

Des probabilités aux aires

Densité, répartition et loi normale – avant le TD6

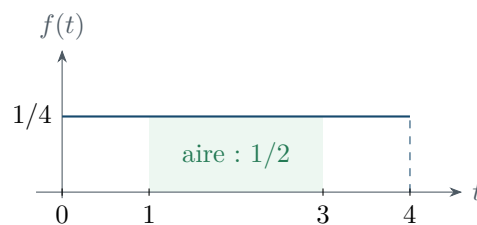
Le changement essentiel

Aux TD4 et TD5, nous additionnions les probabilités de valeurs isolées. Dans les modèles à densité du TD6, une probabilité se lit comme une **aire sous une courbe**. Les intégrales permettent de calculer ces aires.

1. Une densité n'est pas une probabilité

On modélise un temps d'attente U compris entre 0 et 4 minutes par une loi uniforme : des intervalles de même longueur ont la même probabilité. La densité est

$$f(t) = \begin{cases} 1/4 & 0 \leq t \leq 4, \\ 0 & \text{sinon.} \end{cases}$$



L'aire totale est $4 \times (1/4) = 1$. Entre 1 et 3 minutes, l'aire du rectangle vaut

$$\mathbb{P}(1 < U < 3) = (3 - 1) \times \frac{1}{4} = \frac{1}{2}.$$

La hauteur $f(2) = 1/4$ n'est donc pas « la probabilité d'attendre exactement 2 minutes ».

Les deux conditions pour une densité

Une fonction f est une densité si

$$f(t) \geq 0 \text{ pour tout } t, \quad \int_{-\infty}^{+\infty} f(t) dt = 1.$$

Pour une variable X de densité f et $a < b$,

$$\mathbb{P}(a < X < b) = \int_a^b f(t) dt.$$

Dans ce modèle, un point isolé a une probabilité nulle : $\mathbb{P}(U = 2) = 0$. Cela ne signifie pas que 2 est une valeur impossible. Les bornes incluses ou exclues ne changent pas la probabilité d'un intervalle.

Point de vigilance

Une densité peut dépasser 1 : la fonction égale à 2 sur $[0; 1/2]$ et nulle ailleurs est une densité, car son aire totale est 1. **C'est l'aire, et non la hauteur, qui est une probabilité.**

2. Les intégrales dont on a besoin

Une primitive permet de calculer une intégrale

Si A est une primitive de h sur $[a; b]$, c'est-à-dire $A' = h$, alors

$$\int_a^b h(t) dt = [A(t)]_a^b = A(b) - A(a).$$

La lettre t est la variable de calcul ; a et b sont les bornes.

Fonction à intégrer	Une primitive
1	t
t^n (n entier, $n \geq 0$)	$t^{n+1}/(n+1)$
$e^{-\alpha t}$ ($\alpha > 0$)	$-e^{-\alpha t}/\alpha$

Exemple guidé : régler l'aire totale

On cherche c pour que $f(t) = ct$ sur $[0; 2]$, et $f(t) = 0$ ailleurs, soit une densité. La positivité demande $c \geq 0$. Ensuite :

$$\int_0^2 ct dt = c \left[\frac{t^2}{2} \right]_0^2 = 2c.$$

L'aire doit valoir 1, donc $c = 1/2$. **Nous garderons cet exemple :**

$$f(t) = t/2 \text{ sur } [0; 2], \quad f(t) = 0 \text{ ailleurs.}$$

Quand il y a plusieurs morceaux ou une borne infinie

Plusieurs expressions : couper le calcul au point de changement.

$$\int_a^b f(t) dt = \int_a^r f(t) dt + \int_r^b f(t) dt \quad (a < r < b).$$

Une borne infinie : calculer d'abord jusqu'à B , puis passer à la limite. Par exemple,

$$\int_1^{+\infty} \frac{1}{t^2} dt = \lim_{B \rightarrow +\infty} \left[-\frac{1}{t} \right]_1^B = \lim_{B \rightarrow +\infty} \left(1 - \frac{1}{B} \right) = 1.$$

Un ordre de vérification efficace

Vérifier d'abord le **signe**, puis l'**aire totale**. Une fonction positive sur un intervalle ne devient pas automatiquement une densité. La fonction égale à 1 sur $[0; 2]$, et nulle ailleurs, a par exemple une aire totale de 2.

Point de vigilance

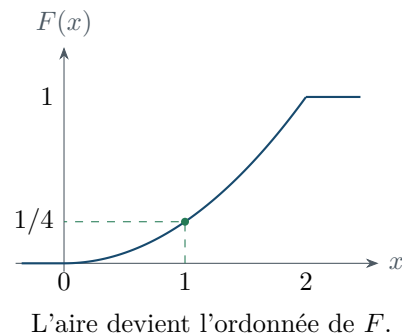
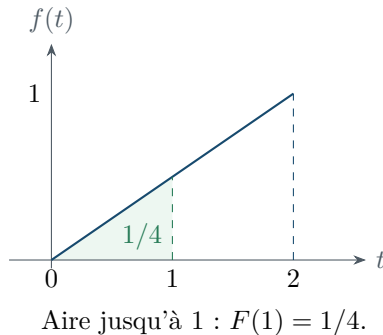
Une aire totale de 1 ne suffit pas non plus : $4t - 1$ sur $[0; 1]$, nulle ailleurs, a pour intégrale 1 mais est négative pour $0 \leq t < 1/4$. Les deux conditions sont nécessaires.

3. La répartition : cumuler l'aire jusqu'à x

La fonction de répartition est définie pour toute variable aléatoire par

$$F(x) = \mathbb{P}(X \leq x).$$

Si X possède une densité f , alors $F(x) = \int_{-\infty}^x f(t) dt$.



Construire F dans notre exemple

Avant 0, aucune aire ; entre 0 et 2, on intègre ; après 2, toute l'aire a été cumulée :

$$F(x) = \begin{cases} 0 & x < 0, \\ \int_0^x \frac{t}{2} dt = \frac{x^2}{4} & 0 \leq x \leq 2, \\ 1 & x > 2. \end{cases}$$

On vérifie les raccords : $F(0) = 0$ et $F(2) = 1$.

Passer de l'aire cumulée à une probabilité d'intervalle

Pour une variable à densité, $\mathbb{P}(a < X < b) = F(b) - F(a)$. Ici,

$$\mathbb{P}(0,5 < X < 1,5) = \frac{1,5^2}{4} - \frac{0,5^2}{4} = \frac{1}{2}.$$

Dans les exemples à densité du TD, on retrouve $f = F'$ à l'intérieur des morceaux où F est dérivable. Les valeurs de f aux points de raccord ne changent pas les intégrales.

Une fonction de répartition peut avoir un saut

Toute fonction de répartition est croissante, continue à droite, de limites 0 et 1 aux deux infinis. Elle n'est pas toujours continue : un saut représente une probabilité portée par un point. Une telle loi n'a pas de densité pour toute sa masse. **Pour une loi à densité, F est continue.** Cette distinction compte dans l'exercice 3 du TD6.

4. Espérance et variance : remplacer les sommes

Le sens des moments reste celui du préliminaire du TD4. Pour une variable à densité, les sommes pondérées sont remplacées par des intégrales, lorsque les moments existent.

	Cas discret	Cas à densité
Espérance	$\sum_k x_k \mathbb{P}(X = x_k)$	$\int_{\mathbb{R}} t f(t) dt$
Second moment	$\sum_k x_k^2 \mathbb{P}(X = x_k)$	$\int_{\mathbb{R}} t^2 f(t) dt$

La formule de variance reste la même

$$\text{Var}(X) = E(X^2) - [E(X)]^2, \quad \sigma(X) = \sqrt{\text{Var}(X)}.$$

Calculer avec $f(t) = t/2$ sur $[0; 2]$

$$E(X) = \int_0^2 t \times \frac{t}{2} dt = \left[\frac{t^3}{6} \right]_0^2 = \frac{4}{3},$$

$$E(X^2) = \int_0^2 t^2 \times \frac{t}{2} dt = \left[\frac{t^4}{8} \right]_0^2 = 2.$$

Ainsi

$$\text{Var}(X) = 2 - \left(\frac{4}{3} \right)^2 = \frac{2}{9}, \quad \sigma(X) = \frac{\sqrt{2}}{3} \approx 0,471.$$

La densité est plus élevée vers 2 : la moyenne $4/3$ se trouve à droite du milieu 1. Ce contrôle graphique aide à repérer une erreur de calcul.

La symétrie peut simplifier un calcul

Si la densité est symétrique autour de 0, donc $f(-t) = f(t)$, et si l'espérance existe, les contributions négatives et positives à $E(X)$ se compensent : $E(X) = 0$. Pour le second moment, $t^2 f(t)$ est alors paire :

$$E(X^2) = 2 \int_0^{+\infty} t^2 f(t) dt.$$

Si le support est $[-a; a]$, la borne supérieure devient a . Pour une densité définie par morceaux, on sépare les intégrales au changement d'expression.

Point de vigilance

$E(X^2)$ n'est pas $[E(X)]^2$. Une moyenne nulle n'implique pas une variance nulle. L'écart-type donne un ordre de grandeur de la dispersion autour de la moyenne ; il n'est pas la moyenne des distances à celle-ci.

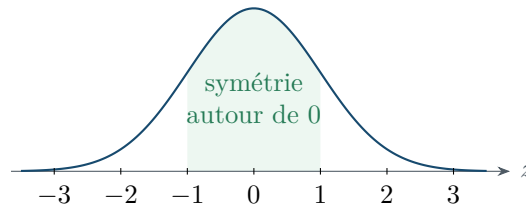
5. La loi normale : se ramener à une seule table

La loi normale possède une densité en cloche, symétrique autour de sa moyenne. Dans cette feuille, la notation est

$$X \sim \mathcal{N}(\mu; \sigma^2), \quad \sigma > 0.$$

μ est la **moyenne** ; σ est l'**écart-type** ; σ^2 est la **variance**. La normale centrée réduite a pour moyenne 0 et pour variance 1 :

$$Z \sim \mathcal{N}(0; 1), \quad \varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$



Centrer et réduire

Si $X \sim \mathcal{N}(\mu; \sigma^2)$, alors

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

Soustraire μ centre la variable ; diviser par σ mesure les écarts en nombre d'écarts-types. Comme $\sigma > 0$, le sens des inégalités est conservé.

Exemple : une masse mesurée en grammes

On modélise une masse T par $\mathcal{N}(50; 4^2)$. La moyenne est 50 g et l'écart-type 4 g. La variable $Z = (T - 50)/4$ suit $\mathcal{N}(0; 1)$.

Seuil pour T	46	50	54
Seuil pour Z	-1	0	1

Par exemple, $\mathbb{P}(T < 54) = \mathbb{P}(Z < 1)$. Le seuil 54 g est situé un écart-type au-dessus de la moyenne.

Point de vigilance

On divise par l'**écart-type** 4, et non par la variance 16. Une densité normale ne s'intègre pas à l'aide des primitives élémentaires de la page 2 : on utilise la table de sa répartition, une calculatrice ou un logiciel.

6. Lire la table et utiliser la symétrie

On note Φ la fonction de répartition de $Z \sim \mathcal{N}(0; 1)$:

$$\Phi(z) = \mathbb{P}(Z \leq z) = \int_{-\infty}^z \varphi(t) dt.$$

Un extrait pour apprendre à lire la table

Ligne + colonne	0,00	0,04	0,06	0,08
0,0	0,5000	0,5160	0,5239	0,5319
1,0	0,8413	0,8508	0,8554	0,8599
1,2	0,8849	0,8925	0,8962	0,8997
1,9	0,9713	0,9738	0,9750	0,9761

Pour $1,28 = 1,2 + 0,08$, on lit la ligne 1,2 et la colonne 0,08 :

$$\Phi(1,28) \approx 0,8997, \quad \mathbb{P}(Z > 1,28) \approx 1 - 0,8997 = 0,1003.$$

Attention au titre de la table du TD6

Malgré le titre « $\mathbb{P}(0 \leq Z \leq z)$ », les valeurs fournies sont celles de $\Phi(z)$. On le voit à la valeur 0,5000 en 0. Une aire entre 0 et 0 vaudrait 0. Pour $z \geq 0$, l'aire entre 0 et z vaut $\Phi(z) - 1/2$.

Un seuil négatif : utiliser la symétrie

L'aire à gauche de $-z$ égale l'aire à droite de z , donc

$$\Phi(-z) = 1 - \Phi(z).$$

Pour le justifier par changement de variable, poser $u = -t$, donc $dt = -du$, et utiliser $\varphi(-u) = \varphi(u)$:

$$\Phi(-z) = \int_{-\infty}^{-z} \varphi(t) dt = \int_{+\infty}^z \varphi(-u)(-du) = \int_z^{+\infty} \varphi(u) du = 1 - \Phi(z).$$

Pour un intervalle : aire cumulée à droite moins aire cumulée à gauche

$$\mathbb{P}(a < Z < b) = \Phi(b) - \Phi(a) \quad (a < b).$$

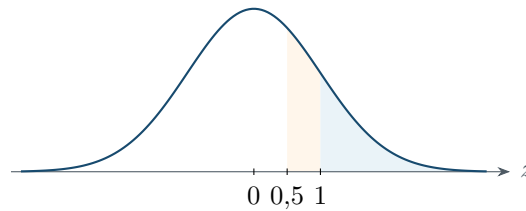
Par exemple,

$$\mathbb{P}(-1,2 < Z < 1,28) = \Phi(1,28) + \Phi(1,2) - 1 \approx 0,7846.$$

7. Conditionner et trouver un intervalle central

« Parmi ceux qui... » : une probabilité conditionnelle

On reprend $T \sim \mathcal{N}(50; 4^2)$. Parmi les objets de masse supérieure à 52 g, quelle proportion dépasse 54 g ? Le groupe de référence est maintenant $B = \{T > 52\}$.



Sur l'axe réduit, 52 g correspond à 0,5 et 54 g à 1. L'aire orange et bleue représente B ; l'aire bleue représente $A = \{T > 54\}$.

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Ici $A \subset B$, donc $A \cap B = A$:

$$\mathbb{P}(T > 54 | T > 52) = \frac{1 - \Phi(1)}{1 - \Phi(0,5)} \approx \frac{0,1587}{0,3085} \approx 0,514.$$

Environ 51,4 % des objets dépassant 52 g dépassent aussi 54 g. On **divise les aires**, comme au TD3 ; on ne les soustrait pas.

Un intervalle central contenant environ 95 % des valeurs

La symétrie partage les 5 % restants en deux queues de 2,5 %. Pour $q \geq 0$,

$$\mathbb{P}(-q < Z < q) = 2\Phi(q) - 1.$$

Pour une aire centrale de 0,95, il faut $\Phi(q) = 0,975$. La table donne $q \approx 1,96$.

Revenir aux unités de départ

Pour $T \sim \mathcal{N}(50; 4^2)$,

$$\mathbb{P}(50 - 1,96 \times 4 < T < 50 + 1,96 \times 4) \approx 0,95.$$

L'intervalle central est $[42,16 ; 57,84]$ g. Plus généralement, on utilise $[\mu - 1,96\sigma ; \mu + 1,96\sigma]$.

Point de vigilance

Cet intervalle décrit la dispersion d'une **mesure individuelle** dans le modèle. Un intervalle de confiance concerne un paramètre inconnu estimé à partir d'un échantillon : c'est une autre question.

8. Choisir le bon chemin dans le TD6

Ce qui est demandé	Le premier réflexe
Est-ce une densité ?	Signe, puis aire totale 1
Trouver une constante	Écrire que l'intégrale totale vaut 1 et vérifier le signe
Construire F à partir de f	Cumuler l'aire jusqu'à x ; distinguer les morceaux
Trouver f à partir de F	Vérifier le cadre à densité et les raccords, puis dériver sur les morceaux
Calculer l'espérance et la variance	Calculer $E(X)$ et $E(X^2)$, puis les combiner
Une probabilité sous une normale	Centrer et réduire, dessiner la zone, utiliser Φ
« Parmi les valeurs qui... »	Identifier le conditionnement et former un quotient

S'entraîner avant de lire les réponses

- La fonction égale à 2 sur $[0; 1]$ et à 0 ailleurs est-elle une densité ?
- Si $f(t) = ct$ sur $[0; 3]$, nulle ailleurs, quelle valeur de c donne une densité ?
- Une variable de densité f vérifie $f(1) = 0,8$. Peut-on conclure que $\mathbb{P}(X = 1) = 0,8$?
- Pour notre exemple $f(t) = t/2$ sur $[0; 2]$, que vaut $\mathbb{P}(X > 1)$?
- Si $W \sim \mathcal{N}(10; 3^2)$, quel seuil réduit correspond à $W = 7$?
- Une table donne $\Phi(1,28) = 0,8997$. Que valent $\mathbb{P}(Z > 1,28)$ et $\Phi(-1,28)$?
- Une fonction de répartition possède un saut en 0. Peut-elle être celle d'une variable à densité ?

Réponses pour se vérifier

- a) Non : aire totale 2. b) $c \times 9/2 = 1$, donc $c = 2/9$.
 c) Non : $\mathbb{P}(X = 1) = 0$ dans un modèle à densité.
 d) $1 - F(1) = 1 - 1/4 = 3/4$. e) $(7 - 10)/3 = -1$.
 f) Les deux valent $1 - 0,8997 = 0,1003$.
 g) Non : le saut représente une masse ponctuelle qu'une densité ne peut porter.

Où revenir pendant les exercices ?

Exercice 1 : pages 1–2 ; exercice 2 : pages 2–3 ; exercice 3 : page 3 ; exercices 4–5 : page 4 ; exercices 6–8 : pages 5–7. Le sens de l'espérance et de la dispersion a été étudié avant le TD4 ; les probabilités conditionnelles, au TD3.