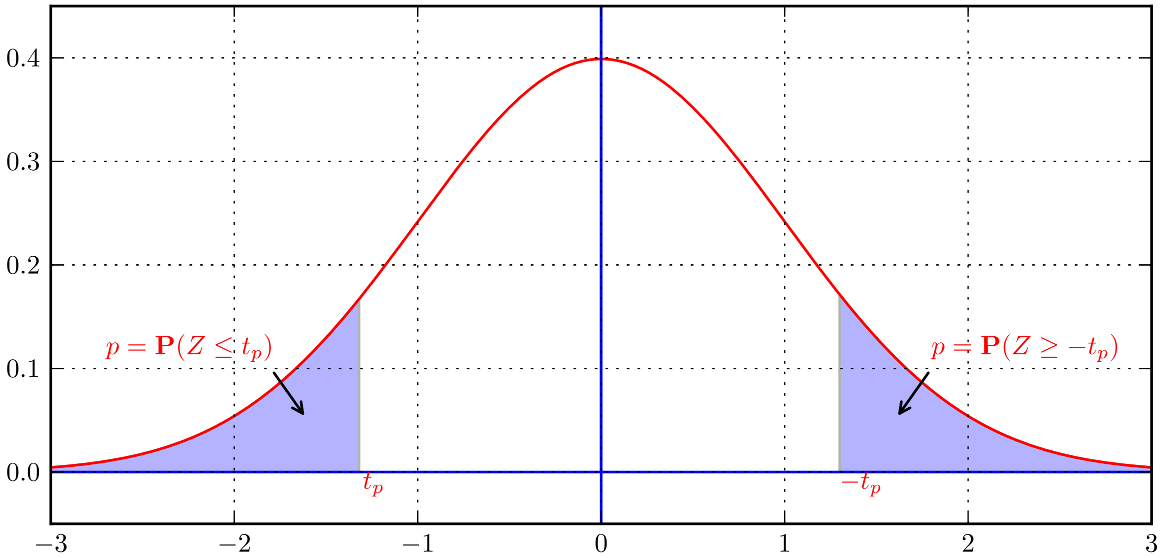


On obtient par exemple que $t_{0,65} \simeq 0,39$.

$\boxed{\textbf{Si } p < \frac{1}{2} \textbf{ :}}$ le quantile recherché est nécessairement négatif. Par l'argument de symétrie, le quantile t_p et son opposé sont liés par la relation

$$p = \mathbf{P}(Z \leq t_p) = 1 - \mathbf{P}(Z \leq -t_p)$$

c'est–à–dire $t_p = -t_{1-p}$.



On est alors ramené au cas précédent en recherchant le quantile d'ordre $1 - p \geq \frac{1}{2}$, dont on prendra l'opposé.

On lit par exemple que $t_{0,27} = -t_{0,73} = -0,61$.

4.4 Loi normale $\mathcal{N}(\mu,\, \sigma)$

La loi normale que nous venons d'étudier est centrée–réduite c'est-à-dire que son espérance est nulle et sa variance 1. On opère maintenant une transformation affine pour obtenir une espérance et une variance quelconque.

Soient $\mu \in \mathbb{R}$ et σ un réel **strictement positif**. On s'intéresse à la variable $X = \mu + \sigma\, Z$ où $Z \sim \mathcal{N}(0,\, 1)$. On se rappelle que la densité d'une variable aléatoire peut s'obtenir comme la dérivée de sa fonction de répartition. On calcule donc la fonction de répartition F_X de X en fonction de F_Z, celle de Z.

$$\begin{aligned} F_X(x) = \mathbf{P}(X \leq x) &= \mathbf{P}(\mu + \sigma\, Z \leq x) \\ &= \mathbf{P}\left(Z \leq \frac{x-\mu}{\sigma}\right) \\ &= F_Z\left(\frac{x-\mu}{\sigma}\right) \ . \end{aligned}$$

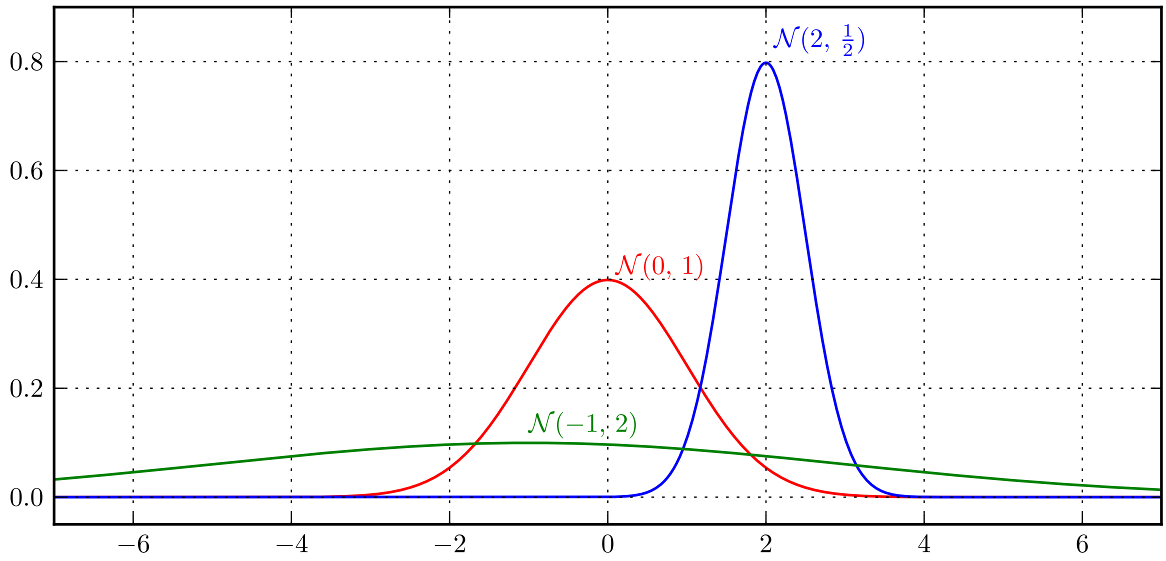
Il suffit alors de dériver cette dernière composition de fonction et d'utiliser le fait que la dérivée de F_Z est précisément la densité de la loi normale centrée réduite.

$$\begin{aligned} f_X(x) &= F_Z'\left(\frac{x-\mu}{\sigma}\right)\left(\frac{x-\mu}{\sigma}\right)' \\ &= f_Z\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma} \\ &= \frac{1}{\sqrt{2\,\pi}\,\sigma}\,\mathrm{e}^{-\dfrac{(x-\mu)^2}{2\,\sigma^2}} \ . \end{aligned}$$

Cette dernière expression est appelée *densité de la loi normale d'espérance μ et d'écart–type σ*, notée $\mathcal{N}(\mu,\, \sigma)$. Le calcul de l'espérance et de la variance est en effet immédiat :

$$\begin{aligned} \mathbf{E}(X) &= \mathbf{E}(\mu + \sigma\, Z) = \mu + \sigma\, \mathbf{E}(Z) = \mu \\ \mathbf{V}(X) &= \mathbf{V}(\mu + \sigma\, Z) = \sigma^2\, \mathbf{V}(Z) = \sigma^2 \ . \end{aligned}$$

En étudiant la densité de la loi normale, on peut remarquer que l'allure générale de la courbe reste la même quels que soient l'espérance et l'écart–type. Un déplacement de l'espérance induit un décalage (translation) de la courbe et un changement d'écart–type induit une dilatation. Plus l'écart–type est grand, plus la courbe «s'étale». À l'inverse la courbe devient très «pointue» si l'écart–type est petit.



4.4.1 Fonction de répartition

Il est hors de question de produire de nombreuses tables pour les fonctions de répartition de plusieurs lois normales d'espérance et de variance quelconque. C'est de plus inutile : en effet, puisque la loi $\mathcal{N}(\mu,\,\sigma)$ a été obtenue par une transformation affine de la loi $\mathcal{N}(0,\,1)$, la transformation affine inverse permet d'obtenir la loi $\mathcal{N}(0,\,1)$ à partir de la loi $\mathcal{N}(\mu,\,\sigma)$.

Proposition 4.2
Si $X \sim \mathcal{N}(\mu,\,\sigma)$, alors

$$Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,\,1).$$

Il suffit de reprendre à l'envers le calcul qui a permis d'obtenir $\mathcal{N}(\mu,\,\sigma)$ au moyen de $\mathcal{N}(0,\,1)$.

Corollaire 4.3
La fonction de répartition de la loi $\mathcal{N}(\mu,\,\sigma)$ s'exprime au moyen de la fonction de répartition de la loi $\mathcal{N}(0,\,1)$ grâce à la relation :

$$\begin{aligned} F_X(t) = \mathbf{P}(X \le t) &= \mathbf{P}\left(\frac{X-\mu}{\sigma} \le \frac{t-\mu}{\sigma}\right) \\ &= \mathbf{P}\left(Z \le \frac{t-\mu}{\sigma}\right) \\ &= F_Z\left(\frac{t-\mu}{\sigma}\right) \ . \end{aligned}$$

On remarque que l'espérance est encore la médiane.

4.5 Somme de variable aléatoires normales

Proposition 4.4
*Si $X_1 \sim \mathcal{N}(\mu_1,\,\sigma_1)$ et $X_2 \sim \mathcal{N}(\mu_2,\,\sigma_2)$ sont deux variables aléatoires définies sur le même espace et **indépendantes**, alors $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2,\,\sqrt{\sigma_1^2 + \sigma_2^2})$.*

Il faut remarquer que l'espérance et la variance d'une somme de variables aléatoires sont déjà connues. Ce qui est remarquable dans ce résultat, est que la loi de la somme reste normale. Ce n'est pas le cas pour toutes les lois.

Par ailleurs, une conséquence de cette proposition est que toute combinaison linéaire non–nulle de normales indépendantes est de loi normale.

4.6 Autres densités

4.6.1 Densité de la loi du chi-deux χ^2

La loi du χ^2 est une loi à densité de probabilité. Cette loi est caractérisée par un paramètre dit degrés de liberté à valeur dans l'ensemble des entiers naturels (non nuls).
Soit Z_1, Z_2, ..., Z_k, k v.a. **indépendantes de loi normale centrée et réduite**. Alors :

$$X = \sum_{i=1}^{k} Z_i^2 \text{ suit une loi du } \chi^2 \text{ à } k \text{ degrés de liberté que l'on notera } \chi^2(k).$$

L'espérance mathématique de X vaut k et sa variance vaut $2k$.

4.6.2 Densité de la loi de Student

La loi de Student est une loi de probabilité, faisant intervenir le quotient entre une variable suivant une loi normale centrée réduite et la racine carrée d'une variable distribuée suivant la loi du χ^2.
Soit Z une variable aléatoire de loi **normale centrée et réduite** et soit U une variable **indépendante** de Z et distribuée suivant la loi du χ^2 à k degrés de liberté.
Par définition, la variable $T = \dfrac{Z}{\sqrt{U/k}}$ suit une loi de Student à k degrés de liberté notée $t(k)$.
Son espérance ne peut pas être définie pour $k = 1$, et est nulle pour $k > 1$. Sa variance est infinie pour $k \leq 2$ et vaut $\dfrac{k}{k-2}$ pour $k > 2$.