

Mes Notes de Lecture

Introduction à la Probabilité

LOU BRUNET

25 octobre 2025

Table des matières

1	Préface	3
2	Probabilités et Dénombrement	4
2.1	Concepts fondamentaux	4
2.2	Définition Naïve de la Probabilité	4
2.3	Permutations (Arrangements)	5
2.4	Le Coefficient Binomial	6
2.5	Identité de Vandermonde	7
2.6	Bose-Einstein (Étoiles et Bâtons)	8
2.7	Principe d'Inclusion-Exclusion	8
3	Probabilité conditionnelle	12
3.1	Définition de la Probabilité Conditionnelle	12
3.2	Règle du Produit (Intersection de deux événements)	12
3.3	Règle de la Chaîne (Intersection de n événements)	13
3.4	Règle de Bayes	14
3.5	Formule des Probabilités Totales	14
3.6	Règle de Bayes avec Conditionnement Additionnel	16
3.7	Formule des Probabilités Totales avec Conditionnement Additionnel	16
3.8	Indépendance de Deux Événements	18
3.9	Indépendance Conditionnelle	18
3.10	Le Problème de Monty Hall	19
4	Variables Aléatoires Discrètes	21
4.1	Variable Aléatoire	21
4.2	Variable Aléatoire Discrète	21
4.3	Fonction de Masse (PMF)	22
4.4	Loi de Bernoulli	22
4.5	Loi Binomiale	23
4.6	Loi Hypergéométrique	24
4.7	Loi Géométrique	24
4.8	Loi de Poisson	25
4.9	Fonction de Répartition (CDF)	27
4.10	Variable Aléatoire Indicatrice	28
5	Espérance et autres notions associées aux variables aléatoires discrètes	29
5.1	Espérance d'une variable aléatoire discrète	29
5.2	Linéarité de l'espérance	29
5.3	Espérance de la loi binomiale	30
5.4	Espérance de la loi géométrique	31
5.5	Loi du statisticien inconscient (LOTUS)	32
5.6	Variance	33
6	Distributions Multivariées et Concepts Associés	34
6.1	Distributions Jointes et Marginales	34
6.2	Espérance d'une fonction de deux variables	35
6.3	Covariance et Corrélation	36
6.4	Linéarité de la Covariance	38
6.5	Résultats sur la Corrélation	38
6.6	Standardisation et Non-Corrélation	39
6.7	Variance d'une Somme de Variables Aléatoires	41
6.8	Théorème sur la somme de lois de Poisson	42

7	Variables Aléatoires Continues	44
7.1	Fonction de Densité de Probabilité (PDF)	44
7.2	Fonction de Répartition (CDF)	45
7.3	Espérance et Variance (Cas Continu)	46
7.4	Loi Uniforme	47
7.5	Loi Exponentielle	48
7.6	Distributions Conjointes (Cas Continu)	49
7.7	Espérance, Indépendance et Covariance (Cas Conjoint)	50
8	La Loi Normale (ou Gaussienne)	52
8.1	Introduction et Fonction de Densité (PDF)	52
8.2	La Loi Normale Centrée Réduite $\mathcal{N}(0, 1)$	55
8.3	Standardisation : Le Score Z	56
8.4	Propriétés Importantes de la Loi Normale	57
8.5	La Règle Empirique (68-95-99.7)	58
8.6	Calcul de Probabilités Normales	59
9	Moments d'une distribution	60
9.1	Définitions fondamentales des moments	60
9.2	Asymétrie (Skewness)	60
9.3	Propriétés de symétrie	61
9.4	Aplatissement (Kurtosis)	62
9.5	Exemples de distributions	63
10	Appendice A : Séries de Taylor et Maclaurin	66
10.1	Construction pas à pas d'une série de Taylor	66
10.2	Intuition de la série de Taylor en un point quelconque a	68
10.3	La Fonction Exponentielle (e^x)	69
10.4	La Fonction Sinus ($\sin(x)$)	70
10.5	La Fonction Cosinus ($\cos(x)$)	71
10.6	Le Logarithme Népérien ($\ln(1 + x)$)	72
10.7	La Série Géométrique ($\frac{1}{1-x}$)	74

1 Préface

En bien

2 Probabilités et Dénombrement

2.1 Concepts fondamentaux

Intuition : Nécessité d'un Cadre Formel

Avant de calculer des probabilités, il est crucial de définir les règles du jeu :

Qu'est-ce qui peut arriver ?

On définit l'ensemble de tous les résultats possibles de l'expérience.

À quoi s'intéresse-t-on ?

On identifie les sous-ensembles de résultats spécifiques qui nous intéressent.

Ces deux idées nous conduisent aux notions d'Univers et d'Événement, qui sont les piliers de toute théorie des probabilités.

Définition : Concepts Fondamentaux

Univers (ou Espace Échantillon), S :

L'ensemble de tous les résultats possibles d'une expérience aléatoire.

Événement, A :

Un sous-ensemble de l'univers ($A \subseteq S$). C'est un ensemble de résultats auxquels on s'intéresse.

Exemple : Univers et Événement

Pour l'expérience du "lancer d'un dé à six faces" :

L'**univers** est $S = \{1, 2, 3, 4, 5, 6\}$. "Obtenir un nombre impair" est un événement, représenté par le sous-ensemble $A = \{1, 3, 5\}$.

2.2 Définition Naïve de la Probabilité

Définition : Probabilité Naïve

Pour une expérience où chaque issue dans un espace échantillon fini S est équiprobable, la probabilité d'un événement A est le rapport du nombre d'issues favorables à A sur le nombre total d'issues :

$$P(A) = \frac{\text{Nombre d'issues favorables}}{\text{Nombre total d'issues}} = \frac{|A|}{|S|}$$

Exemple : Applications de la définition naïve

1. **Lancer une pièce équilibrée** : L'espace échantillon est $S = \{\text{Pile}, \text{Face}\}$, donc $|S| = 2$. Si l'événement A est "obtenir Pile", alors $A = \{\text{Pile}\}$ et $|A| = 1$. La probabilité est $P(A) = \frac{1}{2}$.
2. **Lancer un dé à six faces non pipé** : L'espace échantillon est $S = \{1, 2, 3, 4, 5, 6\}$, donc $|S| = 6$. Si l'événement B est "obtenir un nombre pair", alors $B = \{2, 4, 6\}$ et $|B| = 3$. La probabilité est $P(B) = \frac{3}{6} = \frac{1}{2}$.
3. **Tirer une carte d'un jeu de 52 cartes** : L'espace échantillon S contient 52 cartes, donc $|S| = 52$. Si l'événement C est "tirer un Roi", il y a 4 Rois dans le jeu, donc $|C| = 4$. La probabilité est $P(C) = \frac{4}{52} = \frac{1}{13}$.

2.3 Permutations (Arrangements)

Définition : Permutation de k objets parmi n

Le nombre de façons d'arranger k objets choisis parmi n objets distincts (où l'ordre compte et il n'y a pas de répétition) est noté $P(n, k)$ ou A_n^k et est défini par :

$$P(n, k) = \frac{n!}{(n-k)!}$$

où $n!$ est la factorielle de n , et par convention $0! = 1$.

Intuition : Permutations de k parmi n

Pour placer k objets dans un ordre spécifique en les choisissant parmi n objets disponibles, on a n choix pour la première position, $(n-1)$ choix pour la deuxième, ..., et $(n-k+1)$ choix pour la k -ième position. Cela donne $n \times (n-1) \times \cdots \times (n-k+1)$ arrangements. Ce produit contient k termes. Il est égal à $\frac{n!}{(n-k)!}$, car cela revient à diviser la suite complète $n!$ par les facteurs non utilisés $(n-k) \times (n-k-1) \times \cdots \times 1$.

Exemple : Permutations de k parmi n

Podium d'une course : Une course réunit 8 coureurs. Combien y a-t-il de podiums (1er, 2e, 3e) possibles ?

On cherche le nombre de façons d'ordonner 3 coureurs parmi 8 : $P(8, 3)$.

$$P(8, 3) = \frac{8!}{(8-3)!} = \frac{8!}{5!} = 8 \times 7 \times 6 = 336$$

Il y a 336 podiums possibles.

2.4 Le Coefficient Binomial

Théorème : Formule du Coefficient Binomial

Le nombre de façons de choisir k objets parmi un ensemble de n objets distincts (sans remise et sans ordre) est donné par le coefficient binomial :

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Intuition

L'idée est de relier $\binom{n}{k}$ à quelque chose de plus facile à compter : les **permutations** de k objets parmi n , c'est-à-dire les listes ordonnées. On sait qu'il y en a :

$$P(n, k) = \frac{n!}{(n-k)!}.$$

D'un autre côté, on peut construire chaque permutation en deux étapes :

1. Choisir un **sous-ensemble** de k objets (sans ordre), il y a $\binom{n}{k}$ façons de le faire.
2. Ordonner ces k objets, il y a $k!$ façons de le faire.

Donc, le nombre total de permutations est aussi $\binom{n}{k} \cdot k!$.

En égalisant les deux expressions :

$$\binom{n}{k} \cdot k! = \frac{n!}{(n-k)!} \implies \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Pour rendre cela concret, voici le cas $\binom{5}{3}$. Il y a 10 sous-ensembles de 3 éléments parmi $\{a, b, c, d, e\}$. Chacun donne lieu à $3! = 6$ permutations. Le tableau ci-dessous montre **toutes les 60 permutations**, regroupées par sous-ensemble :

$\{a, b, c\}$	$\{a, b, d\}$	$\{a, b, e\}$	$\{a, c, d\}$	$\{a, c, e\}$	$\{a, d, e\}$	$\{b, c, d\}$	$\{b, c, e\}$	$\{b, d, e\}$	$\{c, d, e\}$
abc	abd	abe	acd	ace	ade	bcd	bce	bde	cde
acb	adb	aeb	adc	aec	aed	bdc	bec	bed	ced
bac	bad	bae	cad	cae	dae	cbd	ceb	dbe	dce
bca	bda	bea	cda	cea	dea	cdb	ceb	deb	dec
cab	dab	eab	dac	eac	ead	dbc	ebc	edb	ecd
cba	dba	eba	dca	eca	eda	dcb	ebc	edb	edc

Chaque colonne correspond à **un seul et même choix non ordonné** (par exemple $\{a, b, c\}$), mais à 6 listes différentes selon l'ordre. Ainsi, pour obtenir le nombre de *choix non ordonnés*, on divise le nombre total de listes (60) par le nombre d'ordres par groupe (6) :

$$\binom{5}{3} = \frac{60}{6} = 10.$$

C'est exactement ce que fait la formule :

$$\binom{n}{k} = \frac{\text{nombre de permutations de } k \text{ parmi } n}{k!} = \frac{n!}{k!(n-k)!}.$$

Exemple : Utilisation du Coefficient Binomial

Comité d'étudiants : De combien de manières peut-on former un comité de 3 étudiants à partir d'une classe de 10 ? L'ordre ne compte pas.

$$\binom{10}{3} = \frac{10!}{3!(10-3)!} = \frac{10 \times 9 \times 8}{3 \times 2 \times 1} = 120 \text{ comités possibles.}$$

2.5 Identité de Vandermonde

Théorème : Identité de Vandermonde

Cette identité offre une relation remarquable entre les coefficients binomiaux. Pour des entiers non négatifs m, n et k , on a :

$$\binom{m+n}{k} = \sum_{j=0}^k \binom{m}{j} \binom{n}{k-j}$$

Intuition

C'est le "principe du diviser pour régner". Imaginez que vous devez choisir un comité de k personnes à partir d'un groupe contenant m hommes et n femmes. Le côté gauche, $\binom{m+n}{k}$, compte directement le nombre total de comités possibles. Le côté droit arrive au même résultat en additionnant toutes les compositions possibles du comité : choisir 0 homme et k femmes, PLUS 1 homme et $k-1$ femmes, PLUS 2 hommes et $k-2$ femmes, etc., jusqu'à choisir k hommes et 0 femme. La somme de toutes ces possibilités doit être égale au total.

Exemple : Application de l'Identité de Vandermonde

On veut former un comité de 3 personnes ($k=3$) à partir d'un groupe de 5 hommes ($m=5$) et 4 femmes ($n=4$). **Méthode directe (côté gauche) :**

On choisit 3 personnes parmi les $5+4=9$ au total.

$$\binom{9}{3} = \frac{9 \times 8 \times 7}{3 \times 2 \times 1} = 84$$

Méthode par cas (côté droit) :

La somme est $\binom{5}{0}\binom{4}{3} + \binom{5}{1}\binom{4}{2} + \binom{5}{2}\binom{4}{1} + \binom{5}{3}\binom{4}{0} = 84$. Les deux méthodes donnent bien le même résultat.

2.6 Bose-Einstein (Étoiles et Bâtons)

Théorème : Combinaisons avec répétition

Le nombre de façons de distribuer k objets indiscernables dans n boîtes discernables (ou de choisir k objets parmi n avec remise, où l'ordre ne compte pas) est donné par la formule :

$$\binom{n+k-1}{k} = \binom{n+k-1}{n-1}$$

Intuition : Étoiles et Bâtons

Imaginez que les k objets sont des étoiles (\star) et que nous avons besoin de $n-1$ bâtons ($|$) pour les séparer en n groupes. Par exemple, pour distribuer $k = 7$ étoiles dans $n = 4$ boîtes, une configuration possible serait :

$$\star \star \star | \star || \star \star \star$$

Cela correspond à 3 objets dans la première boîte, 1 dans la deuxième, 0 dans la troisième et 3 dans la quatrième. Le problème revient à trouver le nombre de façons d'arranger ces k étoiles et $n-1$ bâtons. Nous avons un total de $n+k-1$ positions, et nous devons choisir les k positions pour les étoiles (ou les $n-1$ positions pour les bâtons). Le nombre de manières de le faire est précisément $\binom{n+k-1}{k}$.

Exemple : Distribution de biens identiques

De combien de manières peut-on distribuer 10 croissants identiques à 4 enfants ? Ici, $k = 10$ (les croissants, objets indiscernables) et $n = 4$ (les enfants, boîtes discernables). Le nombre de distributions possibles est :

$$\binom{4+10-1}{10} = \binom{13}{10} = \binom{13}{3} = \frac{13 \times 12 \times 11}{3 \times 2 \times 1} = 13 \times 2 \times 11 = 286$$

Il y a 286 façons de distribuer les croissants.

2.7 Principe d'Inclusion-Exclusion

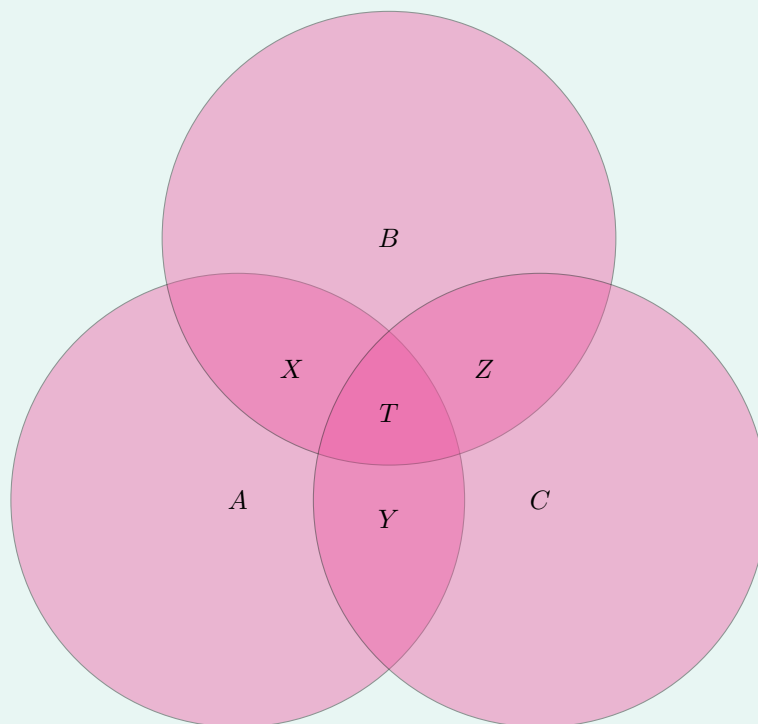
Théorème : Principe d'Inclusion-Exclusion pour 3 ensembles

Pour trois ensembles finis A , B et C , le nombre d'éléments dans leur union est donné par :

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$

Intuition : Visualisation avec 3 ensembles

Le principe d'inclusion-exclusion permet de compter le nombre d'éléments dans une union d'ensembles sans double-comptage. Pour comprendre intuitivement pourquoi on ajoute et soustrait alternativement, considérons trois ensembles A , B et C :



Le problème : Si on additionne simplement $|A| + |B| + |C|$, on compte certaines zones plusieurs fois :

- Les intersections deux à deux (X , Y , Z) sont comptées **deux fois**
- L'intersection triple (T) est comptée **trois fois**

La solution : On corrige en soustrayant les intersections deux à deux, mais alors l'intersection triple est comptée :

- +3 fois dans la somme initiale
- -3 fois dans la soustraction des intersections deux à deux (car elle appartient à chacune)
- Donc 0 fois au total ! Il faut la rajouter.

D'où la formule : $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$

Théorème : Principe d'Inclusion-Exclusion généralisé

Pour n ensembles finis A_1, A_2, \dots, A_n , on a :

$$\begin{aligned} |A_1 \cup A_2 \cup \dots \cup A_n| &= \sum_{i=1}^n |A_i| \\ &\quad - \sum_{1 \leq i < j \leq n} |A_i \cap A_j| \\ &\quad + \sum_{1 \leq i < j < k \leq n} |A_i \cap A_j \cap A_k| \\ &\quad - \dots \\ &\quad + (-1)^{n+1} |A_1 \cap A_2 \cap \dots \cap A_n| \end{aligned}$$

Ce qui s'écrit plus compactement :

$$\left| \bigcup_{i=1}^n A_i \right| = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} |A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}|$$

Intuition : Généralisation

La logique reste la même que pour trois ensembles, mais l'argument clé est de prouver que chaque élément est compté **exactement une fois**, peu importe le nombre d'ensembles auxquels il appartient.

Supposons qu'un élément x est membre d'exactement k ensembles parmi les n ensembles A_1, \dots, A_n . Analysons combien de fois x est compté dans la formule :

- **Première somme** ($\sum |A_i|$) : x est dans k ensembles, donc il est ajouté k fois. Le nombre de fois est $\binom{k}{1}$.
- **Deuxième somme** ($-\sum |A_i \cap A_j|$) : On soustrait x pour chaque paire d'ensembles auxquels il appartient. Il y a $\binom{k}{2}$ telles paires.
- **Troisième somme** ($+\sum |A_i \cap A_j \cap A_k|$) : On ajoute de nouveau x pour chaque triplet d'ensembles auxquels il appartient. Il y en a $\binom{k}{3}$.
- **Et ainsi de suite...**

Au total, l'élément x est compté :

$$\binom{k}{1} - \binom{k}{2} + \binom{k}{3} - \dots + (-1)^{k-1} \binom{k}{k} \text{ fois.}$$

Pour voir que cette somme vaut exactement 1, rappelons une identité fondamentale issue du binôme de Newton :

$$(1-1)^k = \sum_{j=0}^k (-1)^j \binom{k}{j} = \binom{k}{0} - \binom{k}{1} + \binom{k}{2} - \dots + (-1)^k \binom{k}{k} = 0$$

En réarrangeant cette équation, sachant que $\binom{k}{0} = 1$:

$$\binom{k}{0} = \binom{k}{1} - \binom{k}{2} + \binom{k}{3} - \dots - (-1)^k \binom{k}{k}$$

$$1 = \binom{k}{1} - \binom{k}{2} + \binom{k}{3} - \dots + (-1)^{k-1} \binom{k}{k}$$

Cela prouve que n'importe quel élément, qu'il soit dans un seul ensemble ($k = 1$) ou dans plusieurs ($k > 1$), contribue précisément pour 1 au décompte final. Le principe d'inclusion-exclusion est donc une méthode infaillible pour corriger les comptages multiples de manière systématique.

Exemple : Application probabiliste

On lance trois dés équilibrés. Quelle est la probabilité d'obtenir au moins un 6 ?

Solution avec inclusion-exclusion :

Soit A = "le premier dé montre 6", B = "le deuxième dé montre 6", C = "le troisième dé montre 6".

On veut $P(A \cup B \cup C)$.

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad + P(A \cap B \cap C) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} - \frac{1}{36} - \frac{1}{36} - \frac{1}{36} + \frac{1}{216} \\ &= \frac{3}{6} - \frac{3}{36} + \frac{1}{216} = \frac{1}{2} - \frac{1}{12} + \frac{1}{216} \\ &= \frac{108 - 18 + 1}{216} = \frac{91}{216} \approx 0.421 \end{aligned}$$

Vérification par la méthode complémentaire : La probabilité de n'obtenir aucun 6 est $\left(\frac{5}{6}\right)^3 = \frac{125}{216}$, donc la probabilité d'au moins un 6 est $1 - \frac{125}{216} = \frac{91}{216}$.

3 Probabilité conditionnelle

Intuition : Question Fondamentale

La probabilité conditionnelle est le concept qui répond à la question fondamentale : comment devons-nous mettre à jour nos croyances à la lumière des nouvelles informations que nous observons ?

3.1 Définition de la Probabilité Conditionnelle

Définition : Probabilité Conditionnelle

Si A et B sont deux événements avec $P(B) > 0$, alors la probabilité conditionnelle de A sachant B , notée $P(A|B)$, est définie comme :

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

Intuition

Imaginez que l'ensemble de tous les résultats possibles est un grand terrain. Savoir que l'événement B s'est produit, c'est comme si on vous disait que le résultat se trouve dans une zone spécifique de ce terrain. La probabilité conditionnelle $P(A|B)$ ne s'intéresse plus au terrain entier, mais seulement à la proportion de la zone B qui est également occupée par A . On "zoome" sur le monde où B est vrai, et on recalcule les probabilités dans ce nouveau monde plus petit.

3.2 Règle du Produit (Intersection de deux événements)

Théorème : Probabilité de l'intersection de deux événements

Pour tous événements A et B avec des probabilités positives, nous avons :

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$

Cela découle directement de la définition de la probabilité conditionnelle.

Intuition

Pour que deux événements se produisent, le premier doit se produire, PUIS le second doit se produire, sachant que le premier a eu lieu. Cette formule exprime mathématiquement cette idée séquentielle.

Exemple

Quelle est la probabilité de tirer deux As d'un jeu de 52 cartes sans remise? Soit A l'événement "le premier tirage est un As", avec $P(A) = \frac{4}{52}$. Soit B l'événement "le deuxième tirage est un As". Nous cherchons $P(A \cap B)$, que l'on calcule avec la formule $P(A \cap B) = P(A) \times P(B|A)$. La probabilité $P(B|A)$ correspond à tirer un As sachant que la première carte était un As. Il reste alors 51 cartes, dont 3 As. Donc, $P(B|A) = \frac{3}{51}$. Finalement, la probabilité de l'intersection est $P(A \cap B) = \frac{4}{52} \times \frac{3}{51} = \frac{12}{2652} \approx 0.0045$.

3.3 Règle de la Chaîne (Intersection de n événements)

Théorème : Probabilité de l'intersection de n événements

Pour tous événements A_1, \dots, A_n avec $P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$, nous avons :

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap \dots \cap A_{n-1})$$

Intuition

Ceci est une généralisation de l'idée précédente, souvent appelée "règle de la chaîne" (chain rule). Pour qu'une séquence d'événements se produise, chaque événement doit se réaliser tour à tour, en tenant compte de tous les événements précédents qui se sont déjà produits.

Exemple

On tire 3 cartes sans remise. Quelle est la probabilité d'obtenir la séquence Roi, Dame, Valet? La probabilité de tirer un Roi en premier (A_1) est $P(A_1) = \frac{4}{52}$. Ensuite, la probabilité de tirer une Dame (A_2) sachant qu'un Roi a été tiré est $P(A_2|A_1) = \frac{4}{51}$. Enfin, la probabilité de tirer un Valet (A_3) sachant qu'un Roi et une Dame ont été tirés est $P(A_3|A_1 \cap A_2) = \frac{4}{50}$. La probabilité totale de la séquence est donc le produit de ces probabilités : $P(A_1 \cap A_2 \cap A_3) = \frac{4}{52} \times \frac{4}{51} \times \frac{4}{50} \approx 0.00048$.

3.4 Règle de Bayes

Théorème : Règle de Bayes

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

Intuition

La règle de Bayes est la formule pour "inverser" une probabilité conditionnelle. Souvent, il est facile de connaître la probabilité d'un effet étant donné une cause ($P(\text{symptôme}|\text{maladie})$), mais ce qui nous intéresse vraiment, c'est la probabilité de la cause étant donné l'effet observé ($P(\text{maladie}|\text{symptôme})$). La règle de Bayes nous permet de faire ce retournement en utilisant notre connaissance initiale de la probabilité de la cause ($P(\text{maladie})$). C'est le fondement mathématique de la mise à jour de nos croyances.

Exemple : Dépistage médical

Une maladie touche 1% de la population ($P(M) = 0.01$). Un test de dépistage est fiable à 95% : il est positif pour 95% des malades ($P(T|M) = 0.95$) et négatif pour 95% des non-malades, ce qui implique un taux de faux positifs de $P(T|\neg M) = 0.05$. Une personne est testée positive. Quelle est la probabilité qu'elle soit réellement malade, $P(M|T)$? On cherche $P(M|T) = \frac{P(T|M)P(M)}{P(T)}$. D'abord, on calcule $P(T)$ avec la formule des probabilités totales : $P(T) = P(T|M)P(M) + P(T|\neg M)P(\neg M) = (0.95 \times 0.01) + (0.05 \times 0.99) = 0.0095 + 0.0495 = 0.059$. Ensuite, on applique la règle de Bayes : $P(M|T) = \frac{0.95 \times 0.01}{0.059} \approx 0.161$. Malgré un test positif, il n'y a que 16.1% de chance que la personne soit malade.

3.5 Formule des Probabilités Totales

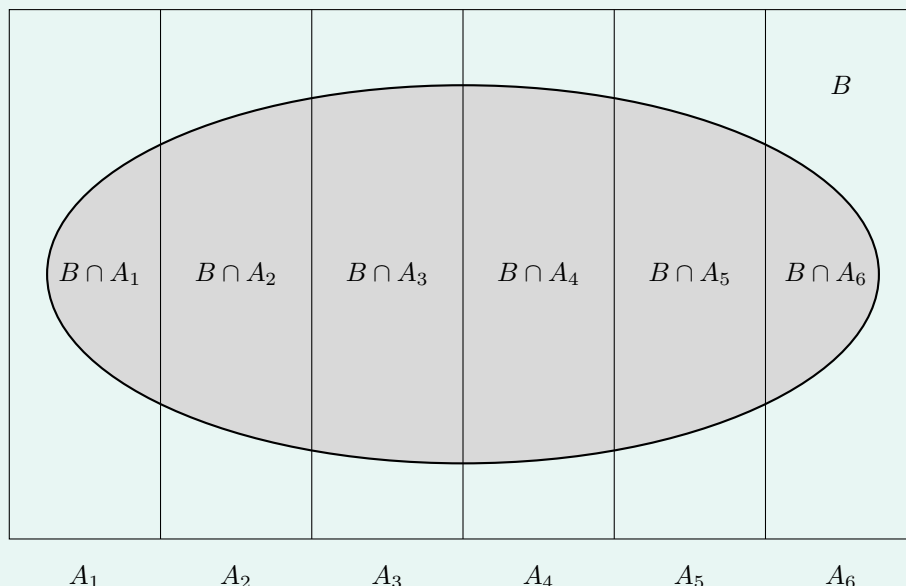
Théorème : Formule des probabilités totales

Soit A_1, \dots, A_n une partition de l'espace échantillon S (c'est-à-dire que les A_i sont des événements disjoints et leur union est S), avec $P(A_i) > 0$ pour tout i . Alors pour tout événement B :

$$P(B) = \sum_{i=1}^n P(B|A_i)P(A_i)$$

Intuition

C'est une stratégie de "diviser pour régner". Pour calculer la probabilité totale d'un événement B , on peut décomposer le monde en plusieurs scénarios mutuellement exclusifs (la partition A_i). On calcule ensuite la probabilité de B dans chacun de ces scénarios ($P(B|A_i)$), on pondère chaque résultat par la probabilité du scénario en question ($P(A_i)$), et on additionne le tout.



Exemple

Une usine possède trois machines, M1, M2, et M3, qui produisent respectivement 50%, 30% et 20% des articles. Leurs taux de production défectueuse sont de 4%, 2% et 5%. Quelle est la probabilité qu'un article choisi au hasard soit défectueux ? Soit D l'événement "l'article est défectueux". Les machines forment une partition avec $P(M1) = 0.5$, $P(M2) = 0.3$, et $P(M3) = 0.2$. Les probabilités conditionnelles de défaut sont $P(D|M1) = 0.04$, $P(D|M2) = 0.02$, et $P(D|M3) = 0.05$. En appliquant la formule, on obtient : $P(D) = P(D|M1)P(M1) + P(D|M2)P(M2) + P(D|M3)P(M3) = (0.04 \times 0.5) + (0.02 \times 0.3) + (0.05 \times 0.2) = 0.02 + 0.006 + 0.01 = 0.036$. La probabilité qu'un article soit défectueux est de 3.6%.

Preuve : Démonstration de la formule des probabilités totales

Puisque les A_i forment une partition de S , on peut décomposer B comme :

$$B = (B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_n)$$

Comme les A_i sont disjoints, les événements $(B \cap A_i)$ le sont aussi. On peut donc sommer leurs probabilités :

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \dots + P(B \cap A_n)$$

En appliquant le théorème de l'intersection des probabilités à chaque terme, on obtient :

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \dots + P(B|A_n)P(A_n) = \sum_{i=1}^n P(B|A_i)P(A_i)$$

3.6 Règle de Bayes avec Conditionnement Additionnel

Théorème : Règle de Bayes avec conditionnement additionnel

À condition que $P(A \cap E) > 0$ et $P(B \cap E) > 0$, nous avons :

$$P(A|B, E) = \frac{P(B|A, E)P(A|E)}{P(B|E)}$$

Intuition

Cette formule est simplement la règle de Bayes standard, mais appliquée à l'intérieur d'un univers que l'on a déjà "rétréci".

Imaginez que vous recevez une information **E** qui élimine une grande partie des possibilités. C'est votre nouveau point de départ, votre monde est plus petit. Toutes les probabilités que vous calculez désormais sont relatives à ce monde restreint.

Dans ce nouveau monde, vous recevez une autre information, l'évidence **B**. La règle de Bayes conditionnelle vous permet alors de mettre à jour votre croyance sur un événement **A**, en utilisant exactement la même logique que la règle de Bayes classique, mais en vous assurant que chaque calcul reste confiné à l'intérieur des frontières de l'univers défini par **E**.

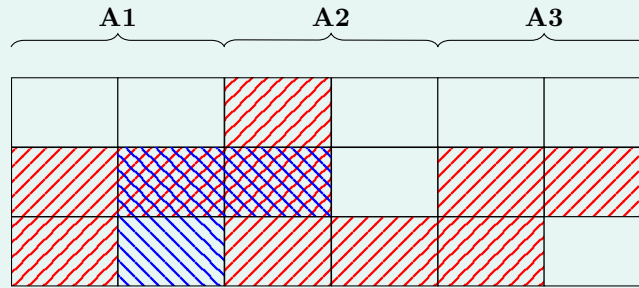
3.7 Formule des Probabilités Totales avec Conditionnement Additionnel

Théorème : Formule des probabilités totales avec conditionnement additionnel

Soit A_1, \dots, A_n une partition de S . À condition que $P(A_i \cap E) > 0$ pour tout i , nous avons :

$$P(B|E) = \sum_{i=1}^n P(B|A_i, E)P(A_i|E)$$

Intuition



Imaginez que le graphique ci-dessus représente la carte d'un trésor. La carte est partitionnée en trois grandes régions : **A1**, **A2**, et **A3**. Sur cette carte, on a identifié deux types de terrains : une **zone marécageuse** (événement E, hachures rouges) qui s'étend sur **10 parcelles**, et une **zone près d'un vieux chêne** (événement B, hachures bleues) qui couvre **3 parcelles**.

On vous donne un premier indice : "Le trésor est dans la zone marécageuse (E)". Votre univers de recherche se réduit instantanément à ces 10 parcelles rouges. Puis, on vous donne un second indice : "Le trésor est aussi près d'un chêne (B)". Votre recherche se concentre alors sur les parcelles qui sont à la fois marécageuses et proches d'un chêne (les cases violettes, $B \cap E$).

La question est : "Sachant que le trésor est dans une parcelle violette, quelle est la probabilité qu'il se trouve dans la région A2?". On cherche donc $P(A_2|B, E)$. La règle de Bayes nous permet de le calculer.

Calcul des termes nécessaires : D'abord, nous devons évaluer les probabilités à l'intérieur du "monde marécageux" (sachant E).

La **vraisemblance** est $P(B|A_2, E)$. En se limitant aux 4 parcelles marécageuses de la région A2, une seule est aussi près d'un chêne. Donc, $P(B|A_2, E) = 1/4$.

La **probabilité a priori** est $P(A_2|E)$. Sur les 10 parcelles marécageuses, 4 sont dans la région A2. Donc, $P(A_2|E) = 4/10$.

L'**évidence**, $P(B|E)$, est la probabilité de trouver un chêne dans l'ensemble de la zone marécageuse. On peut la calculer avec la formule des probabilités totales :

$$P(B|E) = P(B|A_1, E)P(A_1|E) + P(B|A_2, E)P(A_2|E) + P(B|A_3, E)P(A_3|E)$$

$$P(B|E) = \left(\frac{1}{3} \times \frac{3}{10}\right) + \left(\frac{1}{4} \times \frac{4}{10}\right) + \left(0 \times \frac{3}{10}\right) = \frac{1}{10} + \frac{1}{10} = \frac{2}{10}$$

Application de la règle de Bayes : Maintenant, nous assemblons le tout.

$$P(A_2|B, E) = \frac{P(B|A_2, E)P(A_2|E)}{P(B|E)} = \frac{(1/4) \times (4/10)}{2/10} = \frac{1/10}{2/10} = \frac{1}{2}$$

L'intuition confirme le calcul : sachant que le trésor est sur une parcelle violette, et qu'il n'y en a que deux (une en A1, une en A2), il y a bien une chance sur deux qu'il se trouve dans la région A2.

3.8 Indépendance de Deux Événements

Définition : Indépendance de deux événements

Les événements A et B sont indépendants si :

$$P(A \cap B) = P(A)P(B)$$

Si $P(A) > 0$ et $P(B) > 0$, cela est équivalent à :

$$P(A|B) = P(A)$$

Intuition

L'indépendance est l'absence d'information. Si deux événements sont indépendants, apprendre que l'un s'est produit ne change absolument rien à la probabilité de l'autre. Savoir qu'il pleut à Tokyo (B) ne modifie pas la probabilité que vous obteniez pile en lançant une pièce (A).

3.9 Indépendance Conditionnelle

Définition : Indépendance Conditionnelle

Les événements A et B sont dits conditionnellement indépendants étant donné E si :

$$P(A \cap B|E) = P(A|E)P(B|E)$$

Intuition

L'indépendance peut apparaître ou disparaître quand on observe un autre événement. Par exemple, vos notes en maths (A) et en physique (B) ne sont probablement pas indépendantes. Mais si l'on sait que vous avez beaucoup travaillé (E), alors vos notes en maths et en physique pourraient devenir indépendantes. L'information "vous avez beaucoup travaillé" explique la corrélation ; une fois qu'on la connaît, connaître votre note en maths n'apporte plus d'information sur votre note en physique.

3.10 Le Problème de Monty Hall

Remarque : Le problème de Monty Hall

Imaginez que vous êtes à un jeu télévisé. Face à vous se trouvent trois portes fermées. Derrière l'une d'elles se trouve une voiture, et derrière les deux autres, des chèvres.

1. Vous choisissez une porte (disons, la porte n°1).
2. L'animateur, qui sait où se trouve la voiture, ouvre une autre porte (par exemple, la n°3) derrière laquelle se trouve une chèvre.
3. Il vous demande alors : "Voulez-vous conserver votre choix initial (porte n°1) ou changer pour l'autre porte restante (la n°2) ?"

Question : Avez-vous intérêt à changer de porte ? Votre probabilité de gagner la voiture est-elle plus grande si vous changez, si vous ne changez pas, ou est-elle la même dans les deux cas ?

Correction : Solution du problème de Monty Hall

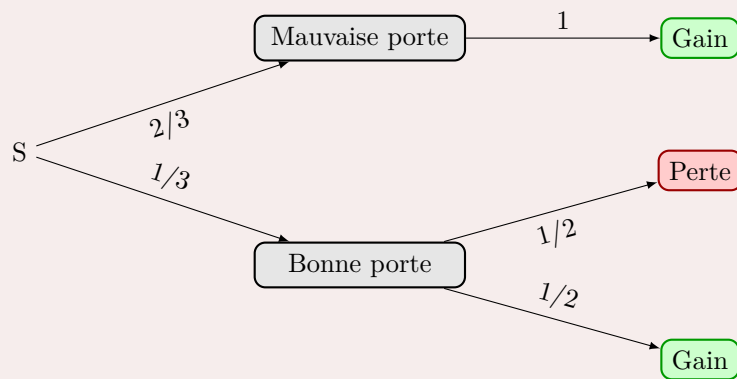
La réponse est sans équivoque : il faut **toujours changer de porte**. Cette stratégie fait passer la probabilité de gagner de $1/3$ à $2/3$. L'intuition et la preuve ci-dessous détaillent ce résultat surprenant.

Intuition : Le secret : l'information de l'animateur

L'erreur commune est de supposer qu'il reste deux portes avec une chance égale de $1/2$. Cela ignore une information capitale : le choix de l'animateur n'est **pas aléatoire**. Il sait où se trouve la voiture et ouvrira toujours une porte perdante. Le raisonnement correct se déroule en deux temps. D'abord, votre choix initial a $1/3$ de chance d'être correct. Cela implique qu'il y a $2/3$ de chance que la voiture soit derrière l'une des *deux autres portes*. Ensuite, lorsque l'animateur ouvre l'une de ces deux portes, il ne fait que vous montrer où la voiture n'est *pas* dans cet ensemble. La probabilité de $2/3$ se **concentre** alors entièrement sur la seule porte qu'il a laissée fermée. Changer de porte revient à miser sur cette probabilité de $2/3$.

Preuve : Preuve par l'arbre de décision

L'analyse de la meilleure stratégie peut être visualisée à l'aide de l'arbre de décision ci-dessous. Il décompose le problème en deux scénarios initiaux : avoir choisi la bonne porte (probabilité $1/3$) ou une mauvaise porte (probabilité $2/3$).



Analyse de l'arbre :

Branche du bas (cas le plus probable) :

Avec une probabilité de $\frac{2}{3}$, votre choix initial se porte sur une "Mauvaise porte". L'animateur est alors obligé de révéler l'autre porte perdante. La seule porte restante est donc la bonne. L'arbre montre que cela mène à un "Gain" avec une probabilité de 1. Ce chemin correspond au résultat de la stratégie "**Changer**".

Branche du haut (cas le moins probable) :

Avec une probabilité de $\frac{1}{3}$, vous avez choisi la "Bonne porte" du premier coup. L'arbre se divise alors en deux issues équiprobables ($\frac{1}{2}$ chacune). L'issue "Gain" correspond à la stratégie "**Garder**" votre choix initial, tandis que l'issue "Perte" correspond à la stratégie "**Changer**" pour la porte perdante restante.

Conclusion :

Pour évaluer la meilleure stratégie, il suffit de sommer les probabilités de gain. La **probabilité de gain en changeant** est de $\frac{2}{3}$, car vous gagnez uniquement si votre choix initial était mauvais (branche du bas). La **probabilité de gain en gardant** est de $\frac{1}{3}$, car vous gagnez uniquement si votre choix initial était bon (branche "Gain" du haut). La stratégie optimale est donc bien de toujours changer de porte.

4 Variables Aléatoires Discrètes

4.1 Variable Aléatoire

Définition : Variable Aléatoire

Étant donné une expérience avec un univers S , une variable aléatoire est une fonction de l'univers S vers les nombres réels \mathbb{R} .

Intuition

Une variable aléatoire est une manière de traduire les résultats d'une expérience en nombres. Au lieu de travailler avec des concepts comme "Pile" ou "Face", on leur assigne des valeurs numériques (par exemple, 1 pour Pile, 0 pour Face). Cela nous permet d'utiliser toute la puissance des outils mathématiques (fonctions, calculs, etc.) pour analyser le hasard. C'est un pont entre le monde concret des événements et le monde abstrait des nombres.

Exemple

On lance deux dés. L'univers S est l'ensemble des 36 paires de résultats, comme $(1, 1), (1, 2), \dots, (6, 6)$. On peut définir une variable aléatoire X comme étant la **somme des deux dés**. Pour le résultat $(2, 5)$, la valeur de la variable aléatoire est $X(2, 5) = 2 + 5 = 7$.

4.2 Variable Aléatoire Discrète

Définition : Variable Aléatoire Discrète

Une variable aléatoire X est dite discrète s'il existe une liste finie ou infinie dénombrable de valeurs a_1, a_2, \dots telle que $P(X = a_j \text{ pour un certain } j) = 1$.

Intuition

Une variable aléatoire est "discrète" si on peut lister (compter) toutes les valeurs qu'elle peut prendre, même si cette liste est infinie. Pensez aux "sauts" d'une valeur à l'autre, sans possibilité de prendre une valeur intermédiaire. C'est comme monter un escalier : on peut être sur la marche 1, 2 ou 3, mais jamais sur la marche 2.5. Le nombre de têtes en 10 lancers, le résultat d'un dé, le nombre d'emails que vous recevez en une heure sont des exemples. À l'opposé, une variable continue pourrait

être la taille exacte d'une personne, qui peut prendre n'importe quelle valeur dans un intervalle.

4.3 Fonction de Masse (PMF)

Définition : Probability Mass Function (PMF)

La fonction de masse (PMF) d'une variable aléatoire discrète X est la fonction P_X donnée par $P_X(x) = P(X = x)$.

Intuition

La PMF est la "carte d'identité" probabiliste d'une variable aléatoire discrète. Pour chaque valeur que la variable peut prendre, la PMF nous donne la probabilité exacte associée à cette valeur. C'est comme si chaque résultat possible avait une "étiquette de prix" qui indique sa chance de se produire. La somme de toutes ces probabilités doit bien sûr valoir 1.

Exemple

Soit X le résultat d'un lancer de dé équilibré. La variable X peut prendre les valeurs $\{1, 2, 3, 4, 5, 6\}$. La PMF de X est la fonction qui assigne $1/6$ à chaque valeur : $P(X = 1) = 1/6$, $P(X = 2) = 1/6$, ..., $P(X = 6) = 1/6$. Pour toute autre valeur x (par exemple $x = 2.5$ ou $x = 7$), $P(X = x) = 0$.

4.4 Loi de Bernoulli

Définition : Distribution de Bernoulli

Une variable aléatoire X suit la distribution de Bernoulli avec paramètre p si $P(X = 1) = p$ et $P(X = 0) = 1 - p$, où $0 < p < 1$. On note cela $X \sim \text{Bern}(p)$.

Intuition

La distribution de Bernoulli est le modèle le plus simple pour une expérience aléatoire avec seulement deux issues : "succès" (codé par 1) et "échec" (codé par 0). C'est la brique de base de nombreuses autres distributions. Pensez à un unique lancer de pièce

(Pile/Face), un unique tir au but (Marqué/Manqué), ou la réponse à une question par oui/non. Le paramètre p est simplement la probabilité du "succès".

4.5 Loi Binomiale

Théorème : PMF Binomiale

Si $X \sim \text{Bin}(n, p)$, alors la PMF de X est :

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

pour $k = 0, 1, \dots, n$.

Intuition

La distribution binomiale répond à la question : "Si je répète n fois la même expérience de Bernoulli (qui a une probabilité de succès p), quelle est la probabilité d'obtenir exactement k succès ?" La formule est construite logiquement en multipliant trois composantes. D'abord, p^k représente la probabilité d'obtenir k succès. Ensuite, $(1 - p)^{n-k}$ est la probabilité que les $n - k$ échecs restants se produisent. Finalement, comme les k succès peuvent apparaître n'importe où parmi les n essais, on multiplie par $\binom{n}{k}$, qui compte le nombre de manières distinctes de placer ces succès.

Exemple

On lance une pièce équilibrée 10 fois ($n = 10$, $p = 0.5$). Quelle est la probabilité d'obtenir exactement 6 Piles ($k = 6$) ?

$$P(X = 6) = \binom{10}{6} (0.5)^6 (1 - 0.5)^{10-6} = \frac{10!}{6!4!} (0.5)^{10} = 210 \times (0.5)^{10} \approx 0.205$$

Il y a environ 20.5% de chance d'obtenir exactement 6 Piles.

4.6 Loi Hypergéométrique

Théorème : PMF Hypergéométrique

Si $X \sim \text{HG}(w, b, m)$, alors la PMF de X est :

$$P(X = k) = \frac{\binom{w}{k} \binom{b}{m-k}}{\binom{w+b}{m}}$$

Intuition

La distribution hypergéométrique est la "cousine" de la binomiale pour les tirages **sans remise**. Imaginez une urne avec des boules de deux couleurs (par exemple, w blanches et b noires). Vous tirez m boules d'un coup. Quelle est la probabilité que vous ayez exactement k boules blanches ? La formule est un simple ratio issu du dénombrement. Le **dénominateur**, $\binom{w+b}{m}$, compte le nombre total de façons de tirer m boules parmi toutes celles disponibles. Le **numérateur** compte les issues favorables : c'est le produit du nombre de façons de choisir k blanches parmi les w ($\binom{w}{k}$) ET de choisir les $m-k$ boules restantes parmi les noires ($\binom{b}{m-k}$). La différence clé avec la loi binomiale est que les tirages ne sont pas indépendants.

Exemple

Un comité de 5 personnes est choisi au hasard parmi un groupe de 8 hommes et 10 femmes. Quelle est la probabilité que le comité soit composé de 2 hommes et 3 femmes ? Ici, on tire 5 personnes ($m = 5$) d'une population de 18 personnes. On s'intéresse au nombre d'hommes ($k = 2$) parmi les 8 disponibles ($w = 8$). Le reste du comité sera composé de femmes ($b = 10$).

$$P(X = 2) = \frac{\binom{8}{2} \binom{10}{3}}{\binom{18}{5}} = \frac{28 \times 120}{8568} \approx 0.392$$

Il y a environ 39.2% de chance que le comité ait exactement cette composition.

4.7 Loi Géométrique

Théorème : PMF de la loi géométrique

Une variable aléatoire X suit la loi géométrique de paramètre p , notée $X \sim \text{Geom}(p)$, si elle modélise le nombre d'échecs avant le premier succès dans une série d'épreuves de Bernoulli indépendantes. Sa fonction de masse (PMF) est :

$$P(X = k) = (1 - p)^k p \quad \text{pour } k = 0, 1, 2, \dots$$

où $q = 1 - p$ est la probabilité d'échec.

Intuition

La formule $P(X = k) = q^k p$ décrit la probabilité d'une séquence très spécifique : k échecs consécutifs (chacun avec une probabilité q , donc q^k pour la série), suivis immédiatement d'un succès (avec une probabilité p). C'est la loi de "l'attente du premier succès".

Exemple : Premier 6 au lancer de dé

On lance un dé jusqu'à obtenir un 6. La probabilité de succès est $p = 1/6$, et celle d'échec est $q = 5/6$. Quelle est la probabilité que l'on ait besoin de 3 lancers (donc 2 échecs avant le premier succès) ? Ici, $k = 2$. La probabilité est :

$$P(X = 2) = (5/6)^2 \cdot (1/6) = \frac{25}{216} \approx 0.116$$

4.8 Loi de Poisson

Définition : Distribution de Poisson

Une variable aléatoire X suit la loi de Poisson de paramètre $\lambda > 0$ si sa PMF est donnée par :

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad \text{pour } k = 0, 1, 2, \dots$$

Elle modélise typiquement le nombre d'événements se produisant dans un intervalle de temps ou d'espace fixe.

Intuition

La loi de Poisson est la loi des événements rares. Imaginez que vous comptez le nombre d'appels arrivant à un standard téléphonique en une minute. Il y a de nombreux instants où un appel pourrait arriver, mais la probabilité à chaque instant est infime. La loi de Poisson modélise ce type de scénario, où l'on connaît seulement le taux moyen d'arrivée des événements (λ).

Théorème : La loi de Poisson comme limite de la loi binomiale

Soit $X_n \sim \text{Bin}(n, p_n)$, où $\lambda = np_n$ est une constante positive fixée. Alors, pour tout $k \in \{0, 1, 2, \dots\}$, nous avons :

$$\lim_{n \rightarrow \infty} P(X_n = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

En pratique, la loi de Poisson est une excellente approximation de la loi binomiale quand n est grand et p est petit.

Preuve : Dérivation de la loi de Poisson à partir de la loi Binomiale

Supposons que les bébés naissent à un taux moyen de λ naissances par jour. On peut modéliser ce processus en divisant la journée en n très petits sous-intervalles de temps.

1. Modèle Binomial :

Si n est suffisamment grand, on peut considérer que durant chaque petit sous-intervalle, il y a au plus une naissance. La probabilité d'une naissance durant un de ces sous-intervalles est donc $p = \lambda/n$.

Chaque sous-intervalle peut être vu comme une épreuve de Bernoulli indépendante (soit une naissance, soit pas de naissance). Le nombre total de naissances en une journée, X , est la somme de ces épreuves. Il suit donc une loi binomiale :

$$X \sim \text{Bin}\left(n, p = \frac{\lambda}{n}\right)$$

La fonction de masse de probabilité (PMF) pour obtenir exactement k naissances est :

$$P(X = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

2. Passage à la limite :

Pour modéliser un processus continu, on fait tendre le nombre de sous-intervalles n vers l'infini. On examine alors la limite de la PMF binomiale :

$$\lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

En utilisant les limites connues $\lim_{n \rightarrow \infty} \frac{n(n-1) \cdots (n-k+1)}{n^k} = 1$ et $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$, on peut montrer que cette expression converge vers :

$$\frac{e^{-\lambda} \lambda^k}{k!}$$

Conclusion :

Lorsque le nombre d'essais (n) devient très grand et la probabilité de succès (p) très petite, la loi binomiale converge vers la loi de Poisson. Le nombre de naissances en un jour suit donc une **loi de Poisson** de paramètre λ , où λ est le taux moyen de naissances par jour.

Exemple : Décès par ruade de cheval : Les données de Bortkiewicz

En 1898, le statisticien Ladislaus Bortkiewicz a publié des données célèbres sur le nombre de soldats de la cavalerie prussienne tués par des ruades de cheval. Ces données sont un exemple classique d'application de la loi de Poisson pour modéliser des événements rares.

Contexte et calcul du paramètre λ :

Sur une période de 20 ans, en observant 10 corps d'armée, il a collecté des données sur 200 "corps-années". Durant cette période, il y a eu un total de 122 décès. Le taux moyen de décès par corps-année est donc :

$$\lambda = \frac{\text{Nombre total de décès}}{\text{Nombre total de corps-années}} = \frac{122}{200} = 0.61$$

Le nombre de décès par corps-année, X , est donc modélisé par une loi de Poisson : $X \sim \text{Poisson}(\lambda = 0.61)$.

Comparaison des données observées et des prédictions du modèle :

On peut calculer la probabilité d'observer k décès en une année-corps en utilisant la PMF de Poisson : $P(X = k) = \frac{e^{-0.61}(0.61)^k}{k!}$. En multipliant cette probabilité par le nombre total d'observations (200), on obtient le nombre de cas attendus (nombre de corps d'armes dans lesquels il y a k décès).

Nombre de décès (k)	Observé	Probabilité de Poisson	Attendu
0	109	$P(X = 0) \approx 0.543$	108.7
1	65	$P(X = 1) \approx 0.331$	66.3
2	22	$P(X = 2) \approx 0.101$	20.2
3	3	$P(X = 3) \approx 0.021$	4.1
4	1	$P(X = 4) \approx 0.003$	0.6
5+	0	$P(X \geq 5) \approx 0.000$	0.0

L'adéquation remarquable entre les fréquences observées et les valeurs attendues par le modèle de Poisson a contribué à populariser cette distribution pour l'analyse d'événements rares.

4.9 Fonction de Répartition (CDF)

Définition : Cumulative Distribution Function (CDF)

La fonction de répartition (CDF) d'une variable aléatoire X est la fonction F_X donnée par $F_X(x) = P(X \leq x)$.

Intuition

Alors que la PMF répond à la question "Quelle est la probabilité d'obtenir *exactement* x ?", la CDF répond à la question "Quelle est la probabilité d'obtenir *au plus* x ?". C'est

une fonction cumulative : pour une valeur x donnée, elle additionne les probabilités de tous les résultats inférieurs ou égaux à x . La CDF a toujours une forme d'escalier pour les variables discrètes. Elle commence à 0 (très loin à gauche) et monte par "sauts" à chaque valeur possible de la variable, pour finalement atteindre 1 (très loin à droite). La hauteur de chaque saut correspond à la valeur de la PMF à ce point.

Exemple

Reprenons le lancer d'un dé équilibré (X). Calculons quelques valeurs de la CDF, notée $F(x)$.

$$F(0.5) = P(X \leq 0.5) = 0$$

$$F(1) = P(X \leq 1) = P(X = 1) = 1/6$$

$$F(1.5) = P(X \leq 1.5) = P(X = 1) = 1/6$$

$$F(2) = P(X \leq 2) = P(X = 1) + P(X = 2) = 2/6$$

$$F(5.9) = P(X \leq 5.9) = P(X = 1) + \dots + P(X = 5) = 5/6$$

$$F(6) = P(X \leq 6) = 1$$

$$F(100) = P(X \leq 100) = 1$$

4.10 Variable Aléatoire Indicatrice

Définition : Variable Aléatoire Indicatrice

La variable aléatoire indicatrice d'un événement A est la variable aléatoire qui vaut 1 si A se produit et 0 sinon. Nous la noterons I_A . Notez que $I_A \sim \text{Bern}(p)$ avec $p = P(A)$.

Intuition

Une variable indicatrice est un interrupteur. Elle est sur "ON" (valeur 1) si un événement qui nous intéresse se produit, et sur "OFF" (valeur 0) sinon. C'est un outil extrêmement puissant car il transforme les questions sur les probabilités des événements en questions sur les espérances des variables aléatoires, ce qui simplifie souvent les calculs.

5 Espérance et autres notions associées aux variables aléatoires discrètes

5.1 Espérance d'une variable aléatoire discrète

Définition : Espérance

L'espérance (ou valeur attendue) d'une variable aléatoire discrète X , qui prend les valeurs distinctes x_1, x_2, \dots , est définie par :

$$E(X) = \sum_j x_j P(X = x_j)$$

Intuition

L'espérance représente la valeur moyenne que l'on obtiendrait si l'on répétait l'expérience un très grand nombre de fois. C'est le **centre de gravité** de la distribution de probabilité. Si les probabilités étaient des masses placées sur une tige aux positions x_j , l'espérance serait le point d'équilibre.

Exemple : Lancer d'un dé

Soit X le résultat d'un lancer de dé équilibré. Chaque face a une probabilité de $1/6$. L'espérance est :

$$E(X) = 1 \left(\frac{1}{6} \right) + 2 \left(\frac{1}{6} \right) + 3 \left(\frac{1}{6} \right) + 4 \left(\frac{1}{6} \right) + 5 \left(\frac{1}{6} \right) + 6 \left(\frac{1}{6} \right) = \frac{21}{6} = 3.5$$

Même si 3.5 n'est pas un résultat possible, c'est la valeur moyenne sur un grand nombre de lancers.

5.2 Linéarité de l'espérance

Théorème : Linéarité de l'espérance

Pour toutes variables aléatoires X et Y , et pour toute constante c , on a :

$$\begin{aligned} E(X + Y) &= E(X) + E(Y) \\ E(cX) &= cE(X) \end{aligned}$$

Cette propriété est extrêmement puissante car elle ne requiert pas que X et Y soient indépendantes.

Intuition

Cette propriété formalise une idée très simple : "la moyenne d'une somme est la somme des moyennes". Si vous jouez à deux jeux de hasard, votre gain moyen total est simplement la somme de ce que vous gagnez en moyenne à chaque jeu, que les jeux soient liés ou non.

Exemple : Somme de deux dés

Soit X_1 le résultat du premier dé et X_2 celui du second. On sait que $E(X_1) = 3.5$ et $E(X_2) = 3.5$. Soit $S = X_1 + X_2$ la somme des deux dés. Grâce à la linéarité, on peut calculer l'espérance de la somme sans avoir à lister les 36 résultats possibles :

$$E(S) = E(X_1 + X_2) = E(X_1) + E(X_2) = 3.5 + 3.5 = 7$$

5.3 Espérance de la loi binomiale

Théorème : Espérance de la loi binomiale

Si $X \sim \text{Bin}(n, p)$, alors son espérance est $E(X) = np$.

Intuition

Ce résultat est très naturel. Si vous lancez une pièce 100 fois ($n = 100$) avec une probabilité de 50% d'obtenir Pile ($p = 0.5$), vous vous attendez en moyenne à obtenir $100 \times 0.5 = 50$ Piles. La formule np généralise cette idée.

Preuve

Le calcul direct de l'espérance avec la PMF binomiale est possible, mais long. En utilisant la linéarité de l'espérance, on obtient une preuve beaucoup plus courte et élégante.

On peut voir une variable binomiale X comme la somme de n variables de Bernoulli indépendantes, $X = I_1 + I_2 + \dots + I_n$, où chaque I_j représente le succès (1) ou l'échec (0) du j -ième essai.

Chaque I_j a pour espérance $E(I_j) = 1 \cdot p + 0 \cdot (1 - p) = p$.

Par linéarité de l'espérance, on a :

$$E(X) = E(I_1) + E(I_2) + \dots + E(I_n) = \underbrace{p + p + \dots + p}_{n \text{ fois}} = np$$

5.4 Espérance de la loi géométrique

Théorème : Espérance de la loi géométrique

L'espérance d'une variable aléatoire $X \sim \text{Geom}(p)$ (comptant le nombre d'échecs) est :

$$E(X) = \frac{1-p}{p} = \frac{q}{p}$$

Intuition

Si un événement a 1 chance sur 10 de se produire ($p = 0.1$), il est logique de penser qu'il faudra en moyenne 9 échecs ($q/p = 0.9/0.1 = 9$) avant qu'il ne se produise. L'espérance du nombre total d'essais (échecs + 1 succès) serait alors $1/p$.

Preuve : Démonstration de l'espérance géométrique via les séries entières

Soit $X \sim \text{Geom}(p)$, où X compte le nombre d'échecs avant le premier succès. La PMF est $P(X = k) = q^k p$ pour $k = 0, 1, 2, \dots$, avec $q = 1 - p$. Par définition, l'espérance est :

$$E(X) = \sum_{k=0}^{\infty} k \cdot P(X = k) = \sum_{k=0}^{\infty} k q^k p$$

Le terme pour $k = 0$ est nul, on peut donc commencer la somme à $k = 1$:

$$E(X) = p \sum_{k=1}^{\infty} k q^k$$

L'astuce consiste à reconnaître que la somme ressemble à la dérivée d'une série géométrique. Rappelons la formule de la série géométrique pour $|q| < 1$:

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$$

En dérivant les deux côtés par rapport à q , on obtient :

$$\frac{d}{dq} \left(\sum_{k=0}^{\infty} q^k \right) = \frac{d}{dq} \left(\frac{1}{1-q} \right)$$

$$\sum_{k=1}^{\infty} k q^{k-1} = \frac{1}{(1-q)^2}$$

Pour faire apparaître ce terme dans notre formule d'espérance, on factorise q dans la somme :

$$E(X) = p \cdot q \sum_{k=1}^{\infty} k q^{k-1}$$

On peut maintenant remplacer la somme par son expression analytique :

$$E(X) = p \cdot q \cdot \frac{1}{(1-q)^2}$$

Puisque $p = 1 - q$, on a :

$$E(X) = p \cdot q \cdot \frac{1}{p^2} = \frac{q}{p}$$

Ce qui démontre que l'espérance du nombre d'échecs avant le premier succès est $\frac{q}{p}$.

5.5 Loi du statisticien inconscient (LOTUS)

Théorème : Théorème de Transfert (LOTUS)

Si X est une variable aléatoire discrète et $g(x)$ est une fonction de \mathbb{R} dans \mathbb{R} , alors l'espérance de la variable aléatoire $g(X)$ est donnée par :

$$E[g(X)] = \sum_x g(x)P(X = x)$$

La somme porte sur toutes les valeurs possibles de X . Ce théorème est utile car il évite d'avoir à trouver la PMF de $g(X)$.

Intuition

Pour trouver la valeur moyenne d'une fonction d'une variable aléatoire (par exemple, le carré du résultat d'un dé), vous n'avez pas besoin de déterminer d'abord la distribution de ce carré. Vous pouvez simplement prendre chaque valeur possible du résultat original, lui appliquer la fonction, et pondérer ce nouveau résultat par la probabilité du résultat original.

Exemple : Calcul de $E(X^2)$ pour un dé

Soit X le résultat d'un lancer de dé. Calculons l'espérance de $Y = X^2$. La fonction est $g(x) = x^2$.

$$\begin{aligned} E(X^2) &= \sum_{k=1}^6 k^2 P(X = k) \\ &= 1^2 \left(\frac{1}{6}\right) + 2^2 \left(\frac{1}{6}\right) + 3^2 \left(\frac{1}{6}\right) + 4^2 \left(\frac{1}{6}\right) + 5^2 \left(\frac{1}{6}\right) + 6^2 \left(\frac{1}{6}\right) \\ &= \frac{1 + 4 + 9 + 16 + 25 + 36}{6} = \frac{91}{6} \approx 15.17 \end{aligned}$$

5.6 Variance

Définition : Variance et écart-type

La **variance** d'une variable aléatoire X mesure la dispersion de sa distribution autour de son espérance. Elle est définie par :

$$\text{Var}(X) = E[(X - E(X))^2]$$

La racine carrée de la variance est appelée l' **écart-type** :

$$\text{SD}(X) = \sqrt{\text{Var}(X)}$$

Intuition

La variance est la "distance carrée moyenne à la moyenne". On prend l'écart de chaque valeur par rapport à la moyenne, on le met au carré (pour que les écarts positifs et négatifs ne s'annulent pas), puis on en calcule la moyenne. L'écart-type est souvent plus interprétable car il ramène cette mesure de dispersion dans les mêmes unités que la variable aléatoire elle-même.

Théorème : Formule de calcul de la variance

Pour toute variable aléatoire X , une formule plus pratique pour le calcul de la variance est :

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

Exemple : Variance d'un lancer de dé

Nous avons déjà calculé pour un dé que $E(X) = 3.5$ et $E(X^2) = 91/6$. On peut maintenant trouver la variance facilement :

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{91}{6} - (3.5)^2 = \frac{91}{6} - 12.25 = 15.166... - 12.25 \approx 2.917$$

L'écart-type est $\text{SD}(X) = \sqrt{2.917} \approx 1.708$.

Voici la section réécrite avec un ordre plus logique.

J'ai regroupé les concepts de base (distributions et espérance), puis tous les outils liés à la covariance et à la corrélation (définitions, propriétés, standardisation), et enfin les applications de ces concepts (variance d'une somme) et les théorèmes spécifiques.

6 Distributions Multivariées et Concepts Associés

6.1 Distributions Jointes et Marginales

Définition : Distribution Jointe (Cas Discret)

Pour deux variables aléatoires discrètes X et Y , la **distribution jointe** (ou loi jointe) spécifie la probabilité de chaque paire d'issues. La fonction de masse de probabilité jointe (joint PMF) est :

$$P(X = x, Y = y)$$

Si X prend ses valeurs dans un ensemble S et Y dans un ensemble T , alors la somme de toutes les probabilités jointes est égale à 1 :

$$\sum_{x \in S} \sum_{y \in T} P(X = x, Y = y) = 1$$

Intuition

La distribution jointe est la "carte" complète de toutes les issues possibles. Elle répond à la question : "Quelle est la probabilité que X prenne cette valeur ET que Y prenne cette autre valeur en même temps?". Si vous imaginez un tableau à double entrée pour X et Y , la loi jointe est l'ensemble de toutes les probabilités à l'intérieur du tableau.

Définition : Distribution Marginale

À partir de la distribution jointe, on peut obtenir la distribution **marginale** (ou loi marginale) de chaque variable. Pour obtenir la probabilité que X prenne une valeur x , on somme sur toutes les valeurs possibles de Y :

$$P(X = x) = \sum_{y \in T} P(X = x, Y = y)$$

Intuition

Les distributions marginales sont les "ombres" ou "projections" de la carte jointe sur un seul axe. Si la loi jointe est un tableau, les lois marginales sont les totaux de chaque ligne et de chaque colonne, que l'on écrirait "dans la marge" du tableau. Elles nous disent la probabilité d'une issue pour X sans se soucier de ce qu'il advient de Y .

Exemple : Loïs jointe et marginale

On lance un dé rouge (X) et un dé bleu (Y). Il y a 36 issues, chacune avec une probabilité de $1/36$. **Loi jointe** : $P(X = x, Y = y) = 1/36$ pour tout $x, y \in \{1, \dots, 6\}$. Par exemple, $P(X = 2, Y = 5) = 1/36$.

Loi marginale de X : Cherchons $P(X = 2)$. C'est la probabilité d'obtenir 2 sur le dé rouge, quel que soit le résultat du bleu.

$$P(X = 2) = \sum_{y=1}^6 P(X = 2, Y = y)$$

$$P(X = 2) = P(X = 2, Y = 1) + \dots + P(X = 2, Y = 6)$$

$$P(X = 2) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{6}{36} = \frac{1}{6}$$

Ceci est bien la loi d'un seul dé.

6.2 Espérance d'une fonction de deux variables

Définition : Espérance d'une fonction $g(X, Y)$

L'espérance d'une fonction $g(X, Y)$ de deux variables aléatoires discrètes X et Y est une généralisation du théorème de transfert (LOTUS) :

$$E[g(X, Y)] = \sum_{x \in S} \sum_{y \in T} g(x, y) P(X = x, Y = y)$$

Intuition

C'est la valeur moyenne attendue de la fonction g . Pour la calculer, on prend chaque résultat possible de $g(x, y)$, on le pondère par la probabilité que cette combinaison (x, y) se produise (donnée par la loi jointe), et on somme le tout.

Exemple

Espérance de $E[X + Y]$ Avec nos deux dés, calculons l'espérance de la somme $S = X + Y$. La fonction est $g(X, Y) = X + Y$.

$$E[X + Y] = \sum_{x=1}^6 \sum_{y=1}^6 (x + y) P(X = x, Y = y)$$

$$E[X + Y] = \sum_{x=1}^6 \sum_{y=1}^6 (x + y) \frac{1}{36}$$

Plutôt que de faire ce long calcul, on peut utiliser la linéarité de l'espérance (qui est un cas particulier de ce théorème) :

$$E[X + Y] = E[X] + E[Y] = 3.5 + 3.5 = 7.$$

6.3 Covariance et Corrélation

Définition : Covariance

La **covariance** entre deux variables aléatoires X et Y , avec pour moyennes respectives μ_X et μ_Y , mesure la façon dont elles varient ensemble.

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

Intuition

La covariance est positive si les variables ont tendance à "bouger" dans la même direction (quand X est au-dessus de sa moyenne, Y a tendance à l'être aussi). Elle est négative si elles bougent en sens opposé (quand X est au-dessus de sa moyenne, Y a tendance à être en dessous). Si elle est nulle, il n'y a pas de tendance linéaire entre elles.

Théorème : Formule de calcul de la covariance

Une formule computationnelle plus simple pour la covariance est :

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

Exemple : Calcul de covariance

Cas 1 : Dés indépendants. X et Y sont les résultats de deux dés. $E[X] = 3.5, E[Y] = 3.5$. Calculons $E[XY]$. Puisqu'ils sont indépendants, $E[XY] = E[X]E[Y] = 3.5 \times 3.5 = 12.25$. $\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 12.25 - 12.25 = 0$. La covariance est nulle, ce qui est attendu pour des variables indépendantes.

Cas 2 : Variables dépendantes. Soit X un lancer de dé, et $Y = 2X$. $E[X] = 3.5$. $E[Y] = E[2X] = 2E[X] = 7$. $E[XY] = E[X \cdot 2X] = E[2X^2] = 2E[X^2]$. On sait que $E[X^2] = \frac{1^2 + \dots + 6^2}{6} = 91/6$. $E[XY] = 2(91/6) = 91/3$. $\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = \frac{91}{3} - (3.5)(7) = \frac{91}{3} - 24.5 = 30.33\ldots - 24.5 \approx 5.833$. La covariance est positive, ce qui est logique : si X est grand, Y l'est aussi.

Définition : Corrélation

La **corrélation** (ou coefficient de corrélation de Pearson, r) est une version normalisée de la covariance, qui se situe toujours entre -1 et 1.

$$\text{Corr}(X, Y) = r = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Intuition

Le problème de la covariance est qu'elle dépend des unités de X et Y (par ex., kg·cm). Si vous changez les unités (grammes et mètres), la valeur de la covariance change, même si la relation est identique. La corrélation résout ce problème : elle est sans unité. Un coefficient de +1 indique une relation linéaire positive parfaite, -1 une relation linéaire négative parfaite, et 0 une absence de relation linéaire.

Intuition : Interprétation de la formule

On peut voir la corrélation de Pearson comme un processus en 3 étapes :

1. **Centrer les variables** : On calcule l'écart de chaque valeur à sa moyenne ($x_i - \bar{x}$ et $y_i - \bar{y}$). Cela élimine "l'effet de base" (ex : une personne de 180cm vs 170cm ; la moyenne change mais les écarts relatifs restent les mêmes).
2. **Normaliser les variables** : On divise chaque écart par l'écart-type de sa variable ($z_{xi} = (x_i - \bar{x})/\sigma_X$ et $z_{yi} = (y_i - \bar{y})/\sigma_Y$). Ces nouvelles variables Z_X et Z_Y sont **standardisées** : elles ont une moyenne de 0, un écart-type de 1, et sont sans unité.
3. **Calculer la covariance des variables standardisées** : La corrélation n'est rien d'autre que la covariance de ces deux nouvelles variables standardisées : $r = \text{Cov}(Z_X, Z_Y)$.

Parce que les deux variables sont maintenant sur la même échelle (écart-type de 1), leur covariance (la corrélation) ne peut pas dépasser 1 en valeur absolue.

Exemple : Calcul de corrélation

Reprenons l'exemple $Y = 2X$, où X est un lancer de dé. On a $\text{Cov}(X, Y) = 5.833... = 35/6$. $\text{Var}(X) = E[X^2] - E[X]^2 = 91/6 - (3.5)^2 = 35/12$. $\text{Var}(Y) = \text{Var}(2X) = 2^2 \text{Var}(X) = 4(35/12) = 35/3$. $\sigma_X \sigma_Y = \sqrt{35/12} \cdot \sqrt{35/3} = \sqrt{(35 \cdot 35)/(12 \cdot 3)} = \sqrt{35^2/36} = 35/6$.

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{35/6}{35/6} = 1$$

La corrélation est de 1, ce qui est parfait : Y est une fonction linéaire parfaite de X .

6.4 Linéarité de la Covariance

Définition : Linéarité de la Covariance

Pour des variables aléatoires X, Y, Z et des constantes a, b, c :

$$\text{Cov}(aX + bY + c, Z) = a\text{Cov}(X, Z) + b\text{Cov}(Y, Z)$$

$$\text{Cov}(X, aY + bZ + c) = a\text{Cov}(X, Y) + b\text{Cov}(X, Z)$$

La covariance est linéaire pour chaque argument (elle est **bilinéaire**). Les constantes additives disparaissent.

6.5 Résultats sur la Corrélation

Théorème : Bornes du Coefficient de Corrélation de Pearson

Pour toutes variables aléatoires X et Y , le coefficient de corrélation $\text{Corr}(X, Y)$ est borné :

$$-1 \leq \text{Corr}(X, Y) \leq 1$$

De plus, si $\text{Corr}(X, Y) = \pm 1$, alors il existe des constantes a et b telles que $Y = aX + b$, indiquant une relation linéaire parfaite.

Preuve : Démonstration des Bornes de la Corrélation

La preuve repose sur le fait que la variance d'une variable aléatoire est toujours positive ou nulle.

Étape 1 : Variables Standardisées

On définit les versions standardisées de X et Y :

$$X^* = \frac{X - \mu_X}{\sigma_X} \quad ; \quad Y^* = \frac{Y - \mu_Y}{\sigma_Y}$$

Par construction, $E[X^*] = E[Y^*] = 0$ et $\text{Var}(X^*) = \text{Var}(Y^*) = 1$.

Étape 2 : Covariance des variables standardisées

Calculons la covariance de X^* et Y^* , qui est, par définition, la corrélation de X et Y .

$$\begin{aligned} \text{Cov}(X^*, Y^*) &= \text{Cov}\left(\frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y}\right) \\ &= \frac{1}{\sigma_X \sigma_Y} \text{Cov}(X - \mu_X, Y - \mu_Y) \\ &= \frac{1}{\sigma_X \sigma_Y} \text{Cov}(X, Y) \\ &= \text{Corr}(X, Y) \end{aligned}$$

Étape 3 : Variance de la somme et de la différence

Considérons la variance de la somme et de la différence de ces variables standardisées.

$$\text{Var}(X^* + Y^*) = \text{Var}(X^*) + \text{Var}(Y^*) + 2\text{Cov}(X^*, Y^*)$$

$$\text{Var}(X^* + Y^*) = 1 + 1 + 2\text{Corr}(X, Y) = 2 + 2\text{Corr}(X, Y)$$

De même :

$$\text{Var}(X^* - Y^*) = \text{Var}(X^*) + \text{Var}(Y^*) - 2\text{Cov}(X^*, Y^*)$$

$$\text{Var}(X^* - Y^*) = 1 + 1 - 2\text{Corr}(X, Y) = 2 - 2\text{Corr}(X, Y)$$

Étape 4 : La variance est toujours ≥ 0

La variance d'une variable aléatoire ne peut pas être négative.

$$\text{Var}(X^* + Y^*) \geq 0 \implies 2 + 2\text{Corr}(X, Y) \geq 0 \implies \text{Corr}(X, Y) \geq -1$$

$$\text{Var}(X^* - Y^*) \geq 0 \implies 2 - 2\text{Corr}(X, Y) \geq 0 \implies \text{Corr}(X, Y) \leq 1$$

Ceci nous donne le résultat final :

$$-1 \leq \text{Corr}(X, Y) \leq 1$$

6.6 Standardisation et Non-Corrélation

Définition : Variable Centrée Réduite

Soit X une variable aléatoire avec :

- moyenne $\mu_X = E[X]$
- écart-type $\sigma_X = \sqrt{\text{Var}(X)} > 0$

On définit sa version **centrée réduite** (standardisée) Z par :

$$Z = \frac{X - \mu_X}{\sigma_X}$$

Alors, Z a les propriétés suivantes :

1. **Centrée (moyenne nulle) :**

$$\begin{aligned} E[Z] &= E\left[\frac{X - \mu_X}{\sigma_X}\right] \\ &= \frac{1}{\sigma_X} E[X - \mu_X] \quad (\text{par linéarité, } \sigma_X \text{ est une constante}) \\ &= \frac{1}{\sigma_X} (E[X] - E[\mu_X]) \\ &= \frac{1}{\sigma_X} (E[X] - \mu_X) \quad (\text{car } \mu_X \text{ est une constante}) \\ &= \frac{\mu_X - \mu_X}{\sigma_X} = 0 \end{aligned}$$

2. **Réduite (écart-type égal à 1) :**

$$\begin{aligned} \text{Var}(Z) &= \text{Var}\left(\frac{X - \mu_X}{\sigma_X}\right) \\ &= \left(\frac{1}{\sigma_X}\right)^2 \text{Var}(X - \mu_X) \quad (\text{propriété } \text{Var}(aY) = a^2 \text{Var}(Y)) \\ &= \frac{1}{\sigma_X^2} \text{Var}(X) \quad (\text{propriété } \text{Var}(Y + b) = \text{Var}(Y)) \\ &= \frac{1}{\sigma_X^2} \cdot \sigma_X^2 = 1 \end{aligned}$$

L'écart-type est donc $\sigma_Z = \sqrt{\text{Var}(Z)} = \sqrt{1} = 1$.

Intuition : Que signifie centrer-réduire ?

Standardiser une variable se fait en deux temps, comme le montre la formule $Z = \frac{X - \mu_X}{\sigma_X}$:

1. **Centrer** ($X - \mu_X$) : C'est la première étape. On soustrait la moyenne μ_X . Cela revient à "déplacer" la distribution pour que son centre de gravité (sa moyenne) soit maintenant à 0. On ne regarde plus les valeurs brutes X , mais leurs **écarts** par rapport à la moyenne. (Propriété 1 : $E[Z] = 0$)
2. **Réduire** (\dots/σ_X) : C'est la deuxième étape. On divise ces écarts par l'écart-type σ_X . Cela revient à changer d'unité de mesure. L'ancienne unité (kg, cm, points...) est remplacée par une nouvelle unité universelle : "le nombre d'écarts-types". (Propriété 2 : $\text{Var}(Z) = 1$)

Au final, une variable Z avec une valeur de 1.5 signifie "cette observation est 1.5 écarts-types au-dessus de la moyenne de sa distribution d'origine", peu importe ce que X mesurait.

Intuition : Analogie simple

Imaginons 2 élèves :

- Alice a des notes entre 80 et 100 (moyenne 90, écart-type 5).
- Bob a des notes entre 0 et 20 (moyenne 10, écart-type 4).

Comparer leurs notes brutes n'a pas de sens. Mais si on les standardise, on peut se demander : "quand Alice est 1 écart-type au-dessus de sa moyenne (une note de 95), Bob est-il aussi 1 écart-type au-dessus de sa propre moyenne (une note de 14) ?". La standardisation permet cette comparaison.

Exemple : Centrer-réduire un dé

Pour un lancer de dé X , on a $\mu_X = 3.5$ et $\sigma_X = \sqrt{35/12} \approx 1.708$. Si on obtient $X = 6$: $Z = (6 - 3.5)/1.708 \approx 1.46$. Si on obtient $X = 1$: $Z = (1 - 3.5)/1.708 \approx -1.46$. Obtenir 6 est à 1.46 écarts-types au-dessus de la moyenne.

Définition : Variables Non Corrélées

On dit que deux variables aléatoires X et Y sont **non corrélées** si leur covariance est nulle :

$$\text{Cov}(X, Y) = 0$$

Cela est équivalent à dire que $E[XY] = E[X]E[Y]$.

Intuition

"Non corrélées" signifie qu'il n'y a **pas de relation linéaire** entre les variables. C'est plus faible que l'indépendance. Si X et Y sont indépendantes, elles sont forcément non corrélées. Mais l'inverse n'est pas vrai : X et Y peuvent être non corrélées ($\text{Cov}=0$) mais quand même dépendantes (par exemple si $Y = X^2$ pour un X centré).

6.7 Variance d'une Somme de Variables Aléatoires

Théorème : Formules pour la variance d'une somme de deux variables

Pour deux variables aléatoires X et Y :

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

Intuition

La "volatilité" (variance) d'une somme n'est pas juste la somme des volatilités. Il faut ajouter le terme d'interaction (covariance). Si $\text{Cov}(X, Y) > 0$ (elles bougent ensemble), la somme est **plus** volatile que la somme des parties. Si $\text{Cov}(X, Y) < 0$ (elles bougent en sens inverse), elles s'amortissent mutuellement. La somme est **moins** volatile. C'est le principe de la diversification en finance.

Théorème : Cas Particulier : Variables Non Corrélées

Si X et Y sont non corrélées ($\text{Cov}=0$), la formule se simplifie :

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Exemple : Variance d'une somme de dés

Soit $S = X + Y$ la somme de deux dés indépendants. Puisqu'ils sont indépendants,

ils sont non corrélés ($\text{Cov}(X, Y) = 0$). On sait $\text{Var}(X) = 35/12$ et $\text{Var}(Y) = 35/12$.

$$\text{Var}(S) = \text{Var}(X) + \text{Var}(Y) = \frac{35}{12} + \frac{35}{12} = \frac{70}{12} = \frac{35}{6} \approx 5.833$$

C'est bien plus simple que de calculer $E[S^2]$ et $E[S]$.

Théorème : Variance d'une somme de N variables

La formule générale pour la somme de N variables aléatoires X_1, \dots, X_n est :

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j)$$

Intuition

La variance totale d'un système (comme un portefeuille d'actions) est la somme de toutes les variances individuelles ("risques propres") plus la somme de **toutes** les paires de covariances ("risques d'interaction"). Dans un grand portefeuille, le nombre de termes de covariance (environ n^2) est bien plus grand que le nombre de termes de variance (n), donc le risque total est dominé par la façon dont les actifs interagissent.

6.8 Théorème sur la somme de lois de Poisson

Théorème : La Somme de v.a. de Poisson Indépendantes est Poisson

Soit X_1, \dots, X_k une séquence de variables aléatoires de Poisson indépendantes, avec des paramètres respectifs $\lambda_1, \dots, \lambda_k$.

$$X_i \sim \text{Poisson}(\lambda_i) \quad \text{pour } i = 1, \dots, k$$

Alors leur somme $Y = X_1 + \dots + X_k$ suit également une loi de Poisson, dont le paramètre est la somme des paramètres :

$$Y \sim \text{Poisson}(\lambda_1 + \dots + \lambda_k)$$

Intuition

Si des événements rares se produisent indépendamment à des taux constants, le nombre total d'événements se produisant est aussi un événement rare se produisant au taux total. Si les emails arrivent à $\lambda_1 = 5/\text{heure}$ et les appels à $\lambda_2 = 10/\text{heure}$, les

"communications totales" arrivent simplement à $\lambda = 5 + 10 = 15$ /heure.

Exemple : Centre d'appels

Un centre d'appels reçoit des appels "Ventes" selon $X_1 \sim \text{Poisson}(10 \text{ appels/heure})$ et des appels "Support" selon $X_2 \sim \text{Poisson}(15 \text{ appels/heure})$. Les deux types d'appels sont indépendants. Le nombre total d'appels $Y = X_1 + X_2$ suit une loi $Y \sim \text{Poisson}(10 + 15 = 25 \text{ appels/heure})$. La probabilité de recevoir exactement 20 appels en une heure est :

$$P(Y = 20) = \frac{e^{-25} 25^{20}}{20!}$$

7 Variables Aléatoires Continues

7.1 Fonction de Densité de Probabilité (PDF)

Définition : Fonction de Densité de Probabilité (PDF)

Soit X une variable aléatoire continue. Une fonction f est une **fonction de densité de probabilité** (Probability Density Function, ou PDF) de X si, pour tout x :

1. $f(x) \geq 0$, pour tout $-\infty < x < \infty$
2. $\int_{-\infty}^{\infty} f(x) dx = 1$ (l'aire totale sous la courbe vaut 1)

Intuition

Dans le cas discret, la PMF donnait une "masse" de probabilité à chaque point. Dans le cas continu, la probabilité en un point exact est nulle ($P(X = x) = 0$). La PDF, $f(x)$, n'est **pas** une probabilité.

Il faut voir $f(x)$ comme une **densité** : elle décrit la "concentration" de probabilité autour de x . Pour obtenir une probabilité (une "masse"), il faut intégrer cette densité sur un intervalle. La probabilité que X tombe dans un intervalle $[a, b]$ est l'aire sous la courbe de la PDF entre a et b :

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

Remarque : PDF vs Probabilité

Une erreur fréquente est de confondre la valeur $f(x)$ avec $P(X = x)$. Pour une variable continue, $P(X = x)$ est **toujours zéro**. La PDF $f(x)$ peut être supérieure à 1 (contrairement à une probabilité), tant que l'aire totale sous la courbe reste égale à 1. Pensez-y comme à une densité de population : elle peut être très élevée en un point, mais la "population" (probabilité) exacte en ce point infinitésimal est nulle.

Exemple : Une PDF simple

Soit X une v.a. avec la PDF $f(x) = 2x$ pour $x \in [0, 1]$, et $f(x) = 0$ sinon.

1. Est-ce une PDF valide ?
 - (1) $f(x) \geq 0$ pour tout x dans $[0, 1]$.
 - (2) $\int_{-\infty}^{\infty} f(x) dx = \int_0^1 2x dx = [x^2]_0^1 = 1 - 0 = 1$.
Oui, c'est une PDF valide.
2. Quelle est la probabilité $P(X \leq 0.5)$?

$$P(X \leq 0.5) = \int_0^{0.5} 2x dx = [x^2]_0^{0.5} = (0.5)^2 - 0 = 0.25$$

7.2 Fonction de Répartition (CDF)

Définition : Fonction de Répartition Continue (CDF)

Soit X une variable aléatoire continue. La **fonction de répartition** (Cumulative Distribution Function, ou CDF) de X est la fonction F définie par :

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$

Pour être une CDF valide, la fonction F doit respecter les propriétés suivantes :

1. $\lim_{x \rightarrow \infty} F(x) = 1$
2. $\lim_{x \rightarrow -\infty} F(x) = 0$
3. F est continue et non décroissante.

Intuition

La CDF est "l'accumulateur" de probabilité. Elle part de 0 (à $-\infty$) et "accumule" l'aire sous la PDF à mesure qu'on avance sur l'axe des x , pour finalement atteindre 1 (à $+\infty$).

Le lien fondamental est que la PDF est la dérivée de la CDF :

$$f(x) = F'(x)$$

Cela signifie que la valeur de la PDF $f(x)$ représente le **taux d'accumulation** de la probabilité au point x .

Remarque : Calcul de Probabilités via la CDF

La CDF est très pratique pour calculer des probabilités sur des intervalles :

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

Pour les variables continues, les inégalités strictes ou larges ne changent rien ($P(X = a) = 0$).

Exemple : CDF de l'exemple précédent

Pour $f(x) = 2x$ sur $[0, 1]$, la CDF $F(x)$ est :

- Si $x < 0$: $F(x) = \int_{-\infty}^x 0 dt = 0$.
- Si $0 \leq x \leq 1$: $F(x) = \int_{-\infty}^0 f(t) dt + \int_0^x 2t dt = 0 + [t^2]_0^x = x^2$.
- Si $x > 1$: $F(x) = \int_{-\infty}^1 f(t) dt + \int_1^x 0 dt = \int_0^1 2t dt = 1$.

$$\text{Donc, } F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x^2 & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

7.3 Espérance et Variance (Cas Continu)

Définition : Espérance et Variance (Cas Continu)

Pour une variable aléatoire X de fonction de densité f :
L'**espérance** de X est le centre de gravité de la densité :

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

La **variance** de X est l'espérance du carré de l'écart à la moyenne :

$$\text{Var}(X) = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2 f(x) dx$$

Théorème : Formule de calcul de la Variance

Une formule plus simple pour le calcul de la variance est :

$$\text{Var}(X) = E[X^2] - (E[X])^2$$

où $E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx$. (Ceci est une application de LOTUS).

Théorème : Théorème de Transfert (LOTUS)

Si X est une v.a. continue de densité $f(x)$, et g une fonction, alors :

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Remarque : Linéarité de l'Espérance

Comme dans le cas discret, l'espérance reste linéaire pour les variables continues :
 $E[aX + bY] = aE[X] + bE[Y]$.

Exemple : Espérance et Variance de l'exemple précédent

Pour $f(x) = 2x$ sur $[0, 1]$:

$$E[X] = \int_0^1 x \cdot (2x) dx = \int_0^1 2x^2 dx = \left[\frac{2x^3}{3} \right]_0^1 = \frac{2}{3}.$$

$$E[X^2] = \int_0^1 x^2 \cdot (2x) dx = \int_0^1 2x^3 dx = \left[\frac{2x^4}{4} \right]_0^1 = \frac{1}{2}.$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{2} - \frac{4}{9} = \frac{9-8}{18} = \frac{1}{18}.$$

7.4 Loi Uniforme

Définition : Loi Uniforme

Une variable aléatoire X est **uniformément distribuée** sur un intervalle $[a, b]$ si sa densité est une constante sur cet intervalle. Pour que l'aire totale soit 1, cette constante doit être $\frac{1}{b-a}$.

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{pour } x \in [a, b] \\ 0 & \text{sinon} \end{cases}$$

On note cela $X \sim \text{Unif}(a, b)$.

Intuition

C'est la distribution du "hasard pur" dans un intervalle borné. La probabilité de tomber dans un sous-intervalle ne dépend que de la **longueur** de ce sous-intervalle, pas de sa position (tant qu'il est dans $[a, b]$).

Théorème : Propriétés de la Loi Uniforme

Si $X \sim \text{Unif}(a, b)$:

- **CDF** : $F(x) = \frac{x-a}{b-a}$ pour $x \in [a, b]$.
- **Espérance** : $E[X] = \frac{a+b}{2}$ (le point milieu de l'intervalle).
- **Variance** : $\text{Var}(X) = \frac{(b-a)^2}{12}$.

7.5 Loi Exponentielle

Définition : Loi Exponentielle

Une variable aléatoire X suit une **loi exponentielle** de paramètre $\lambda > 0$ si sa fonction de densité a la forme :

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{pour } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

On note $X \sim \text{Exp}(\lambda)$.

Intuition : Lien entre les lois de Poisson et Exponentielle

La loi exponentielle modélise le temps d'attente *avant* le prochain événement dans un processus de Poisson.

Posons la question : « Si je commence à observer maintenant, combien de temps T vais-je devoir attendre avant de voir le prochain événement ? »

1. Dans un processus de Poisson de taux λ , le nombre d'événements $N(t)$ dans un intervalle de temps t suit une loi de Poisson de paramètre λt :

$$P(N(t) = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

2. La probabilité de ne voir **aucun** événement ($k = 0$) pendant une durée t est :

$$P(N(t) = 0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}$$

3. Mais ne voir aucun événement pendant un temps t , c'est exactement dire que le temps d'attente T du premier événement est *plus grand* que t .

$$P(T > t) = P(N(t) = 0) = e^{-\lambda t}$$

4. À partir de là, on déduit la fonction de répartition (CDF) de T :

$$F_T(t) = P(T \leq t) = 1 - P(T > t) = 1 - e^{-\lambda t} \quad (\text{pour } t \geq 0)$$

5. En dérivant la CDF pour obtenir la densité (PDF) :

$$f_T(t) = F'_T(t) = \frac{d}{dt}(1 - e^{-\lambda t}) = -(-\lambda e^{-\lambda t}) = \lambda e^{-\lambda t}$$

C'est exactement la densité de la loi exponentielle de paramètre λ .

Théorème : Propriétés de la Loi Exponentielle

Si $X \sim \text{Exp}(\lambda)$:

- **CDF** : $F(x) = 1 - e^{-\lambda x}$ pour $x \geq 0$.

- **Espérance** : $E[X] = \frac{1}{\lambda}$.
- **Variance** : $\text{Var}(X) = \frac{1}{\lambda^2}$.
- **Propriété de non-mémoire** : Pour $s, t \geq 0$, $P(X > s + t \mid X > s) = P(X > t)$.

Remarque : Interprétation du paramètre λ

Le paramètre λ représente le **taux** moyen d'occurrence des événements dans le processus de Poisson sous-jacent (par exemple, nombre moyen d'appels par minute). L'espérance $1/\lambda$ est alors le **temps moyen entre les événements**.

Intuition : La Propriété de Non-Mémoire

C'est la propriété la plus contre-intuitive et la plus importante de la loi exponentielle. Elle signifie que le processus "oublie" le passé. [...]

7.6 Distributions Conjointes (Cas Continu)

Définition : Fonction de Densité Conjointe

Pour des variables aléatoires continues X et Y , la **fonction de densité conjointe** $f(x, y)$ décrit la densité de probabilité sur le plan (x, y) . Elle doit respecter :

1. $f(x, y) \geq 0$, pour tous x, y .
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = 1$.

Intuition : Volume = Probabilité

La probabilité que le couple (X, Y) tombe dans une région A du plan xy est le **volume** sous la surface $z = f(x, y)$ au-dessus de cette région A .

$$P((X, Y) \in A) = \iint_A f(x, y) \, dA$$

Définition : Densités Marginales

On peut retrouver les densités individuelles (marginales) en "écrasant" le volume 3D sur un seul axe. Pour obtenir la PDF de X seul, on intègre $f(x, y)$ sur toutes les valeurs possibles de y :

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) \, dy$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) \, dx$$

Définition : CDF Conjointe

La **fonction de répartition conjointe** (CDF) est :

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^y \int_{-\infty}^x f(s, t) \, ds \, dt$$

Elle représente le volume "au sud-ouest" du point (x, y) .

7.7 Espérance, Indépendance et Covariance (Cas Conjoint)

Théorème : LOTUS pour les v.a. conjointes

Si X et Y ont une densité conjointe $f(x, y)$, et $g(x, y)$ est une fonction :

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) \, dx \, dy$$

Définition : Indépendance et Densité

Les variables aléatoires continues X et Y sont **indépendantes** si et seulement si leur densité conjointe est le produit de leurs densités marginales :

$$f(x, y) = f_X(x) f_Y(y), \quad \text{pour tous } x, y$$

Intuition

Intuitivement, l'indépendance signifie que le "profil" de la densité en x ne change pas quelle que soit la valeur de y (et vice-versa).

Définition : Covariance (cas continu)

La **covariance** de X et Y mesure leur variation linéaire commune :

$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x, y) \, dx \, dy\end{aligned}$$

Théorème : Formule de calcul de la Covariance

Une formule plus simple pour le calcul de la covariance est :

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

où $E[XY]$ est calculé via LOTUS : $E[XY] = \iint xyf(x, y) \, dx \, dy$.

Remarque : Indépendance et Covariance

Si X et Y sont indépendantes, alors $\text{Cov}(X, Y) = 0$. Cependant, la réciproque n'est **pas** toujours vraie pour les variables aléatoires en général (bien qu'elle le soit dans certains cas importants comme pour les variables gaussiennes). Une covariance nulle signifie seulement une absence de *relation linéaire*, mais il peut exister d'autres formes de dépendance.

8 La Loi Normale (ou Gaussienne)

8.1 Introduction et Fonction de Densité (PDF)

Définition : Loi Normale

Une variable aléatoire continue X suit une **loi normale** (ou loi de Gauss) de paramètres μ (l'espérance) et σ^2 (la variance), notée $X \sim \mathcal{N}(\mu, \sigma^2)$, si sa fonction de densité de probabilité (PDF) est donnée par :

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

pour tout $x \in (-\infty, \infty)$, où $\sigma > 0$.

Intuition : La Courbe en Cloche

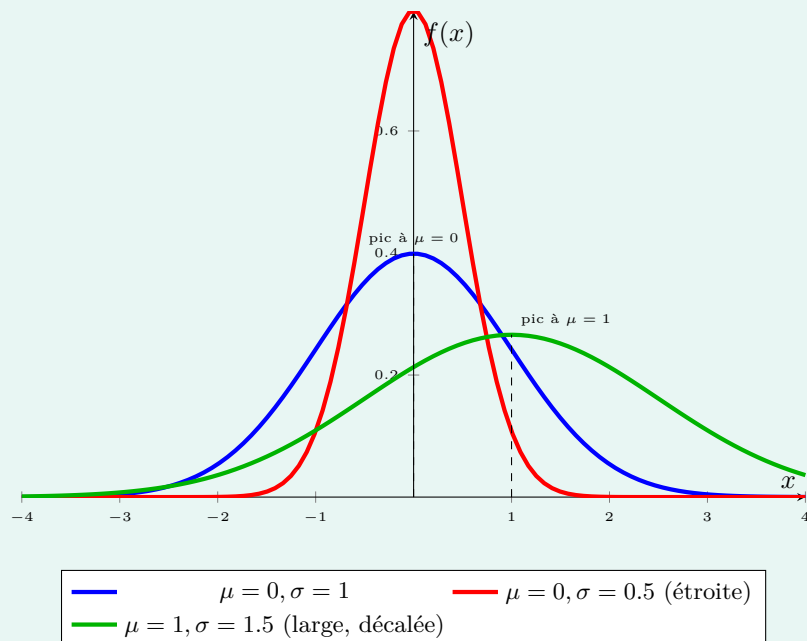
La loi normale est sans doute la distribution la plus importante en probabilités et statistiques. Pourquoi ? Parce qu'elle modélise remarquablement bien de nombreux phénomènes naturels et processus aléatoires où les valeurs tendent à se regrouper autour d'une moyenne, avec des écarts symétriques devenant de plus en plus rares à mesure qu'on s'éloigne de cette moyenne. Pensez à la taille des individus dans une population, aux erreurs de mesure répétées, ou même aux notes d'un grand groupe d'étudiants à un examen bien conçu.

Sa densité a une forme caractéristique de **cloche symétrique** :

- **Le Centre (μ)** : Le paramètre μ représente l'**espérance** (la moyenne) de la distribution. C'est le centre de symétrie de la courbe, là où la cloche atteint son **sommet**. C'est la valeur la plus probable (le mode) et aussi la valeur qui coupe la distribution en deux moitiés égales (la médiane). Changer μ *translate* la cloche horizontalement sans changer sa forme.
- **La Dispersion (σ)** : Le paramètre σ est l'**écart-type** (σ^2 est la variance). Il mesure la **dispersion** des valeurs autour de la moyenne μ . Géométriquement, σ contrôle la **largeur** de la cloche.
 - Un *petit* σ signifie que les données sont très concentrées autour de la moyenne, donnant une cloche **étroite et pointue**.
 - Un *grand* σ signifie que les données sont plus étalées, donnant une cloche **large et aplatie**.

Les points d'inflexion de la courbe (là où la courbure change de sens) se situent exactement à $\mu \pm \sigma$.

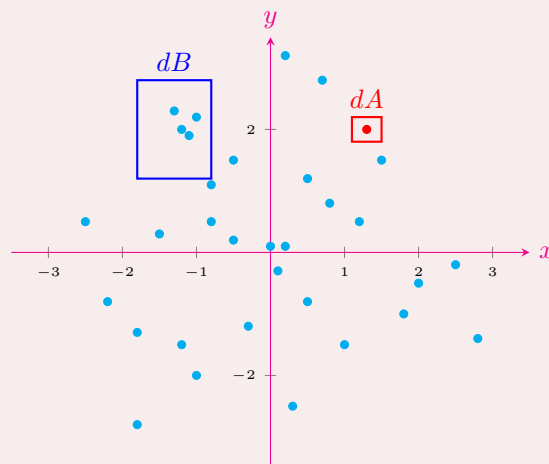
La Courbe en Cloche (PDF de la Loi Normale)



Influence de μ (position) et σ (largeur) sur la forme de la cloche.

Preuve : Dérivation de la Densité Normale à partir des Principes Fondamentaux

Contexte Visuel : Imaginons un nuage de points dispersés autour d'une cible à l'origine $(0, 0)$, comme des impacts de fléchettes. Le graphique ci-dessous illustre cette dispersion. On s'intéresse à la probabilité de tomber dans une petite zone, comme dA , autour d'un point (x, y) .



Objectif : Expliquer comment arriver à la formule mathématique de la courbe en cloche (densité de probabilité normale) en partant de principes fondamentaux sur les erreurs aléatoires.

1. Le Point de Départ : Densité et Aire dA

Dans une distribution continue, la probabilité de tomber *exactement* sur un point (x, y) est nulle. On ne peut donc pas parler de "probabilité d'un point". On parle de la probabilité de tomber *dans une petite zone*, comme un rectangle $dA = dx \cdot dy$ autour du point (x, y) .

Cette probabilité, notée $P(\text{dans } dA)$, est *proportionnelle* à l'aire de la zone dA . La *constante de proportionnalité* est la **fonction de densité de probabilité** $p(x, y)$ évaluée en ce point. En d'autres termes, la densité $p(x, y)$ *représente* localement la concentration de probabilité. Ainsi, la probabilité de tomber dans la zone dA est approximativement :

$$P(\text{dans } dA) \approx p(x, y) \cdot dA$$

Donc, $p(x, y)$ est la *densité de probabilité* en (x, y) , et non la probabilité elle-même multipliée par l'aire. La probabilité totale sur une zone finie est obtenue en intégrant cette densité sur cette zone.

2. Les Hypothèses Fondamentales

On pose deux hypothèses sur la nature de ces erreurs (représentées par la densité $p(x, y)$) :

1. **Indépendance des axes** : L'erreur horizontale (x) est indépendante de l'erreur verticale (y). Cela implique que la densité jointe $p(x, y)$ peut s'écrire comme le produit de la densité marginale sur x , notée $f(x)$, et de la densité marginale sur y , notée $f(y)$. Donc, $p(x, y) = f(x) \cdot f(y)$.
2. **Symétrie de rotation (Isotropie)** : La densité ne dépend que de la distance $r = \sqrt{x^2 + y^2}$ au centre, pas de l'angle. Il existe donc une fonction $\phi(r)$ telle que la densité en (x, y) est $p(x, y) = \phi(\sqrt{x^2 + y^2})$.

3. L'Équation Fonctionnelle

En égalant les deux expressions pour la même densité $p(x, y)$ (à une constante près, car la densité peut être définie à un facteur près avant normalisation), on obtient :

$$f(x) \cdot f(y) = \phi(\sqrt{x^2 + y^2})$$

Pour $y = 0$, on a $f(x) \cdot f(0) = \phi(x)$. Posons $f(0) = \lambda$ (la densité maximale au centre). Alors $\phi(x) = \lambda f(x)$. L'équation devient :

$$f(x) \cdot f(y) = \lambda f(\sqrt{x^2 + y^2})$$

4. Résolution de l'Équation Fonctionnelle

1. **Normalisation** : Posons $g(x) = f(x)/\lambda$, avec $g(0) = 1$. L'équation se simplifie en :

$$g(x)g(y) = g(\sqrt{x^2 + y^2})$$

2. **Changement de variable** : Posons $g(x) = h(x^2)$. L'équation devient $h(x^2)h(y^2) = h(x^2 + y^2)$. Avec $a = x^2$ et $b = y^2$, on a :

$$h(a)h(b) = h(a + b)$$

3. **Solution exponentielle** : La solution continue de cette équation de Cauchy est $h(a) = e^{Aa}$ pour une constante A .
4. **Retour aux fonctions et contrainte physique** : $g(x) = h(x^2) = e^{Ax^2}$. $f(x) = \lambda g(x) = \lambda e^{Ax^2}$. Comme la densité doit diminuer loin du centre, la constante A doit être négative. Posons $A = -k$ avec $k > 0$.

$$f(x) = \lambda e^{-kx^2}$$

5. Normalisation et Identification des Paramètres

1. **Condition** $\int f(x)dx = 1$: $\int_{-\infty}^{\infty} \lambda e^{-kx^2} dx = \lambda \sqrt{\frac{\pi}{k}} = 1 \implies \lambda = \sqrt{\frac{k}{\pi}}$.
2. **Lien avec la Variance** (σ^2) : Pour une distribution centrée, $\sigma^2 = E[X^2] = \int x^2 f(x)dx$.

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 \left(\sqrt{\frac{k}{\pi}} e^{-kx^2} \right) dx = \sqrt{\frac{k}{\pi}} \left(\frac{1}{2k} \sqrt{\frac{\pi}{k}} \right) = \frac{1}{2k}$$

Donc, $k = \frac{1}{2\sigma^2}$.

3. **Substitution Finale** : Remplaçons k dans λ et $f(x)$.

$$\lambda = \sqrt{\frac{1/(2\sigma^2)}{\pi}} = \frac{1}{\sigma\sqrt{2\pi}}$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}x^2} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

4. **Généralisation (Moyenne μ)** : Pour centrer la distribution sur μ , on remplace x par $(x - \mu)$ dans l'exposant :

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

C'est la fonction de densité de la loi normale $\mathcal{N}(\mu, \sigma^2)$, dérivée des hypothèses fondamentales sur les erreurs aléatoires.

8.2 La Loi Normale Centrée Réduite $\mathcal{N}(0, 1)$

Définition : Loi Normale Standard (ou Centrée Réduite)

Un cas particulier extraordinairement utile est la loi normale avec une moyenne $\mu = 0$ et une variance $\sigma^2 = 1$ (donc $\sigma = 1$). On l'appelle la **loi normale standard** ou **centrée réduite**, et on la note souvent Z . Sa PDF est traditionnellement notée $\phi(z)$:

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

Sa fonction de répartition (CDF), qui donne $P(Z \leq z)$, est notée $\Phi(z)$:

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

Intuition : La Référence Universelle et le Changement d'Unités

Pourquoi cette loi $\mathcal{N}(0, 1)$ est-elle si centrale ? Imaginez que vous ayez des mesures en degrés Celsius ($\mathcal{N}(\mu_C, \sigma_C^2)$) et d'autres en degrés Fahrenheit ($\mathcal{N}(\mu_F, \sigma_F^2)$). Comment les comparer ? La loi normale standard fournit un **système d'unités universel**. Toute variable normale $X \sim \mathcal{N}(\mu, \sigma^2)$ peut être transformée ("standardisée") en une

variable $Z \sim \mathcal{N}(0, 1)$ par un simple changement d'échelle et de position : $Z = (X - \mu)/\sigma$.

Cela signifie qu'au lieu de devoir calculer des aires (probabilités) pour une infinité de courbes en cloche différentes (une pour chaque paire μ, σ), on peut tout ramener à **une seule courbe de référence**, $\mathcal{N}(0, 1)$. Les aires sous cette courbe standard ($\Phi(z)$) ont été calculées une fois pour toutes et sont disponibles dans des tables ou des logiciels. On n'a plus qu'à convertir notre problème dans cette "langue" standard, trouver la probabilité, et interpréter le résultat.

Remarque : Notation ϕ et Φ

Les symboles ϕ (phi minuscule) pour la PDF et Φ (phi majuscule) pour la CDF de la loi normale standard sont quasi universels. Il est important de ne pas les confondre. $\phi(z)$ est la *hauteur* de la courbe en z , tandis que $\Phi(z)$ est l'*aire* sous la courbe à gauche de z .

Remarque : Absence de Primitive Simple

L'intégrale $\int e^{-t^2/2} dt$, nécessaire pour calculer $\Phi(z)$, n'a **pas d'expression analytique** en termes de fonctions élémentaires (polynômes, exponentielles, log, sin, cos...). C'est une fonction spéciale, connue sous le nom de **fonction d'erreur** (liée à Φ par une transformation simple). C'est la raison pour laquelle on dépend de tables ou de calculs numériques pour obtenir les valeurs de $\Phi(z)$. Heureusement, ces outils sont omniprésents aujourd'hui.

8.3 Standardisation : Le Score Z

Théorème : Standardisation d'une Variable Normale

Si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors la variable Z définie par :

$$Z = \frac{X - \mu}{\sigma}$$

suit la loi normale standard, $Z \sim \mathcal{N}(0, 1)$.

Intuition : Mesurer en "Unités d'Écart-Type"

Transformer X en Z s'appelle **standardiser** la variable. Le résultat, $z = \frac{x - \mu}{\sigma}$, est appelé le **Score Z** (ou cote Z). Ce score Z est une mesure *sans unité* qui indique **à combien d'écarts-types** une valeur observée x se situe par rapport à la moyenne

μ de sa distribution.

- $z = 0$: x est exactement à la moyenne ($\mathbf{x} = \mu$).
- $z = +1$: x est un écart-type *au-dessus* de la moyenne ($\mathbf{x} = \mu + \sigma$).
- $z = -2$: x est deux écarts-types *en dessous* de la moyenne ($\mathbf{x} = \mu - 2\sigma$).

Cette transformation est extrêmement utile pour :

1. **Comparer des valeurs** issues de distributions normales différentes. Un score Z de $+1.5$ a toujours la même signification relative, que l'on parle de QI, de taille, ou de température.
2. **Calculer des probabilités** en utilisant la table unique de la loi $\mathcal{N}(0, 1)$.

Exemple : Comparaison de Performances

Un étudiant A obtient 80 points à un examen où la moyenne est $\mu_A = 70$ et l'écart-type $\sigma_A = 5$. Un étudiant B obtient 85 points à un autre examen où $\mu_B = 75$ et $\sigma_B = 10$. Qui a le mieux réussi relativement à son groupe ?

Calculons les Z-scores :

$$Z_A = \frac{80 - 70}{5} = \frac{10}{5} = +2.0$$

$$Z_B = \frac{85 - 75}{10} = \frac{10}{10} = +1.0$$

L'étudiant A a un score Z plus élevé ($+2.0$ contre $+1.0$), ce qui signifie qu'il se situe plus d'écarts-types au-dessus de la moyenne de son groupe que l'étudiant B. L'étudiant A a donc relativement mieux réussi.

8.4 Propriétés Importantes de la Loi Normale

Théorème : Stabilité par Transformation Linéaire

Si $X \sim \mathcal{N}(\mu, \sigma^2)$ et $Y = aX + b$ (avec $a \neq 0$), alors Y suit aussi une loi normale :

$$Y \sim \mathcal{N}(a\mu + b, (a\sigma)^2)$$

L'espérance est transformée linéairement ($E[aX + b] = aE[X] + b$), et la variance est multipliée par a^2 ($\text{Var}(aX + b) = a^2\text{Var}(X)$).

Exemple : Changement d'Unités

Si la température en Celsius T_C suit $\mathcal{N}(20, 5^2)$, quelle est la loi de la température en Fahrenheit $T_F = \frac{9}{5}T_C + 32$?

$a = 9/5$, $b = 32$.

Nouvelle moyenne : $E[T_F] = \frac{9}{5}(20) + 32 = 36 + 32 = 68$.

Nouvel écart-type : $\sigma_{T_F} = |a|\sigma_{T_C} = \frac{9}{5}(5) = 9$. Nouvelle variance : $\sigma_{T_F}^2 = 9^2 = 81$.

Donc, $T_F \sim \mathcal{N}(68, 9^2)$.

Théorème : Stabilité par Addition (Indépendance)

Si $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ et $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ sont des variables aléatoires **indépendantes**, alors leur somme $S = X + Y$ suit aussi une loi normale :

$$S \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

Les moyennes s'ajoutent, et (grâce à l'indépendance) les variances s'ajoutent.

Remarque : Attention à l'Indépendance

La propriété d'addition des variances ($\sigma_S^2 = \sigma_X^2 + \sigma_Y^2$) est cruciale et ne tient **que si X et Y sont indépendantes**. Si elles ne le sont pas, la variance de la somme inclut un terme de covariance : $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$. Cependant, la somme de variables normales (même dépendantes) reste normale (si elles sont conjointement normales).

Exemple : Poids Total

Le poids d'une pomme suit $\mathcal{N}(150g, 10^2)$. Le poids d'une orange suit $\mathcal{N}(200g, 15^2)$. On suppose les poids indépendants. Quel est la loi du poids total d'une pomme et d'une orange ?

Soit P le poids de la pomme, O celui de l'orange. $T = P + O$.

$$E[T] = E[P] + E[O] = 150 + 200 = 350g.$$

$$\text{Var}(T) = \text{Var}(P) + \text{Var}(O) = 10^2 + 15^2 = 100 + 225 = 325.$$

Donc, $T \sim \mathcal{N}(350, 325)$. L'écart-type du poids total est $\sqrt{325} \approx 18.03g$.

8.5 La Règle Empirique (68-95-99.7)

Théorème : Règle Empirique

Pour toute variable $X \sim \mathcal{N}(\mu, \sigma^2)$:

- $P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.6827$ (Environ **68%** des valeurs dans $\mu \pm \sigma$).
- $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.9545$ (Environ **95%** des valeurs dans $\mu \pm 2\sigma$).
- $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.9973$ (Environ **99.7%** des valeurs dans $\mu \pm 3\sigma$).

Intuition : Repères Essentiels sur la Cloche

Cette règle, dérivée directement des aires sous la courbe $\mathcal{N}(0, 1)$ entre $z = \pm 1$, $z = \pm 2$ et $z = \pm 3$, fournit des repères extrêmement utiles pour interpréter l'écart-type σ . Elle nous dit où se trouve la grande majorité des données.

Une observation qui tombe en dehors de l'intervalle $\mu \pm 3\sigma$ est très inhabituelle (elle n'a que 0.3pct de chances de se produire). C'est souvent considéré comme une *valeur aberrante* (outlier) potentielle.

8.6 Calcul de Probabilités Normales

Exemple : Utilisation du Z-score

Supposons que le QI d'une population suit $\mathcal{N}(100, 15^2)$. Quelle est la probabilité $P(X > 130)$?

1. **Standardiser** : $z = \frac{130-100}{15} = 2$. On cherche $P(Z > 2)$. 2. **Utiliser la CDF Standard** : $P(Z > 2) = 1 - P(Z \leq 2) = 1 - \Phi(2)$. 3. **Chercher dans la table / Calculer** : $\Phi(2) \approx 0.9772$. 4. **Résultat** : $P(X > 130) = 1 - 0.9772 = 0.0228$.

Exemple : Probabilité entre deux valeurs

Quelle est la probabilité $P(85 \leq X \leq 115)$?

1. **Standardiser** : $z_1 = \frac{85-100}{15} = -1$, $z_2 = \frac{115-100}{15} = 1$. On cherche $P(-1 \leq Z \leq 1)$. 2. **Utiliser la CDF Standard** : $P(-1 \leq Z \leq 1) = \Phi(1) - \Phi(-1)$. 3. **Utiliser la symétrie** : $\Phi(-z) = 1 - \Phi(z)$. Donc $\Phi(-1) = 1 - \Phi(1)$. $P(-1 \leq Z \leq 1) = \Phi(1) - (1 - \Phi(1)) = 2\Phi(1) - 1$. 4. **Chercher dans la table / Calculer** : $\Phi(1) \approx 0.8413$. 5. **Résultat** : $P(85 \leq X \leq 115) \approx 2(0.8413) - 1 = 1.6826 - 1 = 0.6826$.

Exemple : Trouver une valeur pour une probabilité donnée (Problème Inverse)

Quel est le QI minimum requis pour être dans le top 10% de la population ? ($\mu = 100, \sigma = 15$).

1. **Trouver le Z-score correspondant** : On cherche x tel que $P(X > x) = 0.10$. Cela équivaut à $P(Z > z) = 0.10$, où $z = (x - 100)/15$. Si $P(Z > z) = 0.10$, alors $P(Z \leq z) = \Phi(z) = 1 - 0.10 = 0.90$. 2. **Chercher dans la table inverse / Calculer** : On cherche la valeur z pour laquelle l'aire à gauche est 0.90. On trouve $z \approx 1.28$. 3. **Convertir en X** : On utilise la relation $z = (x - \mu)/\sigma$ pour trouver x : $1.28 = \frac{x-100}{15}$ $x = 100 + 1.28 \times 15 = 100 + 19.2 = 119.2$. Il faut un QI d'environ 119.2 pour être dans le top 10%.

9 Moments d'une distribution

9.1 Définitions fondamentales des moments

Définition : Types de Moments

Soit X une variable aléatoire ayant une espérance μ et une variance σ^2 . Pour tout entier positif m , on définit les moments suivants :

- **m -ième moment (non centré)** : $E[X^m]$.
- **m -ième moment centré** : $E[(X - \mu)^m]$.
- **m -ième moment standardisé** : $E\left[\left(\frac{X - \mu}{\sigma}\right)^m\right]$.

Les moments centrés et standardisés permettent d'étudier les propriétés de la distribution indépendamment de sa position (μ) et de son échelle (σ).

9.2 Asymétrie (Skewness)

Définition : Asymétrie (Skewness)

L'**asymétrie** (ou *skewness*) d'une variable aléatoire X de moyenne μ et d'écart-type σ est définie comme le **troisième moment standardisé** :

$$\text{Skew}(X) = E\left[\left(\frac{X - \mu}{\sigma}\right)^3\right].$$

Intuition : Comprendre la Formule du Skewness

Pour une variable aléatoire X de moyenne μ et d'écart-type σ , le **skewness** est défini comme :

$$\text{Skew}(X) = \frac{E[(X - \mu)^3]}{\sigma^3}$$

Logique du numérateur : le moment centré d'ordre 3

- Le terme $(X - \mu)^3$ est le **cube de l'écart à la moyenne**
- Contrairement à $(X - \mu)^2$ (toujours positif), le cube **conserve le signe** de l'écart
- Il pondère différemment les observations à gauche et à droite de la moyenne

Interprétation intuitive

- **Skewness = 0** : Distribution symétrique, écarts positifs et négatifs s'annulent
- **Skewness > 0** : Queue longue à droite, grandes valeurs positives amplifiées par le cube
- **Skewness < 0** : Queue longue à gauche, écarts négatifs dominent

Pourquoi σ^3 au dénominateur ?

- Le moment d'ordre 3 est homogène à des unités au cube
- On divise par σ^3 pour obtenir un coefficient **sans dimension**
- Permet la comparaison entre distributions de différentes échelles

Type	Forme	Interprétation
Skew = 0	Symétrique	Moyenne = Médiane = Mode
Skew > 0	Queue à droite	Valeurs extrêmes tirent la moyenne vers la droite
Skew < 0	Queue à gauche	Valeurs extrêmes tirent la moyenne vers la gauche

Remarque : Pourquoi Standardiser ?

En standardisant d'abord $(\frac{X-\mu}{\sigma})$, la définition de $\text{Skew}(X)$ ne dépend ni de la position (μ) ni de l'échelle (σ) de la distribution, ce qui est raisonnable puisque ces informations sont déjà fournies par la moyenne et l'écart-type. De plus, cette standardisation garantit que l'asymétrie est invariante par changement d'unité de mesure (par exemple, passer des pouces aux mètres n'affecte pas la valeur de l'asymétrie).

9.3 Propriétés de symétrie

Définition : Symétrie d'une Variable Aléatoire

On dit qu'une variable aléatoire X a une distribution **symétrique** autour de μ si la variable $X - \mu$ a la même distribution que $\mu - X$. On dit aussi que X est symétrique ou que sa distribution est symétrique. Ces trois formulations ont le même sens.

Théorème : Symétrie en Termes de Fonction de Densité

Soit X une variable aléatoire continue de fonction de densité de probabilité (PDF) f . Alors, X est symétrique autour de μ si et seulement si :

$$f(x) = f(2\mu - x) \quad \text{pour tout } x.$$

Preuve : Preuve du Théorème de Symétrie

Soit F la fonction de répartition (CDF) de X . Si la symétrie tient, alors :

$$F(x) = P(X \leq x) = P(X - \mu \leq x - \mu) = P(\mu - X \leq x - \mu) = P(X \geq 2\mu - x) = 1 - F(2\mu - x).$$

En prenant la dérivée des deux côtés par rapport à x , on obtient :

$$f(x) = \frac{d}{dx} F(x) = \frac{d}{dx} [1 - F(2\mu - x)] = f(2\mu - x).$$

Cela démontre que la condition $f(x) = f(2\mu - x)$ est nécessaire et suffisante pour la symétrie.

9.4 Aplatissement (Kurtosis)

Définition : Kurtosis (Aplatissement)

Pour une variable aléatoire X de moyenne μ et d'écart-type σ , le **kurtosis** est défini comme le **quatrième moment standardisé** :

$$\text{Kurtosis}(X) = E \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right].$$

Dans la pratique, on utilise plus souvent le **kurtosis excessif** (ou excès de kurtosis), défini comme :

$$\text{Excess Kurtosis}(X) = E \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right] - 3.$$

La soustraction de 3 fait en sorte que le kurtosis d'une loi normale soit égal à 0.

Intuition : Comprendre la Kurtosis

Pour une variable aléatoire X , le **kurtosis** est défini comme :

$$\text{Kurt}(X) = \frac{E[(X - \mu)^4]}{\sigma^4}$$

et l'**excess kurtosis** (kurtosis excédentaire) comme : $\text{Excess Kurtosis} = \text{Kurt}(X) - 3$.

Pourquoi le moment d'ordre 4 ?

- Comme la variance, on utilise une puissance paire (pas d'effet de signe)
- La puissance 4 **amplifie énormément les écarts extrêmes**
- Mesure le **poids des queues** et la **concentration autour de la moyenne**

Interprétation intuitive

- **Excess Kurtosis = 0** (Kurtosis = 3) : Distribution normale (mésocurtique)
- **Excess Kurtosis < 0** (Kurtosis < 3) : Distribution aplatie (platykurtique) — queues légères
- **Excess Kurtosis > 0** (Kurtosis > 3) : Distribution pointue avec queues épaisses (leptokurtique)

Application en finance

- Les rendements financiers ont souvent un excès de kurtosis positif
- Indique une probabilité plus élevée d'événements extrêmes que la loi normale
- Justifie le "vol smile" dans les options

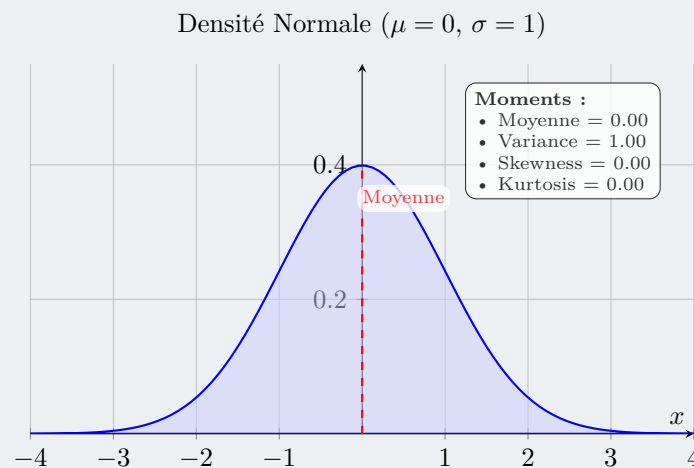
Pourquoi σ^4 au dénominateur ?

- Le moment d'ordre 4 est homogène à des unités⁴
- On divise par σ^4 pour un coefficient **sans dimension**

Type	Kurtosis	Excess	Caractéristiques
Platykurtique	< 3	< 0	Queues légères, centre large
Mésokurtique	$= 3$	$= 0$	Référence normale
Leptokurtique	> 3	> 0	Queues épaisses, pic pointu

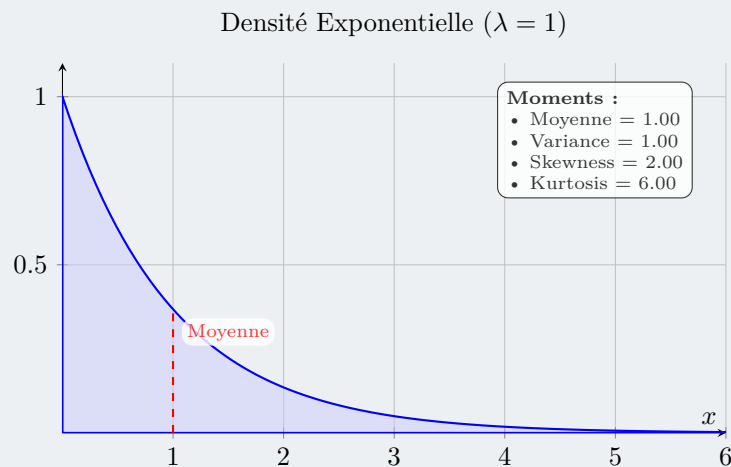
9.5 Exemples de distributions

Exemple : La Distribution Normale (Mésokurtique)



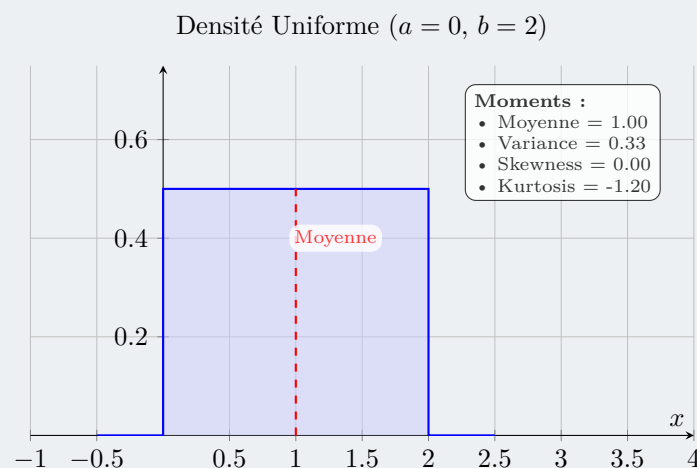
La distribution normale est l'archétype de la courbe en cloche. Imaginez une cible : la majorité des flèches touchent le centre, et plus on s'éloigne du centre, moins il y a de chances d'être touché. C'est une distribution parfaitement symétrique, ce qui se traduit par un **skewness nul (0.00)**. Son pic est ni trop pointu, ni trop plat : c'est notre point de référence, on dit qu'elle est **mésokurtique**, d'où son kurtosis de **0.00**. C'est la base de nombreuses analyses statistiques car elle modélise naturellement beaucoup de phénomènes.

Exemple : La Distribution Exponentielle (Asymétrique à Droite)



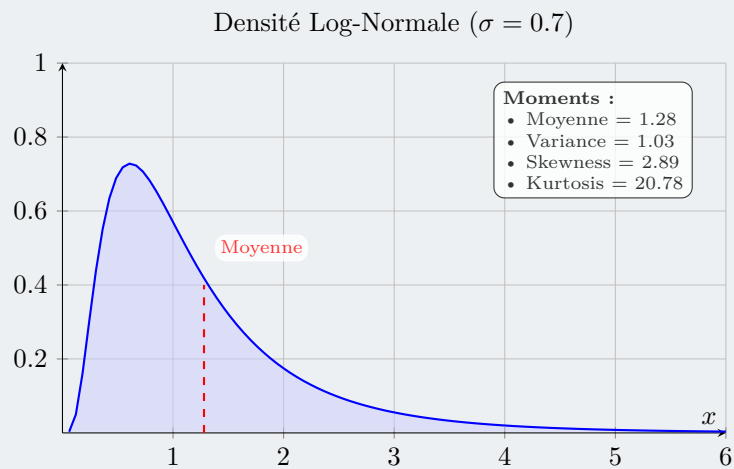
Imaginez le temps d'attente avant un événement rare, comme un appel téléphonique. La plupart du temps, l'appel arrive vite, mais il peut parfois y avoir de longues attentes. C'est exactement ce que modélise la distribution exponentielle : un pic à gauche et une longue queue à droite. Cela se traduit par un **skewness positif élevé (2.00)**, indiquant une asymétrie marquée. Elle est aussi **leptokurtique (kurtosis = 6.00)** : son pic est pointu, et la longue queue droite signifie qu'il y a une probabilité non négligeable de valeurs extrêmes.

Exemple : La Distribution Uniforme (Platykurtique)



La distribution uniforme, c'est le "tirage au sort parfait" : chaque valeur sur un intervalle a la même chance d'être tirée. Visuellement, c'est un rectangle, donc aucune valeur n'est privilégiée. Elle est symétrique (**skewness = 0.00**), mais contrairement à la normale, elle est "plate", sans pic central. Cela se traduit par un **kurtosis négatif (-1.20)**, ce qui signifie qu'elle est **platykurtique**. Elle est donc très différente des distributions avec un pic central comme la normale.

Exemple : La Distribution Log-Normale (Fortement Leptokurtique)



La log-normale est une distribution très asymétrique. Imaginez la richesse d'une population : la majorité est modeste, mais il existe une petite proportion de très riches, ce qui "étire" la droite de la courbe. Cela donne un **skewness très élevé (2.89)**. Elle est extrêmement **leptokurtique (kurtosis = 20.78)** : un pic très aigu et une queue droite très lourde. Cela signifie qu'il y a un risque élevé de valeurs extrêmement grandes, ce qui la rend très utile pour modéliser des phénomènes avec de rares événements extrêmes.

10 Appendice A : Séries de Taylor et Maclaurin

Définition : Séries de Taylor et Maclaurin

Si une fonction f est indéfiniment dérivable au voisinage d'un point a , sa **série de Taylor** centrée en a est définie par :

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-a)^k$$

où $f^{(k)}(a)$ est la k -ième dérivée de f évaluée en a .

Dans le cas particulier où $a = 0$, la série est appelée une **série de Maclaurin**. C'est la forme la plus courante, car elle approxime les fonctions autour de l'origine.

10.1 Construction pas à pas d'une série de Taylor

Intuition : La logique de la correspondance des dérivées

L'objectif fondamental d'une série de Taylor est de construire un polynôme, $P(x)$, qui soit une "copie conforme" d'une fonction $f(x)$ autour d'un point a . Pour ce faire, on force le polynôme à avoir exactement les mêmes propriétés locales que la fonction : même valeur, même pente, même courbure, etc. Cela se traduit mathématiquement par une exigence : **la n -ième dérivée du polynôme en a doit être égale à la n -ième dérivée de la fonction en a** , et ce pour tous les ordres n .

Prenons l'exemple de $f(x) = e^x$ et construisons sa série de Maclaurin (centrée en $a = 0$), où $f^{(k)}(0) = 1$ pour tout k .

1. Ordre 0 : Faire correspondre la valeur

Objectif : Le polynôme $P_0(x)$ doit avoir la même valeur que $f(x)$ en $x = 0$. On veut $P_0(0) = f(0)$.

Solution : On choisit le polynôme le plus simple, une constante : $P_0(x) = f(0)$. Pour e^x , $f(0) = 1$, donc $P_0(x) = 1$.

Vérification : $P_0(0) = 1$. L'objectif est atteint.

2. Ordre 1 : Faire correspondre la première dérivée

Objectif : On veut un nouveau polynôme $P_1(x)$ qui préserve la correspondance précédente ($P_1(0) = f(0)$) ET qui a la même pente, c'est-à-dire $P_1'(0) = f'(0)$.

Solution : On ajoute un terme en x à notre polynôme précédent : $P_1(x) = P_0(x) + c_1x = 1 + c_1x$.

Vérification :

- $P_1(0) = 1 + c_1(0) = 1$. La valeur correspond toujours, car le nouveau terme s'annule en 0.
- On dérive : $P_1'(x) = c_1$. Pour que les pentes correspondent en 0, il faut $P_1'(0) = c_1 = f'(0)$. Comme $f'(0) = 1$, on doit choisir $c_1 = 1$.

Notre polynôme est maintenant $P_1(x) = 1 + x$.

3. Ordre 2 : Faire correspondre la deuxième dérivée

Objectif : On veut $P_2(x)$ tel que $P_2(0) = f(0)$, $P_2'(0) = f'(0)$ ET $P_2''(0) = f''(0)$.

Solution : On ajoute un terme en x^2 : $P_2(x) = P_1(x) + c_2x^2 = 1 + x + c_2x^2$.

Vérification :

- Les dérivées d'ordre 0 et 1 en $x = 0$ ne sont pas affectées, car la dérivée de c_2x^2 (soit $2c_2x$) et le terme lui-même s'annulent en 0. Les objectifs précédents sont préservés.
- On dérive deux fois : $P_2'(x) = 1 + 2c_2x$ et $P_2''(x) = 2c_2$.
- Pour que les courbures correspondent, il faut $P_2''(0) = 2c_2 = f''(0)$. Comme $f''(0) = 1$, on doit choisir $c_2 = 1/2$.

Notre polynôme est $P_2(x) = 1 + x + \frac{1}{2}x^2$.

4. Le schéma général : L'importance de la factorielle

Pour faire correspondre la k -ième dérivée, on ajoute un terme c_kx^k .

Quand on dérive c_kx^k exactement k fois, on obtient $c_k \times k!$.

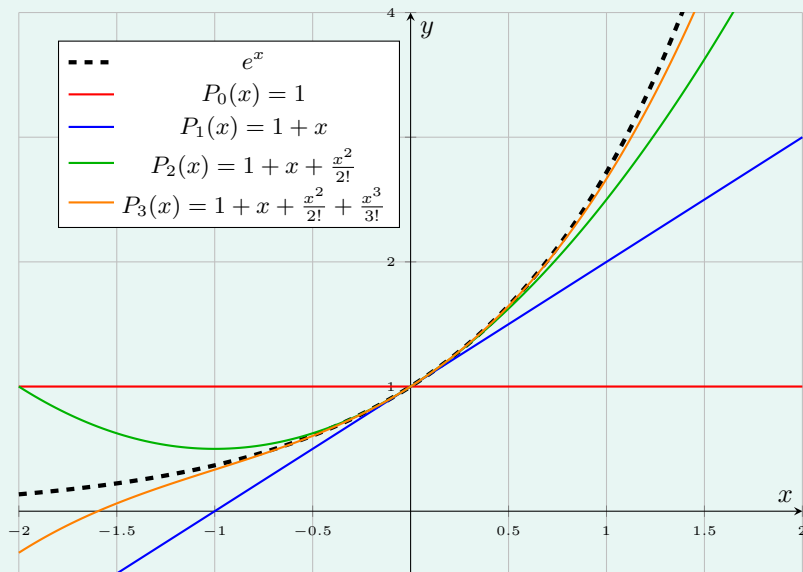
Toutes les dérivées d'ordre inférieur s'annulent en $x = 0$. On doit donc avoir :

$$P_k^{(k)}(0) = c_k \cdot k! = f^{(k)}(0)$$

Cela nous donne la règle pour trouver chaque coefficient :

$$c_k = \frac{f^{(k)}(0)}{k!}$$

C'est précisément le coefficient qui apparaît dans la formule de Taylor, et il est choisi pour cette unique raison : forcer la k -ième dérivée du polynôme à correspondre parfaitement à celle de la fonction au point de développement.



Visualisation de la construction progressive de la série de Maclaurin pour e^x .

10.2 Intuition de la série de Taylor en un point quelconque a

Intuition : Construire une approximation loin de l'origine

La série de Maclaurin est puissante, mais elle nous contraint à approximer une fonction uniquement autour de $x = 0$. Que faire si l'on s'intéresse au comportement d'une fonction ailleurs, par exemple $f(x) = \ln(x)$ autour de $x = 1$ (puisque $\ln(0)$ n'est pas défini) ? C'est là qu'intervient la série de Taylor générale.

L'objectif reste le même : construire un polynôme $P(x)$ qui est une "copie conforme" de $f(x)$ au point a . Pour cela, on force les dérivées du polynôme à correspondre à celles de la fonction en ce point a . La seule différence est que notre "variable" de base n'est plus x , mais l'écart par rapport au centre, c'est-à-dire $(x - a)$.

Prenons l'exemple de $f(x) = \ln(x)$ et construisons sa série centrée en $\mathbf{a} = \mathbf{1}$.

1. Ordre 0 : Faire correspondre la valeur

Objectif : $P_0(a) = f(a)$.

Solution : On calcule $f(1) = \ln(1) = 0$. Le polynôme est la constante $\mathbf{P}_0(\mathbf{x}) = \mathbf{0}$.

2. Ordre 1 : Faire correspondre la pente

Objectif : $P_1(a) = f(a)$ et $P'_1(a) = f'(a)$.

Solution : On ajoute un terme proportionnel à l'écart $(x - a)$: $P_1(x) = f(a) + c_1(x - a)$.

Vérification :

- $P_1(1) = 0 + c_1(1 - 1) = 0$. La valeur correspond.
- On dérive : $P'_1(x) = c_1$. On veut $P'_1(1) = c_1 = f'(1)$.
- La dérivée de $f(x) = \ln(x)$ est $f'(x) = 1/x$, donc $f'(1) = 1$. On doit choisir $\mathbf{c}_1 = \mathbf{1}$.

Notre polynôme est $\mathbf{P}_1(\mathbf{x}) = (\mathbf{x} - \mathbf{1})$. C'est la tangente à $\ln(x)$ en $x = 1$.

3. Ordre 2 : Faire correspondre la courbure

Objectif : Les dérivées jusqu'à l'ordre 2 doivent correspondre en $a = 1$.

Solution : On ajoute un terme en $(x - a)^2$: $P_2(x) = (x - 1) + c_2(x - 1)^2$.

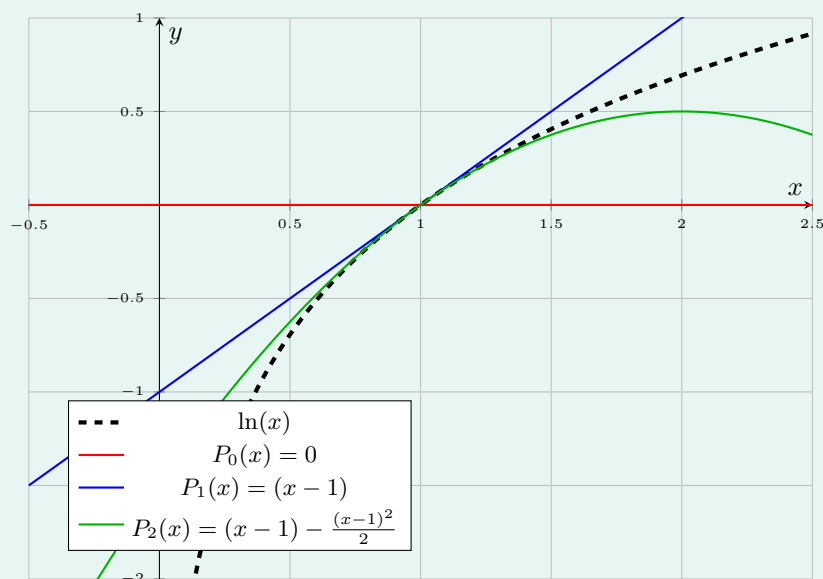
Vérification :

- Les correspondances d'ordre 0 et 1 sont préservées.
- On dérive deux fois : $P'_2(x) = 1 + 2c_2(x - 1)$ et $P''_2(x) = 2c_2$.
- On veut $P''_2(1) = 2c_2 = f''(1)$.
- La dérivée seconde de $f(x)$ est $f''(x) = -1/x^2$, donc $f''(1) = -1$. On choisit $\mathbf{c}_2 = -\mathbf{1}/\mathbf{2}$.

Notre polynôme est $\mathbf{P}_2(\mathbf{x}) = (\mathbf{x} - \mathbf{1}) - \frac{1}{2}(\mathbf{x} - \mathbf{1})^2$.

4. Le schéma général

Le coefficient c_k du terme $(x - a)^k$ est choisi pour faire correspondre la k -ième dérivée. La dérivation de $c_k(x - a)^k$ k fois donne $c_k \cdot k!$. On impose donc $c_k \cdot k! = f^{(k)}(a)$, ce qui mène directement à la formule générale $c_k = \frac{f^{(k)}(a)}{k!}$.



Approximation de $\ln(x)$ autour de $a = 1$. Le polynôme "colle" à la fonction près de $x = 1$.

10.3 La Fonction Exponentielle (e^x)

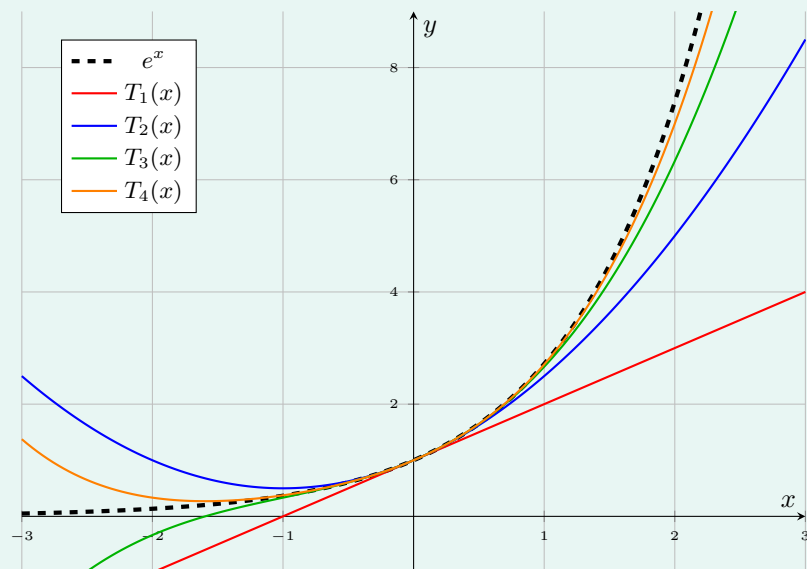
Théorème : Série de Maclaurin pour e^x

Pour tout nombre réel x , la fonction exponentielle peut s'écrire :

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$$

Intuition : Visualiser la Croissance Exponentielle

La fonction exponentielle est unique car elle est sa propre dérivée. Cela signifie que toutes ses informations locales (valeur, pente, courbure) en $a = 0$ sont égales à **1**. La série pour e^x est donc le polynôme le plus « pur », où chaque terme x^k est simplement normalisé par $k!$. Le graphique ci-dessous montre comment les polynômes de Taylor convergent rapidement vers la véritable courbe exponentielle, illustrant sa croissance puissante.



Approximation de e^x par ses polynômes de Maclaurin.

Preuve

Soit $f(x) = e^x$. Pour tout entier $k \geq 0$, la k -ième dérivée est $f^{(k)}(x) = e^x$. En évaluant en $a = 0$, on obtient $f^{(k)}(0) = e^0 = 1$ pour tout k . En appliquant la formule de Maclaurin :

$$e^x = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k = \sum_{k=0}^{\infty} \frac{1}{k!} x^k = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \dots$$

10.4 La Fonction Sinus ($\sin(x)$)

Théorème : Série de Maclaurin pour $\sin(x)$

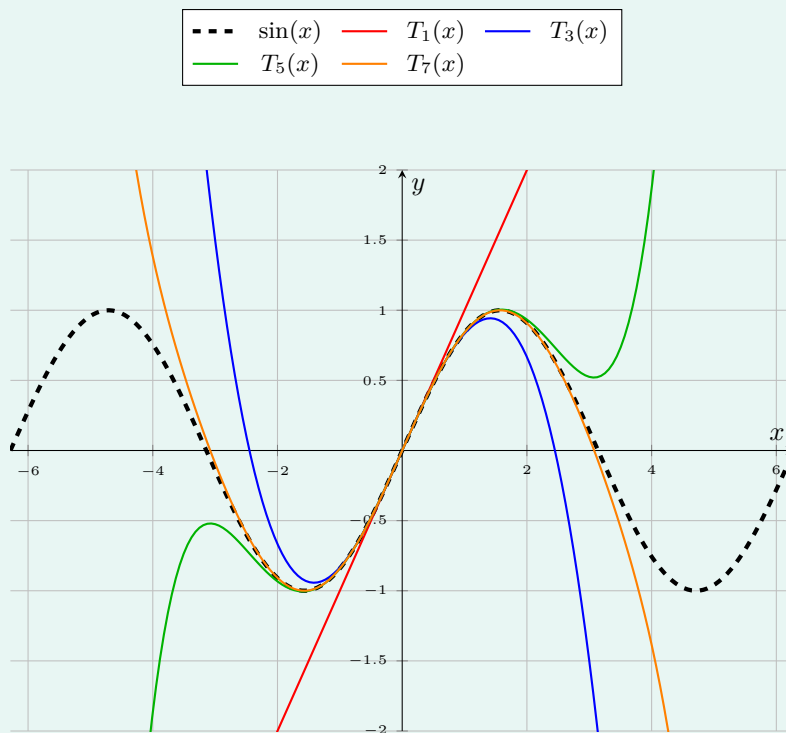
Pour tout nombre réel x :

$$\sin(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots$$

Intuition : Visualiser l'Oscillation du Sinus

La série du sinus reflète ses propriétés fondamentales. En tant que fonction **impair** ($\sin(-x) = -\sin(x)$), son développement ne contient que des puissances **impaires** de

x . Les signes alternés capturent sa nature oscillatoire. Le graphique ci-dessous montre comment l'ajout de termes permet au polynôme d'« épouser » la courbe du sinus sur un plus grand nombre de périodes.



Approximation de $\sin(x)$ par ses polynômes de Maclaurin.

Preuve

Soit $f(x) = \sin(x)$. Les dérivées en $a = 0$ suivent un cycle $(0, 1, 0, -1, \dots)$. Seuls les termes d'ordre impair $(2k+1)$ sont non nuls, avec des valeurs de $(-1)^k$, ce qui donne la formule.

10.5 La Fonction Cosinus ($\cos(x)$)

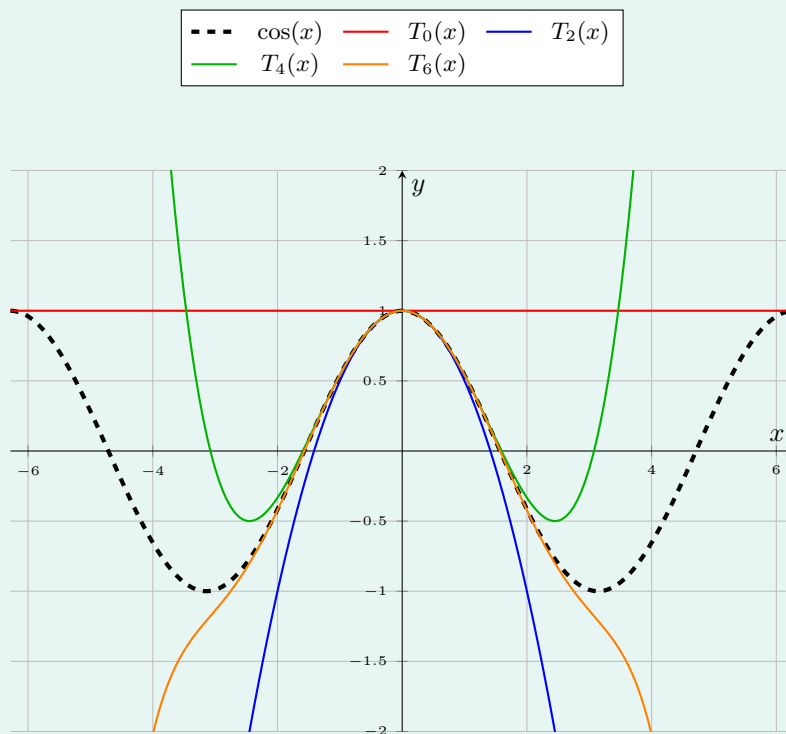
Théorème : Série de Maclaurin pour $\cos(x)$

Pour tout nombre réel x :

$$\cos(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots$$

Intuition : Visualiser la Symétrie du Cosinus

En tant que fonction **paire** ($\cos(-x) = \cos(x)$), la série du cosinus ne contient, de manière appropriée, que des puissances **paire**s de x . Elle commence à 1 (son maximum) puis oscille, un comportement capturé par les signes alternés.



Approximation de $\cos(x)$ par ses polynômes de Maclaurin.

Preuve

Soit $g(x) = \cos(x)$. Les dérivées en $a = 0$ suivent un cycle $(1, 0, -1, 0, \dots)$. Seuls les termes d'ordre pair $(2k)$ sont non nuls, avec des valeurs de $(-1)^k$, ce qui donne la formule.

10.6 Le Logarithme Népérien ($\ln(1+x)$)

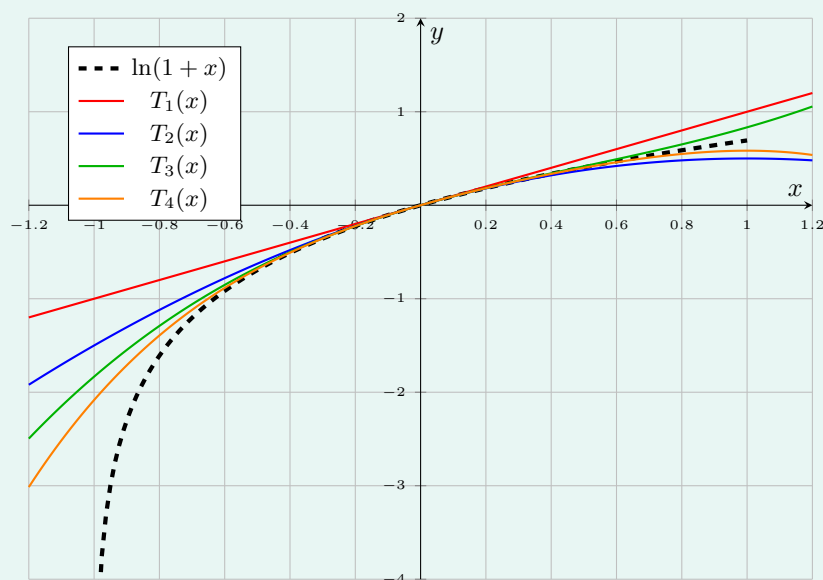
Théorème : Série de Maclaurin pour $\ln(1+x)$

Pour $|x| < 1$:

$$\ln(1+x) = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{x^k}{k} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$$

Intuition : Visualiser l'Approximation Logarithmique

Cette série est essentielle pour approximer les logarithmes près de 1. Contrairement aux fonctions précédentes, elle ne converge que pour $|x| < 1$. Le graphique montre que l'approximation est excellente près de $x = 0$ mais diverge rapidement lorsque x s'approche de la frontière de convergence à $x = 1$.



Approximation de $\ln(1+x)$ par ses polynômes de Maclaurin.

Preuve

Soit $f(x) = \ln(1+x)$. Pour $k \geq 1$, la k -ième dérivée en $a = 0$ est $f^{(k)}(0) = (-1)^{k-1}(k-1)!$. En substituant cela dans la formule de Maclaurin, le $(k-1)!$ au numérateur annule partiellement le $k!$ au dénominateur, laissant un k en bas.

10.7 La Série Géométrique ($\frac{1}{1-x}$)

Théorème : Série de Maclaurin pour $\frac{1}{1-x}$

Pour $|x| < 1$:

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k = 1 + x + x^2 + x^3 + \dots$$

Intuition : Le Fondement de Nombreuses Séries

Cette série, connue sous le nom de série géométrique, est l'un des développements en série de puissances les plus fondamentaux. Elle converge uniquement lorsque la valeur absolue de x est inférieure à 1. Chaque coefficient est simplement 1, ce qui en fait la série de Maclaurin la plus simple. De nombreuses autres séries, comme celle de $\ln(1+x)$ ou de $\arctan(x)$, peuvent être dérivées de celle-ci par intégration ou substitution.

Preuve

Soit $f(x) = (1-x)^{-1}$. Les dérivées successives sont $f'(x) = 1(1-x)^{-2}$, $f''(x) = 2(1-x)^{-3}$, $f'''(x) = 6(1-x)^{-4}$, et ainsi de suite. La formule générale pour la k -ième dérivée est $f^{(k)}(x) = k!(1-x)^{-(k+1)}$. En évaluant en $a = 0$, on obtient $f^{(k)}(0) = k!$. En substituant dans la formule de Maclaurin :

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k = \sum_{k=0}^{\infty} \frac{k!}{k!} x^k = \sum_{k=0}^{\infty} x^k$$