

Mes Notes de Lecture

Introduction à la Probabilité

LOU BRUNET

4 novembre 2025

Table des matières

Sommaire	1
1 Probabilités et Dénombrement	4
1.1 Concepts fondamentaux	4
1.2 Définition Formelle (Axiomatique) de la Probabilité	4
1.3 Définition Naïve de la Probabilité	4
1.4 Permutations (Arrangements)	5
1.5 Le Coefficient Binomial	5
1.6 Identité de Vandermonde	8
1.7 Bose-Einstein (Étoiles et Bâtons)	9
1.8 Principe d'Inclusion-Exclusion	9
1.9 Exercices	12
1.10 Corrections des Exercices	14
1.11 Exercices Python	17
2 Probabilité conditionnelle	19
2.1 Définition de la Probabilité Conditionnelle	19
2.2 Règle du Produit (Intersection de deux événements)	19
2.3 Règle de la Chaîne (Intersection de n événements)	20
2.4 Règle de Bayes	21
2.5 Formule des Probabilités Totales	21
2.6 Règle de Bayes avec Conditionnement Additionnel	22
2.7 Formule des Probabilités Totales avec Conditionnement Additionnel	23
2.8 Indépendance de Deux Événements	25
2.9 Indépendance Conditionnelle	25
2.10 Le Problème de Monty Hall	25
2.11 Exercices	27
2.12 Corrections des Exercices	29
2.13 Exercices Python	32
3 Variables Aléatoires Discrètes	34
3.1 Variable Aléatoire	34
3.2 Variable Aléatoire Discrète	34
3.3 Fonction de Masse (PMF)	34
3.4 Loi de Bernoulli	35
3.5 Loi Binomiale	35
3.6 Loi Hypergéométrique	36
3.7 Loi Géométrique	37
3.8 Loi de Poisson	38
3.9 Fonction de Répartition (CDF)	40
3.10 Variable Aléatoire Indicatrice	41
3.11 Espérance d'une variable aléatoire discrète	41
3.12 Linéarité de l'espérance	42
3.13 Espérance de la loi binomiale	43
3.14 Espérance de la loi géométrique	43
3.15 Loi du statisticien inconscient (LOTUS)	44
3.16 Variance	45
3.17 Exercices	46
3.18 Corrections des Exercices	49
3.19 Exercices Python	52
4 Variables Aléatoires Continues	55
4.1 Fonction de Densité de Probabilité (PDF)	55
4.2 Fonction de Répartition (CDF)	55
4.3 Espérance et Variance (Cas Continu)	56
4.4 Loi Uniforme	58
4.5 Loi Exponentielle	59

4.6	Distributions Conjointes (Cas Continu)	61
4.7	Espérance, Indépendance et Covariance (Cas Conjoint)	62
4.8	Espérance d'une variable aléatoire continue	63
4.9	Linéarité de l'espérance	63
4.10	Loi du statisticien inconscient (LOTUS)	64
4.11	Variance	65
4.12	Changement de Variable pour les Variables Aléatoires Continues	66
4.13	Exercices	68
4.14	Correction	70
4.15	Exercices Python	72
5	La Loi Normale (ou Gaussienne)	75
5.1	Introduction et Fonction de Densité (PDF)	75
5.2	La Loi Normale Centrée Réduite $\mathcal{N}(0, 1)$	78
5.3	Standardisation : Le Score Z	78
5.4	Propriétés Importantes de la Loi Normale	80
5.5	La Règle Empirique (68-95-99.7)	81
5.6	Calcul de Probabilités Normales	82
5.7	Exercices	82
5.8	Corrections des Exercices	84
5.9	Exercices Python	86
6	La Loi Log-Normale	90
6.1	Introduction et Définition	90
6.2	Intuition : Des Sommes aux Produits	90
6.3	Fonction de Densité (PDF)	91
6.4	Comment obtenir l'espérance d'une variable log-normale	92
6.5	Propriétés : Moments et Mesures Centrales	94
6.6	Calcul de Probabilités	94
6.7	Évolution d'un Actif et Émergence de la Loi Log-Normale	95
6.8	Exercices	98
6.9	Corrections des Exercices	100
6.10	Exercices Python	102
7	Distributions Multivariées et Concepts Associés	105
7.1	Distributions Jointes et Marginales	105
7.2	Espérance d'une fonction de deux variables	106
7.3	Covariance et Corrélation	106
7.4	Linéarité de la Covariance	108
7.5	Résultats sur la Corrélation	108
7.6	Standardisation et Non-Corrélation	109
7.7	Variance d'une Somme de Variables Aléatoires	111
7.8	Théorème sur la somme de lois de Poisson	112
7.9	Exercices	113
7.10	Corrections des Exercices	115
7.11	Exercices Python	118
8	Moments d'une distribution	121
8.1	Définitions fondamentales des moments	121
8.2	Asymétrie (Skewness)	121
8.3	Propriétés de symétrie	122
8.4	Aplatissement (Kurtosis)	122
8.5	Exemples de distributions	123
8.6	Moments d'échantillon (Sample Moments)	125
8.7	Fonctions génératrices des moments (MGF)	126
8.8	Génération des moments via les MGF	127
8.9	Sommes de variables aléatoires indépendantes via les MGF	127
8.10	Exercices	128
8.11	Corrections des Exercices	130

8.12	Exercices Python	132
9	Les Loïs des Grands Nombres (LLN)	136
9.1	L'Inégalité de Chebyshev	136
9.2	La Loi Faible des Grands Nombres (LFGN / WLLN)	138
9.3	La Loi Forte des Grands Nombres (LFGN / SLLN)	139
9.4	Différence : Faible vs. Forte	139
9.5	Application : La Méthode de Monte-Carlo	139
9.6	Exercices	140
9.7	Corrections des Exercices	143
9.8	Exercices Python	145
10	Le Théorème Central Limite (TCL)	148
10.1	Introduction : L'omniprésence de la loi normale	148
10.2	L'illustration : la somme des "Pile ou Face"	148
10.3	Distribution de la population vs. Distribution d'échantillonnage	149
10.4	Énoncé formel du Théorème Central Limite	149
10.5	Applications Pratiques du TCL	150
10.6	Exercices	153
10.7	Corrections des Exercices	155
10.8	Exercices Python	157

1 Probabilités et Dénombrement

1.1 Concepts fondamentaux

Avant de pouvoir calculer des probabilités, il est essentiel d'établir un vocabulaire commun pour décrire les expériences aléatoires.

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.

Cette intuition se traduit formellement par deux définitions clés :

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.

Un exemple simple permet de solidifier ces concepts :

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 $A = \{1, 3, 5\}$.

1.2 Définition Formelle (Axiomatique) de la Probabilité

Maintenant que nous avons défini les événements, nous pouvons définir formellement les règles que doit suivre une fonction de probabilité. L'image que vous avez fournie contient cette définition axiomatique fondamentale.

Définition : Fonction de Probabilité (Axiomes)

Une **fonction de probabilité** P sur un univers (espace échantillon) fini S est une fonction qui associe à chaque événement $A \subseteq S$ un nombre $P(A) \in [0, 1]$ tel que :

1. $P(S) = 1$, et
2. $P(A \cup B) = P(A) + P(B)$ si A et B sont disjoints (c'est-à-dire $A \cap B = \emptyset$).

Le nombre $P(A)$ est appelé la **probabilité** que A se produise.

Cette définition est très générale. Le premier cas que nous allons étudier, la "définition naïve", est l'application la plus simple de ces axiomes, valable dans le cas particulier de l'équiprobabilité.

1.3 Définition Naïve de la Probabilité

Pour de nombreuses expériences simples, comme lancer un dé non pipé, chaque résultat possible est "équiprobable". Cette hypothèse est la base de la première définition formelle 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 :

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

Appliquons cette formule à quelques cas classiques :

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 $P(C) = \frac{4}{52} = \frac{1}{13}$.

1.4 Permutations (Arrangements)

Le dénombrement, qui est l'art de compter les tailles $|A|$ et $|S|$, est fondamental pour appliquer la définition naïve. Le premier outil que nous verrons est la permutation, qui compte les arrangements ordonnés.

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 $0! = 1$.

Cette formule peut sembler abstraite, mais elle provient d'un raisonnement logique simple par "cases" :

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 $(n-k) \times (n-k-1) \times \cdots \times 1$.

Voyons une application classique de ce principe :

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.

1.5 Le Coefficient Binomial

Que se passe-t-il si l'ordre ne compte pas ? Au lieu de compter des podiums, nous voulons compter des comités. C'est le rôle du 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 :

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

La preuve de cette formule repose sur un argument combinatoire élégant : nous allons compter la même chose (les permutations) de deux façons différentes.

Preuve

Considérons le nombre de permutations de k objets parmi n , noté $P(n, k)$.

1. **Méthode 1** : Par définition (vue ci-dessus), nous savons que $P(n, k) = \frac{n!}{(n-k)!}$.
2. **Méthode 2** : Nous pouvons construire une telle permutation en deux étapes successives :
 - D'abord, **choisir un sous-ensemble** de k objets parmi n (l'ordre ne compte pas). C'est le nombre que nous cherchons, notons-le $\binom{n}{k}$.
 - Ensuite, **ordonner** ces k objets choisis. Il y a $k!$ façons de les arranger.

Le nombre total de permutations est donc le produit de ces étapes : $P(n, k) = \binom{n}{k} \times k!$.

En égalisant les deux méthodes, on obtient :

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

En divisant par $k!$, on trouve bien la formule :

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

L'intuition visuelle derrière cette preuve est de voir comment chaque "choix" (une colonne du tableau) génère $k!$ "ordres" (les lignes de cette colonne).

Intuition

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) :

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

L'application la plus directe est le tirage d'un groupe où l'ordre n'importe pas :

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.

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

1.6 Identité de Vandermonde

Les coefficients binomiaux obéissent à de nombreuses identités. L'identité de Vandermonde est l'une des plus utiles, car elle montre comment décomposer un problème de comptage complexe en sous-problèmes.

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 :

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

La preuve la plus intuitive est une "preuve par l'histoire" (proof by story), qui consiste à trouver un scénario de dénombrement que les deux côtés de l'équation résolvent.

Preuve : Preuve combinatoire

Imaginons un groupe composé de m hommes et n femmes. Nous souhaitons former un comité de k personnes. Nous allons compter le nombre de comités possibles de deux façons.

Côté gauche : $\binom{m+n}{k}$ Le groupe total contient $m+n$ personnes. Le nombre de façons de choisir un comité de k personnes parmi ce total est, par définition, $\binom{m+n}{k}$.

Côté droit : $\sum_{j=0}^k \binom{m}{j} \binom{n}{k-j}$ Nous pouvons compter le même nombre en conditionnant sur le nombre d'hommes (noté j) dans le comité. Un comité de k personnes doit contenir j hommes ET $k-j$ femmes, où j peut aller de 0 à k .

- Pour $j = 0$: Choisir 0 homme ($\binom{m}{0}$) ET k femmes ($\binom{n}{k}$).
- Pour $j = 1$: Choisir 1 homme ($\binom{m}{1}$) ET $k-1$ femmes ($\binom{n}{k-1}$).
- ...
- Pour $j = k$: Choisir k hommes ($\binom{m}{k}$) ET 0 femme ($\binom{n}{0}$).

Puisque ces cas (0 homme, 1 homme, etc.) sont mutuellement exclusifs, le nombre total de comités est la somme de toutes ces possibilités :

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

Puisque les deux côtés comptent exactement la même chose (le nombre total de comités), ils doivent être égaux.

Vérifions cette identité avec un exemple numérique concret, en reprenant l'analogie du comité :

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.

1.7 Bose-Einstein (Étoiles et Bâtons)

Jusqu'à présent, nous avons supposé un "tirage sans remise". La statistique de Bose-Einstein, ou plus visuellement la méthode des "étoiles et bâtons", s'attaque au problème du **tirage avec remise** où l'ordre ne compte pas.

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 :

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

La preuve de cette formule est l'un des résultats les plus élégants du dénombrement. L'astuce consiste à transformer le problème de distribution en un problème d'arrangement de symboles.

Preuve : Par les "Étoiles et Bâtons"

Nous cherchons à distribuer k objets indiscernables (\star) dans n boîtes discernables. Nous pouvons représenter n'importe quelle distribution comme une séquence de symboles. Nous avons besoin de k étoiles (les objets) et de $n-1$ bâtons ($|$) pour servir de séparateurs entre les n boîtes.

Par exemple, pour distribuer $k=7$ étoiles dans $n=4$ boîtes, la séquence :

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

correspond à : 3 étoiles dans la boîte 1, 1 étoile dans la boîte 2, 0 étoile dans la boîte 3 (l'espace entre deux bâtons), et 3 étoiles dans la boîte 4.

Chaque arrangement unique de ces symboles correspond à une distribution unique. Le problème revient donc à trouver le nombre de façons d'arranger ces k étoiles et ces $n-1$ bâtons.

Nous avons un total de $n+k-1$ positions à remplir. Le nombre de façons de le faire est simplement le nombre de manières de choisir les k positions pour les étoiles (les autres positions étant automatiquement remplies par des bâtons). C'est exactement :

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

(Ce qui est aussi égal à $\binom{n+k-1}{n-1}$, le nombre de façons de choisir les positions des $n-1$ bâtons).

C'est la méthode parfaite pour tout problème de distribution d'objets identiques :

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.

1.8 Principe d'Inclusion-Exclusion

Comment compter le nombre d'éléments dans l'union de plusieurs ensembles ? Si on additionne simplement leurs tailles, on compte les intersections plusieurs fois. Le principe d'inclusion-exclusion corrige systématiquement ce sur-comptage.

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 :

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

La preuve pour 3 ensembles se fait en appliquant la formule pour 2 ensembles de manière répétée.

Preuve

Nous utilisons la formule pour deux ensembles, $|X \cup Y| = |X| + |Y| - |X \cap Y|$, de manière imbriquée. Posons $X = A \cup B$ et $Y = C$.

$$\begin{aligned} |A \cup B \cup C| &= |(A \cup B) \cup C| \\ &= |A \cup B| + |C| - |(A \cup B) \cap C| \end{aligned}$$

Nous devons maintenant développer les deux termes compliqués :

1. $|A \cup B| = |A| + |B| - |A \cap B|$
2. Par distributivité de l'intersection sur l'union, $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$.

Appliquons la formule pour 2 ensembles à ce deuxième terme :

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

Ce qui se simplifie en $|A \cap C| + |B \cap C| - |A \cap B \cap C|$.

Finalement, en substituant tout dans l'équation de départ :

$$\begin{aligned} |A \cup B \cup C| &= \underbrace{(|A| + |B| - |A \cap B|)}_{|A \cup B|} + |C| \\ &\quad - \underbrace{(|A \cap C| + |B \cap C| - |A \cap B \cap C|)}_{|(A \cup B) \cap C|} \end{aligned}$$

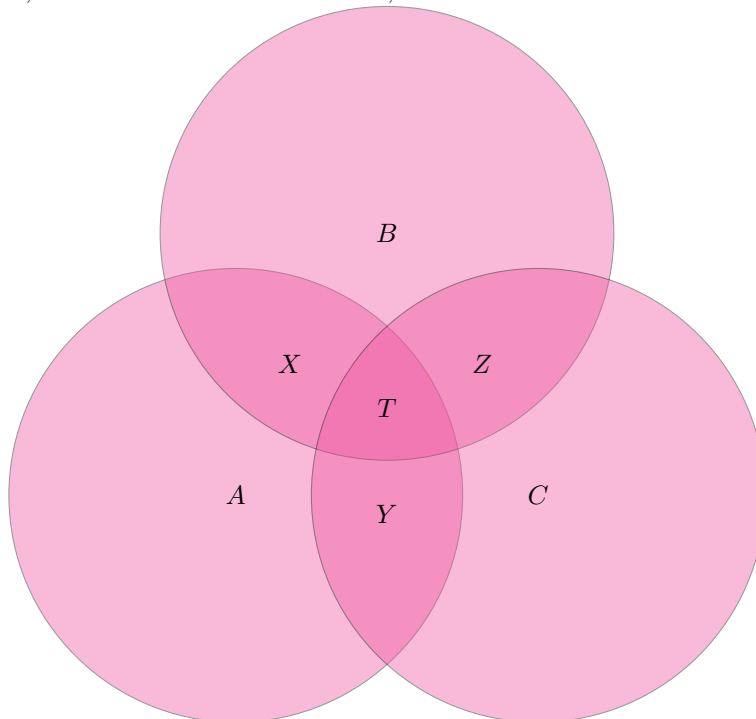
En réarrangeant les termes, on obtient la formule voulue :

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

La formule devient évidente lorsque l'on utilise un diagramme de Venn pour visualiser le sur-comptage et sa correction.

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

Ce que nous avons fait visuellement pour 3 ensembles peut être généralisé par récurrence à n ensembles. La formule générale suit le même principe d'alternance des signes :

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 :

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

La preuve formelle que cette formule gigantesque fonctionne est fascinante. Il suffit de montrer que n'importe quel élément x de l'union, peu importe à combien d'ensembles il appartient, est compté **exactement une fois** au final.

Preuve : Preuve par comptage d'un élément

Considérons un élément x qui appartient à exactement k ensembles parmi les n ensembles A_1, \dots, A_n (où $k \geq 1$). Nous devons montrer que x est compté exactement 1 fois par la formule.

Analysons combien de fois x est compté dans chaque somme de 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** ($-\sum |A_i \cap A_j|$) : x est compté (et soustrait) pour chaque *paire* d'ensembles auxquels il appartient. Comme il appartient à k ensembles, il y a $\binom{k}{2}$ telles paires.
- **Troisième somme** ($+\sum |A_i \cap A_j \cap A_k|$) : x est ajouté 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 :

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

Pour évaluer cette somme, rappelons l'identité fondamentale du binôme de Newton :

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

Si nous posons $x = -1$, nous obtenons :

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

Sachant que $\binom{k}{0} = 1$, on a :

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

En réarrangeant, on trouve :

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

Cela prouve que n'importe quel élément de l'union est compté exactement une fois.

Ce principe est très utile en probabilité, car il permet de calculer $P(A \cup B \cup \dots)$ en se basant sur les probabilités des intersections, qui sont souvent plus faciles à trouver.

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 $1 - \frac{125}{216} = \frac{91}{216}$.

1.9 Exercices

Cette série d'exercices vise à renforcer votre compréhension des concepts fondamentaux du dénombrement et de la probabilité naïve. La difficulté augmente progressivement.

Exercice 1 : Univers et Événements

On lance deux dés à 6 faces, un rouge et un bleu.

1. Décrivez l'univers S de cette expérience. Quelle est sa taille $|S|$?
2. Soit A l'événement "la somme des dés est égale à 7". Listez les issues appartenant à A . Calculez $P(A)$.
3. Soit B l'événement "le dé rouge montre un 3". Listez les issues appartenant à B . Calculez

$P(B)$.

4. Décrivez l'événement $A \cap B$ et calculez sa probabilité.

Exercice 2 : Tirage de Cartes (Prob. Naïve)

On tire une carte au hasard d'un jeu standard de 52 cartes.

1. Quelle est la probabilité de tirer un Roi ?
2. Quelle est la probabilité de tirer une carte rouge (Cœur ou Carreau) ?
3. Quelle est la probabilité de tirer une figure (Valet, Dame, Roi) ?
4. Quelle est la probabilité de tirer un As rouge ?

Exercice 3 : Urne Simple (Prob. Naïve)

Une urne contient 5 boules rouges, 3 boules bleues et 2 boules vertes. On tire une boule au hasard.

1. Quelle est la probabilité qu'elle soit bleue ?
2. Quelle est la probabilité qu'elle ne soit pas verte ?

Exercice 4 : Anagrammes (Permutation Simple)

Combien d'anagrammes distinctes peut-on former avec les lettres du mot "MATHS" ?

Exercice 5 : Course (Arrangement)

Dix athlètes participent à une course. Combien y a-t-il de classements possibles pour les 3 premières places (médaillon d'or, d'argent, de bronze) ?

Exercice 6 : Anagrammes (Permutation avec Répétition)

Combien d'anagrammes distinctes peut-on former avec les lettres du mot "PROBABILITE" ?

Exercice 7 : Choix d'un Comité (Combinaison)

Une classe compte 15 étudiants. De combien de manières peut-on choisir un comité de 4 étudiants ?

Exercice 8 : Mains de Poker (Combinaison)

Dans un jeu de 52 cartes, combien de "mains" de 5 cartes différentes peut-on former ?

Exercice 9 : Comité Mixte (Combinaison)

À partir d'un groupe de 6 hommes et 4 femmes, combien de comités de 3 personnes peut-on former contenant exactement 2 hommes et 1 femme ?

Exercice 10 : Probabilité avec Combinaisons

On tire simultanément 3 cartes d'un jeu de 52 cartes. Quelle est la probabilité d'obtenir exactement 2 Rois ?

Exercice 11 : Distribution de Bonbons (Étoiles et Bâtons)

De combien de manières peut-on distribuer 8 bonbons identiques à 3 enfants ? (Certains enfants peuvent ne rien recevoir).

Exercice 12 : Solutions d'Équation (Étoiles et Bâtons)

Combien y a-t-il de solutions entières non négatives ($x_i \geq 0$) à l'équation $x_1 + x_2 + x_3 + x_4 = 10$?

Exercice 13 : Distribution avec Minimum (Étoiles et Bâtons avec Contrainte)

De combien de manières peut-on distribuer 12 pommes identiques à 4 enfants, si chaque enfant doit recevoir au moins une pomme ?

Exercice 14 : Divisibilité (Inclusion-Exclusion 2 Ensembles)

Parmi les entiers de 1 à 100, combien sont divisibles par 2 OU par 3 ?

Exercice 15 : Langues (Inclusion-Exclusion 2 Ensembles)

Dans un groupe de 50 étudiants, 30 étudient l'anglais, 25 étudient l'espagnol et 10 étudient les deux langues. Combien d'étudiants étudient au moins une de ces deux langues ? Combien n'en étudient aucune ?

Exercice 16 : Divisibilité (Inclusion-Exclusion 3 Ensembles)

Parmi les entiers de 1 à 100, combien sont divisibles par 2, 3 OU 5 ?

Exercice 17 : Chemins sur un Grillage (Combinaison)

Sur un grillage, combien y a-t-il de chemins pour aller du point (0,0) au point (4,3) en se déplaçant uniquement vers la droite (D) ou vers le haut (H) ?

Exercice 18 : Probabilité Hypergéométrique

Une urne contient 7 boules blanches et 5 boules noires. On tire successivement et sans remise 4 boules. Quelle est la probabilité d'obtenir 2 blanches et 2 noires ?

Exercice 19 : Arrangement Circulaire

De combien de manières 6 personnes peuvent-elles s'asseoir autour d'une table ronde ? (Deux arrangements sont considérés identiques si chaque personne a les mêmes voisins).

Exercice 20 : Problème des Dérangements (Inclusion-Exclusion)

Quatre lettres sont adressées à quatre personnes différentes, avec les enveloppes correspondantes. On met chaque lettre au hasard dans une enveloppe. Quelle est la probabilité qu'aucune lettre ne soit dans la bonne enveloppe ?

1.10 Corrections des Exercices

Correction Exercice 1 : Univers et Événements

- 1) L'univers S est l'ensemble de toutes les paires (r, b) où r est le résultat du dé rouge et b celui du dé bleu. $S = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), \dots, (6, 6)\}$. La taille de l'univers est $|S| = 6 \times 6 = 36$.
- 2) L'événement A (somme égale à 7) est $A = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$. Il y a $|A| = 6$ issues favorables. La probabilité est $P(A) = |A|/|S| = 6/36 = 1/6$.
- 3) L'événement B (dé rouge montre 3) est $B = \{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\}$. Il y a $|B| = 6$ issues favorables. La probabilité est $P(B) = |B|/|S| = 6/36 = 1/6$.
- 4) L'événement $A \cap B$ est l'ensemble des issues où la somme est 7 ET le dé rouge est 3. La seule issue possible est $(3, 4)$. Donc $A \cap B = \{(3, 4)\}$. La probabilité est $P(A \cap B) = |A \cap B|/|S| = 1/36$.

Correction Exercice 2 : Tirage de Cartes (Prob. Naïve)

Le nombre total d'issues est $|S| = 52$.

- a) Il y a 4 Rois. $P(\text{Roi}) = 4/52 = 1/13$.
- b) Il y a 26 cartes rouges (13 Cœurs + 13 Carreaux). $P(\text{Rouge}) = 26/52 = 1/2$.
- c) Il y a 12 figures (4 Valets + 4 Dames + 4 Rois). $P(\text{Figure}) = 12/52 = 3/13$.
- d) Il y a 2 As rouges (As de Cœur, As de Carreau). $P(\text{As Rouge}) = 2/52 = 1/26$.

Correction Exercice 3 : Urne Simple (Prob. Naïve)

Le nombre total de boules est $5 + 3 + 2 = 10$.

- a) Il y a 3 boules bleues. $P(\text{Bleue}) = 3/10$.
- b) L'événement "ne pas être verte" est le complémentaire de "être verte". Il y a 2 boules vertes, donc $P(\text{Verte}) = 2/10$. La probabilité cherchée est $P(\text{Non Verte}) = 1 - P(\text{Verte}) = 1 - 2/10 = 8/10 = 4/5$. (Alternativement, il y a $5 + 3 = 8$ boules non vertes, donc $P = 8/10$).

Correction Exercice 4 : Anagrammes (Permutation Simple)

Le mot "MATHS" a 5 lettres distinctes. Le nombre d'anagrammes est le nombre de permutations de ces 5 lettres, soit $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$.

Correction Exercice 5 : Course (Arrangement)

On cherche le nombre de façons d'ordonner 3 athlètes parmi 10. C'est un arrangement (permutation de k parmi n) : $P(10, 3) = \frac{10!}{(10-3)!} = \frac{10!}{7!} = 10 \times 9 \times 8 = 720$. Il y a 720 podiums possibles.

Correction Exercice 6 : Anagrammes (Permutation avec Répétition)

Le mot "PROBABILITE" a 11 lettres. Les répétitions sont : B (2 fois), I (2 fois). Les autres lettres (P, R, O, A, L, T, E) apparaissent une fois. Le nombre d'anagrammes distinctes est :

$$\frac{11!}{2! \times 2!} = \frac{39,916,800}{2 \times 2} = \frac{39,916,800}{4} = 9,979,200$$

Correction Exercice 7 : Choix d'un Comité (Combinaison)

L'ordre ne compte pas, c'est donc une combinaison de 4 étudiants parmi 15 :

$$\binom{15}{4} = \frac{15!}{4!(15-4)!} = \frac{15!}{4!11!} = \frac{15 \times 14 \times 13 \times 12}{4 \times 3 \times 2 \times 1} = 15 \times 7 \times 13 \times 1 = 1365$$

Il y a 1365 comités possibles.

Correction Exercice 8 : Mains de Poker (Combinaison)

On choisit 5 cartes parmi 52, sans ordre. C'est une combinaison :

$$\binom{52}{5} = \frac{52!}{5!(52-5)!} = \frac{52!}{5!47!} = \frac{52 \times 51 \times 50 \times 49 \times 48}{5 \times 4 \times 3 \times 2 \times 1} = 2,598,960$$

Il y a 2,598,960 mains de poker possibles.

Correction Exercice 9 : Comité Mixte (Combinaison)

Il faut choisir 2 hommes parmi 6 ET 1 femme parmi 4. On multiplie les possibilités pour chaque choix : Nombre de façons = (choix des hommes) \times (choix des femmes)

$$= \binom{6}{2} \times \binom{4}{1} = \frac{6 \times 5}{2 \times 1} \times \frac{4}{1} = 15 \times 4 = 60$$

Il y a 60 comités possibles.

Correction Exercice 10 : Probabilité avec Combinaisons

L'univers S est l'ensemble de toutes les mains de 3 cartes. $|S| = \binom{52}{3}$. L'événement A est "obtenir exactement 2 Rois". Pour cela, il faut choisir 2 Rois parmi les 4 Rois ET 1 carte qui n'est pas un Roi parmi les 48 autres cartes. $|A| = \binom{4}{2} \times \binom{48}{1}$. La probabilité est $P(A) = \frac{|A|}{|S|} = \frac{\binom{4}{2}\binom{48}{1}}{\binom{52}{3}}$.

$$P(A) = \frac{\frac{4 \times 3}{2 \times 1} \times 48}{\frac{52 \times 51 \times 50}{3 \times 2 \times 1}} = \frac{6 \times 48}{22100} = \frac{288}{22100} \approx 0.013$$

Correction Exercice 11 : Distribution de Bonbons (Étoiles et Bâtons)

C'est un problème de distribution de $k = 8$ objets identiques (bonbons) dans $n = 3$ boîtes distinctes (enfants). On utilise la formule $\binom{n+k-1}{k}$. Nombre de manières = $\binom{3+8-1}{8} = \binom{10}{8} = \binom{10}{2} = \frac{10 \times 9}{2 \times 1} = 45$.

Correction Exercice 12 : Solutions d'Équation (Étoiles et Bâtons)

Cela revient à distribuer $k = 10$ unités identiques dans $n = 4$ variables distinctes. Nombre de solutions = $\binom{n+k-1}{k} = \binom{4+10-1}{10} = \binom{13}{10} = \binom{13}{3} = \frac{13 \times 12 \times 11}{3 \times 2 \times 1} = 286$.

Correction Exercice 13 : Distribution avec Minimum (Étoiles et Bâtons avec Contrainte)

On doit distribuer $k = 12$ pommes à $n = 4$ enfants, avec $x_i \geq 1$. On commence par donner une pomme à chaque enfant. Il reste $12 - 4 = 8$ pommes à distribuer sans contrainte (les x'_i peuvent être nuls). Le problème devient : distribuer $k' = 8$ pommes à $n = 4$ enfants. Nombre de manières $= \binom{n+k'-1}{k'} = \binom{4+8-1}{8} = \binom{11}{8} = \binom{11}{3} = \frac{11 \times 10 \times 9}{3 \times 2 \times 1} = 165$.

Correction Exercice 14 : Divisibilité (Inclusion-Exclusion 2 Ensembles)

Soit A l'ensemble des entiers ≤ 100 divisibles par 2, et B l'ensemble des entiers ≤ 100 divisibles par 3. On cherche $|A \cup B|$. $|A| = \lfloor 100/2 \rfloor = 50$. $|B| = \lfloor 100/3 \rfloor = 33$. $|A \cap B|$ = ensemble des entiers divisibles par $2 \times 3 = 6$. $|A \cap B| = \lfloor 100/6 \rfloor = 16$. Par inclusion-exclusion : $|A \cup B| = |A| + |B| - |A \cap B| = 50 + 33 - 16 = 67$.

Correction Exercice 15 : Langues (Inclusion-Exclusion 2 Ensembles)

Soit E l'ensemble des étudiants étudiant l'anglais, S l'ensemble de ceux étudiant l'espagnol. $|E| = 30$, $|S| = 25$, $|E \cap S| = 10$. Nombre d'étudiants étudiant au moins une langue : $|E \cup S| = |E| + |S| - |E \cap S| = 30 + 25 - 10 = 45$. Nombre total d'étudiants = 50. Nombre d'étudiants n'étudiant aucune de ces langues = Total - $|E \cup S| = 50 - 45 = 5$.

Correction Exercice 16 : Divisibilité (Inclusion-Exclusion 3 Ensembles)

Soit A_2, A_3, A_5 les ensembles des entiers ≤ 100 divisibles respectivement par 2, 3, 5. On cherche $|A_2 \cup A_3 \cup A_5|$. $|A_2| = 50$, $|A_3| = 33$, $|A_5| = 20$. $|A_2 \cap A_3| = |A_6| = \lfloor 100/6 \rfloor = 16$. $|A_2 \cap A_5| = |A_{10}| = \lfloor 100/10 \rfloor = 10$. $|A_3 \cap A_5| = |A_{15}| = \lfloor 100/15 \rfloor = 6$. $|A_2 \cap A_3 \cap A_5| = |A_{30}| = \lfloor 100/30 \rfloor = 3$. Par inclusion-exclusion : $|A_2 \cup A_3 \cup A_5| = (|A_2| + |A_3| + |A_5|) - (|A_2 \cap A_3| + |A_2 \cap A_5| + |A_3 \cap A_5|) + |A_2 \cap A_3 \cap A_5| = (50 + 33 + 20) - (16 + 10 + 6) + 3 = 103 - 32 + 3 = 74$.

Correction Exercice 17 : Chemins sur un Grillage (Combinaison)

Pour aller de $(0,0)$ à $(4,3)$, il faut faire un total de $4 + 3 = 7$ déplacements. Parmi ces 7 déplacements, il faut choisir les 4 moments où l'on va à droite (les 3 autres seront obligatoirement vers le haut), ou choisir les 3 moments où l'on va vers le haut. Le nombre de chemins est $\binom{7}{4} = \binom{7}{3} = \frac{7 \times 6 \times 5}{3 \times 2 \times 1} = 35$.

Correction Exercice 18 : Probabilité Hypergéométrique

C'est un tirage sans remise. On peut utiliser la loi hypergéométrique ou le dénombrement. Population totale = $7 + 5 = 12$ boules. On en tire $m = 4$. On veut $k = 2$ blanches (parmi $w = 7$) et $m - k = 2$ noires (parmi $b = 5$). Probabilité = $\frac{\binom{w}{k} \binom{b}{m-k}}{\binom{w+b}{m}} = \frac{\binom{7}{2} \binom{5}{2}}{\binom{12}{4}}$.

$$P = \frac{\binom{7 \times 6}{2} \times \binom{5 \times 4}{2}}{\binom{12 \times 11 \times 10 \times 9}{4 \times 3 \times 2 \times 1}} = \frac{21 \times 10}{495} = \frac{210}{495} = \frac{14}{33} \approx 0.424$$

Correction Exercice 19 : Arrangement Circulaire

Pour n objets distincts, le nombre d'arrangements circulaires est $(n-1)!$. Ici, $n = 6$. Le nombre de manières est $(6-1)! = 5! = 120$. L'idée est de fixer une personne, puis d'arranger les 5 autres par rapport à elle.

Correction Exercice 20 : Problème des Dérangements (Inclusion-Exclusion)

On cherche le nombre de dérangements de 4 éléments, noté D_4 ou $!4$. La probabilité sera $D_4/4!$. La formule générale des dérangements (obtenue par inclusion-exclusion) est $D_n = n! \sum_{i=0}^n \frac{(-1)^i}{i!}$. Pour $n = 4$: $D_4 = 4!(1/0! - 1/1! + 1/2! - 1/3! + 1/4!)$ $D_4 = 24(1 - 1 + 1/2 - 1/6 + 1/24)$ $D_4 = 24(1/2 - 1/6 + 1/24) = 24(12/24 - 4/24 + 1/24) = 24(9/24) = 9$. Il y a 9 dérangements possibles sur un total de $4! = 24$ permutations. La probabilité est $P(\text{aucun match}) = D_4/4! = 9/24 = 3/8 = 0.375$.

1.11 Exercices Python

Les exercices suivants appliquent les concepts de dénombrement et de probabilité au célèbre jeu de données "Titanic". Ce dataset, chargé via la bibliothèque `seaborn`, contient des informations démographiques et de voyage sur les passagers du navire.

Le bloc de code ci-dessous initialise notre environnement en chargeant les données dans un DataFrame Pandas `df`. Pour garantir la consistance de nos calculs, nous définirons notre univers S comme l'ensemble des passagers pour lesquels l'âge est connu (en supprimant les lignes avec un âge manquant).

```
import pandas as pd
import seaborn as sns
import math

# Charger le dataset Titanic
df = sns.load_dataset("titanic")

# On retire les lignes où l'âge est inconnu pour simplifier les calculs
# C'est notre Univers S.
df = df.dropna(subset=["age"])
```

Exercice 1 : Probabilité Naïve (Filtre multiple)

Quelle est la probabilité qu'un passager, sélectionné au hasard dans l'univers `df`, soit un homme de plus de 40 ans ET voyageant en troisième classe ?

Votre tâche :

1. Trouver $|S|$, la taille totale de l'univers `df`.
2. Trouver $|A|$, le nombre de passagers remplissant les trois conditions (`sex == 'male'`, `age > 40`, `pclass == 3`).
3. Calculer $P(A) = |A|/|S|$.

Exercice 2 : Dénombrement par Combinaisons ($\binom{n}{k}$)

Pour une enquête de satisfaction, on veut créer un groupe de discussion (un "comité") composé de 5 personnes. Ces 5 personnes doivent être choisies exclusivement parmi les passagers ayant embarqué à Southampton (`embark_town == 'Southampton'`).

Combien de comités uniques de 5 personnes est-il possible de former ?

Votre tâche :

1. Trouver n , le nombre de passagers ayant embarqué à Southampton.
2. Définir $k = 5$.
3. Calculer $\binom{n}{k}$ (par ex., avec `math.comb`).

Exercice 3 : Dénombrement par Permutations

Lors d'un exercice de sécurité, on demande à 4 enfants (passagers de 12 ans ou moins) de s'aligner pour une photo de communication.

En supposant que l'on choisisse 4 enfants au hasard parmi tous les enfants du navire, et que l'ordre dans lequel ils sont alignés pour la photo est important, combien d'alignements différents sont possibles ?

Votre tâche :

1. Trouver n , le nombre total d'enfants ($\text{âge} \leq 12$) à bord.
2. Définir $k = 4$.
3. Calculer $P(n, k)$ (par ex., avec `math.perm`).

Exercice 4 : Principe d'Inclusion-Exclusion

Quelle est la probabilité qu'un passager sélectionné au hasard soit **soit un survivant** (ensemble A), **soit un passager de première classe** (ensemble B) (ou les deux) ?

Votre tâche :

1. Trouver $|S|$.
2. Trouver $|A|$ (nombre de survivants).
3. Trouver $|B|$ (nombre de passagers en 1ère classe).
4. Trouver $|A \cap B|$ (survivants de 1ère classe).
5. Appliquer la formule : $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Exercice 5 : Probabilité (Tirage sans remise)

On sélectionne au hasard un échantillon de 10 passagers de l'univers `df`.

Quelle est la probabilité que cet échantillon contienne exactement **4 survivants** et **6 non-survivants** ?

Votre tâche :

1. Trouver $N = |S|$, le nombre total de passagers.
2. Trouver m , le nombre total de survivants dans S .
3. Trouver p , le nombre total de non-survivants dans S .
4. Calculer le dénominateur : $\binom{N}{10}$ (façons de choisir 10 passagers).
5. Calculer le numérateur : $\binom{m}{4} \times \binom{p}{6}$ (façons de choisir 4 survivants ET 6 non-survivants).
6. Calculer la probabilité (numérateur / dénominateur).

Exercice 6 : Probabilité Complémentaire

On sélectionne au hasard un groupe de 5 passagers. Quelle est la probabilité que ce groupe contienne **au moins un** passager voyageant seul (`alone == True`) ?

Votre tâche :

1. Calculer $P(E^c)$, la probabilité de l'événement complémentaire "aucun passager ne voyage seul".
2. Trouver $N = |S|$, le nombre total de passagers.
3. Trouver $n_{\text{non-seul}}$, le nombre de passagers qui ne voyagent *pas* seuls.
4. Dénominateur $D = \binom{N}{5}$ (façons de choisir 5 passagers).
5. Numérateur $N_c = \binom{n_{\text{non-seul}}}{5}$ (façons de choisir 5 passagers non-seuls).
6. $P(E^c) = N_c / D$.
7. Calculer le résultat final : $P(E) = 1 - P(E^c)$.

Exercice 7 : Probabilité et Analyse de Données

Calculez la probabilité qu'un passager, choisi au hasard, ait payé un tarif (`fare`) supérieur au **tarif moyen** de l'ensemble du navire (`df`).

Votre tâche :

1. Calculer le tarif moyen de tous les passagers dans `df`.
2. Trouver $|S|$, la taille totale de l'univers `df`.
3. Trouver $|A|$, le nombre de passagers dont le `fare` est strictement supérieur à ce tarif moyen.
4. Calculer $P(A) = |A| / |S|$.

2 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 ?

Ce concept de "mise à jour des croyances" est le cœur de la statistique moderne. Il s'agit de quantifier comment une nouvelle information B affecte la probabilité d'un événement A .

2.1 Définition de la Probabilité Conditionnelle

Commençons par la définition formelle.

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 :

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

Cette formule n'est pas sortie de nulle part. Elle représente une "réduction de l'univers" :

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.

2.2 Règle du Produit (Intersection de deux événements)

En réarrangeant simplement les termes de la définition, nous obtenons une règle fondamentale pour calculer la probabilité que deux événements se produisent *ensemble*.

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.

La preuve est une simple réorganisation algébrique :

Preuve

La preuve est une simple réorganisation algébrique de la définition de la probabilité conditionnelle. Par définition, nous avons :

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

En multipliant les deux côtés par $P(B)$, on obtient :

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

De même, en partant de $P(B|A) = \frac{P(B \cap A)}{P(A)}$, on obtient :

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

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

Cette formule exprime mathématiquement l'idée séquentielle suivante :

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 règle est particulièrement utile pour les tirages sans remise, où la probabilité du second événement dépend du résultat du premier.

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 $P(A \cap B) = \frac{4}{52} \times \frac{3}{51} = \frac{12}{2652} \approx 0.0045$.

2.3 Règle de la Chaîne (Intersection de n événements)

On peut logiquement étendre cette règle de deux à 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 :

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

La preuve se fait par une simple récurrence :

Preuve : Preuve par récurrence

Nous pouvons prouver cela par une application répétée de la règle du produit pour deux événements.

Cas de base ($n=2$) : $P(A_1 \cap A_2) = P(A_1)P(A_2|A_1)$. C'est le théorème précédent.

Étape ($n=3$) : Traitons $(A_1 \cap A_2)$ comme un seul événement :

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P((A_1 \cap A_2) \cap A_3) \\ &= P(A_1 \cap A_2) \times P(A_3|A_1 \cap A_2) \\ &= (P(A_1)P(A_2|A_1)) \times P(A_3|A_1 \cap A_2) \end{aligned}$$

Généralisation : En continuant ce processus, on voit que pour ajouter A_n , on multiplie par la probabilité de A_n conditionnée par l'intersection de tous les événements précédents $(A_1 \cap \dots \cap A_{n-1})$.

Cette "règle de la chaîne" (chain rule) est cruciale pour les processus stochastiques :

Intuition

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.

Reprenons l'exemple des cartes, mais en continuant le tirage :

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 : $P(A_1 \cap A_2 \cap A_3) = \frac{4}{52} \times \frac{4}{51} \times \frac{4}{50} \approx 0.00048$.

2.4 Règle de Bayes

La règle du produit est aussi la pierre angulaire de la formule la plus célèbre des probabilités conditionnelles, qui nous permet d'inverser la condition.

Théorème : Règle de Bayes

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

La preuve est élégante car elle utilise simplement la symétrie de l'intersection :

Preuve

La preuve découle de l'égalité de la règle du produit. Nous savons que :

1. $P(A \cap B) = P(A|B)P(B)$
2. $P(A \cap B) = P(B|A)P(A)$

En égalisant ces deux expressions, on a :

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

En supposant $P(B) > 0$ et en divisant par $P(B)$, on obtient :

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

L'importance de cette formule ne peut être sous-estimée :

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.

2.5 Formule des Probabilités Totales

Le dénominateur $P(B)$ dans la règle de Bayes est souvent inconnu. Pour le trouver, nous avons besoin d'un autre outil puissant.

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 :

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

La démonstration repose sur la décomposition de l'événement B sur la partition A_i .

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 :

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

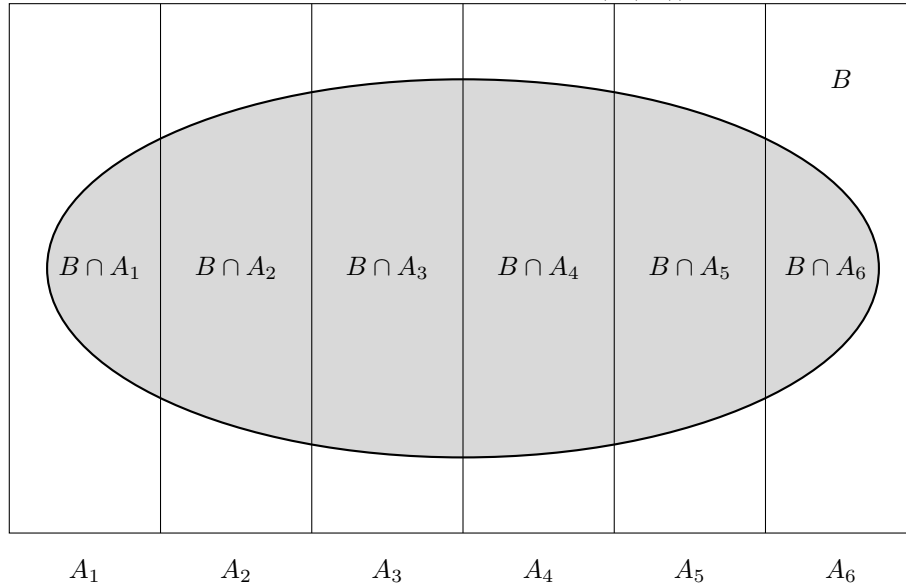
En appliquant le théorème de l'intersection des probabilités à chaque terme, on obtient :

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

Visuellement, cette formule consiste à "découper" l'événement B et à additionner les morceaux :

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.



L'exemple de l'usine est un cas d'école pour cette formule :

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%.

Maintenant, nous pouvons combiner la Règle de Bayes et la Formule des Probabilités Totales pour résoudre des problèmes complexes, comme celui du dépistage médical.

Exemple : Application Combinée : Bayes et Probabilités Totales

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 (la partition est $\{M, \neg M\}$) : $P(T) = P(T|M)P(M) + P(T|\neg M)P(\neg M) = (0.95 \times 0.01) + (0.05 \times 0.99) = 0.0095 + 0.0495 = 0.059$.

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.

2.6 Règle de Bayes avec Conditionnement Additionnel

Les règles que nous venons de voir (Bayes, Probabilités Totales) fonctionnent aussi si nous avons déjà une information de base E .

Théorème : Règle de Bayes avec conditionnement additionnel

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

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

La preuve consiste à appliquer la définition de la probabilité conditionnelle à un univers déjà restreint par E .

Preuve

La preuve est identique à celle de la règle de Bayes standard, mais en appliquant la définition de la probabilité conditionnelle à un univers restreint E .

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

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

De la première équation : $P(A \cap B \cap E) = P(A|B, E)P(B \cap E)$. De la seconde : $P(A \cap B \cap E) = P(B|A, E)P(A \cap E)$. En égalisant : $P(A|B, E)P(B \cap E) = P(B|A, E)P(A \cap E)$. D'où : $P(A|B, E) = \frac{P(B|A, E)P(A \cap E)}{P(B \cap E)}$. En utilisant $P(X \cap Y) = P(X|Y)P(Y)$, on a $P(A \cap E) = P(A|E)P(E)$ et $P(B \cap E) = P(B|E)P(E)$.

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

Cette formule peut sembler intimidante, mais elle signifie simplement que nous appliquons la même logique dans un "sous-monde" :

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**.

2.7 Formule des Probabilités Totales avec Conditionnement Additionnel

De même, la loi des probabilités totales s'adapte à ce nouvel univers restreint.

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 :

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

La démonstration est une application directe de la formule standard, mais à l'intérieur de l'univers E .

Preuve

La preuve suit celle de la formule des probabilités totales standard, mais tout est conditionné par E . Soit $P_E(\cdot)$ une mesure de probabilité définie par $P_E(X) = P(X|E)$. Les A_i forment une partition de S , donc les $(A_i \cap E)$ forment une partition de E . On applique la formule standard

à $B \cap E$:

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

Par la définition de la probabilité conditionnelle :

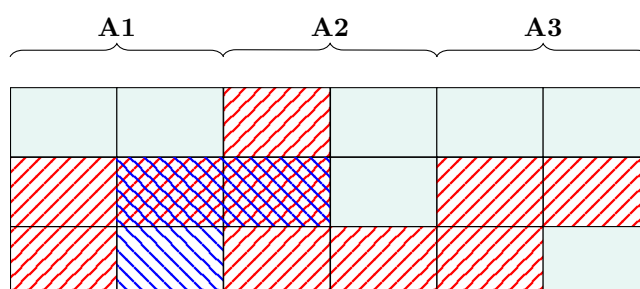
$$P(B \cap A_i|E) = \frac{P(B \cap A_i \cap E)}{P(E)}$$

Et $P(B|A_i, E)P(A_i|E) = \frac{P(B \cap A_i \cap E)}{P(A_i \cap E)} \times \frac{P(A_i \cap E)}{P(E)} = \frac{P(B \cap A_i \cap E)}{P(E)}$ Les deux termes sont égaux, donc :

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

L'exemple visuel de la carte au trésor illustre parfaitement cette double-conditionnalité :

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.

2.8 Indépendance de Deux Événements

Le concept d'indépendance est un cas spécial de probabilité conditionnelle où l'information B n'a aucun effet sur la probabilité de A .

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 :

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

En d'autres termes :

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).

2.9 Indépendance Conditionnelle

Attention : l'indépendance n'est pas la même chose que l'exclusion mutuelle. Il faut aussi se méfier de l'indépendance qui n'est qu'apparente, ou qui dépend d'une autre condition.

Définition : Indépendance Conditionnelle

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

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

C'est un concept subtil mais crucial :

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.

2.10 Le Problème de Monty Hall

Pour tester notre compréhension de tous ces concepts, le problème de Monty Hall est un exercice incontournable. Il met en lumière à quel point notre intuition sur la mise à jour des probabilités peut être faussée.

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 ?

La réponse est contre-intuitive pour la plupart des gens, mais mathématiquement claire.

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.

Pourquoi ? L'erreur est de penser que l'animateur agit au hasard.

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

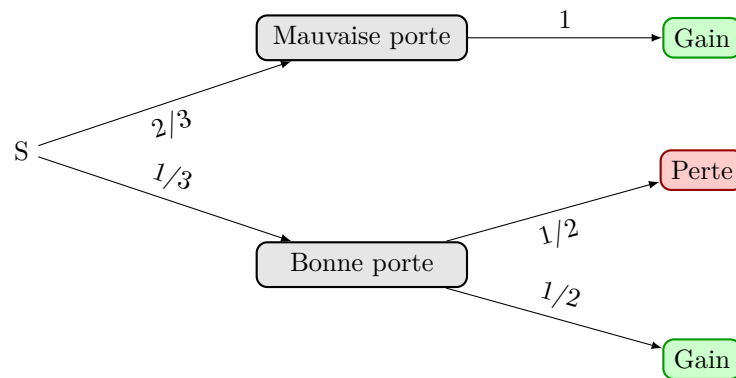
L'erreur commune est de supposer qu'il reste deux portes avec une chance égale de $1/2$. Cela ignore une information capitale : le choix de l'animateur n'est **pas aléatoire**. Il sait où se trouve la voiture et ouvrira toujours une porte perdante.

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 $2/3$.

La preuve la plus claire est de suivre les stratégies :

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 $2/3$, 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 $1/3$, vous avez choisi la "Bonne porte" du premier coup. L'arbre se divise alors en deux issues équiprobables ($1/2$ 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 $2/3$, car vous gagnez uniquement si votre choix initial était mauvais (branche du bas). La **probabilité de gain en gardant** est de $1/3$, 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.

2.11 Exercices

Exercice 1 : Dés et Probabilité Conditionnelle Simple

On lance deux dés équilibrés à 6 faces.

1. Quelle est la probabilité que la somme des dés soit 8 ?
2. Sachant que le premier dé a donné un 3, quelle est la probabilité que la somme soit 8 ?
3. Sachant que la somme est 8, quelle est la probabilité que le premier dé ait donné un 3 ?

Exercice 2 : Tirage de Cartes (Sans Remise)

On tire deux cartes successivement et sans remise d'un jeu standard de 52 cartes.

1. Quelle est la probabilité que la deuxième carte soit un Roi, sachant que la première était un Roi ?
2. Quelle est la probabilité de tirer deux Rois ?

Exercice 3 : Urne (Règle du Produit)

Une urne contient 7 boules rouges et 3 boules bleues. On tire deux boules successivement et sans remise.

1. Quelle est la probabilité que la première boule soit rouge ?
2. Quelle est la probabilité que la deuxième boule soit bleue, sachant que la première était rouge ?
3. Quelle est la probabilité de tirer une boule rouge puis une boule bleue ?

Exercice 4 : Famille (Condition Simple)

Une famille a deux enfants. On suppose que la probabilité d'avoir un garçon (G) ou une fille (F) est la même (0.5) et que les naissances sont indépendantes.

1. Quel est l'univers S des possibilités ?
2. Sachant que l'aîné est un garçon, quelle est la probabilité que la famille ait deux garçons ?

Exercice 5 : Famille (Condition "Au Moins")

En utilisant le même scénario que l'exercice 4 (famille de deux enfants) : Sachant qu'il y a *au moins un* garçon dans la famille, quelle est la probabilité que la famille ait deux garçons ?

Exercice 6 : Indépendance (Dés)

On lance deux dés équilibrés. Soit A l'événement "le premier dé donne 3" et B l'événement "la somme des deux dés est 7". Les événements A et B sont-ils indépendants ? Justifiez par le calcul.

Exercice 7 : Indépendance (Cartes)

On tire une carte d'un jeu de 52 cartes. Soit A l'événement "la carte est un Roi" et B l'événement "la carte est un Cœur". Les événements A et B sont-ils indépendants ?

Exercice 8 : Indépendance vs Exclusion Mutuelle

Soient A et B deux événements avec $P(A) = 0.5$ et $P(B) = 0.3$.

1. Si A et B sont mutuellement exclusifs (disjoints), sont-ils indépendants ?
2. Si A et B sont indépendants, quelle est $P(A \cup B)$?

Exercice 9 : LTP (Deux Urnes)

L'urne U_1 contient 2 boules noires et 3 boules blanches. L'urne U_2 contient 4 boules noires et 1 boule blanche. On choisit une urne au hasard (chaque urne a 50% de chance d'être choisie), puis on tire une boule de cette urne. Quelle est la probabilité de tirer une boule blanche ?

Exercice 10 : LTP (Usine)

Une usine utilise deux machines, M_1 et M_2 , pour produire des pièces. M_1 produit 40% des pièces et M_2 produit 60%. 5% des pièces de M_1 sont défectueuses, et 2% des pièces de M_2

sont défectueuses. Si l'on choisit une pièce au hasard dans la production totale, quelle est la probabilité qu'elle soit défectueuse ?

Exercice 11 : LTP (Pièce de Monnaie Inconnue)

On a deux pièces. La pièce A est équilibrée ($P(\text{Pile}) = 0.5$). La pièce B est truquée ($P(\text{Pile}) = 0.8$). On choisit une pièce au hasard et on la lance. Quelle est la probabilité d'obtenir Pile ?

Exercice 12 : Bayes (Test Médical)

Une maladie touche 1 personne sur 1000 ($P(M) = 0.001$). Un test de dépistage donne un résultat positif chez 98% des personnes malades ($P(T|M) = 0.98$). Il donne aussi un résultat positif (un "faux positif") chez 3% des personnes non malades ($P(T|\neg M) = 0.03$). Une personne reçoit un test positif. Quelle est la probabilité qu'elle soit réellement malade ?

Exercice 13 : Bayes (Inversion d'Urnnes)

Reprenons le scénario de l'exercice 9 (U1 avec 2N/3B, U2 avec 4N/1B). On a tiré une boule et on constate qu'elle est blanche. Quelle est la probabilité qu'elle provienne de l'urne U1 ?

Exercice 14 : Bayes (Spam)

Dans une boîte de réception, 60% des emails sont des spams. 70% des spams contiennent le mot "gratuit". Seuls 10% des emails légitimes contiennent le mot "gratuit". Vous recevez un email qui contient le mot "gratuit". Quelle est la probabilité que ce soit un spam ?

Exercice 15 : Bayes (Usine Inversée)

Reprenons le scénario de l'exercice 10 (M1 : 40% prod, 5% défaut ; M2 : 60%). On trouve une pièce défectueuse. Quelle est la probabilité qu'elle ait été produite par la machine M1 ?

Exercice 16 : Règle de la Chaîne (3 Cartes)

On tire 3 cartes successivement et sans remise d'un jeu de 52 cartes. Quelle est la probabilité de tirer 3 Piques ?

Exercice 17 : Problème de Monty Hall (Calcul)

En utilisant la formalisation du problème de Monty Hall (vous choisissez la Porte 1, la voiture est en $V \in \{1, 2, 3\}$, l'animateur ouvre $H \in \{2, 3\}$) : Calculez $P(V = 1|H = 3)$ (la probabilité que la voiture soit derrière votre porte, sachant que l'animateur a ouvert la 3). Supposez que $P(V = i) = 1/3$ pour $i = 1, 2, 3$.

Exercice 18 : Bayes avec Mise à Jour (Pièce Truquée)

Reprenons l'exercice 11 (Pièce A équilibrée, Pièce B truquée $P(\text{Pile}) = 0.8$). On choisit une pièce au hasard. On la lance deux fois et on obtient Pile, puis Pile (PP). Quelle est la probabilité que l'on ait choisi la pièce truquée (Pièce B) ?

Exercice 19 : Indépendance Conditionnelle (Dés)

On lance deux dés, D_1 et D_2 . Soit $S = D_1 + D_2$ leur somme. Soit A l'événement " $D_1 = 1$ ", B l'événement " $D_2 = 1$ ". A et B sont indépendants. Sont-ils indépendants conditionnellement à l'événement $C = \{S = 2\}$?

Exercice 20 : Jeu Séquentiel

Alice et Bob jouent à un jeu. Ils lancent un dé à tour de rôle, en commençant par Alice. Le premier qui obtient un 6 gagne. Quelle est la probabilité qu'Alice gagne ?

2.12 Corrections des Exercices

Correction Exercice 1 : Dés et Probabilité Conditionnelle Simple

L'univers S a $|S| = 6 \times 6 = 36$ issues.

1. Soit A l'événement "la somme est 8". $A = \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}$. $|A| = 5$, donc $P(A) = 5/36$.
2. Soit B l'événement "le premier dé donne 3". $B = \{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\}$. On cherche $P(A|B)$. Sachant B , l'univers est réduit à ces 6 issues. Parmi celles-ci, seule l'issue $(3, 5)$ donne une somme de 8. Donc, $P(A|B) = 1/6$. *Par formule :* $A \cap B = \{(3, 5)\}$, $P(A \cap B) = 1/36$. $P(B) = 6/36 = 1/6$. $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/36}{1/6} = 1/6$.
3. On cherche $P(B|A)$. Sachant A , l'univers est réduit aux 5 issues de A . Parmi celles-ci, seule l'issue $(3, 5)$ a 3 sur le premier dé. Donc, $P(B|A) = 1/5$. *Par formule :* $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1/36}{5/36} = 1/5$.

Correction Exercice 2 : Tirage de Cartes (Sans Remise)

Soit K_1 l'événement "Roi au 1er tirage" et K_2 "Roi au 2e tirage".

1. On cherche $P(K_2|K_1)$. Si K_1 s'est produit, il reste 51 cartes dans le jeu, dont $4 - 1 = 3$ Rois. $P(K_2|K_1) = 3/51 = 1/17$.
2. On cherche $P(K_1 \cap K_2)$. On utilise la règle du produit : $P(K_1 \cap K_2) = P(K_1) \times P(K_2|K_1)$. $P(K_1) = 4/52 = 1/13$. $P(K_1 \cap K_2) = (4/52) \times (3/51) = (1/13) \times (1/17) = 1/221$.

Correction Exercice 3 : Urne (Règle du Produit)

Urne avec 7 Rouges (R) et 3 Bleues (B). Total = 10. Soit R_1 "Rouge au 1er tirage" et B_2 "Bleue au 2e tirage".

1. $P(R_1) = 7/10$.
2. On cherche $P(B_2|R_1)$. Si R_1 s'est produit, il reste 9 boules (6R, 3B). $P(B_2|R_1) = 3/9 = 1/3$.
3. On cherche $P(R_1 \cap B_2)$. $P(R_1 \cap B_2) = P(R_1) \times P(B_2|R_1) = (7/10) \times (1/3) = 7/30$.

Correction Exercice 4 : Famille (Condition Simple)

1. L'univers est $S = \{GG, GF, FG, FF\}$, où le premier enfant est l'aîné. $|S| = 4$, chaque issue a une probabilité de $1/4$.
2. Soit A l'événement "l'aîné est un garçon" : $A = \{GG, GF\}$. $P(A) = 2/4 = 1/2$. Soit B l'événement "la famille a deux garçons" : $B = \{GG\}$. $P(B) = 1/4$. On cherche $P(B|A)$. L'événement $A \cap B = \{GG\}$. $P(A \cap B) = 1/4$. $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1/4}{1/2} = 1/2$.

Correction Exercice 5 : Famille (Condition "Au Moins")

Soit B l'événement "la famille a deux garçons" : $B = \{GG\}$. Soit C l'événement "il y a au moins un garçon" : $C = \{GG, GF, FG\}$. $P(C) = 3/4$. On cherche $P(B|C)$. L'événement $B \cap C = \{GG\}$. $P(B \cap C) = 1/4$. $P(B|C) = \frac{P(B \cap C)}{P(C)} = \frac{1/4}{3/4} = 1/3$. *Intuition :* L'univers de C est $\{GG, GF, FG\}$. Parmi ces 3 issues équiprobables, une seule est GG .

Correction Exercice 6 : Indépendance (Dés)

A = "premier dé = 3". $P(A) = 6/36 = 1/6$. B = "somme = 7". $B = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$. $P(B) = 6/36 = 1/6$. $A \cap B$ = "premier dé = 3 ET somme = 7" = $\{(3, 4)\}$. $P(A \cap B) = 1/36$. On teste si $P(A \cap B) = P(A)P(B)$. $P(A)P(B) = (1/6) \times (1/6) = 1/36$. Puisque $P(A \cap B) = P(A)P(B)$, les événements A et B sont indépendants.

Correction Exercice 7 : Indépendance (Cartes)

A = "Roi". $P(A) = 4/52 = 1/13$. B = "Cœur". $P(B) = 13/52 = 1/4$. $A \cap B$ = "Roi de Cœur". $P(A \cap B) = 1/52$. On teste si $P(A \cap B) = P(A)P(B)$. $P(A)P(B) = (1/13) \times (1/4) = 1/52$. Puisque $P(A \cap B) = P(A)P(B)$, les événements A et B sont indépendants.

Correction Exercice 8 : Indépendance vs Exclusion Mutuelle

$P(A) = 0.5, P(B) = 0.3$.

1. Si A et B sont mutuellement exclusifs, $A \cap B = \emptyset$, donc $P(A \cap B) = 0$. Pour qu'ils soient indépendants, il faudrait $P(A \cap B) = P(A)P(B) = 0.5 \times 0.3 = 0.15$. Puisque $0 \neq 0.15$, ils ne sont pas indépendants. (Deux événements non impossibles ne peuvent pas être à la fois mutuellement exclusifs et indépendants).

2. Si A et B sont indépendants, $P(A \cap B) = P(A)P(B) = 0.15$. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.3 - 0.15 = 0.65$.

Correction Exercice 9 : LTP (Deux Urnes)

Soit U_1 et U_2 les événements "choisir l'urne 1" et "choisir l'urne 2". $P(U_1) = 0.5, P(U_2) = 0.5$. Soit W l'événement "tirer une boule blanche". On a $P(W|U_1) = 3/(2+3) = 3/5 = 0.6$. On a $P(W|U_2) = 1/(4+1) = 1/5 = 0.2$. Par la formule des probabilités totales : $P(W) = P(W|U_1)P(U_1) + P(W|U_2)P(U_2)$ $P(W) = (0.6 \times 0.5) + (0.2 \times 0.5) = 0.3 + 0.1 = 0.4$.

Correction Exercice 10 : LTP (Usine)

Soit M_1 et M_2 les machines. $P(M_1) = 0.4, P(M_2) = 0.6$. Soit D l'événement "la pièce est défectueuse". On a $P(D|M_1) = 0.05$ et $P(D|M_2) = 0.02$. Par la formule des probabilités totales : $P(D) = P(D|M_1)P(M_1) + P(D|M_2)P(M_2)$ $P(D) = (0.05 \times 0.4) + (0.02 \times 0.6) = 0.020 + 0.012 = 0.032$. La probabilité est de 3.2%.

Correction Exercice 11 : LTP (Pièce de Monnaie Inconnue)

Soit A "choisir pièce A" et B "choisir pièce B". $P(A) = 0.5, P(B) = 0.5$. Soit H l'événement "obtenir Pile". On a $P(H|A) = 0.5$ et $P(H|B) = 0.8$. Par la formule des probabilités totales : $P(H) = P(H|A)P(A) + P(H|B)P(B)$ $P(H) = (0.5 \times 0.5) + (0.8 \times 0.5) = 0.25 + 0.40 = 0.65$.

Correction Exercice 12 : Bayes (Test Médical)

Soit M "Malade" et T "Test Positif". $P(M) = 0.001$, donc $P(\neg M) = 0.999$. $P(T|M) = 0.98$. $P(T|\neg M) = 0.03$. On cherche $P(M|T)$. Par la règle de Bayes : $P(M|T) = \frac{P(T|M)P(M)}{P(T)}$.

1. Calculer $P(T)$ (dénominateur) avec la LTP : $P(T) = P(T|M)P(M) + P(T|\neg M)P(\neg M)$ $P(T) = (0.98 \times 0.001) + (0.03 \times 0.999) = 0.00098 + 0.02997 = 0.03095$.

2. Appliquer la règle de Bayes : $P(M|T) = \frac{0.00098}{0.03095} \approx 0.03166$. Il n'y a que 3.17% de chance que la personne soit malade, même avec un test positif.

Correction Exercice 13 : Bayes (Inversion d'Urnes)

D'après l'exercice 9, on a : $P(W) = 0.4$ (prob. totale de tirer une blanche). $P(W|U_1) = 0.6$. $P(U_1) = 0.5$. On cherche $P(U_1|W)$. Par la règle de Bayes : $P(U_1|W) = \frac{P(W|U_1)P(U_1)}{P(W)} = \frac{0.6 \times 0.5}{0.4} = \frac{0.3}{0.4} = 0.75$. Sachant que la boule est blanche, il y a 75% de chance qu'elle vienne de l'urne U_1 .

Correction Exercice 14 : Bayes (Spam)

Soit S "Spam" et G "Contient 'gratuit'". $P(S) = 0.6$, donc $P(\neg S) = 0.4$. $P(G|S) = 0.7$. $P(G|\neg S) = 0.1$. On cherche $P(S|G)$. Par la règle de Bayes : $P(S|G) = \frac{P(G|S)P(S)}{P(G)}$.

1. Calculer $P(G)$ (dénominateur) avec la LTP : $P(G) = P(G|S)P(S) + P(G|\neg S)P(\neg S)$ $P(G) = (0.7 \times 0.6) + (0.1 \times 0.4) = 0.42 + 0.04 = 0.46$.

2. Appliquer la règle de Bayes : $P(S|G) = \frac{0.42}{0.46} \approx 0.913$. Il y a 91.3% de chance que l'email soit un spam.

Correction Exercice 15 : Bayes (Usine Inversée)

D'après l'exercice 10, on a : $P(D) = 0.032$ (prob. totale d'être défectueux). $P(D|M_1) = 0.05$. $P(M_1) = 0.4$. On cherche $P(M_1|D)$. Par la règle de Bayes : $P(M_1|D) = \frac{P(D|M_1)P(M_1)}{P(D)} = \frac{0.05 \times 0.4}{0.032} = \frac{0.02}{0.032} = 0.625$. Sachant que la pièce est défectueuse, il y a 62.5% de chance qu'elle vienne de M_1 .

Correction Exercice 16 : Règle de la Chaîne (3 Cartes)

Soit P_i l'événement "tirer un Pique au i -ème tirage". Il y a 13 Piques sur 52 cartes. On cherche $P(P_1 \cap P_2 \cap P_3)$. On utilise la règle de la chaîne : $P(P_1 \cap P_2 \cap P_3) = P(P_1) \times P(P_2|P_1) \times P(P_3|P_1 \cap P_2)$. $P(P_1) = 13/52$. $P(P_2|P_1) = 12/51$ (il reste 12 Piques sur 51 cartes). $P(P_3|P_1 \cap P_2) = 11/50$ (il reste 11 Piques sur 50 cartes). $P = (13/52) \times (12/51) \times (11/50) = \frac{1}{4} \times \frac{4}{17} \times \frac{11}{50} = \frac{11}{17 \times 50} = 11/850 \approx 0.0129$.

Correction Exercice 17 : Problème de Monty Hall (Calcul)

On cherche $P(V = 1|H = 3)$. On utilise la règle de Bayes : $P(V = 1|H = 3) = \frac{P(H=3|V=1)P(V=1)}{P(H=3)}$.

*Numérateur : $P(V = 1) = 1/3$. $P(H = 3|V = 1)$ est la probabilité que l'animateur ouvre la 3, sachant que vous avez choisi la 1 et que la voiture est en 1. Il peut ouvrir la 2 ou la 3 (deux chèvres). On suppose qu'il choisit au hasard : $P(H = 3|V = 1) = 1/2$. Numérateur = $(1/2) \times (1/3) = 1/6$.

*Dénominateur $P(H = 3)$ par LTP (partition sur V) : $P(H = 3) = P(H = 3|V = 1)P(V = 1) + P(H = 3|V = 2)P(V = 2) + P(H = 3|V = 3)P(V = 3)$. $P(H = 3|V = 1) = 1/2$ (calculé ci-dessus). $P(H = 3|V = 2) = 1$ (l'animateur doit ouvrir la 3, car vous avez choisi 1 et la voiture est en 2). $P(H = 3|V = 3) = 0$ (l'animateur ne peut pas ouvrir la porte 3 car elle contient la voiture). $P(H = 3) = (1/2 \times 1/3) + (1 \times 1/3) + (0 \times 1/3) = 1/6 + 1/3 + 0 = 1/2$.

*Résultat : $P(V = 1|H = 3) = \frac{1/6}{1/2} = 1/3$. (La probabilité que la voiture soit derrière votre porte reste 1/3. La probabilité qu'elle soit derrière l'autre porte fermée (la 2) est $P(V = 2|H = 3) = 1 - P(V = 1|H = 3) = 2/3$. Il faut donc changer.)

Correction Exercice 18 : Bayes avec Mise à Jour (Pièce Truquée)

Soit A "pièce A (équil.)" et B "pièce B (truquée, $p=0.8$)". $P(A) = P(B) = 0.5$. Soit E l'événement "obtenir Pile, Pile" (PP). On cherche $P(B|E) = \frac{P(E|B)P(B)}{P(E)}$.

1. Probabilités conditionnelles de l'évidence E : $P(E|A) = P(PP|A) = 0.5 \times 0.5 = 0.25$ (indépendance des lancers). $P(E|B) = P(PP|B) = 0.8 \times 0.8 = 0.64$.

2. Calculer $P(E)$ (dénominateur) avec la LTP : $P(E) = P(E|A)P(A) + P(E|B)P(B)$. $P(E) = (0.25 \times 0.5) + (0.64 \times 0.5) = 0.125 + 0.320 = 0.445$.

3. Appliquer la règle de Bayes : $P(B|E) = \frac{P(E|B)P(B)}{P(E)} = \frac{0.64 \times 0.5}{0.445} = \frac{0.32}{0.445} \approx 0.719$. Après avoir observé PP, la probabilité que ce soit la pièce truquée passe de 50% à 71.9%.

Correction Exercice 19 : Indépendance Conditionnelle (Dés)

$A = \{D_1 = 1\}$, $B = \{D_2 = 1\}$, $C = \{S = 2\}$. On teste si $P(A \cap B|C) = P(A|C)P(B|C)$.

L'événement $C = \{S = 2\}$ ne peut se produire que d'une seule façon : $C = \{(1, 1)\}$. Donc, C est l'événement $A \cap B$. $C \subseteq A$ et $C \subseteq B$.

Calculons les termes : - $P(A|C) = P(A \cap C)/P(C)$. Puisque $C \subseteq A$, $A \cap C = C$. $P(A|C) = P(C)/P(C) = 1$. - $P(B|C) = P(B \cap C)/P(C)$. Puisque $C \subseteq B$, $B \cap C = C$. $P(B|C) = P(C)/P(C) = 1$. - $P(A \cap B|C) = P((A \cap B) \cap C)/P(C)$. Puisque $A \cap B = C$, $(A \cap B) \cap C = C$. $P(A \cap B|C) = P(C)/P(C) = 1$.

Test d'indépendance : $P(A \cap B|C) = 1$. $P(A|C)P(B|C) = 1 \times 1 = 1$. Puisque $1 = 1$, les événements A et B sont bien indépendants conditionnellement à C . *Intuition : * Sachant que la somme est 2, nous savons avec certitude que $D_1 = 1$ et $D_2 = 1$. Il n'y a plus d'aléa.

Correction Exercice 20 : Jeu Séquentiel

Soit $p = 1/6$ la probabilité de gagner (obtenir un 6) et $q = 5/6$ la probabilité de rater. Alice gagne si elle réussit au tour 1, OU si (elle rate ET Bob rate) et elle réussit au tour 3, OU si (A rate, B rate, A rate, B rate) et elle réussit au tour 5, etc.

$P(\text{A gagne}) = P(\text{A au tour 1}) + P(\text{A au tour 3}) + P(\text{A au tour 5}) + \dots$. $P(\text{A gagne}) = p + (q \times q)p + (q \times q \times q \times q)p + \dots$. $P(\text{A gagne}) = p + q^2p + q^4p + \dots$. $P(\text{A gagne}) = p \times (1 + q^2 + q^4 + \dots)$. $P(\text{A gagne}) = p \sum_{k=0}^{\infty} (q^2)^k$.

C'est une série géométrique de premier terme p et de raison $r = q^2 = (5/6)^2 = 25/36$. La somme est $\frac{\text{premier terme}}{1 - \text{raison}} = \frac{p}{1 - q^2}$. $P(\text{A gagne}) = \frac{1/6}{1 - 25/36} = \frac{1/6}{11/36} = \frac{1}{6} \times \frac{36}{11} = 6/11$.

2.13 Exercices Python

Ces exercices appliquent les concepts de probabilité conditionnelle, de la règle de Bayes et de la formule des probabilités totales au jeu de données "Titanic".

```
import pandas as pd
import seaborn as sns
import math

# Charger le dataset Titanic
df = sns.load_dataset("titanic")

# On retire les lignes où l'âge est inconnu pour simplifier les calculs
# C'est notre Univers S.
df = df.dropna(subset=["age"])
```

Exercice 1 : Définition de $P(A|B)$

Calculez la probabilité qu'un passager ait survécu (A), **sachant que** ce passager était en première classe (B).

Votre tâche :

1. Soit A = "le passager a survécu" (`survived == 1`).
2. Soit B = "le passager est en première classe" (`pclass == 1`).
3. Trouver $|B|$ (le nombre de passagers en 1ère classe).
4. Trouver $|A \cap B|$ (le nombre de survivants de 1ère classe).
5. Calculer $P(A|B) = \frac{|A \cap B|}{|B|}$.

Exercice 2 : Règle du Produit (Tirage sans remise)

On tire au hasard 2 passagers de l'univers `df` sans remise. Calculez la probabilité que le premier passager soit un survivant (A_1) ET que le second passager soit aussi un survivant (A_2).

Utilisez la Règle du Produit : $P(A_1 \cap A_2) = P(A_1)P(A_2|A_1)$.

Votre tâche :

1. Trouver $|S|$ (total passagers) et $|A_1|$ (total survivants).
2. Calculer $P(A_1) = |A_1|/|S|$.
3. Calculer $P(A_2|A_1)$. (Indice : après avoir tiré un survivant, combien de passagers restent ? Combien de survivants restent ?).
4. Calculer le produit $P(A_1) \times P(A_2|A_1)$.

Exercice 3 : Formule des Probabilités Totales

Calculez la probabilité totale qu'un passager ait survécu (B) en utilisant la formule des probabilités totales. Utilisez la partition des trois classes de passagers (A_1 =1ère, A_2 =2e, A_3 =3e classe).

La formule est : $P(B) = \sum_{i=1}^3 P(B|A_i)P(A_i)$.

Votre tâche :

1. Pour $i = 1, 2, 3$, calculer $P(A_i)$, la probabilité d'appartenir à chaque classe (ex : $P(A_1) = |pclass 1|/|S|$).
2. Pour $i = 1, 2, 3$, calculer $P(B|A_i)$, la probabilité de survie sachant la classe (cf. Exercice 1).
3. Appliquer la formule : $P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)$.
4. (Vérification) Comparez votre résultat au calcul direct $P(B) = |\text{survivants}|/|S|$.

Exercice 4 : Règle de Bayes

En utilisant les résultats de l'exercice précédent, appliquez la Règle de Bayes.

On observe qu'un passager a survécu (B). Quelle est la probabilité qu'il s'agisse d'un passager de première classe (A_1) ?

Votre tâche :

1. On cherche $P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B)}$.
2. Récupérer $P(B|A_1)$ (la probabilité de survie en 1ère classe) de l'exercice 3.
3. Récupérer $P(A_1)$ (la probabilité d'être en 1ère classe) de l'exercice 3.
4. Récupérer $P(B)$ (la probabilité totale de survie) de l'exercice 3.
5. Effectuer le calcul.

Exercice 5 : Indépendance de deux événements

Les événements A = "être une femme" (`sex == 'female'`) et B = "avoir survécu" (`survived == 1`) sont-ils indépendants ?

Votre tâche :

1. Prouver ou réfuter l'indépendance en vérifiant si $P(A \cap B) = P(A)P(B)$.
2. Calculer $P(A) = |\text{femmes}|/|S|$.
3. Calculer $P(B) = |\text{survivants}|/|S|$.
4. Calculer $P(A \cap B) = |\text{femmes survivantes}|/|S|$.
5. Comparer $P(A \cap B)$ au produit $P(A) \times P(B)$.
6. (Alternative) Comparer $P(B|A)$ à $P(B)$. L'information "être une femme" change-t-elle la probabilité de survie ?

3 Variables Aléatoires Discrètes

3.1 Variable Aléatoire

Jusqu'à présent, nous avons parlé d'événements (comme "obtenir Pile" ou "tirer un Roi"). Pour analyser ces phénomènes avec des outils mathématiques plus puissants, nous devons traduire ces résultats concrets en nombres. C'est le rôle de la 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} .

Cette définition formelle masque une idée très simple :

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.

Prenons un exemple classique :

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 $X(2, 5) = 2 + 5 = 7$.

3.2 Variable Aléatoire Discrète

Les variables aléatoires peuvent être de différents types. Nous commençons par le type le plus simple à "compter".

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 $P(X = a_j \text{ pour un certain } j) = 1$.

L'analogie la plus simple pour comprendre le terme "discret" est celle d'un escalier.

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.

3.3 Fonction de Masse (PMF)

Maintenant que nous avons une variable aléatoire qui produit des nombres discrets, nous avons besoin d'une fonction pour décrire la probabilité de chacun de ces nombres.

Définition : Probability Mass Function (PMF)

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

C'est la "carte d'identité" probabiliste de la variable :

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.

Un exemple très simple est le lancer de dé :

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 $x = 7$), $P(X = x) = 0$.

3.4 Loi de Bernoulli

Commençons par la loi de probabilité discrète la plus simple.

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 $X \sim \text{Bern}(p)$.

C'est la brique fondamentale de nombreuses autres distributions.

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".

3.5 Loi Binomiale

Que se passe-t-il si nous répétons une expérience de Bernoulli n fois et que nous comptons le nombre total de succès ?

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

La preuve de cette formule est un argument combinatoire direct.

Preuve

Nous voulons trouver la probabilité d'obtenir exactement k succès au cours de n essais indépendants.

1. **Probabilité d'une séquence spécifique** : Considérons d'abord une séquence spécifique contenant k succès (S) et $n - k$ échecs (E), par exemple $S, S, \dots, S, E, E, \dots, E$. Puisque les essais sont indépendants, la probabilité de cette séquence est le produit des probabilités individuelles :

$$\underbrace{p \times p \times \dots \times p}_{k \text{ fois}} \times \underbrace{(1 - p) \times \dots \times (1 - p)}_{n-k \text{ fois}} = p^k (1 - p)^{n-k}$$

2. **Nombre de séquences possibles** : La séquence ci-dessus n'est qu'une des nombreuses façons d'obtenir k succès. Le nombre total de façons d'arranger k succès parmi n positions (essais) est donné par le coefficient binomial $\binom{n}{k}$.

3. **Probabilité totale :** Chacune de ces $\binom{n}{k}$ séquences a la même probabilité $p^k(1-p)^{n-k}$. Puisque toutes ces séquences sont des événements disjoints, la probabilité totale d'obtenir k succès (dans n'importe quel ordre) est la somme de leurs probabilités :

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

Chaque partie de cette formule a une signification logique claire.

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.

Appliquons cela à un exemple classique :

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.

3.6 Loi Hypergéométrique

La loi binomiale suppose que les essais sont indépendants, ce qui est vrai si l'on tire *avec remise*. Que se passe-t-il si l'on tire *sans remise* ?

Théorème : PMF Hypergéométrique

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

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

La preuve de cette formule est un argument de dénombrement pur, basé sur la définition naïve de la probabilité.

Preuve

Nous utilisons la définition naïve $P(A) = |A|/|S|$. Nous tirons m boules d'une urne contenant w blanches et b noires, soit $w+b$ boules au total.

1. **Taille de l'univers ($|S|$) :** Le nombre total de façons de choisir m boules parmi $w+b$ est $\binom{w+b}{m}$.
2. **Taille de l'événement favorable ($|A|$) :** Nous voulons l'événement $A =$ "obtenir exactement k boules blanches ET $m-k$ boules noires".
 - Le nombre de façons de choisir k blanches parmi w est $\binom{w}{k}$.
 - Le nombre de façons de choisir $m-k$ noires parmi b est $\binom{b}{m-k}$.

Par le principe de la multiplication (dénombrement), le nombre total de façons de réaliser A est $|A| = \binom{w}{k} \binom{b}{m-k}$.

3. **Probabilité :** En divisant le nombre d'issues favorables par le nombre total d'issues, on obtient :

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

Chaque terme de cette fraction a un sens très concret :

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.

Un exemple typique est la formation de comités à partir d'un groupe.

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.

3.7 Loi Géométrique

Revenons aux essais de Bernoulli (indépendants). Au lieu de fixer le nombre d'essais n , demandons-nous : combien d'essais faut-il avant d'obtenir notre premier succès ?

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.

La preuve de cette formule est une application directe de l'indépendance des essais.

Preuve

Soit S_i l'événement "succès au i -ème essai" et E_i l'événement "échec au i -ème essai". L'événement $\{X = k\}$ signifie que nous avons observé exactement k échecs, suivis d'un succès au $(k+1)$ -ème essai. C'est la séquence d'événements : $E_1 \cap E_2 \cap \dots \cap E_k \cap S_{k+1}$.

Puisque tous les essais sont indépendants, la probabilité de cette intersection est le produit des probabilités individuelles :

$$\begin{aligned} P(X = k) &= P(E_1) \times P(E_2) \times \dots \times P(E_k) \times P(S_{k+1}) \\ &= \underbrace{(1-p) \times (1-p) \times \dots \times (1-p)}_{k \text{ fois}} \times p \\ &= (1-p)^k p \end{aligned}$$

La formule est donc très littérale :

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".

Un exemple classique est l'attente d'un résultat spécifique sur un dé.

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 :

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

3.8 Loi de Poisson

Introduisons maintenant une loi utilisée pour modéliser le nombre d'événements se produisant dans un intervalle de temps ou d'espace fixe.

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.

Cette loi est souvent appelée la loi des événements rares.

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).

Mais d'où vient cette formule avec e et une factorielle ? Elle vient d'une approximation de la loi binomiale lorsque n est très grand et p très petit.

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.

Intuition : Convergence Binomiale vers Poisson : L'Exemple des Naissances

Supposons que les bébés naissent dans une grande ville à un taux moyen de $\lambda = 10$ naissances par jour. Comment modéliser le nombre X de naissances un jour donné ?

1. Approche Binomiale (Découpage du Temps) : On peut diviser la journée (24h) en n très petits intervalles de temps (par exemple, $n = 24 \times 60 \times 60 = 86400$ secondes).

- Si n est très grand, la chance p qu'une naissance se produise *exactement* pendant une seconde donnée est minuscule. On peut calculer cette probabilité p comme le taux moyen divisé par le nombre d'intervalles : $p = \lambda/n = 10/86400$.
- On peut aussi supposer que la probabilité d'avoir *deux* naissances ou plus dans la même seconde est négligeable. Chaque seconde est donc comme un mini-essai de Bernoulli : soit 1 naissance (avec probabilité p), soit 0 naissance (avec probabilité $1 - p$).
- Le nombre total de naissances X sur la journée est la somme de ces n essais de Bernoulli quasi-indépendants. X suit donc approximativement une loi binomiale : $X \approx \text{Bin}(n, p = \lambda/n)$.

La probabilité d'avoir exactement k naissances serait $P(X = k) \approx \binom{n}{k} p^k (1 - p)^{n-k}$.

2. Le Passage à la Limite (Modèle Continu) : Que se passe-t-il si on rend les intervalles de temps infiniment petits ($n \rightarrow \infty$) ? C'est là que la magie opère :

- Le terme $\binom{n}{k}$ (combien de façons de choisir k secondes parmi n) se comporte comme $n^k/k!$ pour n grand.
- Le terme $p^k = (\lambda/n)^k$ devient λ^k/n^k .
- Le terme $(1-p)^{n-k} = (1-\lambda/n)^{n-k}$. Comme k est petit par rapport à n , ceci est très proche de $(1-\lambda/n)^n$, qui tend vers $e^{-\lambda}$.

En combinant ces approximations (expliquées plus en détail dans la preuve formelle), on trouve que la probabilité $P(X = k)$ tend vers $\frac{n^k}{k!} \frac{\lambda^k}{n^k} e^{-\lambda} = \frac{e^{-\lambda} \lambda^k}{k!}$.

Conclusion : La loi de Poisson apparaît naturellement comme la limite d'un processus binomial où l'on a un très grand nombre d'opportunités (n) pour qu'un événement rare (probabilité p) se produise, tout en maintenant un taux moyen constant ($\lambda = np$).

La preuve formelle montre comment les termes de la formule binomiale se transforment en ceux de la formule de Poisson lorsque $n \rightarrow \infty$.

Preuve : Dérivation de la loi de Poisson à partir de la loi Binomiale (Détailée)

On part de la fonction de masse (PMF) d'une variable aléatoire X_n suivant une loi binomiale $\text{Bin}(n, p)$, où l'on pose $p = \lambda/n$. L'objectif est de trouver la limite de cette PMF lorsque n tend vers l'infini, tout en gardant $\lambda = np$ constant (ce qui implique que p doit tendre vers 0).

La PMF binomiale est :

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

Substituons $p = \lambda/n$:

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

Maintenant, développons le coefficient binomial $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}$:

$$P(X_n = k) = \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Réorganisons les termes pour isoler ceux qui dépendent de n :

$$P(X_n = k) = \frac{\lambda^k}{k!} \times \frac{n(n-1)\cdots(n-k+1)}{n^k} \times \left(1 - \frac{\lambda}{n}\right)^n \times \left(1 - \frac{\lambda}{n}\right)^{-k}$$

Nous allons maintenant examiner la limite de chaque partie lorsque $n \rightarrow \infty$, pour k et λ fixés.

1. $\frac{\lambda^k}{k!}$: Ce terme est constant par rapport à n , donc sa limite est lui-même.
2. $\frac{n(n-1)\cdots(n-k+1)}{n^k}$: Ce terme est un produit de k facteurs divisé par n^k . On peut le réécrire comme :

$$\begin{aligned} & \frac{n}{n} \times \frac{n-1}{n} \times \frac{n-2}{n} \times \cdots \times \frac{n-k+1}{n} \\ &= 1 \times \left(1 - \frac{1}{n}\right) \times \left(1 - \frac{2}{n}\right) \times \cdots \times \left(1 - \frac{k-1}{n}\right) \end{aligned}$$

Lorsque $n \rightarrow \infty$, chacun des termes $\frac{1}{n}, \frac{2}{n}, \dots, \frac{k-1}{n}$ tend vers 0 (car k est fixe). Donc, chaque parenthèse tend vers $(1-0) = 1$. Puisqu'il y a un nombre fixe k de termes dans le produit, la limite du produit est le produit des limites :

$$\lim_{n \rightarrow \infty} \frac{n(n-1)\cdots(n-k+1)}{n^k} = 1 \times 1 \times \cdots \times 1 = 1$$

Intuition : Pour n très grand par rapport à k , les k termes $n, n-1, \dots, n-k+1$ sont tous "presque" égaux à n . Leur produit est donc "presque" n^k , et le ratio est "presque" 1.

3. $\left(1 - \frac{\lambda}{n}\right)^n$: C'est une limite fondamentale en analyse. On sait que pour tout réel x , $\lim_{n \rightarrow \infty} (1 + x/n)^n = e^x$. Ici, nous avons $x = -\lambda$. Donc :

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

Intuition : C'est la définition même de l'exponentielle comme limite d'intérêts composés continus (ici, avec un taux négatif).

4. $\left(1 - \frac{\lambda}{n}\right)^{-k}$: Lorsque $n \rightarrow \infty$, le terme λ/n tend vers 0. L'expression à l'intérieur de la parenthèse tend donc vers $(1 - 0) = 1$. Puisque k est un exposant fixe :

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

Intuition : Pour n très grand, $(1 - \lambda/n)$ est très proche de 1. Élever ce nombre très proche de 1 à une puissance fixe k le laisse très proche de 1.

Finalement, en multipliant les limites de chaque partie (puisque la limite d'un produit est le produit des limites) :

$$\lim_{n \rightarrow \infty} P(X_n = k) = \left(\frac{\lambda^k}{k!}\right) \times (1) \times (e^{-\lambda}) \times (1) = \frac{e^{-\lambda} \lambda^k}{k!}$$

Ceci est exactement la fonction de masse de probabilité d'une loi de Poisson de paramètre λ .

Un ensemble de données historiques célèbres illustre parfaitement cette loi.

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.

3.9 Fonction de Répartition (CDF)

Nous avons la PMF, qui donne $P(X = x)$. Une autre fonction tout aussi importante est la fonction de répartition (CDF), qui "accumule" ces probabilités.

Définition : Cumulative Distribution Function (CDF)

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

Cette fonction répond à une question différente de celle de la PMF.

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.

Traçons cette fonction "en escalier" pour notre exemple du dé.

Exemple

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

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

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

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

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

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

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

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

3.10 Variable Aléatoire Indicatrice

Enfin, nous introduisons un outil simple mais qui s'avérera extraordinairement puissant pour les preuves, notamment celles concernant l'espérance.

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 $p = P(A)$.

C'est un simple interrupteur "on/off".

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.

3.11 Espérance d'une variable aléatoire discrète

Maintenant que nous avons défini les variables aléatoires discrètes et leur distribution (PMF), l'étape suivante est de résumer ces distributions. La mesure la plus importante est leur "centre", ou leur valeur moyenne.

Définition : Espérance (cas discret)

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

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

Cette formule est une moyenne pondérée de toutes les valeurs possibles.

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.

L'exemple le plus simple est le lancer d'un dé.

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.

3.12 Linéarité de l'espérance

Le calcul de l'espérance deviendrait très fastidieux si nous devions toujours utiliser la définition. Heureusement, l'espérance possède une propriété fondamentale qui simplifie énormément les calculs.

Théorème : Linéarité de l'espérance

Pour toutes variables aléatoires X et Y (discrètes ou continues), 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.

Preuve

La première propriété est directe.

$$\bullet \text{ Cas discret : } E(cX) = \sum_x (cx)P(X = x) = c \sum_x xP(X = x) = cE(X)$$

Pour la seconde, $E(X + Y) = E(X) + E(Y)$:

Cas discret : Soit $S = X + Y$. L'espérance $E(S)$ se calcule en sommant sur toutes les paires possibles (x, y) avec la PMF jointe $P(X = x, Y = y)$:

$$\begin{aligned} E(X + Y) &= \sum_x \sum_y (x + y)P(X = x, Y = y) \\ &= \sum_x \sum_y xP(X = x, Y = y) + \sum_x \sum_y yP(X = x, Y = y) \\ &= \sum_x x \left(\sum_y P(X = x, Y = y) \right) + \sum_y y \left(\sum_x P(X = x, Y = y) \right) \end{aligned}$$

Par la loi des probabilités marginales, la somme interne $\sum_y P(X = x, Y = y)$ est $P(X = x)$, et de même $\sum_x P(X = x, Y = y) = P(Y = y)$.

$$E(X + Y) = \sum_x xP(X = x) + \sum_y yP(Y = y) = E(X) + E(Y)$$

Notez que l'indépendance n'a jamais été requise pour cette preuve.

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.

Cette propriété rend le calcul de l'espérance d'une somme trivial, comme le montre l'exemple des deux dés.

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 :

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

3.13 Espérance de la loi binomiale

Nous pouvons maintenant utiliser cette puissante propriété de linéarité pour trouver l'espérance de nos distributions de référence, en évitant des sommes complexes.

Théorème : Espérance de la loi binomiale

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

Ce résultat est profondément intuitif.

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.

La preuve formelle est un exemple parfait de l'élégance de la linéarité, utilisant les variables indicatrices.

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 :

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

3.14 Espérance de la loi géométrique

Calculons maintenant l'espérance pour la loi qui modélise le temps d'attente.

Théorème : Espérance de la loi géométrique

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

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

L'intuition est aussi très forte ici :

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 $1/p$.

Contrairement à la loi binomiale, la preuve la plus directe ne repose pas sur la linéarité mais sur une manipulation de séries.

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 $\frac{q}{p}$.

3.15 Loi du statisticien inconscient (LOTUS)

Souvent, nous ne sommes pas intéressés par l'espérance de X elle-même, mais par l'espérance d'une fonction de X , par exemple $E(X^2)$ ou $E(e^X)$.

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 :

$$\bullet \text{ Cas discret : } E[g(X)] = \sum_x g(x) P(X = x)$$

Ce théorème est utile car il évite d'avoir à trouver la distribution (PMF) de $g(X)$.

La preuve dans le cas discret consiste simplement à regrouper les termes.

Preuve

Nous montrons la preuve pour le cas discret.

Soit $Y = g(X)$. Par définition, l'espérance de Y est $E(Y) = \sum_y y P(Y = y)$. L'ensemble des valeurs y que Y peut prendre est $\{g(x) \mid x \in \text{support de } X\}$. Pour une valeur y donnée,

l'événement $\{Y = y\}$ est l'union de tous les événements $\{X = x\}$ tels que $g(x) = y$.

$$P(Y = y) = P(g(X) = y) = \sum_{x:g(x)=y} P(X = x)$$

En substituant cela dans la définition de $E(Y)$:

$$E(Y) = \sum_y y \left(\sum_{x:g(x)=y} P(X = x) \right)$$

On peut réécrire y comme $g(x)$ à l'intérieur de la seconde somme :

$$E(g(X)) = \sum_y \sum_{x:g(x)=y} g(x) P(X = x)$$

Cette double somme parcourt toutes les valeurs de y , et pour chaque y , elle parcourt tous les x correspondants. Cela revient à simplement sommer sur tous les x possibles dès le départ :

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

Ce théorème justifie son nom : c'est ce que l'on ferait "inconsciemment".

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.

Utilisons ce théorème pour calculer $E(X^2)$ pour notre dé.

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

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

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

3.16 Variance

L'espérance nous donne le centre d'une distribution, mais elle ne dit rien sur sa "largeur" ou sa "dispersion". C'est le rôle de la variance.

Définition : Variance et écart-type

La **variance** d'une variable aléatoire X mesure la dispersion de sa distribution autour de son espérance $\mu = E(X)$. Elle est définie par :

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

Concrètement, cela se traduit par (en utilisant LOTUS avec $g(x) = (x - \mu)^2$) :

$$\cdot \text{ Cas discret : } \text{Var}(X) = \sum_x (x - \mu)^2 P(X = x)$$

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

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

L'idée est de mesurer l'écart quadratique moyen à l'espérance.

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.

La définition $E[(X - \mu)^2]$ est excellente pour l'interprétation, mais pénible pour le calcul. Une formule alternative est presque toujours utilisée.

Théorème : Formule de calcul de la variance

Pour toute variable aléatoire X (discrète ou continue), une formule plus pratique pour le calcul de la variance est :

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

La preuve est une simple expansion algébrique utilisant la linéarité de l'espérance.

Preuve

Soit $\mu = E(X)$. On part de la définition de la variance :

$$\begin{aligned}\text{Var}(X) &= E[(X - \mu)^2] \\ &= E[X^2 - 2X\mu + \mu^2] \quad (\text{On développe le carré}) \\ &= E(X^2) - E(2\mu X) + E(\mu^2) \quad (\text{Par linéarité de l'espérance}) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \quad (\text{Car } 2\mu \text{ et } \mu^2 \text{ sont des constantes}) \\ &= E(X^2) - 2\mu(\mu) + \mu^2 \quad (\text{Car } E(X) = \mu) \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 = E(X^2) - [E(X)]^2\end{aligned}$$

Nous pouvons maintenant calculer la variance de notre lancer de dé.

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 :

$$\begin{aligned}\text{Var}(X) &= E(X^2) - [E(X)]^2 = \frac{91}{6} - (3.5)^2 \\ &= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} \\ &= \frac{182}{12} - \frac{147}{12} = \frac{35}{12} \approx 2.917\end{aligned}$$

L'écart-type est $\text{SD}(X) = \sqrt{35/12} \approx 1.708$.

3.17 Exercices

Exercice 1 : Loi Binomiale (Quiz)

Un étudiant répond au hasard à un QCM de 10 questions. Chaque question a 4 choix de réponse, dont un seul est correct. Soit X le nombre de bonnes réponses.

1. Quelle loi suit X ? Précisez ses paramètres.
2. Quelle est la probabilité que l'étudiant ait exactement 5 bonnes réponses ?
3. Quelle est la probabilité que l'étudiant ait au moins une bonne réponse ?

Exercice 2 : Loi Binomiale (Contrôle Qualité)

Une usine produit des ampoules. 5% des ampoules sont défectueuses. On prélève un lot de 20 ampoules. Soit X le nombre d'ampoules défectueuses dans le lot.

1. Quelle loi suit X ? (On suppose le prélèvement "avec remise" ou d'une production très grande).
2. Quelle est la probabilité qu'il n'y ait aucune ampoule défectueuse ?
3. Quelle est la probabilité qu'il y ait exactement deux ampoules défectueuses ?

Exercice 3 : Espérance et Variance (Binomiale)

Un archer touche la cible avec une probabilité $p = 0.8$ à chaque tir. Il tire $n = 40$ flèches. Soit X le nombre de tirs réussis.

1. Calculer l'espérance $E(X)$.
2. Calculer la variance $\text{Var}(X)$ et l'écart-type $\text{SD}(X)$.

Exercice 4 : Loi de Bernoulli (Indicatrice)

Soit A un événement avec $P(A) = p$. Soit I_A la variable indicatrice de A .

1. Écrire la PMF de I_A .
2. Calculer $E(I_A)$.
3. Calculer $\text{Var}(I_A)$ en utilisant $\text{Var}(X) = E(X^2) - [E(X)]^2$. (Indice : $I_A^2 = I_A$).

Exercice 5 : Loi de Poisson (Emails)

Un serveur de messagerie reçoit en moyenne 2 emails "spam" par minute. Soit X le nombre de spams reçus en une minute.

1. Quelle loi suit X ?
2. Quelle est la probabilité de recevoir exactement 3 spams en une minute ?
3. Quelle est la probabilité de recevoir au plus 2 spams en une minute ?

Exercice 6 : Loi de Poisson (Échelle de temps)

Une substance radioactive émet en moyenne $\lambda = 4$ particules par seconde. Soit Y le nombre de particules émises en 3 secondes.

1. Quelle est la loi de Y ? (Indice : ajuster le paramètre λ).
2. Quelle est la probabilité que $Y = 10$?

Exercice 7 : Approximation de Poisson (Binomiale)

Un livre de 500 pages contient 1000 fautes de frappe distribuées au hasard. Soit X le nombre de fautes de frappe sur une page donnée.

1. Quelle est la loi exacte de X ? (On suppose qu'une faute ne peut pas être à cheval sur deux pages).
2. Par quelle loi peut-on approximer X ? Précisez le paramètre.
3. En utilisant l'approximation, calculez la probabilité qu'une page choisie au hasard contienne au moins une faute.

Exercice 8 : Loi Géométrique (Échecs avant succès)

On lance un dé équilibré jusqu'à obtenir un 6. Soit X le nombre d'échecs (lancers qui ne sont pas 6) avant d'obtenir le premier 6.

1. Quelle loi suit X ? Précisez le paramètre p .
2. Quelle est la probabilité d'échouer exactement 3 fois ? (c-à-d, le 6 arrive au 4ème lancer).
3. Quelle est l'espérance du nombre d'échecs $E(X)$?

Exercice 9 : Loi Géométrique (Propriété)

Soit $X \sim \text{Geom}(p)$ (comptant les échecs). Quelle est la probabilité $P(X \geq k)$? C'est-à-dire, la probabilité d'avoir au moins k échecs.

Exercice 10 : Loi Hypergéométrique (Comité)

Un club est composé de 12 hommes et 8 femmes. On choisit un comité de 5 personnes au hasard. Soit X le nombre de femmes dans le comité.

1. Quelle loi suit X ? Précisez les paramètres.
2. Quelle est la probabilité que le comité soit composé d'exactly 2 femmes ?

Exercice 11 : Loi Hypergéométrique (Pêche)

Un lac contient 100 poissons, dont 10 ont été marqués. Un pêcheur attrape 8 poissons (sans remise). Soit X le nombre de poissons marqués parmi les 8 attrapés.

1. Quelle est la loi de X ?
2. Quelle est la probabilité qu'il n'attrape aucun poisson marqué ?

Exercice 12 : Binomiale vs Hypergéométrique

Expliquez la différence fondamentale entre la loi Binomiale et la loi Hypergéométrique. Dans quel cas la loi Binomiale est-elle une bonne approximation de la loi Hypergéométrique ?

Exercice 13 : PMF (Trouver la constante)

Soit X une variable aléatoire discrète dont la PMF est donnée par $P(X = k) = c \cdot k^2$ pour $k \in \{1, 2, 3\}$. Pour toutes les autres valeurs, $P(X = x) = 0$.

1. Trouvez la valeur de la constante c .
2. Calculez $P(X \geq 2)$.

Exercice 14 : Espérance (Jeu de Hasard)

Un joueur paie 5€ pour jouer à un jeu. Il lance deux dés. Il gagne S euros, où S est la somme des deux dés. Soit G son gain net (gain - mise).

1. Rappeler $E(S)$ (espérance de la somme de two dés).
2. Calculer $E(G)$. Le jeu est-il équitable ?

Exercice 15 : LOTUS (Jeu de Hasard 2)

On lance un dé équilibré. Soit X le résultat. Vous gagnez $g(X) = (X - 3)^2$ euros.

1. Calculez l'espérance de votre gain, $E[g(X)]$.

Exercice 16 : CDF (Fonction de Répartition)

Soit X une variable aléatoire avec la PMF suivante : $P(X = 0) = 0.2$, $P(X = 1) = 0.5$, $P(X = 2) = 0.3$.

1. Écrivez la fonction de répartition $F_X(x) = P(X \leq x)$.
2. Calculez $P(0 < X \leq 2)$.
3. Calculez $P(X > 1)$.

Exercice 17 : Variance (Calcul)

Pour la variable X de l'exercice 16 :

1. Calculez $E(X)$.
2. Calculez $E(X^2)$ en utilisant LOTUS.
3. Calculez $\text{Var}(X)$ en utilisant la formule $E(X^2) - [E(X)]^2$.

Exercice 18 : Propriétés de la Variance

Soit X une variable aléatoire avec $E(X) = 10$ et $\text{Var}(X) = 4$. Soit $Y = 5X - 2$.

1. Calculez $E(Y)$.
2. Calculez $\text{Var}(Y)$.

Exercice 19 : Linéarité de l'Espérance (Mélange)

On lance un dé (résultat D) et une pièce (résultat P , 1 pour Pile, 0 for Face). Soit $X = D + 3P$. Calculez $E(X)$ en utilisant la linéarité de l'espérance.

Exercice 20 : Indicateurs (Problème des Chapeaux)

n personnes jettent leur chapeau au centre d'une pièce. Les chapeaux sont mélangés, et chaque personne en reprend un au hasard. Soit X le nombre de personnes qui reprennent leur propre chapeau.

1. Définir X comme une somme de n variables indicatrices I_i . Que représente I_i ?
2. Quelle est la probabilité $P(I_i = 1)$? (C-à-d, la probabilité que la personne i reprenne son chapeau).
3. Calculez $E(X)$ en utilisant la linéarité.

Exercice 21 : Indicateurs (Collectionneurs)

On achète $n = 5$ boîtes de céréales. Chaque boîte contient une figurine au hasard parmi $k = 4$ types de figurines (types A, B, C, D). Soit X le nombre de types de figurines *distincts* que nous avons obtenus.

1. Soit I_A l'indicateur que nous avons obtenu au moins une figurine de type A.
2. Calculez $P(I_A = 0)$. (Probabilité de n'avoir *aucune* figurine A dans les 5 boîtes).
3. Calculez $E(I_A)$.
4. Exprimez X en fonction d'indicateurs I_A, I_B, I_C, I_D et calculez $E(X)$.

Exercice 22 : Variance (Propriétés)

Prouvez que $\text{Var}(aX + b) = a^2\text{Var}(X)$ pour des constantes a et b .

Exercice 23 : Identifier la Loi

Pour chaque scénario, identifiez la loi discrète la plus appropriée (Bernoulli, Binomiale, Hypergéométrique, Géométrique, Poisson).

1. On compte le nombre de "Face" lors de 20 lancers de pièce.
2. On compte le nombre d'accidents à une intersection un jour donné (sachant un taux moyen de 1.5/jour).
3. On compte le nombre de Rois dans une main de 5 cartes tirées d'un jeu de 52 cartes.
4. On compte le nombre de lancers de dé nécessaires avant d'obtenir le premier 3 (en comptant les échecs).
5. On vérifie si un seul composant électronique est défectueux ou non.

Exercice 24 : Espérance (Loi Hypergéométrique)

Soit $X \sim \text{HG}(w, b, m)$ (tirage de m boules parmi w blanches et b noires). En utilisant des variables indicatrices, montrez que $E(X) = m \left(\frac{w}{w+b} \right)$. (Indice : Soit I_j l'indicateur que la j -ème boule tirée est blanche, pour $j = 1 \dots m$. $X = \sum I_j$. Calculez $E(I_j)$.)

Exercice 25 : Variance (Poisson)

On admet que si $X \sim \text{Bin}(n, p)$, $\text{Var}(X) = np(1-p)$. En utilisant l'approximation Poisson $X_n \sim \text{Bin}(n, \lambda/n)$, que devient la variance lorsque $n \rightarrow \infty$?

3.18 Corrections des Exercices

Correction Exercice 1 : Loi Binomiale (Quiz)

1. X suit une loi Binomiale, car c'est la somme de $n = 10$ succès indépendants (bonnes réponses), chacun avec une probabilité $p = 1/4 = 0.25$. $X \sim \text{Bin}(10, 0.25)$. 2. $P(X = 5) = \binom{10}{5} (0.25)^5 (1 - 0.25)^{10-5} = 252 \times (0.25)^5 (0.75)^5 \approx 0.0584$. 3. $P(X \geq 1) = 1 - P(X = 0)$. $P(X = 0) = \binom{10}{0} (0.25)^0 (0.75)^{10} = (0.75)^{10} \approx 0.0563$. $P(X \geq 1) = 1 - 0.0563 = 0.9437$.

Correction Exercice 2 : Loi Binomiale (Contrôle Qualité)

1. $X \sim \text{Bin}(n = 20, p = 0.05)$. 2. $P(X = 0) = \binom{20}{0}(0.05)^0(0.95)^{20} = (0.95)^{20} \approx 0.3585$. 3. $P(X = 2) = \binom{20}{2}(0.05)^2(0.95)^{18} = 190 \times (0.0025) \times (0.95)^{18} \approx 0.1887$.

Correction Exercice 3 : Espérance et Variance (Binomiale)

1. $X \sim \text{Bin}(40, 0.8)$. L'espérance est $E(X) = np = 40 \times 0.8 = 32$. (On s'attend à 32 tirs réussis). 2. $\text{Var}(X) = np(1 - p) = 40 \times 0.8 \times 0.2 = 6.4$. $\text{SD}(X) = \sqrt{\text{Var}(X)} = \sqrt{6.4} \approx 2.53$.

Correction Exercice 4 : Loi de Bernoulli (Indicatrice)

1. I_A suit la loi de Bernoulli $\text{Bern}(p)$. $P(I_A = 1) = p$ et $P(I_A = 0) = 1 - p$. 2. $E(I_A) = 1 \cdot P(I_A = 1) + 0 \cdot P(I_A = 0) = 1 \cdot p + 0 = p$. 3. On utilise LOTUS pour $E(I_A^2)$: $E(I_A^2) = (1^2) \cdot P(I_A = 1) + (0^2) \cdot P(I_A = 0) = 1 \cdot p + 0 = p$. (Note : $I_A^2 = I_A$ car $1^2 = 1$ et $0^2 = 0$, donc $E(I_A^2) = E(I_A) = p$). $\text{Var}(I_A) = E(I_A^2) - [E(I_A)]^2 = p - p^2 = p(1 - p)$.

Correction Exercice 5 : Loi de Poisson (Emails)

1. X suit une loi de Poisson de paramètre $\lambda = 2$. $X \sim \text{Poisson}(2)$. 2. $P(X = 3) = \frac{e^{-2}2^3}{3!} = \frac{e^{-2} \times 8}{6} \approx 0.1804$. 3. $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$. $P(X = 0) = \frac{e^{-2}2^0}{0!} = e^{-2}$. $P(X = 1) = \frac{e^{-2}2^1}{1!} = 2e^{-2}$. $P(X = 2) = \frac{e^{-2}2^2}{2!} = 2e^{-2}$. $P(X \leq 2) = e^{-2}(1 + 2 + 2) = 5e^{-2} \approx 0.6767$.

Correction Exercice 6 : Loi de Poisson (Échelle de temps)

1. Le taux est $\lambda = 4$ par seconde. Sur 3 secondes, le taux moyen est $\lambda' = 4 \times 3 = 12$. $Y \sim \text{Poisson}(12)$. 2. $P(Y = 10) = \frac{e^{-12}12^{10}}{10!} \approx 0.1048$.

Correction Exercice 7 : Approximation de Poisson (Binomiale)

1. C'est un tirage sans remise (une faute ne peut pas être comptée deux fois). C'est Hypergéométrique. (Ou Binomiale si on considère que chaque mot a une prob p d'être une faute, mais n (nb de mots) est inconnu). On peut aussi voir cela comme $n = 1000$ essais (fautes) de Bernoulli où le succès est "tomber sur la page X" (prob $p = 1/500$). $X \sim \text{Bin}(1000, 1/500)$. 2. La loi Binomiale $\text{Bin}(n = 1000, p = 1/500)$ a n grand et p petit. On approxime par Poisson avec $\lambda = np = 1000 \times (1/500) = 2$. $X \approx \text{Poisson}(2)$. 3. $P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{e^{-2}2^0}{0!} = 1 - e^{-2} \approx 1 - 0.1353 = 0.8647$.

Correction Exercice 8 : Loi Géométrique (Échecs avant succès)

1. X compte le nombre d'échecs avant le premier succès. Le succès est "obtenir 6", $p = 1/6$. $X \sim \text{Geom}(p = 1/6)$. 2. On cherche $P(X = 3)$. $P(X = k) = (1 - p)^k p$. $P(X = 3) = (5/6)^3 \times (1/6) = \frac{125}{216} \times \frac{1}{6} = \frac{125}{1296} \approx 0.0965$. 3. $E(X) = \frac{q}{p} = \frac{1-p}{p} = \frac{5/6}{1/6} = 5$. (On s'attend à 5 échecs en moyenne).

Correction Exercice 9 : Loi Géométrique (Propriété)

$P(X \geq k)$ est la probabilité d'avoir au moins k échecs. Cela signifie que les k premiers essais ont *tous* été des échecs. La probabilité d'un échec est $q = 1 - p$. La probabilité de k échecs consécutifs est q^k . (Après ces k échecs, peu importe ce qui se passe, la condition $X \geq k$ est remplie). Donc, $P(X \geq k) = (1 - p)^k$.

Correction Exercice 10 : Loi Hypergéométrique (Comité)

1. Tirage sans remise. Population totale $w + b = 20$. On s'intéresse aux femmes (disons "blanches", $w = 8$). Les hommes sont "noires" ($b = 12$). On tire $m = 5$ personnes. $X \sim \text{HG}(w = 8, b = 12, m = 5)$. 2. $P(X = 2) = \frac{\binom{w}{k} \binom{b}{m-k}}{\binom{w+b}{m}} = \frac{\binom{8}{2} \binom{12}{3}}{\binom{20}{5}} = \frac{28 \times 220}{15504} = \frac{6160}{15504} \approx 0.3973$.

Correction Exercice 11 : Loi Hypergéométrique (Pêche)

1. $w = 10$ (marqués), $b = 90$ (non marqués). $w + b = 100$. $m = 8$ (tirage). $X \sim \text{HG}(w = 10, b = 90, m = 8)$. 2. $P(X = 0) = \frac{\binom{10}{0} \binom{90}{8}}{\binom{100}{8}} = \frac{1 \times \frac{90!}{8!92!}}{\frac{100!}{8!92!}} = \frac{90!92!}{82!100!} = \frac{90 \dots 83}{100 \dots 93} \approx 0.4166$.

Correction Exercice 12 : Binomiale vs Hypergéométrique

La différence est l'indépendance des tirages.

- **Binomiale** : Modélise n tirages *avec remise* (ou indépendants). La probabilité de succès p est constante.
- **Hypergéométrique** : Modélise m tirages *sans remise* d'une population finie. La probabilité de succès change à chaque tirage.

La Binomiale approxime bien l'Hypergéométrique lorsque la taille de la population ($w + b$) est très grande par rapport à la taille de l'échantillon (m). Dans ce cas, $p \approx w/(w + b)$ est presque constant.

Correction Exercice 13 : PMF (Trouver la constante)

1. La somme des probabilités doit valoir 1 : $\sum P(X = k) = 1$. $P(X = 1) + P(X = 2) + P(X = 3) = 1$ $c \cdot 1^2 + c \cdot 2^2 + c \cdot 3^2 = 1$ $c(1 + 4 + 9) = 1$ $\implies 14c = 1 \implies c = 1/14$. 2. $P(X \geq 2) = P(X = 2) + P(X = 3) = c \cdot 2^2 + c \cdot 3^2 = 4c + 9c = 13c = 13/14$.

Correction Exercice 14 : Espérance (Jeu de Hasard)

1. $E(S) = E(D_1 + D_2) = E(D_1) + E(D_2) = 3.5 + 3.5 = 7$. 2. $G = S - 5$ (Gain = Somme - Mise). Par linéarité : $E(G) = E(S - 5) = E(S) - E(5) = E(S) - 5$. $E(G) = 7 - 5 = 2$. L'espérance de gain net est de 2€. Le jeu est très en faveur du joueur (et non équitable).

Correction Exercice 15 : LOTUS (Jeu de Hasard 2)

On cherche $E[g(X)] = E[(X - 3)^2]$. On utilise LOTUS : $E[g(X)] = \sum g(x)P(X = x)$. $E[(X - 3)^2] = \sum_{k=1}^6 (k - 3)^2 P(X = k) = \sum_{k=1}^6 (k - 3)^2 \cdot (1/6) = \frac{1}{6} [(1 - 3)^2 + (2 - 3)^2 + (3 - 3)^2 + (4 - 3)^2 + (5 - 3)^2 + (6 - 3)^2] = \frac{1}{6} [(-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2 + 3^2] = \frac{1}{6} [4 + 1 + 0 + 1 + 4 + 9] = \frac{19}{6} \approx 3.167\text{£}$.

Correction Exercice 16 : CDF (Fonction de Répartition)

1. $F_X(x)$ accumule les probabilités : $F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 0.2 & \text{si } 0 \leq x < 1 \\ 0.2 + 0.5 = 0.7 & \text{si } 1 \leq x < 2 \\ 0.7 + 0.3 = 1.0 & \text{si } x \geq 2 \end{cases}$ 2. $P(0 < X \leq 2) = P(X = 1) + P(X = 2) = 0.5 + 0.3 = 0.8$. (Alternativement : $F_X(2) - F_X(0) = 1.0 - 0.2 = 0.8$). 3. $P(X > 1) = P(X = 2) = 0.3$. (Alternativement : $1 - P(X \leq 1) = 1 - F_X(1) = 1 - 0.7 = 0.3$).

Correction Exercice 17 : Variance (Calcul)

1. $E(X) = \sum xP(X = x) = (0)(0.2) + (1)(0.5) + (2)(0.3) = 0 + 0.5 + 0.6 = 1.1$. 2. $E(X^2) = \sum x^2P(X = x) = (0^2)(0.2) + (1^2)(0.5) + (2^2)(0.3) = 0 + 0.5 + (4)(0.3) = 0.5 + 1.2 = 1.7$. 3. $\text{Var}(X) = E(X^2) - [E(X)]^2 = 1.7 - (1.1)^2 = 1.7 - 1.21 = 0.49$.

Correction Exercice 18 : Propriétés de la Variance

1. $E(Y) = E(5X - 2) = E(5X) - E(2) = 5E(X) - 2 = 5(10) - 2 = 48$. 2. $\text{Var}(Y) = \text{Var}(5X - 2) = \text{Var}(5X) = 5^2 \text{Var}(X) = 25 \times 4 = 100$. (La constante $b = -2$ n'affecte pas la dispersion).

Correction Exercice 19 : Linéarité de l'Espérance (Mélange)

On cherche $E(X) = E(D + 3P)$. Par linéarité : $E(X) = E(D) + E(3P) = E(D) + 3E(P)$. $E(D) = 3.5$ (espérance d'un dé). P est une Bernoulli Bern(0.5). $E(P) = p = 0.5$. $E(X) = 3.5 + 3(0.5) = 3.5 + 1.5 = 5.0$.

Correction Exercice 20 : Indicateurs (Problème des Chapeaux)

1. I_i est la variable indicatrice de l'événement "la personne i reprend son propre chapeau". $X = \sum_{i=1}^n I_i$. 2. Il y a $n!$ permutations (arrangements) possibles des chapeaux. Il y a $(n - 1)!$ permutations où la personne i a son propre chapeau. $P(I_i = 1) = \frac{(n-1)!}{n!} = \frac{1}{n}$. 3. Par linéarité : $E(X) = E(\sum I_i) = \sum E(I_i)$. $E(I_i) = P(I_i = 1) = 1/n$. $E(X) = \sum_{i=1}^n (1/n) = n \times (1/n) = 1$.

En moyenne, quel que soit n , une seule personne reprend son chapeau !

Correction Exercice 21 : Indicateurs (Collectionneurs)

1. $I_A = 1$ si on a au moins une fig. A, $I_A = 0$ sinon. 2. $P(I_A = 0) = P(\text{n'avoir aucune fig. A})$. À chaque boîte, $P(\text{pas A}) = 3/4$. Pour 5 boîtes indépendantes : $P(I_A = 0) = (3/4)^5$. 3. $E(I_A) = P(I_A = 1) = 1 - P(I_A = 0) = 1 - (3/4)^5$. 4. X est le nombre de types distincts, $X = I_A + I_B + I_C + I_D$. Par linéarité : $E(X) = E(I_A) + E(I_B) + E(I_C) + E(I_D)$. Par symétrie, $E(I_A) = E(I_B) = E(I_C) = E(I_D) = 1 - (3/4)^5$. $E(X) = 4 \times (1 - (3/4)^5) \approx 4 \times (1 - 0.2373) = 4 \times 0.7627 = 3.0508$.

Correction Exercice 22 : Variance (Propriétés)

Soit $\mu = E(X)$. Alors $E(aX + b) = aE(X) + b = a\mu + b$. Par définition de la variance : $\text{Var}(aX + b) = E[(aX + b) - E[aX + b]]^2 = E[(aX + b) - (a\mu + b)]^2 = E[(aX - a\mu)]^2 = E[a(X - \mu)]^2 = E[a^2(X - \mu)^2] = a^2 E[(X - \mu)^2] = a^2 \text{Var}(X)$ (car a^2 est une constante).

Correction Exercice 23 : Identifier la Loi

1. **Binomiale** ($n=20$ essais de Bernoulli indépendants). 2. **Poisson** (comptage d'événements rares dans un intervalle fixe). 3. **Hypergéométrique** (tirage sans remise d'une population finie). 4. **Géométrique** (comptage d'échecs avant le premier succès). 5. **Bernoulli** (un seul essai, deux issues).

Correction Exercice 24 : Espérance (Loi Hypergéométrique)

Soit $I_j = 1$ si la j -ème boule tirée (pour $j = 1 \dots m$) est blanche, $I_j = 0$ sinon. Le nombre total de blanches est $X = \sum_{j=1}^m I_j$. Par linéarité, $E(X) = \sum_{j=1}^m E(I_j)$. $E(I_j) = P(I_j = 1)$. Quelle est la probabilité que la j -ème boule tirée soit blanche ? Par symétrie (ou en considérant un tirage aléatoire), n'importe quelle boule a la même probabilité d'être à la j -ème position. Il y a w blanches sur $w + b$ boules au total. $P(I_j = 1) = \frac{w}{w+b}$ (c'est vrai pour $j = 1, j = 2, \dots, j = m$). $E(X) = \sum_{j=1}^m \frac{w}{w+b} = m \left(\frac{w}{w+b} \right)$.

Correction Exercice 25 : Variance (Poisson)

$X_n \sim \text{Bin}(n, \lambda/n)$. $\text{Var}(X_n) = np(1-p) = n(\lambda/n)(1 - \lambda/n) = \lambda(1 - \lambda/n)$. Lorsque $n \rightarrow \infty$: $\lim_{n \rightarrow \infty} \text{Var}(X_n) = \lim_{n \rightarrow \infty} \lambda(1 - \lambda/n)$ Puisque $\lambda/n \rightarrow 0$, la limite est $\lambda(1 - 0) = \lambda$. On en déduit que pour $X \sim \text{Poisson}(\lambda)$, la variance est $\text{Var}(X) = \lambda$. (L'espérance et la variance sont égales pour la loi de Poisson).

3.19 Exercices Python

Les exercices suivants appliquent les concepts de variables aléatoires discrètes, de leurs lois de probabilité (PMF, CDF) et de leurs caractéristiques (espérance, variance) au jeu de données "Taxis" de Seaborn.

```
import pandas as pd
import seaborn as sns
import math

# Charger le dataset Taxis
df = sns.load_dataset("taxis")

# 'df' est maintenant notre Univers S.
# S_total = len(df)
```

Exercice 1 : PMF (Fonction de Masse)

Soit X la variable aléatoire discrète représentant le nombre de passagers (**passengers**) dans un taxi.

Votre tâche :

- Calculer la Fonction de Masse (PMF) de X . Trouvez $P(X = 1)$, $P(X = 2)$, $P(X = 3)$, etc., pour toutes les valeurs non nulles.

2. Vérifier que $\sum_k P(X = k) = 1$.

Exercice 2 : CDF (Fonction de Répartition)

En utilisant la variable X (`passengers`) de l'exercice 1 :

Votre tâche :

1. Calculer la valeur de la Fonction de Répartition (CDF) au point 2, c'est-à-dire $F_X(2) = P(X \leq 2)$.
2. Calculer $P(X > 3)$ en utilisant la CDF.

Exercice 3 : Espérance d'une VA Discrète

En utilisant la variable X (`passengers`) et sa PMF $P_X(k)$ calculée à l'exercice 1, calculez l'espérance de X .

L'espérance est $E[X] = \sum_k k \cdot P(X = k)$.

Votre tâche :

1. Lister les valeurs k possibles pour `passengers`.
2. Pour chaque k , calculer le produit $k \times P(X = k)$.
3. Sommer ces produits pour obtenir $E[X]$.
4. (Vérification) Comparez votre résultat à la moyenne directe `df['passengers'].mean()`.

Exercice 4 : Variance d'une VA Discrète

Calculez la variance de X (`passengers`) en utilisant la formule $\text{Var}(X) = E[X^2] - (E[X])^2$.

Votre tâche :

1. Utiliser l'espérance $E[X]$ calculée à l'exercice 3.
2. Calculer l'espérance du carré, $E[X^2] = \sum_k k^2 \cdot P(X = k)$.
3. Appliquer la formule de la variance.
4. (Vérification) Comparez votre résultat à `df['passengers'].var()`.

Exercice 5 : Variable Aléatoire Indicatrice et Espérance

Soit I_A une variable aléatoire indicatrice pour l'événement $A = \text{"le passager a donné un pour-boire"}$.

Votre tâche :

1. Créer une nouvelle colonne `got_tip` dans le DataFrame.
2. Assigner $I_A = 1$ si `tip > 0`, et $I_A = 0$ sinon.
3. Calculer l'espérance de cette variable, $E[I_A]$.
4. Constatez que $E[I_A]$ est exactement la probabilité $P(A)$ que le passager donne un pour-boire.

Exercice 6 : Loi de Bernoulli

Soit X une variable aléatoire de Bernoulli modélisant le type de paiement. $X = 1$ si le paiement est "credit_card" (Succès) et $X = 0$ si c'est "cash" (Échec).

Votre tâche :

1. Calculer p , la probabilité de succès, $P(X = 1)$.
2. Calculer $1 - p$, la probabilité d'échec, $P(X = 0)$.
3. Quelle est l'espérance $E[X]$ et la variance $\text{Var}(X)$ de cette variable ? (Utilisez les formules p et $p(1 - p)$).

Exercice 7 : Loi Binomiale (PMF et Espérance)

Nous utilisons le paramètre p (probabilité de payer par carte) de l'exercice 6. On observe un échantillon de $n = 10$ courses indépendantes. Soit Y le nombre de courses payées par carte dans cet échantillon. Y suit une loi binomiale $Y \sim \text{Bin}(n = 10, p)$.

Votre tâche :

1. En utilisant la PMF de la loi binomiale, calculer la probabilité d'avoir exactement $k = 4$

paiements par carte, $P(Y = 4)$.

2. Calculer l'espérance $E[Y]$ et la variance $\text{Var}(Y)$ de cette variable binomiale.

Exercice 8 : Loi Géométrique

Nous observons les trajets un par un (supposés indépendants) jusqu'à trouver notre premier "succès". Le "succès" est défini comme "trouver un trajet avec un pourboire (**tip**) de plus de 5 dollars".

Soit X le nombre d'échecs (trajets avec **tip** ≤ 5) avant le premier succès. X suit une loi géométrique $X \sim \text{Geom}(p)$.

Votre tâche :

1. Calculer p , la probabilité de "succès" (trouver un **tip** > 5).
2. En utilisant la PMF de la loi géométrique, calculer la probabilité d'avoir exactement $k = 10$ échecs avant le premier succès.

Exercice 9 : Loi Hypergéométrique

Considérons le tirage **sans remise**. Notre population est l'ensemble du **df**. Soit w le nombre total de paiements par "credit_card" et b le nombre total de paiements "cash". On tire un échantillon de $m = 20$ trajets. Soit Z le nombre de paiements par carte dans cet échantillon. Z suit une loi Hypergéométrique.

Votre tâche :

1. Trouver w , b et $m = 20$.
2. En utilisant la PMF de la loi hypergéométrique, calculer la probabilité d'avoir exactement $k = 15$ paiements par carte dans l'échantillon, $P(Z = 15)$.

Exercice 10 : Loi de Poisson

La loi de Poisson modélise le nombre d'événements dans un intervalle de temps. Nous voulons modéliser le nombre de courses par heure dans un quartier.

Votre tâche :

1. Filtrer le DataFrame pour ne garder que les courses dans le quartier "Manhattan" (`pickup_borough == 'Manhattan'`).
2. Convertir la colonne `pickup` en datetime.
3. Agréger les données pour compter le nombre de courses par heure (vous pouvez "arrondir" l'heure de début de course).
4. Calculer λ , le taux moyen de courses par heure à Manhattan (l'espérance de la loi de Poisson).
5. En utilisant la PMF de la loi de Poisson avec ce λ , calculer la probabilité qu'il y ait exactement $k = 50$ courses lors d'une heure donnée, $P(X = 50)$.

4 Variables Aléatoires Continues

4.1 Fonction de Densité de Probabilité (PDF)

Nous passons maintenant aux variables aléatoires qui peuvent prendre n'importe quelle valeur dans un intervalle, comme la taille d'une personne ou le temps d'attente exact. Pour ces variables, la notion de PMF n'a plus de sens, car la probabilité d'obtenir une valeur *exacte* est nulle. Nous introduisons donc le concept de densité.

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)

Il est crucial de comprendre que $f(x)$ n'est *pas* une probabilité.

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 :

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

Cette distinction est fondamentale.

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.

Vérifions un exemple simple.

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 $P(X \leq 0.5)$?

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

4.2 Fonction de Répartition (CDF)

Comme dans le cas discret, nous pouvons définir une fonction qui accumule la probabilité. Pour le cas continu, cette accumulation se fait via une intégrale.

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.

La CDF est l'intégrale de la PDF, et inversement, la PDF est la dérivée de la CDF.

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 (par le théorème fondamental de l'analyse) :

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

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

La CDF est souvent le moyen le plus simple de calculer des probabilités sur des intervalles.

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 ($P(X = a) = 0$).

Calculons la CDF de notre exemple précédent.

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

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

4.3 Espérance et Variance (Cas Continu)

Les concepts d'espérance et de variance s'étendent naturellement au cas continu, en remplaçant les sommes par des intégrales.

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 :

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

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

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

Comme dans le cas discret, une formule alternative existe pour la variance.

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).

La preuve est identique à celle du cas discret, en utilisant la linéarité de l'espérance.

Preuve

Soit $\mu = E(X)$. On part de la définition de la variance :

$$\begin{aligned} \text{Var}(X) &= E[(X - \mu)^2] \\ &= E[X^2 - 2X\mu + \mu^2] \quad (\text{On développe le carré}) \\ &= E(X^2) - E(2\mu X) + E(\mu^2) \quad (\text{Par linéarité de l'espérance, qui s'applique aussi au cas continu}) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \quad (\text{Car } 2\mu \text{ et } \mu^2 \text{ sont des constantes}) \\ &= E(X^2) - 2\mu(\mu) + \mu^2 \quad (\text{Car } E(X) = \mu) \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 = E(X^2) - [E(X)]^2 \end{aligned}$$

Le calcul de $E[X^2]$ (et plus généralement de $E[g(X)]$) repose sur le théorème de transfert, adapté au cas continu.

Théorème : Théorème de Transfert (LOTUS)

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

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

La preuve formelle est plus avancée, mais l'idée est analogue au cas discret : on pondère chaque valeur $g(x)$ par la densité de probabilité $f(x)$ au voisinage de x .

Preuve : Idée de la preuve

La preuve formelle repose sur la théorie de la mesure ou sur un argument de changement de variable pour l'intégrale, en passant par la fonction de répartition de $Y = g(X)$. Intuitivement, pour un petit intervalle dx autour de x , la "masse" de probabilité est $f(x)dx$. Cette masse correspond à une valeur $g(x)$ pour la nouvelle variable. L'espérance est la somme (intégrale) de ces valeurs pondérées par leur masse : $\int g(x)f(x)dx$.

La propriété la plus importante de l'espérance reste valide.

Remarque : Linéarité de l'Espérance

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

Calculons l'espérance et la variance pour notre exemple.

Exemple : Espérance et Variance de l'exemple précédent

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

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

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

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

4.4 Loi Uniforme

La loi continue la plus simple est celle où la densité est constante sur un intervalle.

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

C'est le modèle du "hasard pur" sur un segment.

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 $[a, b]$).

Les propriétés de cette loi sont faciles à dériver par intégration directe.

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** : $\text{Var}(X) = \frac{(b-a)^2}{12}$.

Preuve : Dérivation des propriétés

Soit $f(x) = \frac{1}{b-a}$ pour $x \in [a, b]$ et 0 sinon.

CDF : Pour $x \in [a, b]$,

$$F(x) = \int_{-\infty}^x f(t) dt = \int_a^x \frac{1}{b-a} dt = \frac{1}{b-a} [t]_a^x = \frac{x-a}{b-a}$$

(Pour $x < a$, $F(x) = 0$. Pour $x > b$, $F(x) = 1$.)

Espérance :

$$E[X] = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}$$

Variance : D'abord, calculons $E[X^2]$.

$$E[X^2] = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b = \frac{1}{b-a} \frac{b^3 - a^3}{3}$$

En utilisant $b^3 - a^3 = (b-a)(b^2 + ab + a^2)$, on obtient $E[X^2] = \frac{b^2 + ab + a^2}{3}$. Maintenant, appliquons

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

$$\begin{aligned}\text{Var}(X) &= \frac{b^2 + ab + a^2}{3} - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{b^2 + ab + a^2}{3} - \frac{a^2 + 2ab + b^2}{4} \\ &= \frac{4(b^2 + ab + a^2) - 3(a^2 + 2ab + b^2)}{12} \\ &= \frac{4b^2 + 4ab + 4a^2 - 3a^2 - 6ab - 3b^2}{12} \\ &= \frac{b^2 - 2ab + a^2}{12} = \frac{(b-a)^2}{12}\end{aligned}$$

4.5 Loi Exponentielle

Passons à une loi fondamentale pour modéliser les temps d'attente.

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

Cette loi est intimement liée au processus de Poisson.

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 .

Cette loi possède des propriétés remarquables.

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 $s, t \geq 0$, $P(X > s + t \mid X > s) = P(X > t)$.

Les preuves de l'espérance et de la variance nécessitent une intégration par parties. La preuve de la non-mémoire est plus directe.

Preuve : Dérivation des propriétés

Soit $f(x) = \lambda e^{-\lambda x}$ pour $x \geq 0$.

CDF : A été dérivée dans l'intuition ci-dessus.

$$F(x) = \int_0^x \lambda e^{-\lambda t} dt = [-e^{-\lambda t}]_0^x = -e^{-\lambda x} - (-e^0) = 1 - e^{-\lambda x}$$

Espérance : On utilise l'intégration par parties ($\int u dv = uv - \int v du$) avec $u = x$ et $dv = \lambda e^{-\lambda x} dx$. Alors $du = dx$ et $v = -e^{-\lambda x}$.

$$\begin{aligned} E[X] &= \int_0^\infty x(\lambda e^{-\lambda x}) dx \\ &= [x(-e^{-\lambda x})]_0^\infty - \int_0^\infty (-e^{-\lambda x}) dx \\ &= (0 - 0) + \int_0^\infty e^{-\lambda x} dx \quad (\text{car } \lim_{x \rightarrow \infty} -xe^{-\lambda x} = 0) \\ &= \left[-\frac{1}{\lambda} e^{-\lambda x} \right]_0^\infty = (0) - \left(-\frac{1}{\lambda} e^0 \right) = \frac{1}{\lambda} \end{aligned}$$

Variance : D'abord $E[X^2]$. Intégration par parties avec $u = x^2$, $dv = \lambda e^{-\lambda x} dx$. $du = 2x dx$, $v = -e^{-\lambda x}$.

$$\begin{aligned} E[X^2] &= \int_0^\infty x^2(\lambda e^{-\lambda x}) dx \\ &= [x^2(-e^{-\lambda x})]_0^\infty - \int_0^\infty (-e^{-\lambda x})(2x dx) \\ &= 0 + \int_0^\infty 2xe^{-\lambda x} dx \\ &= \frac{2}{\lambda} \int_0^\infty x(\lambda e^{-\lambda x}) dx \quad (\text{On fait apparaître } E[X]) \\ &= \frac{2}{\lambda} E[X] = \frac{2}{\lambda} \left(\frac{1}{\lambda} \right) = \frac{2}{\lambda^2} \end{aligned}$$

Donc, $\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda} \right)^2 = \frac{1}{\lambda^2}$.

Propriété de non-mémoire : Rappelons que $P(X > t) = e^{-\lambda t}$.

$$\begin{aligned} P(X > s + t \mid X > s) &= \frac{P(X > s + t \text{ et } X > s)}{P(X > s)} \\ &= \frac{P(X > s + t)}{P(X > s)} \quad (\text{car si } X > s + t, \text{ alors } X > s) \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = \frac{e^{-\lambda s} e^{-\lambda t}}{e^{-\lambda s}} = e^{-\lambda t} \\ &= P(X > t) \end{aligned}$$

Le paramètre λ a une interprétation concrète.

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**.

La propriété de non-mémoire est unique à la loi exponentielle (dans le cas continu).

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. Si vous attendez un bus qui arrive selon un processus de Poisson (et donc le temps d'attente suit une loi exponentielle), et que vous avez déjà attendu 5 minutes ($X > 5$), la probabilité que vous deviez attendre encore au moins 2 minutes ($X > 5 + 2$) est la même que si vous veniez juste d'arriver à l'arrêt et deviez attendre au moins 2 minutes ($X > 2$). L'information "j'ai déjà attendu 5 minutes" est inutile pour prédire l'attente future.

4.6 Distributions Conjointes (Cas Continu)

Comme pour le cas discret, nous pouvons définir des lois conjointes pour plusieurs variables aléatoires continues.

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. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$.

Ici, la probabilité est associée à un volume sous la surface de densité.

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 .

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

On retrouve les lois marginales en intégrant (en "écrasant" le volume).

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 :

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

La CDF conjointe accumule ce volume.

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) .

4.7 Espérance, Indépendance et Covariance (Cas Conjoint)

Les concepts clés s'étendent naturellement au cas conjoint continu.

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 :

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

La preuve est analogue à celle de LOTUS 1D, mais en dimension supérieure.

Preuve : Idée de la preuve

Comme pour LOTUS 1D, la preuve rigoureuse utilise des arguments de théorie de la mesure. L'intuition est que pour un petit rectangle $dx dy$ autour de (x, y) , la "masse" de probabilité est $f(x, y) dx dy$. Cette masse correspond à la valeur $g(x, y)$. L'espérance est la somme (double intégrale) de ces valeurs $g(x, y)$ pondérées par leur masse $f(x, y) dx dy$.

La condition d'indépendance s'exprime via la factorisation de la densité.

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 :

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

Cela signifie que le profil selon x ne dépend pas de 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). La surface de densité $z = f(x, y)$ peut être "séparée" en une fonction de x multipliée par une fonction de y .

La covariance se définit et se calcule de manière similaire.

Définition : Covariance (cas continu)

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

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

La formule de calcul reste la même.

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 : $E[XY] = \iint xy f(x, y) dx dy$.

La preuve est identique à celle du cas discret.

Preuve

La preuve est identique à celle vue pour les variables discrètes, car elle ne repose que sur la

linéarité de l'espérance, qui est vraie aussi dans le cas continu. Soit $\mu_X = E[X]$ et $\mu_Y = E[Y]$.

$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \\ &= E[XY] - E[X\mu_Y] - E[Y\mu_X] + E[\mu_X\mu_Y] \\ &= E[XY] - \mu_Y E[X] - \mu_X E[Y] + \mu_X\mu_Y \\ &= E[XY] - \mu_Y\mu_X - \mu_X\mu_Y + \mu_X\mu_Y \\ &= E[XY] - E[X]E[Y]\end{aligned}$$

La relation entre indépendance et covariance reste la même.

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.

4.8 Espérance d'une variable aléatoire continue

Lorsque la variable aléatoire X est continue, sa distribution est décrite par une fonction de densité de probabilité (PDF), $f(x)$. L'espérance est définie de manière analogue, en remplaçant la somme par une intégrale.

Définition : Espérance (cas continu)

L'espérance (ou valeur attendue) d'une variable aléatoire continue X avec une fonction de densité $f(x)$ est définie par :

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

L'intégrale doit être absolument convergente, c'est-à-dire $\int_{-\infty}^{\infty} |x| f(x) dx < \infty$.

Intuition

L'intuition du **centre de gravité** est toujours valable. Si la fonction de densité $f(x)$ représente la répartition de la masse sur une tige (l'axe des x), alors $E(X)$ est le point d'équilibre où la tige tiendrait en balance.

Exemple : Loi uniforme

Soit $X \sim \mathcal{U}(a, b)$. Sa densité est $f(x) = \frac{1}{b-a}$ pour $x \in [a, b]$, et 0 ailleurs.

$$\begin{aligned}E(X) &= \int_{-\infty}^{\infty} x f(x) dx = \int_a^b x \left(\frac{1}{b-a} \right) dx \\ &= \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{1}{b-a} \left(\frac{b^2 - a^2}{2} \right) \\ &= \frac{1}{b-a} \frac{(b-a)(b+a)}{2} = \frac{a+b}{2}\end{aligned}$$

L'espérance est le point milieu de l'intervalle, ce qui est intuitivement correct.

4.9 Linéarité de l'espérance

Le calcul de l'espérance deviendrait très fastidieux si nous devions toujours utiliser la définition. Heureusement, l'espérance possède une propriété fondamentale qui simplifie énormément les calculs.

Théorème : Linéarité de l'espérance

Pour toutes variables aléatoires X et Y (discrètes ou continues), 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.

Preuve

La première propriété est directe.

· **Cas continu** : $E(cX) = \int (cx)f(x)dx = c \int xf(x)dx = cE(X)$

Pour la seconde, $E(X + Y) = E(X) + E(Y)$:

Cas continu : La preuve est identique en remplaçant les sommes par des intégrales et la PMF jointe par la PDF jointe $f(x, y)$:

$$\begin{aligned}E(X + Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y)f(x, y) dx dy \\&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xf(x, y) dx dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf(x, y) dx dy \\&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f(x, y) dy \right) dx + \int_{-\infty}^{\infty} y \left(\int_{-\infty}^{\infty} f(x, y) dx \right) dy\end{aligned}$$

Les intégrales internes sont les densités marginales $f_X(x) = \int f(x, y)dy$ et $f_Y(y) = \int f(x, y)dx$.

$$E(X + Y) = \int_{-\infty}^{\infty} xf_X(x) dx + \int_{-\infty}^{\infty} yf_Y(y) dy = E(X) + E(Y)$$

Notez que l'indépendance n'a jamais été requise pour cette preuve.

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.

4.10 Loi du statisticien inconscient (LOTUS)

Souvent, nous ne sommes pas intéressés par l'espérance de X elle-même, mais par l'espérance d'une fonction de X , par exemple $E(X^2)$ ou $E(e^X)$.

Théorème : Théorème de Transfert (LOTUS)

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

· **Cas continu** : $E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x) dx$

Ce théorème est utile car il évite d'avoir à trouver la distribution (PDF) de $g(X)$.

Preuve

La preuve pour le cas continu est plus technique (utilisant un changement de variable) et est omise.

Ce théorème justifie son nom : c'est ce que l'on ferait "inconsciemment".

Intuition

Pour trouver la valeur moyenne d'une fonction d'une variable aléatoire, vous n'avez pas besoin de déterminer d'abord la distribution de cette fonction. Vous pouvez simplement prendre chaque valeur possible du résultat original, lui appliquer la fonction, et pondérer ce nouveau résultat par la densité du résultat original.

Exemple : Calcul de $E(X^2)$ pour une loi uniforme (continu)

Soit $X \sim \mathcal{U}(0, 1)$. Sa densité est $f(x) = 1$ sur $[0, 1]$. Calculons l'espérance de $Y = X^2$. La fonction est $g(x) = x^2$.

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} g(x)f(x) dx = \int_0^1 x^2 \cdot 1 dx \\ &= \left[\frac{x^3}{3} \right]_0^1 = \frac{1}{3} \end{aligned}$$

4.11 Variance

L'espérance nous donne le centre d'une distribution, mais elle ne dit rien sur sa "largeur" ou sa "dispersion". C'est le rôle de la variance.

Définition : Variance et écart-type

La **variance** d'une variable aléatoire X mesure la dispersion de sa distribution autour de son espérance $\mu = E(X)$. Elle est définie par :

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

Concrètement, cela se traduit par (en utilisant LOTUS avec $g(x) = (x - \mu)^2$) :

$$\cdot \text{Cas continu : } \text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

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

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

L'idée est de mesurer l'écart quadratique moyen à l'espérance.

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.

La définition $E[(X - \mu)^2]$ est excellente pour l'interprétation, mais pénible pour le calcul. Une formule alternative est presque toujours utilisée.

Théorème : Formule de calcul de la variance

Pour toute variable aléatoire X (discrète ou continue), une formule plus pratique pour le calcul de la variance est :

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

La preuve est une simple expansion algébrique utilisant la linéarité de l'espérance.

Preuve

Soit $\mu = E(X)$. On part de la définition de la variance :

$$\begin{aligned} \text{Var}(X) &= E[(X - \mu)^2] \\ &= E[X^2 - 2X\mu + \mu^2] \quad (\text{On développe le carré}) \\ &= E(X^2) - E(2\mu X) + E(\mu^2) \quad (\text{Par linéarité de l'espérance}) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \quad (\text{Car } 2\mu \text{ et } \mu^2 \text{ sont des constantes}) \\ &= E(X^2) - 2\mu(\mu) + \mu^2 \quad (\text{Car } E(X) = \mu) \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 = E(X^2) - [E(X)]^2 \end{aligned}$$

Exemple : Variance de la loi uniforme

Calculons la variance de $X \sim \mathcal{U}(a, b)$. Nous avons trouvé $E(X) = \frac{a+b}{2}$. Nous devons d'abord calculer $E(X^2)$ en utilisant LOTUS :

$$\begin{aligned} E(X^2) &= \int_a^b x^2 f(x) dx = \int_a^b x^2 \left(\frac{1}{b-a} \right) dx \\ &= \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b = \frac{1}{b-a} \left(\frac{b^3 - a^3}{3} \right) \\ &= \frac{1}{b-a} \frac{(b-a)(a^2 + ab + b^2)}{3} = \frac{a^2 + ab + b^2}{3} \end{aligned}$$

On utilise maintenant la formule de calcul $\text{Var}(X) = E(X^2) - [E(X)]^2$:

$$\begin{aligned} \text{Var}(X) &= \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2} \right)^2 \\ &= \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4} \\ &= \frac{4(a^2 + ab + b^2) - 3(a^2 + 2ab + b^2)}{12} \\ &= \frac{4a^2 + 4ab + 4b^2 - 3a^2 - 6ab - 3b^2}{12} \\ &= \frac{a^2 - 2ab + b^2}{12} = \frac{(b-a)^2}{12} \end{aligned}$$

4.12 Changement de Variable pour les Variables Aléatoires Continues

Un problème courant en probabilité est de déterminer la densité d'une variable aléatoire $Y = g(X)$, où X est une variable aléatoire continue de densité connue $f_X(x)$ et g est une fonction. Cette technique est fondamentale pour transformer des distributions.

Théorème : Changement de Variable - Cas Monotone

Soit X une variable aléatoire continue de densité $f_X(x)$ supportée sur un intervalle I . Soit $g : I \rightarrow J$ une fonction strictement monotone (croissante ou décroissante) et dérivable, avec $g'(x) \neq 0$ sur I . Alors $Y = g(X)$ est une variable aléatoire continue de densité :

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right|$$

pour $y \in J$, et $f_Y(y) = 0$ ailleurs.

Preuve : Preuve du Théorème

La preuve repose sur la relation entre la CDF de Y et celle de X .

Cas 1 : g strictement croissante.

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

En dérivant par rapport à y :

$$f_Y(y) = \frac{d}{dy} F_X(g^{-1}(y)) = f_X(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y)$$

Comme g est croissante, g^{-1} est croissante, donc $\frac{d}{dy} g^{-1}(y) > 0$. On a donc :

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right|$$

Cas 2 : g strictement décroissante.

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y))$$

En dérivant par rapport à y :

$$f_Y(y) = -f_X(g^{-1}(y)) \cdot \frac{d}{dy}g^{-1}(y)$$

Comme g est décroissante, g^{-1} est décroissante, donc $\frac{d}{dy}g^{-1}(y) < 0$. On a donc :

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy}g^{-1}(y) \right|$$

Exemple : Changement d'unité : distance entre véhicules sur une autoroute

Contexte : Supposons que l'on modélise le trafic sur une portion isolée d'autoroute en zone rurale. Les véhicules passent rarement, de manière indépendante les uns des autres, à des instants aléatoires. Dans ce cadre, le nombre de voitures observées sur un intervalle de temps fixé suit une **loi de Poisson**, ce qui implique naturellement que le **temps entre deux passages successifs** suit une loi exponentielle. Par extension, si l'on suppose une vitesse constante moyenne, alors la **distance entre deux véhicules consécutifs** peut également être modélisée par une loi exponentielle.

On s'intéresse précisément à cette distance X , exprimée en **mètres**. On suppose qu'elle suit une loi exponentielle de paramètre $\lambda = 0.001$, ce qui correspond à une distance moyenne de $1/\lambda = 1000$ mètres entre deux voitures — une situation réaliste pour une très faible densité de circulation.

Ainsi, la densité de X est :

$$f_X(x) = \begin{cases} 0.001 e^{-0.001x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Un ingénieur souhaite maintenant exprimer cette distance en **kilomètres** pour des raisons pratiques (rapports, comparaison avec d'autres données). On définit donc :

$$Y = \frac{X}{1000}$$

Notre objectif est de déterminer la densité $f_Y(y)$ de Y en suivant rigoureusement la méthode de changement de variable présentée dans le théorème précédent.

Étape 1 : Identifier la transformation La fonction g est définie par :

$$g(x) = \frac{x}{1000}$$

C'est une fonction **strictement croissante** et dérivable sur $[0, \infty[$, donc le théorème s'applique.

Étape 2 : Calculer la réciproque

$$y = \frac{x}{1000} \Rightarrow x = 1000y \Rightarrow g^{-1}(y) = 1000y$$

Étape 3 : Calculer la dérivée de la réciproque

$$\frac{d}{dy}g^{-1}(y) = \frac{d}{dy}(1000y) = 1000$$

Donc :

$$\left| \frac{d}{dy}g^{-1}(y) \right| = 1000$$

Étape 4 : Appliquer la formule du changement de variable D'après le théorème :

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy}g^{-1}(y) \right| = f_X(1000y) \cdot 1000$$

Substituons $f_X(1000y)$:

$$f_X(1000y) = 0.001 e^{-0.001 \cdot (1000y)} = 0.001 e^{-y}$$

Donc :

$$f_Y(y) = 0.001 e^{-y} \cdot 1000 = e^{-y}, \quad y \geq 0$$

Résultat final :

$$f_Y(y) = \begin{cases} e^{-y} & y \geq 0 \\ 0 & y < 0 \end{cases}$$

Ce qui signifie que $Y \sim \mathcal{E}(1)$, une loi exponentielle de paramètre 1, dont l'espérance est $\mathbb{E}[Y] = 1$ kilomètre.

Vérification probabiliste : Calculons la probabilité que la distance entre deux véhicules soit inférieure à 500 mètres.

- En mètres : $P(X < 500) = 1 - e^{-0.001 \times 500} = 1 - e^{-0.5} \approx 0.3935$ - En kilomètres : $500 \text{ m} = 0.5 \text{ km}$, donc $P(Y < 0.5) = 1 - e^{-1 \times 0.5} = 1 - e^{-0.5} \approx 0.3935$

Les probabilités sont identiques, confirmant la validité du changement de variable.

Conclusion : Le passage d'une unité à une autre (ici, mètres \rightarrow kilomètres) modifie la forme apparente de la densité, mais respecte la structure probabiliste sous-jacente. Le facteur jacobien $\left| \frac{d}{dy} g^{-1}(y) \right| = 1000$ compense exactement la compression de l'échelle, garantissant que la nouvelle densité intègre bien à 1 et que toutes les probabilités restent cohérentes avec la réalité physique du phénomène modélisé.

4.13 Exercices

Exercice 1 :

Soit la fonction $f(x) = cx^2$ pour $x \in [0, 2]$, et $f(x) = 0$ sinon.

1. Calculez la valeur de c pour que $f(x)$ soit une PDF valide.
2. En utilisant la valeur de c trouvée, calculez $P(X > 1)$.

Exercice 2 :

Soit une v.a. X de PDF $f(x) = \frac{1}{2}x$ sur $[0, 2]$, et $f(x) = 0$ sinon.

1. Déterminez la fonction de répartition (CDF) $F(x)$.
2. En utilisant la CDF, calculez $P(1 \leq X \leq 1.5)$.

Exercice 3 :

La CDF d'une v.a. X est donnée par :

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \sin(x) & \text{si } 0 \leq x \leq \pi/2 \\ 1 & \text{si } x > \pi/2 \end{cases}$$

Calculez la PDF $f(x)$ de X .

Exercice 4 :

Soit X une variable aléatoire continue avec une PDF $f(x)$. Démontrez que pour tout $a \in \mathbb{R}$, $P(X = a) = 0$.

Exercice 5 :

Soit X une v.a. de PDF $f(x) = 3x^2$ sur $[0, 1]$.

1. Calculez l'espérance $E[X]$.
2. Calculez $E[X^2]$.
3. Déduisez-en la variance $\text{Var}(X)$.

Exercice 6 :

Soit X une v.a. de PDF $f(x) = 2x$ sur $[0, 1]$. En utilisant le théorème de LOTUS, calculez $E[1/X]$.

Exercice 7 : Variance d'une transformation affine

Soit X une v.a. avec une variance $\text{Var}(X)$ finie. En utilisant les propriétés de l'espérance, démontrez que pour $a, b \in \mathbb{R}$:

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

Exercice 8 :

Le temps d'attente T (en minutes) d'un bus suit une loi $T \sim \text{Unif}(0, 12)$.

1. Quelle est la probabilité d'attendre moins de 4 minutes ?
2. Quelle est la probabilité d'attendre entre 5 et 10 minutes ?

Exercice 9 :

Pour le temps d'attente $T \sim \text{Unif}(0, 12)$ de l'exercice précédent :

1. Calculez le temps d'attente moyen $E[T]$.
2. Calculez la variance $\text{Var}(T)$.

Exercice 10 :

Soit $X \sim \text{Unif}(a, b)$. La PDF est $f(x) = \frac{1}{b-a}$ pour $x \in [a, b]$. Démontrez par un calcul intégral direct que $E[X] = \frac{a+b}{2}$.

Exercice 11 :

Un générateur produit un nombre $X \sim \text{Unif}(a, b)$. On sait que $E[X] = 10$ et $\text{Var}(X) = 12$. Trouvez a et b .

Exercice 12 :

La durée de vie T (en années) d'un composant suit $T \sim \text{Exp}(\lambda)$. Sa durée de vie moyenne $E[T]$ est de 5 ans.

1. Déterminez λ .
2. Calculez la probabilité que le composant tombe en panne avant 2 ans, $P(T \leq 2)$.

Exercice 13 :

En utilisant le scénario de l'exercice 12 ($E[T] = 5$ ans), calculez la probabilité que le composant dure 10 ans de plus, sachant qu'il a déjà duré 3 ans : $P(T > 13 \mid T > 3)$. Comparez ce résultat à $P(T > 10)$.

Exercice 14 :

Soit $X \sim \text{Exp}(\lambda)$. Sa PDF est $f(x) = \lambda e^{-\lambda x}$ pour $x \geq 0$. En utilisant une intégration par parties, prouvez que $E[X] = 1/\lambda$.

Exercice 15 :

Soit $X \sim \text{Exp}(\lambda)$. Démontrez formellement la propriété d'absence de mémoire :

$$P(X > s + t \mid X > s) = P(X > t) \quad \text{pour tous } s, t \geq 0$$

Exercice 16 :

Soit $F(x)$ la CDF d'une v.a. continue X . Démontrez que pour $a < b$, on a :

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

(Indice : Utilisez $P(A) = P(B) - P(B \setminus A)$ avec $A = \{X \leq a\}$ et $B = \{X \leq b\}$).

Exercice 17 :

Soit X une v.a. continue de PDF $f(x)$. Démontrez que $E[aX + b] = aE[X] + b$ pour $a, b \in \mathbb{R}$, en utilisant la définition de l'espérance (LOTUS).

Exercice 18 :

Soit $X \sim \text{Unif}(0, 1)$. Soit $Y = 5X - 2$. Trouvez la PDF $f_Y(y)$ de Y en utilisant la méthode de la CDF ou la formule de changement de variable. Sur quel intervalle Y est-elle définie ?

Exercice 19 :

Soit $X \sim \text{Unif}(-1, 1)$. Soit $Y = X^2$. Trouvez la PDF $f_Y(y)$ de Y . (Attention : la fonction $g(x) = x^2$ n'est pas monotone sur $[-1, 1]$).

Exercice 20 :

Soit $X \sim \text{Exp}(1)$ (donc $f_X(x) = e^{-x}$ pour $x \geq 0$). Soit $Y = \sqrt{X}$. Trouvez la PDF $f_Y(y)$ de Y .

4.14 Correction**Exercice 1 :**

1. Pour que $f(x)$ soit une PDF valide, on doit avoir $\int_{-\infty}^{\infty} f(x) dx = 1$. Alors :

$$\int_0^2 cx^2 dx = c \left[\frac{x^3}{3} \right]_0^2 = c \cdot \frac{8}{3} = 1 \implies c = \frac{3}{8}.$$

2. Avec $c = \frac{3}{8}$, on calcule :

$$P(X > 1) = \int_1^2 \frac{3}{8} x^2 dx = \frac{3}{8} \left[\frac{x^3}{3} \right]_1^2 = \frac{3}{8} \left(\frac{8}{3} - \frac{1}{3} \right) = \frac{3}{8} \cdot \frac{7}{3} = \frac{7}{8}.$$

Exercice 2 :

- La CDF est définie par $F(x) = \int_{-\infty}^x f(t) dt$.
- Si $x < 0$, $F(x) = 0$.
- Si $x \in [0, 2]$, $F(x) = \int_0^x \frac{1}{2} t dt = \frac{1}{2} \cdot \frac{t^2}{2} \Big|_0^x = \frac{x^2}{4}$.
- Si $x > 2$, $F(x) = 1$. Donc :

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

5. $P(1 \leq X \leq 1.5) = F(1.5) - F(1) = \frac{(1.5)^2}{4} - \frac{1^2}{4} = \frac{2.25-1}{4} = \frac{1.25}{4} = 0.3125$.

Exercice 3 :

La PDF est la dérivée de la CDF :

$$f(x) = F'(x) = \begin{cases} 0 & x < 0 \\ \cos(x) & 0 \leq x \leq \pi/2 \\ 0 & x > \pi/2 \end{cases}$$

Donc $f(x) = \cos(x)$ pour $x \in [0, \pi/2]$, et 0 sinon.

Exercice 4 :

Soit X continue avec PDF $f(x)$. Alors :

$$P(X = a) = \int_a^a f(x) dx = 0,$$

car l'intégrale sur un intervalle de longueur nulle est nulle.

Exercice 5 :

- $E[X] = \int_0^1 x \cdot 3x^2 dx = 3 \int_0^1 x^3 dx = 3 \cdot \frac{1}{4} = \frac{3}{4}$.
- $E[X^2] = \int_0^1 x^2 \cdot 3x^2 dx = 3 \int_0^1 x^4 dx = 3 \cdot \frac{1}{5} = \frac{3}{5}$.
- $\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{3}{5} - \left(\frac{3}{4}\right)^2 = \frac{3}{5} - \frac{9}{16} = \frac{48-45}{80} = \frac{3}{80}$.

Exercice 6 :

Par LOTUS : $E[1/X] = \int_0^1 \frac{1}{x} \cdot 2x \, dx = \int_0^1 2 \, dx = 2$. Mais attention : en $x = 0$, $1/x$ n'est pas défini. Toutefois, comme il s'agit d'un point de mesure nulle, l'intégrale converge et vaut 2.

Exercice 7 :

On utilise $\text{Var}(Y) = E[Y^2] - (E[Y])^2$ avec $Y = aX + b$:

$$E[aX + b] = aE[X] + b, \quad E[(aX + b)^2] = E[a^2X^2 + 2abX + b^2] = a^2E[X^2] + 2abE[X] + b^2.$$

Alors :

$$\begin{aligned} \text{Var}(aX + b) &= E[(aX + b)^2] - (E[aX + b])^2 \\ &= a^2E[X^2] + 2abE[X] + b^2 - (aE[X] + b)^2 \\ &= a^2E[X^2] + 2abE[X] + b^2 - (a^2(E[X])^2 + 2abE[X] + b^2) \\ &= a^2E[X^2] - a^2(E[X])^2 = a^2\text{Var}(X). \end{aligned}$$

Exercice 8 :

$T \sim \text{Unif}(0, 12)$, donc $f_T(t) = \frac{1}{12}$ pour $t \in [0, 12]$.

1. $P(T < 4) = \int_0^4 \frac{1}{12} dt = \frac{4}{12} = \frac{1}{3}$.
2. $P(5 \leq T \leq 10) = \int_5^{10} \frac{1}{12} dt = \frac{5}{12}$.

Exercice 9 :

1. $E[T] = \frac{0+12}{2} = 6$ minutes.
2. $\text{Var}(T) = \frac{(12-0)^2}{12} = \frac{144}{12} = 12$.

Exercice 10 :

$$E[X] = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b x \, dx = \frac{1}{b-a} \cdot \frac{b^2 - a^2}{2} = \frac{b^2 - a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}.$$

Exercice 11 :

On sait : $E[X] = \frac{a+b}{2} = 10$, $\text{Var}(X) = \frac{(b-a)^2}{12} = 12$. De la première équation : $a + b = 20$. De la seconde : $(b-a)^2 = 144 \implies b - a = 12$ (puisque $b > a$). En résolvant le système :

$$a + b = 20, \quad b - a = 12 \implies 2b = 32 \implies b = 16, \quad a = 4.$$

Donc $a = 4$, $b = 16$.

Exercice 12 :

1. $E[T] = \frac{1}{\lambda} = 5 \implies \lambda = \frac{1}{5} = 0.2$.
2. $P(T \leq 2) = 1 - e^{-\lambda \cdot 2} = 1 - e^{-0.4} \approx 1 - 0.6703 = 0.3297$.

Exercice 13 :

Par la propriété d'absence de mémoire :

$$P(T > 13 \mid T > 3) = P(T > 10) = e^{-\lambda \cdot 10} = e^{-2} \approx 0.1353.$$

Et $P(T > 10) = e^{-2} \approx 0.1353$. Les deux probabilités sont égales.

Exercice 14 :

$$E[X] = \int_0^\infty x \lambda e^{-\lambda x} dx.$$

Intégration par parties : $u = x$, $dv = \lambda e^{-\lambda x} dx$, donc $du = dx$, $v = -e^{-\lambda x}$.

$$E[X] = [-xe^{-\lambda x}]_0^\infty + \int_0^\infty e^{-\lambda x} dx = 0 + \left[-\frac{1}{\lambda} e^{-\lambda x}\right]_0^\infty = \frac{1}{\lambda}.$$

Exercice 15 :

$$P(X > s + t \mid X > s) = \frac{P(X > s + t)}{P(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t).$$

Exercice 16 :

Soit $A = \{X \leq a\}$, $B = \{X \leq b\}$ avec $a < b$. Alors $A \subset B$, et :

$$P(a < X \leq b) = P(B \setminus A) = P(B) - P(A) = F(b) - F(a).$$

Exercice 17 :

$$E[aX + b] = \int_{-\infty}^{\infty} (ax + b)f_X(x)dx = a \int_{-\infty}^{\infty} xf_X(x)dx + b \int_{-\infty}^{\infty} f_X(x)dx = aE[X] + b.$$

Exercice 18 :

$X \sim \text{Unif}(0, 1)$, $Y = 5X - 2$. Transformation linéaire croissante. CDF : $F_Y(y) = P(Y \leq y) = P(5X - 2 \leq y) = P(X \leq \frac{y+2}{5}) = F_X(\frac{y+2}{5})$. Comme $X \in [0, 1]$, alors $Y \in [-2, 3]$. PDF : $f_Y(y) = f_X(\frac{y+2}{5}) \cdot \left| \frac{d}{dy}(\frac{y+2}{5}) \right| = 1 \cdot \frac{1}{5} = \frac{1}{5}$ pour $y \in [-2, 3]$. Donc $Y \sim \text{Unif}(-2, 3)$.

Exercice 19 :

$X \sim \text{Unif}(-1, 1)$, $f_X(x) = \frac{1}{2}$ sur $[-1, 1]$, $Y = X^2$. Pour $y \in [0, 1]$, on a deux branches : $x = \sqrt{y}$ et $x = -\sqrt{y}$. Formule générale pour transformation non monotone :

$$f_Y(y) = \sum_i f_X(x_i) \left| \frac{dx_i}{dy} \right|, \quad \frac{dx}{dy} = \frac{d}{dy}(\pm\sqrt{y}) = \pm \frac{1}{2\sqrt{y}}.$$

Donc :

$$f_Y(y) = f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} + f_X(-\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} = \frac{1}{2} \cdot \frac{1}{2\sqrt{y}} + \frac{1}{2} \cdot \frac{1}{2\sqrt{y}} = \frac{1}{4\sqrt{y}} + \frac{1}{4\sqrt{y}} = \frac{1}{2\sqrt{y}}.$$

Ainsi, $f_Y(y) = \frac{1}{2\sqrt{y}}$ pour $y \in (0, 1]$, et 0 sinon.

Exercice 20 :

$X \sim \text{Exp}(1)$, $f_X(x) = e^{-x}$ pour $x \geq 0$, $Y = \sqrt{X} \Rightarrow X = Y^2$. Transformation strictement croissante pour $y \geq 0$.

$$f_Y(y) = f_X(y^2) \cdot \left| \frac{d}{dy}(y^2) \right| = e^{-y^2} \cdot 2y, \quad y \geq 0.$$

Donc $f_Y(y) = 2ye^{-y^2}$ pour $y \geq 0$.

4.15 Exercices Python

Les exercices suivants appliquent les concepts de variables aléatoires continues (PDF, CDF, espérance, variance) en utilisant la bibliothèque NumPy pour la simulation numérique afin de vérifier les résultats théoriques.

```
import numpy as np
import math
```

Exercice 1 : PDF CDF et Espérance (Simulation)

Soit X une v.a. continue avec la PDF $f(x) = 2x$ pour $x \in [0, 1]$, et $f(x) = 0$ sinon. Par calcul (que vous pouvez faire à la main), on trouve :

- CDF : $F(x) = x^2$ (pour $x \in [0, 1]$)
- Espérance : $E[X] = 2/3$

Nous pouvons simuler cette variable en utilisant la méthode de la transformée inverse : si $U \sim \text{Unif}(0, 1)$, alors $X = F^{-1}(U) = \sqrt{U}$ suit la loi de X .

Votre tâche (avec NumPy) :

1. Générer $N = 100000$ échantillons U d'une loi Uniforme(0, 1) avec `np.random.rand`.
2. Transformer ces échantillons pour obtenir N échantillons de X (en prenant la racine carrée).
3. Calculer l'espérance empirique $E[X]$ (la moyenne de vos échantillons X) et la comparer à la valeur théorique $2/3$.

Exercice 2 : Variance (Simulation)

En utilisant les échantillons X de l'exercice 1. La valeur théorique (calculée à la main) de la variance est $\text{Var}(X) = 1/18$.

Votre tâche (avec NumPy) :

1. Calculer la variance empirique $\text{Var}(X)$ de vos échantillons X avec `np.var`.
2. Comparer le résultat empirique à la valeur théorique $1/18$.

Exercice 3 : Loi Uniforme (Simulation vs Théorie)

Soit $X \sim \text{Unif}(a = 5, b = 15)$. Les valeurs théoriques sont $E[X] = \frac{a+b}{2}$ et $\text{Var}(X) = \frac{(b-a)^2}{12}$.

Votre tâche (avec NumPy) :

1. Calculer l'espérance et la variance théoriques.
2. Générer $N = 100000$ échantillons aléatoires de X avec `np.random.uniform`.
3. Calculer l'espérance empirique (`np.mean`) et la variance empirique (`np.var`) des échantillons.
4. Comparer les résultats empiriques aux résultats théoriques.

Exercice 4 : Loi Uniforme (Vérification de la PDF)

Pour $X \sim \text{Unif}(5, 15)$, la PDF est $f(x) = \frac{1}{10}$ sur $[5, 15]$. La probabilité $P(7 \leq X \leq 10)$ est $\int_7^{10} \frac{1}{10} dx = \frac{10-7}{10} = 0.3$.

Votre tâche (avec NumPy) :

1. Utiliser les échantillons de X de l'exercice 3.
2. Calculer la probabilité empirique $P(7 \leq X \leq 10)$ en comptant la proportion d'échantillons qui tombent dans cet intervalle.
3. Comparer le résultat empirique à la valeur théorique 0.3.

Exercice 5 : Loi Exponentielle (Simulation vs Théorie)

Soit $X \sim \text{Exp}(\lambda = 0.5)$. Les valeurs théoriques sont $E[X] = \frac{1}{\lambda}$ et $\text{Var}(X) = \frac{1}{\lambda^2}$.

Note : `np.random.exponential` prend un paramètre "scale" $\beta = 1/\lambda$.

Votre tâche (avec NumPy) :

1. Définir λ et calculer $E[X]$ et $\text{Var}(X)$ théoriques.
2. Calculer le paramètre β (scale) pour NumPy.
3. Générer $N = 100000$ échantillons aléatoires de X .
4. Calculer et comparer les espérances et variances empiriques et théoriques.

Exercice 6 : Loi Exponentielle (Vérification de la CDF)

Pour $X \sim \text{Exp}(\lambda = 0.5)$, la CDF est $F(x) = 1 - e^{-\lambda x}$. Calculons $P(X \leq 3) = F(3) = 1 - e^{-0.5 \times 3}$.

Votre tâche (avec NumPy) :

1. Calculer la valeur théorique $F(3)$.
2. Utiliser les échantillons de X de l'exercice 5.
3. Calculer la probabilité empirique $P(X \leq 3)$ en comptant la proportion d'échantillons ≤ 3 .

4. Comparer les deux valeurs.

Exercice 7 : Propriété de Non-Mémoire (Exponentielle)

Nous allons vérifier numériquement la propriété de non-mémoire $P(X > s + t \mid X > s) = P(X > t)$ en utilisant les échantillons de X de l'exercice 5 ($\lambda = 0.5$).

Votre tâche (avec NumPy) :

1. Choisir $s = 1$ et $t = 2$.
2. Calculer $P(X > t)$ (théoriquement $e^{-\lambda t}$). Calculer la probabilité empirique (proportion d'échantillons $> t$).
3. Calculer $P(X > s + t \mid X > s)$ empiriquement :
 - Filtrer les échantillons pour ne garder que ceux où $X > s$.
 - Parmi ce sous-ensemble, calculer la proportion de ceux où $X > s + t$.
4. Comparer les deux probabilités empiriques.

Exercice 8 : Théorème de Transfert (LOTUS)

Soit $X \sim \text{Unif}(0, 2)$. La PDF est $f(x) = 1/2$. Soit $g(X) = X^2$. Nous voulons $E[g(X)] = E[X^2]$. Théoriquement : $E[X^2] = \int_0^2 x^2 f(x) dx = \int_0^2 x^2 (1/2) dx = \frac{1}{2} [\frac{x^3}{3}]_0^2 = \frac{1}{2} (\frac{8}{3}) = 4/3$.

Votre tâche (avec NumPy) :

1. Générer $N = 100000$ échantillons $X \sim \text{Unif}(0, 2)$.
2. Créer les échantillons $Y = g(X) = X^2$.
3. Calculer l'espérance empirique $E[Y]$ (la moyenne de Y).
4. Comparer le résultat empirique à la valeur théorique $4/3$.

Exercice 9 : Linéarité de l'Espérance ($E[aX+b]$)

)] Soit $X \sim \text{Unif}(5, 15)$ (de l'exercice 3). Nous savons que $E[X] = 10$. Soit $Y = 5X - 3$. Par linéarité, l'espérance théorique est $E[Y] = E[5X - 3] = 5E[X] - 3$.

Votre tâche (avec NumPy) :

1. Calculer $E[Y]$ théoriquement en utilisant $E[X] = 10$.
2. Utiliser les échantillons X de l'exercice 3.
3. Créer les échantillons $Y = 5 \times X - 3$.
4. Calculer l'espérance empirique $E[Y]$ (la moyenne de Y).
5. Comparer les deux résultats.

Exercice 10 : Propriétés de la Variance ($\text{Var}(aX+b)$)

Soit $X \sim \text{Unif}(5, 15)$ (de l'exercice 3). $\text{Var}(X) = \frac{(15-5)^2}{12} = 100/12 \approx 8.333$. Soit $Y = 5X - 3$. Théoriquement : $\text{Var}(Y) = \text{Var}(5X - 3) = \text{Var}(5X) = 5^2 \text{Var}(X) = 25 \times \text{Var}(X)$.

Votre tâche (avec NumPy) :

1. Calculer $\text{Var}(Y)$ théoriquement en utilisant $\text{Var}(X) = 100/12$.
2. Utiliser les échantillons Y de l'exercice 9.
3. Calculer la variance empirique $\text{Var}(Y)$ (avec `np.var`).
4. Comparer les deux résultats.

5 La Loi Normale (ou Gaussienne)

5.1 Introduction et Fonction de Densité (PDF)

Après les lois discrètes et les lois continues de base (Uniforme, Exponentielle), nous abordons la distribution la plus célèbre et la plus utilisée en probabilités et statistiques.

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 $\sigma > 0$.

Cette formule, bien qu'imposante, décrit une forme très familière : la courbe en cloche.

Intuition : La Courbe en Cloche

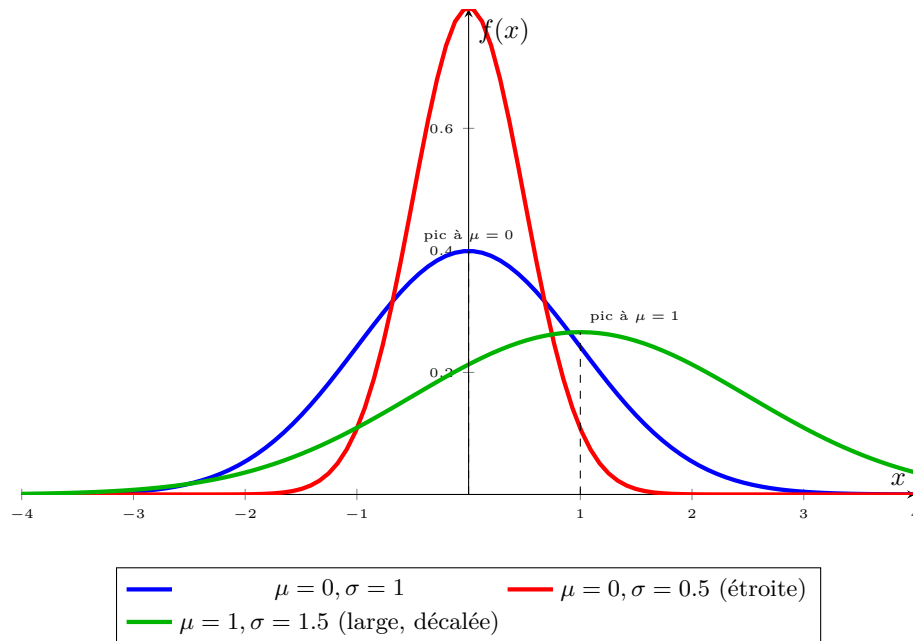
La loi normale est sans doute la distribution la plus importante en probabilités et statistiques. Pourquoi ? Parce qu'elle modélise remarquablement bien de nombreux phénomènes naturels et processus aléatoires où les valeurs tendent à se regrouper autour d'une moyenne, avec des écarts symétriques devenant 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)

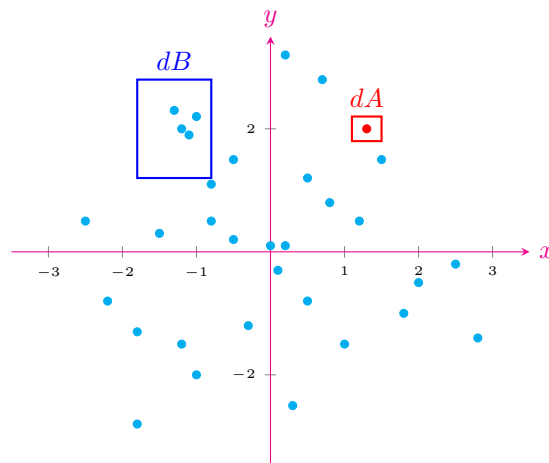


Influence de μ (position) et σ (largeur) sur la forme de la cloche.

Mais d'où vient cette formule spécifique ? Il existe une dérivation fascinante à partir d'hypothèses fondamentales sur les erreurs aléatoires (argument d'Herschel-Maxwell).

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 :

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

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

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

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 :

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

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

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

5. Normalisation et Identification des Paramètres

1. **Condition** $\int_{-\infty}^{\infty} f(x)dx = 1$: L'intégrale Gaussienne $\int_{-\infty}^{\infty} e^{-kx^2} dx = \sqrt{\frac{\pi}{k}}$. Donc, $\int_{-\infty}^{\infty} f(x)dx = \lambda \sqrt{\frac{\pi}{k}} = 1 \implies \lambda = \sqrt{\frac{k}{\pi}}$.
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 $\mathcal{N}(\mu, \sigma^2)$.

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

Avant d'explorer les propriétés de la loi normale générale, concentrons-nous sur son cas le plus simple et le plus fondamental.

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 $\Phi(z)$:

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

Pourquoi cette version standard est-elle si importante ? Elle sert de référence universelle.

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.

La notation est très standardisée pour cette loi.

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 .

Un détail technique important concerne le calcul de $\Phi(z)$.

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.

5.3 Standardisation : Le Score Z

Formalisons cette transformation clé qui relie toute loi normale à la loi standard.

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

La preuve formelle utilise un changement de variable dans la fonction de répartition.

Preuve

Soit $F_X(x)$ la CDF de X et $F_Z(z)$ la CDF de Z . Nous voulons montrer que $F_Z(z) = \Phi(z)$.

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

Par définition de la CDF de X :

$$F_X(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$$

Donc,

$$F_Z(z) = \int_{-\infty}^{\mu+z\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$$

Effectuons le changement de variable $u = (t - \mu)/\sigma$. Alors $t = \mu + u\sigma$ et $dt = \sigma du$. Les bornes d'intégration changent :

- Quand $t \rightarrow -\infty$, $u \rightarrow -\infty$.
- Quand $t = \mu + z\sigma$, $u = ((\mu + z\sigma) - \mu)/\sigma = z$.

L'intégrale devient :

$$\begin{aligned} F_Z(z) &= \int_{-\infty}^z \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}u^2} (\sigma du) \\ F_Z(z) &= \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \end{aligned}$$

C'est exactement la définition de $\Phi(z)$, la CDF de la loi normale standard. Ainsi, $Z \sim \mathcal{N}(0, 1)$.

Cette transformation a une interprétation très concrète.

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

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 $\mathcal{N}(0, 1)$.

Un exemple classique est la comparaison de notes.

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 :

$$\begin{aligned} Z_A &= \frac{80 - 70}{5} = \frac{10}{5} = +2.0 \\ Z_B &= \frac{85 - 75}{10} = \frac{10}{10} = +1.0 \end{aligned}$$

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.

5.4 Propriétés Importantes de la Loi Normale

La loi normale possède des propriétés de stabilité remarquables sous certaines transformations.

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 ($\text{Var}(aX + b) = a^2\text{Var}(X)$).

Preuve

Nous utilisons le fait que si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors $Z = (X - \mu)/\sigma \sim \mathcal{N}(0, 1)$. Exprimons X en fonction de Z : $X = \mu + \sigma Z$. Substituons cela dans l'expression de Y :

$$Y = a(\mu + \sigma Z) + b = (a\mu + b) + (a\sigma)Z$$

Posons $\mu_Y = a\mu + b$ et $\sigma_Y = |a|\sigma$. Alors $Y = \mu_Y + \sigma_Y Z$ (si $a > 0$) ou $Y = \mu_Y - \sigma_Y Z$ (si $a < 0$). Dans les deux cas, Y est une transformation linéaire d'une variable normale standard Z . La CDF de Y peut être exprimée en termes de la CDF Φ de Z . Si $a > 0$:

$$P(Y \leq y) = P(\mu_Y + a\sigma Z \leq y) = P(a\sigma Z \leq y - \mu_Y) = P\left(Z \leq \frac{y - \mu_Y}{a\sigma}\right) = \Phi\left(\frac{y - \mu_Y}{a\sigma}\right)$$

C'est la CDF d'une loi $\mathcal{N}(\mu_Y, (a\sigma)^2)$. Le cas $a < 0$ est similaire et mène au même résultat pour la distribution (la variance dépend de a^2). Ainsi, $Y \sim \mathcal{N}(a\mu + b, (a\sigma)^2)$.

Cette propriété est très utile pour les changements d'unités.

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

$$T_F = \frac{9}{5}T_C + 32 ?$$

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

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

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

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

Une autre propriété cruciale concerne la somme de variables normales indépendantes.

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.

La preuve formelle de ce théorème est plus avancée et utilise généralement les fonctions caractéristiques ou les fonctions génératrices des moments.

Preuve : Idée de la preuve (via Fonctions Caractéristiques)

La fonction caractéristique $\varphi_X(t)$ d'une variable aléatoire X est définie comme $\varphi_X(t) = E[e^{itX}]$. Pour une loi normale $X \sim \mathcal{N}(\mu, \sigma^2)$, sa fonction caractéristique est $\varphi_X(t) = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$. Si X et Y sont indépendantes, la fonction caractéristique de leur somme $S = X + Y$ est le produit

de leurs fonctions caractéristiques : $\varphi_S(t) = \varphi_X(t)\varphi_Y(t)$.

$$\begin{aligned}\varphi_S(t) &= \left(e^{i\mu_X t - \frac{1}{2}\sigma_X^2 t^2}\right) \left(e^{i\mu_Y t - \frac{1}{2}\sigma_Y^2 t^2}\right) \\ &= e^{i(\mu_X + \mu_Y)t - \frac{1}{2}(\sigma_X^2 + \sigma_Y^2)t^2}\end{aligned}$$

On reconnaît ici la fonction caractéristique d'une loi normale avec pour moyenne $\mu_X + \mu_Y$ et pour variance $\sigma_X^2 + \sigma_Y^2$. Comme la fonction caractéristique détermine de manière unique la distribution, on conclut que $S \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

Il est essentiel de se souvenir de la condition d'indépendance pour l'addition des variances.

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).

Appliquons ce théorème à un exemple concret.

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 $\sqrt{325} \approx 18.03g$.

5.5 La Règle Empirique (68-95-99.7)

Une conséquence directe des aires sous la courbe normale standard est une règle approximative très utile.

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 $\mu \pm 3\sigma$).

Preuve : Dérivation à partir de $\Phi(z)$

Ces valeurs sont obtenues en calculant les aires sous la PDF de la loi normale standard $\mathcal{N}(0, 1)$ entre les Z-scores correspondants.

- $P(-1 \leq Z \leq 1) = \Phi(1) - \Phi(-1) = \Phi(1) - (1 - \Phi(1)) = 2\Phi(1) - 1$. Avec $\Phi(1) \approx 0.8413$, on obtient $2(0.8413) - 1 \approx 0.6826$.
- $P(-2 \leq Z \leq 2) = \Phi(2) - \Phi(-2) = 2\Phi(2) - 1$. Avec $\Phi(2) \approx 0.9772$, on obtient $2(0.9772) - 1 \approx 0.9544$.
- $P(-3 \leq Z \leq 3) = \Phi(3) - \Phi(-3) = 2\Phi(3) - 1$. Avec $\Phi(3) \approx 0.99865$, on obtient $2(0.99865) - 1 \approx 0.9973$.

Ces valeurs sont souvent arrondies à 68

Cette règle fournit des repères très pratiques.

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

5.6 Calcul de Probabilités Normales

En pratique, pour calculer une probabilité $P(a \leq X \leq b)$ pour une loi $\mathcal{N}(\mu, \sigma^2)$, on utilise systématiquement la standardisation.

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** : $P(X > 130) = 1 - 0.9772 = 0.0228$. Environ 2.3

Pour les intervalles, on utilise la propriété $P(a \leq Z \leq b) = \Phi(b) - \Phi(a)$.

Exemple : Probabilité entre deux valeurs

Quelle est la probabilité $P(85 \leq X \leq 115)$? ($\mu = 100, \sigma = 15$)

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** : $P(85 \leq X \leq 115) \approx 2(0.8413) - 1 = 1.6826 - 1 = 0.6826$. (On retrouve la règle des 68)

On peut aussi inverser le processus : trouver la valeur x correspondant à une probabilité donnée.

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 (le 90ème percentile). 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%.

5.7 Exercices

Pour tous les exercices de calcul, vous pouvez utiliser les valeurs suivantes pour la fonction de répartition de la loi normale standard $\Phi(z) = P(Z \leq z)$:

- $\Phi(0) = 0.5$
- $\Phi(0.675) \approx 0.75$
- $\Phi(1) \approx 0.8413$
- $\Phi(1.28) \approx 0.90$
- $\Phi(1.5) \approx 0.9332$
- $\Phi(1.96) \approx 0.975$
- $\Phi(2) \approx 0.9772$
- $\Phi(2.5) \approx 0.9938$
- $\Phi(3) \approx 0.9987$

Et rappelez-vous la propriété de symétrie : $\Phi(-z) = 1 - \Phi(z)$.

Exercice 1 : Concepts (PDF)

Soit deux lois normales $A \sim \mathcal{N}(10, 4)$ et $B \sim \mathcal{N}(10, 9)$.

1. Laquelle des deux distributions a le pic le plus élevé ?
2. Laquelle des deux distributions est la plus "aplatie" ?

Exercice 2 : Z-Score (Calcul)

La taille des étudiants suit $X \sim \mathcal{N}(175, 6^2)$, où les unités sont en cm. Un étudiant mesure 184 cm. Calculez son Z-score.

Exercice 3 : Z-Score (Interprétation)

Un score de $Z = -2.5$ est obtenu pour une variable X . Qu'est-ce que cela signifie en termes de moyenne (μ) et d'écart-type (σ) ?

Exercice 4 : Z-Score (Calcul Inverse)

Pour une distribution $\mathcal{N}(50, 100)$, quelle valeur x correspond à un Z-score de $z = 1.5$?

Exercice 5 : Z-Score (Comparaison)

Alice obtient 115 à un test de QI $\mathcal{N}(100, 15^2)$. Bob obtient 24 à un test d'aptitude $\mathcal{N}(20, 2^2)$. Qui a le mieux réussi par rapport à son groupe ?

Exercice 6 : Règle Empirique (68-95-99.7)

Le poids de paquets de café suit $\mathcal{N}(500g, 10^2)$. En utilisant la règle empirique, quel pourcentage approximatif de paquets pèse entre 490g et 510g ?

Exercice 7 : Règle Empirique (Queue)

En utilisant la règle empirique pour $\mathcal{N}(\mu, \sigma^2)$, quelle est la probabilité approximative $P(X > \mu + 2\sigma)$?

Exercice 8 : Loi Standard (Lecture Directe)

Soit $Z \sim \mathcal{N}(0, 1)$. Calculez $P(Z \leq 1.5)$.

Exercice 9 : Loi Standard (Queue Droite)

Soit $Z \sim \mathcal{N}(0, 1)$. Calculez $P(Z > 1)$.

Exercice 10 : Loi Standard (Queue Gauche)

Soit $Z \sim \mathcal{N}(0, 1)$. Calculez $P(Z \leq -2)$.

Exercice 11 : Loi Standard (Intervalle)

Soit $Z \sim \mathcal{N}(0, 1)$. Calculez $P(-1 \leq Z \leq 2)$.

Exercice 12 : Calcul (Standardisation)

Soit $X \sim \mathcal{N}(50, 25)$. (Attention : $\sigma^2 = 25$). Calculez $P(X \leq 60)$.

Exercice 13 : Calcul (Standardisation)

Soit $X \sim \mathcal{N}(100, 225)$. (Attention : $\sigma^2 = 225$). Calculez $P(X > 77.5)$.

Exercice 14 : Calcul (Intervalle)

La durée de vie d'une batterie suit $X \sim \mathcal{N}(40 \text{ heures}, 16)$. Calculez $P(38 \leq X \leq 42)$.

Exercice 15 : Propriété de la CDF

En utilisant les définitions, montrez que $P(Z > z) = P(Z \leq -z)$.

Exercice 16 : Inverse (Z-Score)

Soit $Z \sim \mathcal{N}(0, 1)$. Trouvez la valeur z telle que $P(Z \leq z) = 0.975$.

Exercice 17 : Inverse (Z-Score Queue Droite)

Soit $Z \sim \mathcal{N}(0, 1)$. Trouvez la valeur z telle que $P(Z > z) = 0.1587$. (Indice : $1 - 0.1587 = 0.8413$).

Exercice 18 : Inverse (Z-Score Intervalle Central)

Soit $Z \sim \mathcal{N}(0, 1)$. Trouvez la valeur z (positive) telle que $P(-z \leq Z \leq z) = 0.95$. (Indice : si l'aire centrale est 0.95, combien vaut l'aire à gauche de z ?)

Exercice 19 : Problème Inverse (Application)

Le score à un examen suit $\mathcal{N}(100, 15^2)$. Pour obtenir la note "A", un étudiant doit être dans le top 2.5% (c'est-à-dire $P(X > x) = 0.025$). Quel score x minimum faut-il obtenir ?

Exercice 20 : Problème Inverse (Application)

Une machine remplit des sacs de farine $\mathcal{N}(1000g, 20^2)$. On veut garantir que 99.87% des sacs pèsent *plus* qu'un certain poids x . Quelle est la valeur de x ?

Exercice 21 : Stabilité par Transformation Linéaire

Si $X \sim \mathcal{N}(10, 4)$ (donc $\sigma = 2$), quelle est la loi de la variable $Y = 3X + 5$?

Exercice 22 : Stabilité (Changement d'Unités)

La taille T_{cm} d'une population suit $\mathcal{N}(170, 100)$. Quelle est la loi de la taille T_m en mètres ? (Rappel : $T_m = T_{cm}/100$).

Exercice 23 : Stabilité par Addition (Indépendante)

Soit $X \sim \mathcal{N}(50, 10)$ et $Y \sim \mathcal{N}(30, 6)$ deux variables indépendantes. Quelle est la loi de la somme $S = X + Y$?

Exercice 24 : Stabilité par Différence (Indépendante)

En utilisant les variables X et Y de l'exercice 23, quelle est la loi de la différence $D = X - Y$? (Indice : $D = X + (-1)Y$).

Exercice 25 : Somme de n variables i.i.d.

On prélève 10 pommes d'un lot où le poids d'une pomme suit $\mathcal{N}(150g, 10^2)$. Soit P_{total} le poids total des 10 pommes (supposées indépendantes). Quelle est la loi de P_{total} ?

5.8 Corrections des Exercices

Correction Exercice 1 : Concepts (PDF)

1. La variance de A ($\sigma_A^2 = 4$) est plus petite que celle de B ($\sigma_B^2 = 9$). Une variance plus petite signifie une distribution plus concentrée, donc un pic plus élevé (l'aire totale devant rester 1). C'est la distribution A. 2. La distribution B ($\sigma_B = 3$) a un écart-type plus grand que A ($\sigma_A = 2$), elle est donc plus dispersée, c'est-à-dire plus "large et aplatie".

Correction Exercice 2 : Z-Score (Calcul)

On a $x = 184$, $\mu = 175$, $\sigma = 6$.

$$Z = \frac{x - \mu}{\sigma} = \frac{184 - 175}{6} = \frac{9}{6} = 1.5$$

L'étudiant se situe à 1.5 écarts-types au-dessus de la moyenne.

Correction Exercice 3 : Z-Score (Interprétation)

Cela signifie que la valeur X observée est 2.5 écarts-types **en dessous** de la moyenne μ . (c-à-d, $x = \mu - 2.5\sigma$).

Correction Exercice 4 : Z-Score (Calcul Inverse)

On a $\mu = 50$ et $\sigma^2 = 100 \implies \sigma = 10$. On cherche x . $z = \frac{x - \mu}{\sigma} \implies x = \mu + z\sigma$
 $x = 50 + (1.5)(10) = 50 + 15 = 65$.

Correction Exercice 5 : Z-Score (Comparaison)

On compare les Z-scores : $Z_{Alice} = \frac{115-100}{15} = \frac{15}{15} = 1.0$. $Z_{Bob} = \frac{24-20}{2} = \frac{4}{2} = 2.0$. Bob a un Z-score plus élevé (2.0 contre 1.0), il a donc mieux réussi par rapport à son groupe.

Correction Exercice 6 : Règle Empirique (68-95-99.7)

On a $\mu = 500$ et $\sigma = 10$. L'intervalle $[490, 510]$ correspond à $[\mu - 1\sigma, \mu + 1\sigma]$. Selon la règle empirique, environ **68%** des paquets se trouvent dans cet intervalle.

Correction Exercice 7 : Règle Empirique (Queue)

La règle dit que $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95$. L'aire totale est 1. L'aire en dehors de cet intervalle est $1 - 0.95 = 0.05$. En raison de la symétrie, cette aire de 0.05 est répartie également entre les deux queues (gauche et droite). L'aire de la queue droite $P(X > \mu + 2\sigma)$ est donc $0.05/2 = 0.025$ (soit 2.5%).

Correction Exercice 8 : Loi Standard (Lecture Directe)

$P(Z \leq 1.5) = \Phi(1.5)$. En utilisant la table fournie, $P(Z \leq 1.5) \approx 0.9332$.

Correction Exercice 9 : Loi Standard (Queue Droite)

$P(Z > 1) = 1 - P(Z \leq 1) = 1 - \Phi(1)$. $P(Z > 1) \approx 1 - 0.8413 = 0.1587$.

Correction Exercice 10 : Loi Standard (Queue Gauche)

$P(Z \leq -2) = \Phi(-2)$. Par symétrie, $\Phi(-2) = 1 - \Phi(2)$. $P(Z \leq -2) \approx 1 - 0.9772 = 0.0228$.

Correction Exercice 11 : Loi Standard (Intervalle)

$P(-1 \leq Z \leq 2) = P(Z \leq 2) - P(Z \leq -1) = \Phi(2) - \Phi(-1)$. $\Phi(-1) = 1 - \Phi(1) \approx 1 - 0.8413 = 0.1587$. $P(-1 \leq Z \leq 2) \approx 0.9772 - 0.1587 = 0.8185$.

Correction Exercice 12 : Calcul (Standardisation)

$X \sim \mathcal{N}(50, 25) \implies \mu = 50, \sigma = 5$. On cherche $P(X \leq 60)$. $Z = \frac{60-50}{5} = \frac{10}{5} = 2$. $P(X \leq 60) = P(Z \leq 2) = \Phi(2) \approx 0.9772$.

Correction Exercice 13 : Calcul (Standardisation)

$X \sim \mathcal{N}(100, 225) \implies \mu = 100, \sigma = 15$. On cherche $P(X > 77.5)$. $Z = \frac{77.5-100}{15} = \frac{-22.5}{15} = -1.5$. $P(X > 77.5) = P(Z > -1.5) = 1 - P(Z \leq -1.5) = 1 - \Phi(-1.5)$. $\Phi(-1.5) = 1 - \Phi(1.5) \approx 1 - 0.9332 = 0.0668$. $P(Z > -1.5) = 1 - 0.0668 = 0.9332$. (Logique : $P(Z > -1.5) = P(Z \leq 1.5)$ par symétrie).

Correction Exercice 14 : Calcul (Intervalle)

$X \sim \mathcal{N}(40, 16) \implies \mu = 40, \sigma = 4$. On cherche $P(38 \leq X \leq 42)$. $z_1 = \frac{38-40}{4} = -0.5$. $z_2 = \frac{42-40}{4} = 0.5$. $P(38 \leq X \leq 42) = P(-0.5 \leq Z \leq 0.5) = \Phi(0.5) - \Phi(-0.5)$. $\Phi(-0.5) = 1 - \Phi(0.5)$. $P = \Phi(0.5) - (1 - \Phi(0.5)) = 2\Phi(0.5) - 1$. (La valeur $\Phi(0.5) \approx 0.6915$ n'est pas fournie, mais $P(-1 \leq Z \leq 1) \approx 0.68$, donc on s'attend à une valeur plus petite).

Correction Exercice 15 : Propriété de la CDF

$P(Z > z) = 1 - P(Z \leq z) = 1 - \Phi(z)$. $P(Z \leq -z) = \Phi(-z)$. Par symétrie de la PDF $\phi(z) = \phi(-z)$, l'aire à droite de z est égale à l'aire à gauche de $-z$. Donc $P(Z > z) = P(Z \leq -z)$, ce qui implique $1 - \Phi(z) = \Phi(-z)$.

Correction Exercice 16 : Inverse (Z-Score)

On cherche z tel que $P(Z \leq z) = 0.975$. C'est $\Phi(z) = 0.975$. D'après la table, $z = 1.96$.

Correction Exercice 17 : Inverse (Z-Score Queue Droite)

On cherche z tel que $P(Z > z) = 0.1587$. Cela signifie $P(Z \leq z) = 1 - 0.1587 = 0.8413$. $\Phi(z) = 0.8413$. D'après la table, $z = 1$.

Correction Exercice 18 : Inverse (Z-Score Intervalle Central)

Si $P(-z \leq Z \leq z) = 0.95$, l'aire restante dans les deux queues est $1 - 0.95 = 0.05$. Par symétrie, l'aire dans la queue gauche $P(Z < -z)$ est $0.05/2 = 0.025$. L'aire totale à gauche de z est $P(Z \leq z) = 0.95 + 0.025 = 0.975$. On cherche z tel que $\Phi(z) = 0.975$. D'après la table, $z = 1.96$. (On retrouve $\mu \pm 1.96\sigma$ comme l'intervalle à 95% exact).

Correction Exercice 19 : Problème Inverse (Application)

$X \sim \mathcal{N}(100, 15^2)$. On cherche x tel que $P(X > x) = 0.025$. Standardisation : $P(Z > z) = 0.025$, où $z = (x - 100)/15$. $P(Z \leq z) = 1 - 0.025 = 0.975$. D'après la table, z tel que $\Phi(z) = 0.975$ est $z = 1.96$. On résout : $1.96 = \frac{x-100}{15} \implies x = 100 + 1.96(15) = 100 + 29.4 = 129.4$. Il faut un score minimum de 129.4.

Correction Exercice 20 : Problème Inverse (Application)

$X \sim \mathcal{N}(1000, 20^2)$. On cherche x tel que $P(X > x) = 0.9987$. Standardisation : $P(Z > z) = 0.9987$, où $z = (x - 1000)/20$. $P(Z \leq z) = 1 - 0.9987 = 0.0013$. C'est une valeur très faible. Utilisons la symétrie : $\Phi(-z) = 1 - \Phi(z)$. $\Phi(z) = 0.0013$. On cherche z dans la table. On voit que $\Phi(3) = 0.9987$, donc $\Phi(-3) = 1 - 0.9987 = 0.0013$. Le Z-score est $z = -3$. On résout : $-3 = \frac{x-1000}{20} \implies x = 1000 - 3(20) = 1000 - 60 = 940$. Le poids garanti est 940g.

Correction Exercice 21 : Stabilité par Transformation Linéaire

$X \sim \mathcal{N}(10, 4) \implies \mu_X = 10, \sigma_X^2 = 4$. $Y = aX + b$ avec $a = 3, b = 5$. Y suit une loi normale. $E[Y] = a\mu_X + b = 3(10) + 5 = 35$. $\text{Var}(Y) = a^2\text{Var}(X) = 3^2 \times 4 = 9 \times 4 = 36$. Donc, $Y \sim \mathcal{N}(35, 36)$.

Correction Exercice 22 : Stabilité (Changement d'Unités)

$T_{cm} \sim \mathcal{N}(170, 100) \implies \mu = 170, \sigma^2 = 100$. $T_m = aT_{cm} + b$ avec $a = 1/100 = 0.01$ et $b = 0$. $E[T_m] = a\mu + b = 0.01(170) + 0 = 1.7$. $\text{Var}(T_m) = a^2\text{Var}(T_{cm}) = (0.01)^2 \times 100 = 0.0001 \times 100 = 0.01$. Donc, $T_m \sim \mathcal{N}(1.7, 0.01)$. (L'écart-type est $\sigma_m = \sqrt{0.01} = 0.1$ m, ce qui est logique : 10cm = 0.1m).

Correction Exercice 23 : Stabilité par Addition (Indépendante)

$X \sim \mathcal{N}(50, 10)$, $Y \sim \mathcal{N}(30, 6)$. X, Y indépendantes. $S = X + Y$ suit une loi normale. $E[S] = E[X] + E[Y] = 50 + 30 = 80$. $\text{Var}(S) = \text{Var}(X) + \text{Var}(Y) = 10 + 6 = 16$. Donc, $S \sim \mathcal{N}(80, 16)$.

Correction Exercice 24 : Stabilité par Différence (Indépendante)

$D = X - Y = X + (-1)Y$. C'est une somme de variables normales indépendantes (si X, Y le sont, $X, -Y$ le sont aussi). D suit une loi normale. $E[D] = E[X] + E[-Y] = E[X] - E[Y] = 50 - 30 = 20$. $\text{Var}(D) = \text{Var}(X) + \text{Var}(-1 \cdot Y) = \text{Var}(X) + (-1)^2\text{Var}(Y) = \text{Var}(X) + \text{Var}(Y) = 10 + 6 = 16$. Donc, $D \sim \mathcal{N}(20, 16)$. (Note : les variances s'ajoutent toujours!)

Correction Exercice 25 : Somme de n variables i.i.d.

Soit $P_i \sim \mathcal{N}(150, 100)$ le poids de la i -ème pomme. $P_{total} = P_1 + \dots + P_{10}$. C'est une somme de 10 v.a. normales indépendantes. P_{total} suit une loi normale. $E[P_{total}] = E[P_1] + \dots + E[P_{10}] = 10 \times E[P_1] = 10 \times 150 = 1500$ g. $\text{Var}(P_{total}) = \text{Var}(P_1) + \dots + \text{Var}(P_{10}) = 10 \times \text{Var}(P_1) = 10 \times 100 = 1000$. Donc, $P_{total} \sim \mathcal{N}(1500, 1000)$.

5.9 Exercices Python

Ces exercices appliquent les concepts de la loi normale au jeu de données "Yahoo Finance". Nous allons modéliser les **rendements journaliers** (variation en pourcentage) des actions, qui sont souvent (par approximation) considérés comme suivant une loi normale.

Nous allons travailler avec les rendements de Google (X) et de Microsoft (Y).

```
!pip install yfinance
import yfinance as yf
import pandas as pd
```

```
import numpy as np
from scipy.stats import norm # Moteur pour les calculs de CDF/PDF

# Définir les tickers et la période
tickers = ["GOOG", "MSFT"]
start_date = "2020-01-01"
end_date = "2024-12-31"

# Télécharger les prix de clôture ajustés
data = yf.download(tickers, start=start_date, end=end_date)["Adj Close"]

# Calculer les rendements journaliers en pourcentage
returns = data.pct_change().dropna()

# Renommer les colonnes pour plus de clarté
returns.columns = ["GOOG_Return", "MSFT_Return"]

# 'returns' est notre DataFrame principal.
# X = returns["GOOG_Return"]
# Y = returns["MSFT_Return"]
```

Exercice 1 : Estimation des Paramètres μ et σ^2

Soit X la v.a. "Rendement journalier de GOOG" et Y la v.a. "Rendement journalier de MSFT". Nous supposons $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ et $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$.

Votre tâche :

1. Estimer μ_X et μ_Y (les espérances) en calculant la moyenne empirique (`.mean()`) des deux séries.
2. Estimer σ_X^2 et σ_Y^2 (les variances) en calculant la variance empirique (`.var()`).
3. Estimer σ_X et σ_Y (les écarts-types) en prenant la racine carrée des variances (ou `.std()`).

Exercice 2 : Standardisation (Score Z)

Le Z-score nous dit à combien d'écarts-types une observation se situe de la moyenne.

Votre tâche :

1. Utiliser les μ_X et σ_X (pour GOOG) estimés à l'exercice 1.
2. Trouver le rendement du dernier jour disponible dans vos données pour GOOG.
3. Calculer le Z-score de ce rendement : $Z = (x - \mu_X) / \sigma_X$.
4. Interpréter ce score (ex : "Le rendement de ce jour était à [Z] écarts-types de la moyenne...").

Exercice 3 : Calcul de Probabilité (Utilisation de Φ)

En utilisant les paramètres μ_X et σ_X pour GOOG (Ex 1) et en supposant la distribution normale :

Votre tâche :

1. Calculer la probabilité qu'un jour donné, le rendement de GOOG soit "calme", c'est-à-dire $P(-0.01 \leq X \leq 0.01)$.
2. (Indice : Standardisez $a = -0.01$ et $b = 0.01$ en z_a, z_b , puis calculez $\Phi(z_b) - \Phi(z_a)$. Vous aurez besoin de `scipy.stats.norm.cdf()`).

Exercice 4 : Problème Inverse (Value at Risk)

Le "Value at Risk" (VaR) est une valeur x telle qu'il y a une probabilité p de perdre plus que x .

Votre tâche :

1. Toujours avec les paramètres de GOOG, trouvez la valeur x (le rendement) qui correspond au "top 5%" des pires jours.
2. Autrement dit, trouvez x tel que $P(X \leq x) = 0.05$.
3. (Indice : Trouvez le Z-score z tel que $\Phi(z) = 0.05$ (vous aurez besoin de `scipy.stats.norm.ppf()`), puis "dé-standardisez" : $x = \mu_X + z\sigma_X$).

Exercice 5 : PDF (Calcul de la Densité)

La PDF $f(x)$ n'est pas une probabilité, mais une "densité". La valeur $f(\mu)$ est le point le plus haut de la cloche.

Votre tâche :

1. Utiliser les μ_X et σ_X de GOOG (Ex 1).
2. Calculer la valeur de la densité de probabilité au point $x = \mu_X$ (le pic de la cloche).
3. Calculer la valeur de la densité au point $x = 0.0$.
4. (Indice : Vous aurez besoin de `scipy.stats.norm.pdf(x, loc=mu, scale=sigma)`).

Exercice 6 : Vérification de la Règle Empirique (68-95-99.7)

La théorie dit que $\approx 68\%$ des données devraient être dans $\mu \pm \sigma$. Nous allons vérifier si les rendements boursiers respectent cette règle.

Votre tâche :

1. Utiliser les μ_X et σ_X de GOOG (Ex 1).
2. Calculer la proportion **réelle** des rendements qui tombent dans l'intervalle $[\mu_X - \sigma_X, \mu_X + \sigma_X]$.
3. (Bonus) Faire de même pour $[\mu_X - 2\sigma_X, \mu_X + 2\sigma_X]$ (théorie : 95%).
4. (Conclusion) Les rendements de GOOG semblent-ils suivre parfaitement la règle?

Exercice 7 : Calcul des Quartiles

L'écart interquartile (IQR) est une autre mesure de dispersion, $IQR = Q_3 - Q_1$. Q_1 est la valeur x telle que $P(X \leq x) = 0.25$. Q_3 est la valeur x telle que $P(X \leq x) = 0.75$.

Votre tâche :

1. En utilisant le modèle normal pour GOOG (μ_X, σ_X), trouver les Z-scores z_1 et z_3 pour $p = 0.25$ et $p = 0.75$ (`norm.ppf`).
2. "Dé-standardiser" ces Z-scores pour trouver Q_1 et Q_3 .
3. Calculer l'IQR **théorique** ($Q_3 - Q_1$).
4. Comparer cet IQR théorique à l'IQR **empirique** de la série X (Indice : `X.quantile(0.75) - X.quantile(0.25)`).

Exercice 8 : Stabilité par Transformation Linéaire (Y)

Soit un produit financier (ETF) qui amplifie par 2 les mouvements de GOOG, $Y = 2X$. La théorie prédit $E[Y] = 2E[X]$ et $\text{Var}(Y) = 2^2\text{Var}(X) = 4\text{Var}(X)$.

Votre tâche :

1. Calculer $E[Y]$ et $\text{Var}(Y)$ **théoriquement** en utilisant $E[X]$ et $\text{Var}(X)$ de l'Ex 1.
2. **Vérification empirique :**
 - Créer la série de données $Y_{series} = 2 \times X$.
 - Calculer la moyenne (`.mean()`) et la variance (`.var()`) de Y_{series} .
3. Comparer les résultats théoriques et empiriques.

Exercice 9 : Stabilité par Addition (Portefeuille S)

Théorie : $E[X + Y] = E[X] + E[Y]$ et $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$. *Note : X et Y (GOOG, MSFT) ne sont PAS indépendantes.*

Votre tâche :

1. Créer la série $S = X + Y$.
2. Calculer $E[S]$ (la moyenne de S).
3. Vérifier que $E[S] = E[X] + E[Y]$ (en utilisant les valeurs de l'Ex 1).
4. Calculer $\text{Var}(S)$ (la variance de S).
5. Est-ce que $\text{Var}(S) = \text{Var}(X) + \text{Var}(Y)$? Pourquoi y a-t-il une différence?

Exercice 10 : Covariance et Variance du Portefeuille (Corrigé)

Reprenons l'exercice 9. La différence est due à la covariance.

Votre tâche :

1. Calculer $\text{Cov}(X, Y)$ (vous pouvez utiliser `returns.cov()`).
2. Calculer la variance **théorique** de $S = X + Y$ avec la formule complète : $\text{Var}(S_{theo}) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$.
3. Comparer $\text{Var}(S_{theo})$ à la variance empirique $\text{Var}(S)$ calculée à l'exercice 9.

Exercice 11 : Distribution d'un Portefeuille Pondéré

Un portefeuille P est composé de 60% de GOOG (X) et 40% de MSFT (Y). Donc $P = 0.6X + 0.4Y$. Si X et Y sont (conjointement) normales, P est aussi normale.

Votre tâche :

1. Calculer l'espérance $E[P]$ (en utilisant la linéarité : $0.6E[X] + 0.4E[Y]$).
2. Calculer la variance $\text{Var}(P)$ (en utilisant la formule complète : $\text{Var}(0.6X + 0.4Y) = 0.6^2\text{Var}(X) + 0.4^2\text{Var}(Y) + 2(0.6)(0.4)\text{Cov}(X, Y)$).
3. Énoncer la loi de probabilité approximative du portefeuille : $P \sim \mathcal{N}(\mu_P, \sigma_P^2)$.

Exercice 12 : Distribution de la Différence (Pairs Trading)

Un "spread" D est la différence de rendement $D = X - Y$. Théorie : $E[D] = E[X] - E[Y]$ et $\text{Var}(D) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y)$.

Votre tâche :

1. Calculer $E[D]$ et $\text{Var}(D)$ **théoriquement** en utilisant les valeurs des exercices précédents.
2. **Vérification empirique :**
 - Créer la série de données $D_{series} = X - Y$.
 - Calculer la moyenne (`.mean()`) et la variance (`.var()`) de D_{series} .
3. Comparer les résultats théoriques et empiriques.

6 La Loi Log-Normale

6.1 Introduction et Définition

Après la loi normale, qui modélise des phénomènes résultant de *l'addition* de nombreux petits effets, nous abordons une loi qui modélise le *produit* de nombreux petits facteurs.

Définition : Loi Log-Normale

Une variable aléatoire continue Y suit une **loi log-normale** de paramètres μ et σ^2 , notée $Y \sim \text{Log-}\mathcal{N}(\mu, \sigma^2)$, si son logarithme népérien $X = \ln(Y)$ suit une loi normale $\mathcal{N}(\mu, \sigma^2)$. Puisque $\ln(Y)$ doit être un nombre réel, le support de Y (l'ensemble des valeurs qu'elle peut prendre) est l'intervalle $(0, \infty)$.

Remarque : Interprétation Cruciale des Paramètres

C'est le point le plus important (et le plus confus) de la loi log-normale :

- μ n'est **PAS** l'espérance de Y .
- σ^2 n'est **PAS** la variance de Y .

Au lieu de cela, $\mu = E[\ln(Y)]$ et $\sigma^2 = \text{Var}(\ln(Y))$ sont l'espérance et la variance de la variable normale *sous-jacente* (le log de Y).

6.2 Intuition : Des Sommes aux Produits

La loi log-normale peut sembler complexe, mais elle émerge naturellement de processus multiplicatifs, tout comme la loi normale émerge de processus additifs.

Intuition : Du Multiplicatif à l'Additif (et retour)

1. Le Contraste (Additif \rightarrow Normal) : La loi normale (via le Théorème Central Limite) décrit le résultat d'une **somme** de nombreux petits chocs aléatoires indépendants (comme les ± 1 d'un jeu de pile ou face répété). Le résultat est symétrique autour d'une moyenne.

$$S_n = X_1 + X_2 + \dots + X_n \xrightarrow{\text{TCL}} \mathcal{N}$$

2. Le Problème (Multiplicatif) : De nombreux phénomènes réels ne s'additionnent pas, ils se *multiplient*.

- **Finance :** Un capital C évolue par des rendements en pourcentage. $C_T = C_0 \times (1 + r_1) \times (1 + r_2) \times \dots \times (1 + r_T)$.
- **Biologie :** La taille d'une population de bactéries est $P_t = P_0 \times F_1 \times F_2 \times \dots \times F_t$, où F_i est le facteur de croissance à l'étape i .

Nous voulons décrire la distribution d'un **produit** de facteurs aléatoires :

$$Y = F_1 \times F_2 \times \dots \times F_n$$

3. L'Astuce (Le Logarithme) : Les produits sont difficiles à manipuler. L'astuce mathématique fondamentale est de prendre le logarithme, qui **transforme les produits en sommes** :

$$\ln(Y) = \ln(F_1) + \ln(F_2) + \dots + \ln(F_n)$$

4. L'Apparition de la Loi Normale : Nous avons transformé notre problème ! La variable $\ln(Y)$ est maintenant une *somme* de petits chocs aléatoires (les $\ln(F_i)$). D'après le Théorème Central Limite, si n est grand, la distribution de cette *somme* $\ln(Y)$ va converger vers une loi normale.

$$\ln(Y) = X \sim \mathcal{N}(\mu, \sigma^2)$$

C'est l'origine du nom "log-normal" : le *log* de la variable est *normal*.

5. Le Retour (L'Exponentielle) : Pour trouver la distribution de Y (notre variable d'intérêt), il nous faut inverser l'opération du logarithme. Nous prenons l'exponentielle :

$$Y = e^X \quad \text{où } X \sim \mathcal{N}(\mu, \sigma^2)$$

Cette transformation $x \mapsto e^x$ change radicalement la forme de la distribution.

- La cloche symétrique de X (qui peut être négative) est transformée.
- Puisque $e^x > 0$ pour tout x , la variable Y est **strictement positive**.
- La partie gauche de la cloche (valeurs $x \rightarrow -\infty$) est **compressée** près de 0 ($e^x \rightarrow 0$).
- La partie droite de la cloche (valeurs $x \rightarrow +\infty$) est **étirée** exponentiellement ($e^x \rightarrow \infty$).

Le résultat est une distribution **asymétrique à droite** (biaisée positivement), avec une longue queue vers les valeurs élevées.

Conclusion : La loi log-normale est la loi des **produits** de nombreux petits facteurs aléatoires. Elle modélise des grandeurs qui ne peuvent être négatives et qui ont une probabilité non négligeable d'atteindre des valeurs très grandes.

6.3 Fonction de Densité (PDF)

On peut déduire la forme de la PDF log-normale en utilisant la méthode du changement de variable sur $Y = e^X$.

Théorème : Fonction de Densité Log-Normale

Si $Y \sim \text{Log-}\mathcal{N}(\mu, \sigma^2)$, sa fonction de densité de probabilité est donnée par :

$$f_Y(y) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{(\ln(y)-\mu)^2}{2\sigma^2}}$$

pour tout $y > 0$.

Preuve : Dérivation par Changement de Variable

Nous cherchons la densité $f_Y(y)$ de $Y = g(X) = e^X$, sachant que $X = g^{-1}(Y) = \ln(Y)$ a pour densité $f_X(x) = \phi\left(\frac{x-\mu}{\sigma}\right) \frac{1}{\sigma}$. La formule de changement de variable est $f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d(g^{-1}(y))}{dy} \right|$. Ici, $x = g^{-1}(y) = \ln(y)$. La dérivée est $\frac{dx}{dy} = \frac{1}{y}$. Puisque $y > 0$, on a $\left| \frac{1}{y} \right| = \frac{1}{y}$. Substituons $x = \ln(y)$ dans la densité de X et multiplions par la dérivée :

$$f_Y(y) = f_X(\ln(y)) \cdot \frac{1}{y}$$

$$f_Y(y) = \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{\ln(y)-\mu}{\sigma} \right)^2} \right] \cdot \frac{1}{y}$$

En réarrangeant, on obtient :

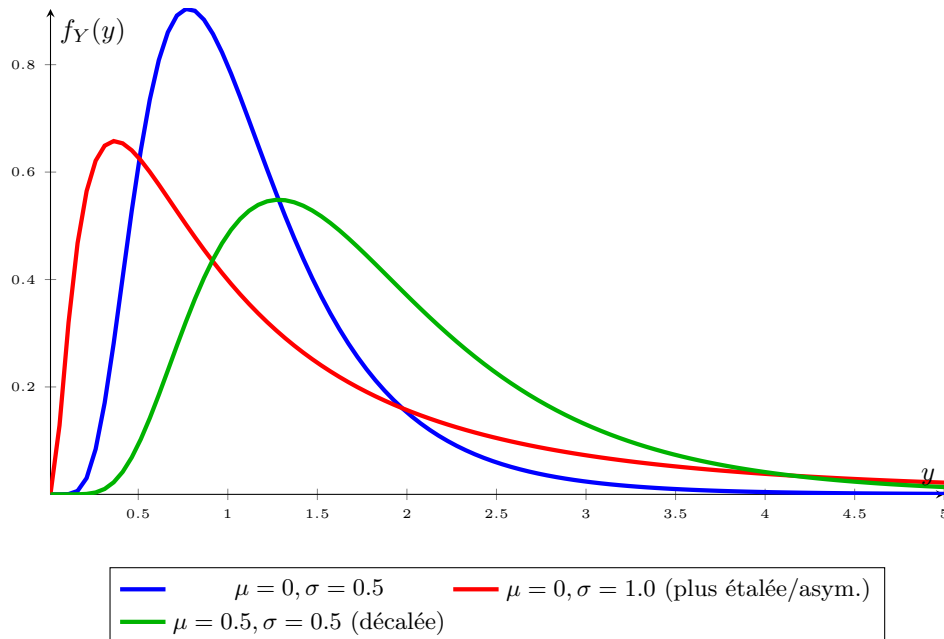
$$f_Y(y) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{(\ln(y)-\mu)^2}{2\sigma^2}}$$

Intuition : Forme de la PDF

La PDF log-normale confirme notre intuition :

- Elle n'est définie que pour $y > 0$.
- Elle est fortement **asymétrique à droite**. Elle commence à $f_Y(0) = 0$, monte rapidement vers un pic (le mode), puis redescend lentement, formant une longue "queue" vers les grandes valeurs.
- Le paramètre σ (l'écart-type de $\ln Y$) contrôle l'**asymétrie** de Y . Un σ plus grand rend la distribution plus étalée et plus asymétrique.

PDF de la Loi Log-Normale



Influence de μ (échelle) et σ (forme/asymétrie) sur la PDF log-normale.

6.4 Comment obtenir l'espérance d'une variable log-normale

L'espérance d'une variable log-normale ne se calcule pas directement à partir de sa densité sans précaution, car la forme de cette densité rend l'intégrale peu intuitive. Heureusement, la définition même de la loi log-normale — via une transformation exponentielle d'une variable normale — fournit une voie élégante et puissante.

Intuition : Stratégie de calcul : exploiter la variable sous-jacente

Puisque $Y \sim \text{Log-}\mathcal{N}(\mu, \sigma^2)$ signifie que $Y = e^X$ avec $X \sim \mathcal{N}(\mu, \sigma^2)$, on peut écrire :

$$\mathbb{E}[Y] = \mathbb{E}[e^X].$$

Il s'agit donc de calculer l'espérance de l'exponentielle d'une variable normale. Ce type d'espérance est exactement ce que la **fonction génératrice des moments** (MGF, pour *Moment Generating Function*) permet de calculer.

Rappelons que pour une variable aléatoire X , la MGF est définie par :

$$M_X(t) = \mathbb{E}[e^{tX}],$$

lorsqu'elle existe. Ainsi, $\mathbb{E}[e^X] = M_X(1)$.

Preuve : Calcul détaillé de $\mathbb{E}[Y]$

Y

par intégration directe

Nous allons maintenant démontrer rigoureusement que :

$$\mathbb{E}[Y] = e^{\mu + \sigma^2/2},$$

en partant de la définition de l'espérance et en effectuant le calcul d'intégrale.

Puisque $Y = e^X$ et $X \sim \mathcal{N}(\mu, \sigma^2)$, on a :

$$\mathbb{E}[Y] = \mathbb{E}[e^X] = \int_{-\infty}^{+\infty} e^x \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx.$$

Regroupons les exponentielles :

$$\mathbb{E}[Y] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(x - \frac{(x-\mu)^2}{2\sigma^2}\right) dx.$$

Développons l'exposant :

$$x - \frac{(x-\mu)^2}{2\sigma^2} = x - \frac{x^2 - 2\mu x + \mu^2}{2\sigma^2} = -\frac{x^2}{2\sigma^2} + \left(\frac{\mu}{\sigma^2} + 1\right)x - \frac{\mu^2}{2\sigma^2}.$$

Complétons le carré en x . Posons :

$$-\frac{1}{2\sigma^2} [x^2 - 2(\mu + \sigma^2)x] - \frac{\mu^2}{2\sigma^2}.$$

Le terme entre crochets devient :

$$x^2 - 2(\mu + \sigma^2)x = (x - (\mu + \sigma^2))^2 - (\mu + \sigma^2)^2.$$

Donc l'exposant total vaut :

$$-\frac{1}{2\sigma^2} (x - (\mu + \sigma^2))^2 + \frac{(\mu + \sigma^2)^2 - \mu^2}{2\sigma^2}.$$

Calculons le terme constant :

$$(\mu + \sigma^2)^2 - \mu^2 = 2\mu\sigma^2 + \sigma^4, \quad \text{donc} \quad \frac{2\mu\sigma^2 + \sigma^4}{2\sigma^2} = \mu + \frac{\sigma^2}{2}.$$

Ainsi :

$$\mathbb{E}[Y] = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\mu + \frac{\sigma^2}{2}\right) \int_{-\infty}^{+\infty} \exp\left(-\frac{(x - (\mu + \sigma^2))^2}{2\sigma^2}\right) dx.$$

Mais l'intégrale est celle de la densité d'une loi normale $\mathcal{N}(\mu + \sigma^2, \sigma^2)$, donc elle vaut $\sigma\sqrt{2\pi}$.
Il reste donc :

$$\mathbb{E}[Y] = \exp\left(\mu + \frac{\sigma^2}{2}\right).$$

Remarque : Pourquoi cette formule est contre-intuitive ?

Il est tentant de penser que si $\ln(Y)$ a pour moyenne μ , alors Y devrait avoir pour moyenne e^μ . Mais à cause de la **convexité** de la fonction exponentielle, l'espérance de e^X est toujours **strictement supérieure** à $e^{\mathbb{E}[X]}$ dès que X a une variance non nulle. C'est une conséquence directe de l'**inégalité de Jensen** :

$$\mathbb{E}[e^X] \geq e^{\mathbb{E}[X]},$$

avec égalité uniquement si X est constante (variance nulle). Le terme supplémentaire $\sigma^2/2$ dans l'exposant quantifie exactement cet « effet de convexité ».

Exemple : Application numérique : comparaison des mesures centrales

Soit $Y \sim \text{Log-}\mathcal{N}(1, 0.25)$ (donc $\mu = 1$, $\sigma = 0.5$).

- Médiane : $e^\mu = e^1 \approx 2.718$
- Mode : $e^{\mu - \sigma^2} = e^{1 - 0.25} = e^{0.75} \approx 2.117$
- Espérance : $e^{\mu + \sigma^2/2} = e^{1 + 0.125} = e^{1.125} \approx 3.080$

On observe clairement :

$$\text{Mode} \approx 2.12 < \text{Médiane} \approx 2.72 < \text{Espérance} \approx 3.08,$$

illustrant l'asymétrie à droite et l'impact de la variance sur la moyenne.

6.5 Propriétés : Moments et Mesures Centrales

L'asymétrie de la loi log-normale implique que ses mesures de tendance centrale (moyenne, médiane, mode) ne coïncident pas.

Théorème : Moments et Mesures Centrales

Si $Y \sim \text{Log-}\mathcal{N}(\mu, \sigma^2)$, alors :

- **Espérance** : $E[Y] = e^{\mu + \sigma^2/2}$
- **Variance** : $\text{Var}(Y) = (e^{\sigma^2} - 1) \cdot e^{2\mu + \sigma^2}$
- **Médiane** : $\text{Med}(Y) = e^\mu$
- **Mode** : $\text{Mode}(Y) = e^{\mu - \sigma^2}$

Preuve : Dérivation de l'Espérance

Nous utilisons le fait que $Y = e^X$ où $X \sim \mathcal{N}(\mu, \sigma^2)$.

$$E[Y] = E[e^X]$$

Ceci est, par définition, la fonction génératrice des moments $M_X(t) = E[e^{tX}]$ de la variable normale X , évaluée en $t = 1$. Rappelons que la MGF d'une loi $\mathcal{N}(\mu, \sigma^2)$ est $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$. En posant $t = 1$, on obtient :

$$E[Y] = M_X(1) = e^{\mu(1) + \frac{1}{2}\sigma^2(1)^2} = e^{\mu + \sigma^2/2}$$

Un calcul similaire pour $E[Y^2] = E[e^{2X}] = M_X(2)$ permet de trouver la variance.

Remarque : Mode < Médiane < Espérance

Pour $\sigma > 0$, on a toujours $\mu - \sigma^2 < \mu < \mu + \sigma^2/2$. En appliquant l'exponentielle (qui est croissante), on a :

$$e^{\mu - \sigma^2} < e^\mu < e^{\mu + \sigma^2/2}$$

$$\text{Mode} < \text{Médiane} < \text{Espérance}$$

C'est la signature mathématique d'une distribution avec une asymétrie à droite (positive). La moyenne est "tirée" vers le haut par les valeurs extrêmes de la queue droite.

6.6 Calcul de Probabilités

Le principal avantage de la définition $X = \ln(Y)$ est qu'elle rend les calculs de probabilité log-normale très simples : il suffit de tout ramener à la loi normale standard.

Théorème : Calcul de la CDF Log-Normale

Si $Y \sim \text{Log-}\mathcal{N}(\mu, \sigma^2)$, sa fonction de répartition $F_Y(y) = P(Y \leq y)$ est donnée par :

$$F_Y(y) = \Phi\left(\frac{\ln(y) - \mu}{\sigma}\right) \quad \text{pour } y > 0$$

où $\Phi(\cdot)$ est la fonction de répartition de la loi normale standard $\mathcal{N}(0, 1)$.

Preuve

Puisque $\ln(\cdot)$ est une fonction strictement croissante, l'événement $Y \leq y$ est équivalent à l'événement $\ln(Y) \leq \ln(y)$.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\ln(Y) \leq \ln(y)) \end{aligned}$$

Puisque $X = \ln(Y) \sim \mathcal{N}(\mu, \sigma^2)$, on standardise cette variable X :

$$= P\left(\frac{\ln(Y) - \mu}{\sigma} \leq \frac{\ln(y) - \mu}{\sigma}\right)$$

Soit $Z = (\ln(Y) - \mu)/\sigma \sim \mathcal{N}(0, 1)$:

$$\begin{aligned} &= P\left(Z \leq \frac{\ln(y) - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{\ln(y) - \mu}{\sigma}\right) \end{aligned}$$

Exemple : Calcul de Probabilité Log-Normale

Le revenu annuel Y des ménages d'une région (en milliers d'euros) est supposé suivre une loi log-normale $\text{Log-}\mathcal{N}(\mu = 3.8, \sigma = 0.5)$. (Note : $\mu = 3.8$ n'est pas la moyenne des revenus, mais la moyenne du *log-revenu*).

1. Quel est le revenu médian ? $\text{Med}(Y) = e^\mu = e^{3.8} \approx 44.7$. (Soit 44 700 €)
2. Quelle est la probabilité qu'un ménage ait un revenu supérieur à 60 000 € ? On cherche $P(Y > 60)$. (Rappel : Y est en milliers). $P(Y > 60) = 1 - P(Y \leq 60)$.
On applique la formule :

$$P(Y \leq 60) = \Phi\left(\frac{\ln(60) - \mu}{\sigma}\right) = \Phi\left(\frac{\ln(60) - 3.8}{0.5}\right)$$

Calculons l'argument :

$$\frac{4.094 - 3.8}{0.5} = \frac{0.294}{0.5} = 0.588$$

On cherche $\Phi(0.588)$. En utilisant une table ou un logiciel, $\Phi(0.588) \approx 0.7217$.
 $P(Y > 60) = 1 - 0.7217 = 0.2783$. Environ 27.8

6.7 Évolution d'un Actif et Émergence de la Loi Log-Normale

Exemple : Évolution d'un Actif et Émergence de la Loi Log-Normale

Considérons un investisseur qui place un capital initial de $C_0 = 1000$ € dans une action cotée en bourse. Le prix de cette action évolue chaque jour selon des rendements aléatoires. Nous allons suivre son évolution sur $t = 4$ jours consécutifs, puis généraliser pour comprendre l'origine des paramètres $t\mu$ et $t\sigma^2$.

1. Données observées : rendements quotidiens

Les rendements (exprimés en décimal) sont les suivants :

Jour	Rendement r_i	Facteur de croissance $(1 + r_i)$	Log-rendement $x_i = \ln(1 + r_i)$
1	+0,05	1,05	$\ln(1,05) \approx 0,04879$
2	-0,02	0,98	$\ln(0,98) \approx -0,02020$
3	+0,03	1,03	$\ln(1,03) \approx 0,02956$
4	-0,01	0,99	$\ln(0,99) \approx -0,01005$

2. Calcul du capital final : la nature multiplicative

Le capital après chaque jour se calcule par multiplication :

$$\begin{aligned} C_1 &= 1000 \times 1,05 = 1050,00 \\ C_2 &= 1050 \times 0,98 = 1029,00 \\ C_3 &= 1029 \times 1,03 = 1059,87 \\ C_4 &= 1059,87 \times 0,99 = 1049,27 \end{aligned}$$

On peut aussi écrire directement :

$$C_4 = C_0 \times \prod_{i=1}^4 (1 + r_i) = 1000 \times (1,05 \times 0,98 \times 1,03 \times 0,99) = 1000 \times 1,04927.$$

Le **facteur de croissance total** est donc :

$$Y = \frac{C_4}{C_0} = 1,04927.$$

> *Pourquoi multiplier ?* Parce que les rendements sont relatifs : une perte de 2% le deuxième jour s'applique au capital du jour 1 (1050 €), pas au capital initial. L'addition des pourcentages (+5 - 2 + 3 - 1 = +5%) serait donc incorrecte.

3. Transformation logarithmique : du produit à la somme

On définit le log-rendement total :

$$X = \ln(Y) = \sum_{i=1}^4 \ln(1 + r_i) = \sum_{i=1}^4 x_i.$$

En utilisant les valeurs du tableau :

$$X \approx 0,04879 - 0,02020 + 0,02956 - 0,01005 = 0,04810.$$

Vérification : $e^{0,04810} \approx 1,04927 = Y$. La transformation est exacte.

Pourquoi faire cela ? Parce que la somme de variables aléatoires est bien comprise en probabilité (moyenne additive, variance additive sous indépendance), contrairement au produit.

4. Modélisation stochastique : hypothèses fondamentales

Dans la réalité, les rendements futurs sont inconnus. On suppose donc que chaque log-rendement journalier x_i est une variable aléatoire indépendante, identiquement distribuée, avec :

- Espérance $\mathbb{E}[x_i] = \mu$: rendement moyen continu par jour,
- Variance $\text{Var}(x_i) = \sigma^2$: volatilité au carré par jour.

Ces hypothèses traduisent l'idée que le marché est « mémoire courte » (indépendance) et que les conditions de volatilité sont stables.

5. Propriétés de la somme sur t périodes

Soit $t = 4$ le nombre de jours. Le log-rendement total est :

$$X = \sum_{i=1}^t x_i.$$

Grâce à l'indépendance :

$$\mathbb{E}[X] = \sum_{i=1}^t \mathbb{E}[x_i] = t\mu,$$

$$\text{Var}(X) = \sum_{i=1}^t \text{Var}(x_i) = t\sigma^2.$$

C'est ici que naissent les termes $t\mu$ et $t\sigma^2$. - La moyenne s'accumule linéairement : en 2 ans, le rendement attendu est le double de celui d'1 an. - La variance s'accumule aussi linéairement : l'incertitude croît avec le temps. - L'écart-type, lui, croît comme \sqrt{t} , ce qui est une signature des processus aléatoires diffusifs (comme le mouvement brownien).

6. Le Théorème Central Limite (TCL) : vers la normalité

Même si la distribution des x_i n'est pas normale, le TCL garantit que, pour t suffisamment grand, la somme X est approximativement normale :

$$X = \ln\left(\frac{S(t)}{S(0)}\right) \xrightarrow[t \text{ grand}]{} \mathcal{N}(t\mu, t\sigma^2).$$

Dans la pratique financière (où t est souvent en centaines de jours), cette approximation est excellente. On l'adopte donc comme modèle, même pour des t modérés.

7. Retour au prix : définition de la loi log-normale

Puisque $X \sim \mathcal{N}(t\mu, t\sigma^2)$, alors :

$$Y = \frac{S(t)}{S(0)} = e^X \sim \text{Log-}\mathcal{N}(t\mu, t\sigma^2).$$

Ainsi, le prix futur est :

$$S(t) = S(0) \cdot e^X, \quad \text{où } X \sim \mathcal{N}(t\mu, t\sigma^2).$$

Conséquences clés :

- $S(t) > 0$ presque sûrement cohérent avec la réalité (un actif ne peut pas avoir un prix négatif).
- La distribution de $S(t)$ est asymétrique à droite : possibilité de fortes hausses, mais baisse limitée à -100% .
- Les paramètres μ et σ sont ceux de la variable *sous-jacente* $X = \ln(S(t)/S(0))$, pas de $S(t)$ elle-même.

8. Illustration numérique avec des paramètres estimés

Supposons qu'à partir de données historiques, on estime : - Rendement annuel moyen continu : $\mu_{\text{annuel}} = 0,08$, - Volatilité annuelle : $\sigma_{\text{annuel}} = 0,20$.

Pour une période de $t = 4$ jours, on convertit en unité annuelle : $t = 4/252 \approx 0,01587$ année.

Alors :

$$\mathbb{E}[X] = t\mu = 0,01587 \times 0,08 \approx 0,00127,$$

$$\text{Var}(X) = t\sigma^2 = 0,01587 \times (0,20)^2 \approx 0,000635,$$

$$\text{Écart-type} = \sqrt{0,000635} \approx 0,0252.$$

Notre observation réelle ($X = 0,04810$) est plus élevée que la moyenne, mais reste à moins de 2 écarts-types $((0,04810 - 0,00127)/0,0252 \approx 1,86)$, donc parfaitement plausible.

9. Probabilité que le prix double en 6 mois

On souhaite maintenant calculer la probabilité que le prix de l'actif double dans les 6 mois, c'est-à-dire :

$$\mathbb{P}(S(0,5) \geq 2S(0)) = \mathbb{P}\left(\frac{S(0,5)}{S(0)} \geq 2\right).$$

Comme $\ln(S(t)/S(0)) \sim \mathcal{N}(t\mu, t\sigma^2)$, on a :

$$\mathbb{P}\left(\ln\left(\frac{S(0,5)}{S(0)}\right) \geq \ln(2)\right) = \mathbb{P}(X \geq \ln(2)), \quad \text{où } X \sim \mathcal{N}(0,5\mu, 0,5\sigma^2).$$

Avec $\mu = 0,08$ et $\sigma = 0,20$ (annuels), on obtient :

$$\mathbb{E}[X] = 0,5 \times 0,08 = 0,04,$$

$$\text{Écart-type}(X) = \sqrt{0,5 \times 0,20} \approx 0,1414,$$

$$\ln(2) \approx 0,6931.$$

On standardise :

$$Z = \frac{\ln(2) - 0,04}{0,1414} \approx \frac{0,6531}{0,1414} \approx 4,62.$$

La probabilité recherchée est donc :

$$\mathbb{P}(Z \geq 4,62) = 1 - \Phi(4,62),$$

où Φ est la fonction de répartition de la loi normale centrée réduite.

À l'aide des tables ou d'un logiciel, $\Phi(4,62) \approx 0,999998$, donc :

$$\mathbb{P}(S(0,5) \geq 2S(0)) \approx 1 - 0,999998 = 0,000002,$$

soit environ $0,0002\%$.

Interprétation : Sous ces hypothèses de rendement et de volatilité, il est extrêmement improbable que l'actif double en 6 mois. Cela illustre la faible probabilité des événements extrêmes dans un modèle log-normal avec volatilité modérée.

10. Conclusion

Cet exemple montre, pas à pas, comment un processus financier réel — la composition multiplicative des rendements — conduit naturellement à modéliser le prix futur comme une variable log-normale. L'astuce du logarithme permet d'utiliser la puissance du Théorème Central Limite, et les paramètres $t\mu$ et $t\sigma^2$ apparaissent inévitablement comme conséquence des propriétés additives de l'espérance et de la variance sous indépendance. C'est sur cette base solide que repose le modèle de Black-Scholes.

6.8 Exercices

Pour tous les exercices de calcul, vous pouvez utiliser les valeurs suivantes pour la fonction de répartition de la loi normale standard $\Phi(z) = P(Z \leq z)$:

- $\Phi(0) = 0.5$
- $\Phi(0.5) \approx 0.6915$
- $\Phi(1) \approx 0.8413$
- $\Phi(1.5) \approx 0.9332$
- $\Phi(1.96) \approx 0.975$
- $\Phi(2) \approx 0.9772$
- $\Phi(2.5) \approx 0.9938$
- $\Phi(3) \approx 0.9987$

Et rappelez-vous la propriété de symétrie : $\Phi(-z) = 1 - \Phi(z)$.

Exercice 1 : Définition (Paramètres)

Soit $Y \sim \text{Log-}\mathcal{N}(3, 16)$.

1. Soit $X = \ln(Y)$. Quelle est la loi de X ?
2. Que valent $E[X]$ et $\text{Var}(X)$?

Exercice 2 : Définition (Inverse)

Soit $X \sim \mathcal{N}(0, 1)$. Si $Y = e^X$, quelle est la notation de la loi de Y ?

Exercice 3 : Intuition (Somme vs Produit)

La taille d'une population de bactéries au temps t est $P(t) = P(0) \times F_1 \times F_2 \times \dots \times F_t$, où les F_i sont des facteurs de croissance aléatoires. Pourquoi est-il plus probable que $P(t)$ suive une loi log-normale plutôt qu'une loi normale ?

Exercice 4 : Intuition (Transformation)

Si $X \sim \mathcal{N}(0, \sigma^2)$, on sait que la distribution de X est symétrique autour de 0. Pourquoi la distribution de $Y = e^X$ n'est-elle pas symétrique ?

Exercice 5 : PDF (Propriétés)

Soit $Y \sim \text{Log-}\mathcal{N}(\mu, \sigma^2)$. Quelle est la probabilité $P(Y \leq 0)$?

Exercice 6 : Espérance

Soit $Y \sim \text{Log-}\mathcal{N}(\mu = 4, \sigma^2 = 2)$. Calculez l'espérance $E[Y]$.

Exercice 7 : Médiane

Soit $Y \sim \text{Log-}\mathcal{N}(\mu = 4, \sigma^2 = 2)$. Calculez la médiane $\text{Med}(Y)$.

Exercice 8 : Mode

Soit $Y \sim \text{Log-}\mathcal{N}(\mu = 4, \sigma^2 = 2)$. Calculez le mode $\text{Mode}(Y)$.

Exercice 9 : Relation d'Ordre

En utilisant les résultats des exercices 6, 7 et 8, vérifiez la relation d'ordre (inégalité) entre le mode, la médiane et l'espérance.

Exercice 10 : Variance

Soit $Y \sim \text{Log-}\mathcal{N}(\mu = 1, \sigma^2 = 1)$. Calculez la variance $\text{Var}(Y)$.

Exercice 11 : Moments (Problème Inverse)

Soit Y une variable log-normale. On sait que sa médiane est e^5 et que son espérance est $e^{5.5}$.

1. Trouvez μ .
2. Trouvez σ^2 .

Exercice 12 : Impact de σ sur l'Espérance

Soient $A \sim \text{Log-}\mathcal{N}(0, 1)$ et $B \sim \text{Log-}\mathcal{N}(0, 2)$. Les deux ont la même médiane ($e^0 = 1$). Laquelle a l'espérance la plus élevée ? Pourquoi ?

Exercice 13 : Calcul de CDF (Simple)

Soit $Y \sim \text{Log-}\mathcal{N}(\mu = 5, \sigma^2 = 1)$. (donc $\sigma = 1$). Calculez $P(Y \leq e^5)$.

Exercice 14 : Calcul de CDF

Soit $Y \sim \text{Log-}\mathcal{N}(\mu = 3, \sigma = 2)$. Calculez $P(Y \leq e^4)$.

Exercice 15 : Calcul de Probabilité (Queue Droite)

Soit $Y \sim \text{Log-}\mathcal{N}(\mu = 0, \sigma = 1)$. Calculez $P(Y > 1)$.

Exercice 16 : Calcul de Probabilité (Queue Gauche)

Soit $Y \sim \text{Log-}\mathcal{N}(\mu = 1, \sigma = 1)$. Calculez $P(Y \leq e^{-1})$.

Exercice 17 : Calcul de Probabilité (Intervalle)

Soit $Y \sim \text{Log-}\mathcal{N}(\mu = 10, \sigma = 3)$. Calculez $P(e \leq Y \leq e^{16})$.

Exercice 18 : Application (Revenu)

Le revenu Y (en milliers) suit $\text{Log-}\mathcal{N}(\mu = 3, \sigma = 1)$. Quelle est la probabilité qu'un individu ait un revenu Y inférieur à e^2 (milliers) ?

Exercice 19 : Problème Inverse (Médiane)

Soit $Y \sim \text{Log-}\mathcal{N}(\mu = 5, \sigma = 2)$. Trouvez la valeur y telle que $P(Y \leq y) = 0.5$.

Exercice 20 : Problème Inverse (Percentile)

Soit $Y \sim \text{Log-}\mathcal{N}(\mu = 0, \sigma = 1)$. Trouvez le 84.13ème percentile de Y .

Exercice 21 : Problème Inverse (Percentile)

Soit $Y \sim \text{Log-}\mathcal{N}(\mu = 2, \sigma = 3)$. Trouvez la valeur y telle que $P(Y \leq y) \approx 0.9772$.

Exercice 22 : Problème Inverse (Quantile bas)

Soit $Y \sim \text{Log-}\mathcal{N}(\mu = 10, \sigma = 2)$. Trouvez la valeur y telle que $P(Y > y) \approx 0.9938$.

Exercice 23 : Modèle Financier (Paramètres)

Le log-rendement journalier x_i d'un actif a $\mu = E[x_i] = 0.01$ et $\sigma^2 = \text{Var}(x_i) = 0.04$. Soit $X = \sum_{i=1}^{16} x_i$ le log-rendement total sur 16 jours. On suppose l'indépendance. Quelle est la loi de X ? (Donnez son nom, $E[X]$ et $\text{Var}(X)$).

Exercice 24 : Modèle Financier (Loi du Prix)

En utilisant l'exercice 23, soit $S(0)$ le prix initial et $S(16)$ le prix après 16 jours. Quelle est la loi du *ratio* de prix $Y = S(16)/S(0)$?

Exercice 25 : Modèle Financier (Calcul de Prob.)

Un actif a $S(0) = 100$. Le log-rendement sur 1 an $X = \ln(S(1)/S(0))$ suit $\mathcal{N}(\mu = 0.05, \sigma^2 = 0.09)$. Quelle est la probabilité que le prix de l'actif après 1 an soit inférieur à 100 ? (c-à-d, $P(S(1) < 100)$).

6.9 Corrections des Exercices

Correction Exercice 1 : Définition (Paramètres)

1. Par définition, si $Y \sim \text{Log-}\mathcal{N}(\mu, \sigma^2)$, alors $X = \ln(Y)$ suit une loi normale $\mathcal{N}(\mu, \sigma^2)$. Ici, $X \sim \mathcal{N}(3, 16)$. 2. $E[X] = \mu = 3$. $\text{Var}(X) = \sigma^2 = 16$.

Correction Exercice 2 : Définition (Inverse)

Si $X \sim \mathcal{N}(0, 1)$, alors $Y = e^X$ suit une loi log-normale avec $\mu = 0$ et $\sigma^2 = 1$. La notation est $Y \sim \text{Log-}\mathcal{N}(0, 1)$.

Correction Exercice 3 : Intuition (Somme vs Produit)

La loi normale résulte de l'addition de nombreux petits effets (TCL). La loi log-normale résulte de la multiplication de nombreux petits facteurs. La croissance d'une population de bactéries est un processus multiplicatif ($P(t) = P(t-1) \times F_t$). En prenant le logarithme, $\ln(P(t)) = \ln(P(0)) + \sum \ln(F_i)$, on obtient une somme qui tend vers une loi normale. Donc $P(t)$ lui-même tend vers une loi log-normale. De plus, $P(t)$ doit être positif, ce que garantit la loi log-normale (support $(0, \infty)$).

Correction Exercice 4 : Intuition (Transformation)

La transformation $Y = e^X$ n'est pas linéaire. Elle est convexe. Elle "étire" la partie droite de la distribution normale ($x > 0 \implies e^x$ croît exponentiellement) et "compresse" la partie gauche ($x < 0 \implies e^x$ s'écroule vers 0). Une distribution symétrique (Normale) transformée par une fonction asymétrique (Exponentielle) donne une distribution asymétrique (Log-Normale).

Correction Exercice 5 : PDF (Propriétés)

Le support de la loi log-normale Y est $(0, \infty)$. La variable ne peut jamais être négative ou nulle. Par conséquent, $P(Y \leq 0) = 0$.

Correction Exercice 6 : Espérance

$Y \sim \text{Log-}\mathcal{N}(\mu = 4, \sigma^2 = 2)$. $E[Y] = e^{\mu + \sigma^2/2} = e^{4+2/2} = e^{4+1} = e^5$.

Correction Exercice 7 : Médiane

$Y \sim \text{Log-}\mathcal{N}(\mu = 4, \sigma^2 = 2)$. $\text{Med}(Y) = e^\mu = e^4$.

Correction Exercice 8 : Mode

$Y \sim \text{Log-}\mathcal{N}(\mu = 4, \sigma^2 = 2)$. $\text{Mode}(Y) = e^{\mu - \sigma^2} = e^{4-2} = e^2$.

Correction Exercice 9 : Relation d'Ordre

Les valeurs sont : Mode $= e^2 \approx 7.39$ Médiane $= e^4 \approx 54.60$ Espérance $= e^5 \approx 148.41$ On vérifie bien Mode $<$ Médiane $<$ Espérance, ce qui est la signature d'une asymétrie à droite.

Correction Exercice 10 : Variance

$Y \sim \text{Log-}\mathcal{N}(\mu = 1, \sigma^2 = 1)$. $\text{Var}(Y) = (e^{\sigma^2} - 1) \cdot e^{2\mu + \sigma^2}$ $\text{Var}(Y) = (e^1 - 1) \cdot e^{2(1)+1} = (e - 1)e^3$.

Correction Exercice 11 : Moments (Problème Inverse)

1. $\text{Med}(Y) = e^\mu$. On nous donne $\text{Med}(Y) = e^5$. Donc $\mu = 5$. 2. $E[Y] = e^{\mu + \sigma^2/2}$. On nous donne $E[Y] = e^{5.5}$. Donc, $e^{5.5} = e^{\mu + \sigma^2/2} = e^{5 + \sigma^2/2}$. En égalant les exposants : $5.5 = 5 + \sigma^2/2$. $0.5 = \sigma^2/2 \implies \sigma^2 = 1$.

Correction Exercice 12 : Impact de σ sur l'Espérance

$E[A] = e^{\mu_A + \sigma_A^2/2} = e^{0+1/2} = e^{0.5}$. $E[B] = e^{\mu_B + \sigma_B^2/2} = e^{0+2/2} = e^1$. $E[B] > E[A]$. L'espérance $E[Y] = e^{\mu + \sigma^2/2}$ croît avec σ^2 . Cela est dû à l'inégalité de Jensen ($\mathbb{E}[e^X] \geq e^{\mathbb{E}[X]}$) : une plus grande variance (incertitude) sur $X = \ln(Y)$ augmente l'espérance de $Y = e^X$ à cause de la convexité de l'exponentielle.

Correction Exercice 13 : Calcul de CDF (Simple)

On cherche $P(Y \leq e^5)$ pour $Y \sim \text{Log-}\mathcal{N}(5, 1)$.

$$P(Y \leq e^5) = P(\ln(Y) \leq \ln(e^5)) = P(X \leq 5)$$

Puisque $X = \ln(Y) \sim \mathcal{N}(\mu = 5, \sigma^2 = 1)$, on cherche la probabilité que X soit inférieure à sa propre moyenne. Par symétrie de la loi normale, c'est 0.5. Formellement : $Z = \frac{5-5}{1} = 0$. $P(Z \leq 0) = \Phi(0) = 0.5$.

Correction Exercice 14 : Calcul de CDF

$Y \sim \text{Log-}\mathcal{N}(\mu = 3, \sigma = 2)$. $P(Y \leq e^4) = P(\ln(Y) \leq \ln(e^4)) = P(X \leq 4)$, où $X \sim \mathcal{N}(3, 4)$. Standardisation : $Z = \frac{X-\mu}{\sigma} = \frac{4-3}{2} = \frac{1}{2} = 0.5$. $P(X \leq 4) = P(Z \leq 0.5) = \Phi(0.5) \approx 0.6915$.

Correction Exercice 15 : Calcul de Probabilité (Queue Droite)

$Y \sim \text{Log-}\mathcal{N}(0, 1)$. $P(Y > 1) = 1 - P(Y \leq 1) = 1 - P(\ln(Y) \leq \ln(1)) = 1 - P(X \leq 0)$. $X \sim \mathcal{N}(0, 1)$. $P(X \leq 0) = \Phi(0) = 0.5$. $P(Y > 1) = 1 - 0.5 = 0.5$. (Logique : $e^0 = 1$ est la médiane, donc 50)

Correction Exercice 16 : Calcul de Probabilité (Queue Gauche)

$Y \sim \text{Log-}\mathcal{N}(1, 1)$. $P(Y \leq e^{-1}) = P(\ln(Y) \leq \ln(e^{-1})) = P(X \leq -1)$, où $X \sim \mathcal{N}(1, 1)$. Standardisation : $Z = \frac{X-\mu}{\sigma} = \frac{-1-1}{1} = -2$. $P(X \leq -1) = P(Z \leq -2) = \Phi(-2) = 1 - \Phi(2) \approx 1 - 0.9772 = 0.0228$.

Correction Exercice 17 : Calcul de Probabilité (Intervalle)

$Y \sim \text{Log-}\mathcal{N}(10, 9)$. ($X \sim \mathcal{N}(10, 9)$, $\sigma = 3$). $P(e \leq Y \leq e^{16}) = P(\ln(e) \leq \ln(Y) \leq \ln(e^{16})) = P(1 \leq X \leq 16)$. Standardisation : $z_1 = \frac{1-10}{3} = -3$. $z_2 = \frac{16-10}{3} = 2$. $P(1 \leq X \leq 16) = P(-3 \leq Z \leq 2) = \Phi(2) - \Phi(-3)$. $\Phi(-3) = 1 - \Phi(3) \approx 1 - 0.9987 = 0.0013$. $P = \Phi(2) - \Phi(-3) \approx 0.9772 - 0.0013 = 0.9759$.

Correction Exercice 18 : Application (Revenu)

$Y \sim \text{Log-}\mathcal{N}(3, 1)$. ($X \sim \mathcal{N}(3, 1)$, $\sigma = 1$). $P(Y < e^2) = P(\ln(Y) < \ln(e^2)) = P(X < 2)$. Standardisation : $Z = \frac{2-3}{1} = -1$. $P(X < 2) = P(Z < -1) = \Phi(-1) = 1 - \Phi(1) \approx 1 - 0.8413 = 0.1587$.

Correction Exercice 19 : Problème Inverse (Médiane)

On cherche y tel que $P(Y \leq y) = 0.5$. C'est la définition de la médiane. $\text{Med}(Y) = e^\mu = e^5$. Donc $y = e^5$.

Correction Exercice 20 : Problème Inverse (Percentile)

$Y \sim \text{Log-}\mathcal{N}(0, 1)$. On cherche y tel que $P(Y \leq y) = 0.8413$. $P(\ln(Y) \leq \ln(y)) = 0.8413$. $P(X \leq \ln(y)) = 0.8413$, où $X \sim \mathcal{N}(0, 1)$. $P(Z \leq \ln(y)) = 0.8413$. On cherche $z = \ln(y)$ tel que $\Phi(z) = 0.8413$. D'après la table, $z = 1$. $\ln(y) = 1 \implies y = e^1 = e$.

Correction Exercice 21 : Problème Inverse (Percentile)

$Y \sim \text{Log-}\mathcal{N}(2, 9)$. ($X \sim \mathcal{N}(2, 9)$, $\sigma = 3$). On cherche y tel que $P(Y \leq y) \approx 0.9772$. $P(\ln(Y) \leq \ln(y)) = P(X \leq \ln(y)) = 0.9772$. Standardisation : $P\left(Z \leq \frac{\ln(y)-2}{3}\right) = 0.9772$. On cherche z tel que $\Phi(z) = 0.9772$. D'après la table, $z = 2$. $\frac{\ln(y)-2}{3} = 2 \implies \ln(y) - 2 = 6 \implies \ln(y) = 8 \implies y = e^8$.

Correction Exercice 22 : Problème Inverse (Quantile bas)

$Y \sim \text{Log-}\mathcal{N}(10, 4)$. ($X \sim \mathcal{N}(10, 4)$, $\sigma = 2$). On cherche y tel que $P(Y > y) \approx 0.9938$. $P(Y \leq y) = 1 - 0.9938 = 0.0062$. $P(X \leq \ln(y)) = 0.0062$. Standardisation : $P\left(Z \leq \frac{\ln(y)-10}{2}\right) = 0.0062$. On cherche z tel que $\Phi(z) = 0.0062$. La table ne donne pas cette valeur, mais $\Phi(2.5) \approx 0.9938$. Par symétrie, $\Phi(-2.5) = 1 - \Phi(2.5) \approx 1 - 0.9938 = 0.0062$. Donc, $z = -2.5$. $\frac{\ln(y)-10}{2} = -2.5 \implies \ln(y) - 10 = -5 \implies \ln(y) = 5 \implies y = e^5$.

Correction Exercice 23 : Modèle Financier (Paramètres)

X est une somme de 16 v.a. i.i.d. $x_i \sim \mathcal{N}(0.01, 0.04)$. X suit une loi normale. $E[X] = \sum E[x_i] = t \cdot \mu = 16 \times 0.01 = 0.16$. $\text{Var}(X) = \sum \text{Var}(x_i) = t \cdot \sigma^2 = 16 \times 0.04 = 0.64$. Donc, $X \sim \mathcal{N}(0.16, 0.64)$.

Correction Exercice 24 : Modèle Financier (Loi du Prix)

$Y = S(16)/S(0)$. $X = \ln(Y) = \ln(S(16)/S(0))$. D'après l'exercice 23, $X \sim \mathcal{N}(0.16, 0.64)$. Par définition, $Y = e^X$ suit une loi log-normale. $Y \sim \text{Log-}\mathcal{N}(\mu_X = 0.16, \sigma_X^2 = 0.64)$.

Correction Exercice 25 : Modèle Financier (Calcul de Prob.)

$X = \ln(S(1)/S(0)) \sim \mathcal{N}(0.05, 0.09)$. (donc $\mu = 0.05, \sigma = 0.3$). On cherche $P(S(1) < 100)$. $P(S(1) < 100) = P(S(1)/S(0) < 100/100) = P(S(1)/S(0) < 1)$. On applique le log : $P(\ln(S(1)/S(0)) < \ln(1)) = P(X < 0)$. On standardise X : $Z = \frac{X-\mu}{\sigma} = \frac{0-0.05}{0.3} = -0.05/0.3 \approx -0.167$. $P(X < 0) = P(Z < -0.167)$. C'est $\Phi(-0.167) = 1 - \Phi(0.167)$. Puisque $\Phi(0) = 0.5$, $\Phi(0.167)$ est légèrement supérieur à 0.5, donc $\Phi(-0.167)$ est légèrement inférieur à 0.5. La probabilité est légèrement inférieure à 50%.

6.10 Exercices Python

La loi log-normale est fondamentale en finance. Elle repose sur l'idée que si les **log-rendements** d'une action $X_i = \ln(P_i/P_{i-1})$ sont (approximativement) normaux, alors le prix futur P_t , qui est un **produit** de ces rendements ($P_t = P_0 \times e^{X_1} \times \dots \times e^{X_t}$), suivra une loi log-normale.

Nous allons estimer les paramètres μ et σ^2 de la loi normale sous-jacente à partir des log-rendements journaliers de Microsoft (MSFT) et Google (GOOG), puis utiliser la théorie log-normale pour modéliser les prix.

```
!pip install yfinance
import yfinance as yf
import pandas as pd
import numpy as np
from scipy.stats import norm # Moteur pour les calculs de CDF/PDF

# Définir les tickers et la période
tickers = ["MSFT", "GOOG"]
start_date = "2020-01-01"
end_date = "2024-12-31"

# Télécharger les prix de clôture ajustés
data = yf.download(tickers, start=start_date, end=end_date)["Adj Close"]

# Calculer les LOG-RENDEMENTS journaliers
log_returns = np.log(data / data.shift(1)).dropna()

# Renommer les colonnes
log_returns.columns = ["MSFT_LogReturn", "GOOG_LogReturn"]

# 'log_returns' est notre DataFrame.
# X_msft = log_returns["MSFT_LogReturn"]
# X_goog = log_returns["GOOG_LogReturn"]
```

Exercice 1 : Estimer les Paramètres μ et σ^2

Soit P_t le prix de MSFT. Le modèle suppose que $X = \ln(P_t/P_{t-1}) \sim \mathcal{N}(\mu, \sigma^2)$. Les paramètres μ et σ^2 sont les paramètres "log-normaux".

Votre tâche :

1. Estimer μ (l'espérance du log-rendement journalier) pour MSFT.
2. Estimer σ^2 (la variance du log-rendement journalier) pour MSFT.
3. Estimer σ (l'écart-type du log-rendement journalier) pour MSFT.

Exercice 2 : Test de Normalité (Règle 68-95-99.7)

La théorie log-normale repose sur la normalité des log-rendements X . Vérifions-le.

Votre tâche :

1. Utiliser μ et σ (pour MSFT) de l'Exercice 1.
2. Calculer la proportion **empirique** des log-rendements de MSFT qui tombent dans l'intervalle $[\mu - \sigma, \mu + \sigma]$.
3. Comparer ce pourcentage à la valeur théorique (68.27%). Le modèle semble-t-il bien s'ajuster ?

Exercice 3 : Asymétrie (Prix vs Log-Rendements)

La théorie dit que les log-rendements X sont symétriques (Normaux), mais que les prix P_t sont asymétriques à droite (Log-Normaux).

Votre tâche :

1. Calculer la moyenne et la médiane de la série des **log-rendements** de MSFT.
2. Calculer la moyenne et la médiane de la série des **prix** de MSFT (la colonne `data['MSFT']`).
3. Comparer les deux paires. Les log-rendements sont-ils symétriques (moyenne \approx médiane) ? Les prix sont-ils asymétriques (moyenne $>$ médiane) ?

Exercice 4 : Espérance vs Médiane (Théorique)

Soit $Y = P_t/P_{t-1} = e^X$ la variable "ratio de prix journalier". $Y \sim \text{Log-}\mathcal{N}(\mu, \sigma^2)$. Théorie : $\text{Med}(Y) = e^\mu$ et $E[Y] = e^{\mu + \sigma^2/2}$.

Votre tâche :

1. Utiliser μ et σ^2 (pour MSFT) de l'Exercice 1.
2. Calculer la médiane **théorique** $\text{Med}(Y)$.
3. Calculer l'espérance **théorique** $E[Y]$.
4. Vérifier que $E[Y] > \text{Med}(Y)$, confirmant l'asymétrie.

Exercice 5 : Espérance Théorique vs Empirique

Vérifions le calcul de $E[Y]$ de l'exercice 4 de manière empirique.

Votre tâche :

1. Créer la série Y (ratio de prix journalier) : $Y = \exp(X_{\text{msft}})$.
2. Calculer l'espérance **empirique** de Y (la moyenne de cette série Y).
3. Comparer cette valeur empirique à l'espérance **théorique** $e^{\mu + \sigma^2/2}$ calculée à l'exercice 4.

Exercice 6 : Variance Théorique vs Empirique

Théorie : $\text{Var}(Y) = (e^{\sigma^2} - 1) \cdot e^{2\mu + \sigma^2}$.

Votre tâche :

1. Utiliser μ et σ^2 (pour MSFT) de l'Exercice 1.
2. Calculer la variance **théorique** $\text{Var}(Y)$ en utilisant la formule ci-dessus.
3. Calculer la variance **empirique** de la série Y (créée à l'Ex 5).
4. Comparer les deux résultats.

Exercice 7 : Modélisation du Prix Futur (Paramètres)

Modélisons le prix de GOOG dans $t = 20$ jours ouvrés (environ 1 mois). Le prix P_{20} est log-normal si l'on suppose $P_{20} = P_0 \cdot e^{X_{20}}$, où P_0 est le prix actuel. Le log-rendement total $X_{20} = \ln(P_{20}/P_0)$ suit $X_{20} \sim \mathcal{N}(t\mu, t\sigma^2)$.

Votre tâche :

1. Estimer μ_G et σ_G^2 (journaliers) pour GOOG (similaire à l'Ex 1).
2. Définir $t = 20$.
3. Calculer $\mu_{20} = t\mu_G$ (l'espérance du log-rendement sur 20 jours).
4. Calculer $\sigma_{20}^2 = t\sigma_G^2$ (la variance du log-rendement sur 20 jours).

Exercice 8 : Calcul de Probabilité (Prix Futur)

En utilisant les paramètres μ_{20} et $\sigma_{20} = \sqrt{\sigma_{20}^2}$ de l'exercice 7 pour GOOG :

Votre tâche :

1. Calculer la probabilité que GOOG ait un rendement positif sur 20 jours.
2. On cherche $P(P_{20} > P_0) \implies P(P_{20}/P_0 > 1) \implies P(\ln(P_{20}/P_0) > \ln(1))$.
3. Calculer $P(X_{20} > 0)$.
4. (Indice : Standardiser 0 avec μ_{20} et σ_{20} , puis utiliser $1 - \Phi(z)$).

Exercice 9 : Calcul de Probabilité (Perte > 5%)

En utilisant les paramètres μ_{20} et σ_{20} de l'exercice 7 pour GOOG :

Votre tâche :

1. Calculer la probabilité que GOOG perde plus de 5% sur 20 jours.
2. On cherche $P(P_{20} < 0.95 \times P_0) \implies P(P_{20}/P_0 < 0.95)$.
3. Calculer $P(X_{20} < \ln(0.95))$.
4. (Indice : Standardiser $\ln(0.95)$ avec μ_{20} et σ_{20} , puis utiliser $\Phi(z)$).

Exercice 10 : Problème Inverse (Intervalle de Confiance)

Trouvons l'intervalle de 95% pour le prix de GOOG dans 20 jours. Nous cherchons les bornes y_1, y_2 telles que $P(y_1 \leq P_{20} \leq y_2) = 0.95$. On suppose un intervalle centré sur la loi normale sous-jacente (entre $z = -1.96$ et $z = +1.96$).

Votre tâche :

1. Trouver $z_{inf} = -1.96$ et $z_{sup} = +1.96$.
2. "Dé-standardiser" ces Z-scores pour trouver les log-rendements x_1 et x_2 : $x = \mu_t + z\sigma_t$ (en utilisant μ_{20} et σ_{20} de l'Ex 7).
3. Convertir ces log-rendements en ratios de prix $y = e^x$.
4. (Conclusion) L'intervalle de 95% pour le ratio de prix est $[y_1, y_2]$.

Exercice 11 : Calcul de la Médiane vs Espérance (Prix Futur)

Pour le prix de GOOG dans 20 jours, $P_{20} = P_0 \cdot Y_{20}$, où $Y_{20} \sim \text{Log-}\mathcal{N}(\mu_{20}, \sigma_{20}^2)$.

Votre tâche :

1. Calculer le ratio de prix **médian** attendu : $\text{Med}(Y_{20}) = e^{\mu_{20}}$.
2. Calculer le ratio de prix **moyen** (espérance) attendu : $E[Y_{20}] = e^{\mu_{20} + \sigma_{20}^2/2}$.
3. (Conclusion) Lequel est le plus élevé ? Pourquoi est-ce important pour un investisseur ?

Exercice 12 : Mode (Prix Futur)

Théorie : Le mode (la valeur la plus probable) du ratio de prix Y_{20} est $\text{Mode}(Y_{20}) = e^{\mu_{20} - \sigma_{20}^2}$.

Votre tâche :

1. Calculer le ratio de prix **modal** (le plus probable) pour GOOG dans 20 jours.
2. Comparer le Mode (Ex 12), la Médiane (Ex 11) et l'Espérance (Ex 11).
3. Vérifier que $\text{Mode} < \text{Médiane} < \text{Espérance}$, confirmant l'asymétrie à droite.

7 Distributions Multivariées et Concepts Associés

7.1 Distributions Jointes et Marginales

Jusqu'à présent, nous avons étudié les variables aléatoires isolément. Nous allons maintenant examiner comment analyser les relations entre *plusieurs* variables aléatoires.

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 :

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

Cette loi jointe est la "carte" complète de toutes les issues possibles.

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.

Cette "carte" complète contient toutes les informations. Si nous ne nous intéressons qu'à une seule variable, nous pouvons la "réduire" en calculant sa distribution marginale.

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 :

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

Visuellement, cela correspond à "écraser" le tableau de probabilités sur un seul de ses axes.

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 .

L'exemple le plus simple est le lancer de deux dés.

Exemple : Lois 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.

7.2 Espérance d'une fonction de deux variables

Maintenant que nous avons la loi jointe (la carte des probabilités), nous pouvons l'utiliser pour calculer l'espérance de n'importe quelle fonction qui dépend des deux variables, $g(X, Y)$.

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) :

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

C'est la moyenne de g , pondérée par les probabilités jointes.

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.

Le cas le plus important de $g(X, Y)$ est la somme $X + Y$.

Exemple

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

$$\begin{aligned} E[X + Y] &= \sum_{x=1}^6 \sum_{y=1}^6 (x + y) P(X = x, Y = y) \\ E[X + Y] &= \sum_{x=1}^6 \sum_{y=1}^6 (x + y) \frac{1}{36} \end{aligned}$$

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

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

7.3 Covariance et Corrélation

La linéarité $E[X + Y] = E[X] + E[Y]$ est un outil puissant. Mais l'espérance ne nous dit rien sur la *relation* entre X et Y . Pour cela, nous introduisons la covariance.

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.

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

Elle mesure la direction de leur relation.

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.

La définition $E[(X - \mu_X)(Y - \mu_Y)]$ est bonne pour l'intuition, mais difficile à calculer. Une formule alternative est presque toujours utilisée.

Théorème : Formule de calcul de la covariance

Une formule computationnelle plus simple pour la covariance est :

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

La preuve est une simple expansion algébrique.

Preuve

Soit $\mu_X = E[X]$ et $\mu_Y = E[Y]$. On part de la définition :

$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \quad (\text{On développe}) \\ &= E[XY] - E[X\mu_Y] - E[Y\mu_X] + E[\mu_X\mu_Y] \quad (\text{Par linéarité}) \\ &= E[XY] - \mu_Y E[X] - \mu_X E[Y] + \mu_X\mu_Y \quad (\text{Les moyennes sont des constantes}) \\ &= E[XY] - \mu_Y\mu_X - \mu_X\mu_Y + \mu_X\mu_Y \\ &= E[XY] - \mu_X\mu_Y \\ &= E[XY] - E[X]E[Y]\end{aligned}$$

Voyons cette formule en action.

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.

La covariance est un bon indicateur de la direction de la relation, mais sa magnitude est difficile à interpréter. Pour cela, nous la normalisons.

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.

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

La corrélation résout le problème des unités.

Intuition

Le problème de la covariance est qu'elle dépend des unités de X et Y (par ex., $\text{kg} \cdot \text{cm}$). 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.

Cette normalisation se comprend mieux en voyant la corrélation comme une covariance de variables standardisées.

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.

Reprenons notre exemple de dépendance parfaite :

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 .

7.4 Linéarité de la Covariance

Tout comme l'espérance, la covariance possède d'importantes propriétés de linéarité qui simplifient les calculs.

Définition : Linéarité de la Covariance

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

$$\begin{aligned}\text{Cov}(aX + bY + c, Z) &= a\text{Cov}(X, Z) + b\text{Cov}(Y, Z) \\ \text{Cov}(X, aY + bZ + c) &= a\text{Cov}(X, Y) + b\text{Cov}(X, Z)\end{aligned}$$

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

7.5 Résultats sur la Corrélation

La propriété la plus importante de la corrélation, qui découle de sa normalisation, est qu'elle est bornée.

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.

La preuve de ces bornes repose sur le fait que la variance est toujours positive.

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 :

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

7.6 Standardisation et Non-Corrélation

Le processus de 'standardisation' utilisé dans la preuve de la corrélation et dans l'intuition est un concept fondamental en soi.

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 $\sigma_Z = \sqrt{\text{Var}(Z)} = \sqrt{1} = 1$.

Cette transformation permet de comparer des variables sur des échelles différentes.

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.

Maintenant, formalisons le concept d'une covariance nulle.

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 $E[XY] = E[X]E[Y]$.

Il est crucial de ne pas confondre "non corrélées" et "indépendantes".

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).

7.7 Variance d'une Somme de Variables Aléatoires

Nous pouvons maintenant combiner nos connaissances de la variance et de la covariance pour répondre à une question cruciale : quelle est la variance d'une somme de variables, $X + Y$?

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

Pour deux variables aléatoires X et Y :

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

La preuve découle de la définition de la variance et de la linéarité de l'espérance.

Preuve

Soit $\mu_X = E[X]$ et $\mu_Y = E[Y]$.

$$\begin{aligned}\text{Var}(X + Y) &= E[(X + Y - E[X + Y])^2] \\ &= E[(X + Y - (\mu_X + \mu_Y))^2] \quad (\text{Par linéarité de } E) \\ &= E[(X - \mu_X + Y - \mu_Y)^2] \quad (\text{On regroupe les termes})\end{aligned}$$

Posons $A = (X - \mu_X)$ et $B = (Y - \mu_Y)$.

$$\begin{aligned}&= E[(A + B)^2] = E[A^2 + 2AB + B^2] \\ &= E[A^2] + 2E[AB] + E[B^2] \quad (\text{Par linéarité de } E)\end{aligned}$$

Or, par définition :

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

$$E[B^2] = E[(Y - \mu_Y)^2] = \text{Var}(Y)$$

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

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

Cette formule est fondamentale en finance et en ingénierie.

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.

Cela mène à un corollaire très important lorsque la covariance est nulle.

Théorème : Cas Particulier : Variables Non Corrélées

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

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

Preuve

Cela découle directement du théorème précédent. Si X et Y sont non corrélées, alors $\text{Cov}(X, Y) = 0$. Le terme $2\text{Cov}(X, Y)$ dans la formule générale $\text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$ devient nul, laissant :

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

C'est le cas pour nos dés indépendants.

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 $E[S]$.

On peut généraliser cette formule à N variables.

Théorème : Variance d'une somme de N variables

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

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

Preuve

On utilise la propriété $\text{Var}(S) = \text{Cov}(S, S)$ et la bilinéarité de la covariance. Soit $S = \sum_{i=1}^n X_i$.

$$\begin{aligned} \text{Var}(S) &= \text{Cov}(S, S) = \text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) \quad (\text{Par bilinéarité}) \end{aligned}$$

On peut séparer cette double somme en deux parties : le cas où $i = j$ et le cas où $i \neq j$.

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

Puisque $\text{Cov}(X_i, X_i) = E[(X_i - \mu_i)(X_i - \mu_i)] = E[(X_i - \mu_i)^2] = \text{Var}(X_i)$, on obtient :

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

Cette formule est au cœur de la théorie moderne du portefeuille.

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.

7.8 Théorème sur la somme de lois de Poisson

Terminons avec un théorème très utile qui combine les idées d'indépendance et de somme de variables aléatoires pour une distribution spécifique.

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 :

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

La preuve pour $k = 2$ (qui se généralise par récurrence) utilise l'indépendance et la formule du binôme de Newton.

Preuve : Preuve pour la somme de deux v.a.

Soit $X \sim \text{Poisson}(\lambda_1)$ et $Y \sim \text{Poisson}(\lambda_2)$, indépendantes. Soit $S = X + Y$. Nous cherchons $P(S = k)$. Pour que $S = k$, il faut que $X = j$ et $Y = k - j$, pour toutes les valeurs possibles de j (de 0 à k).

$$P(S = k) = \sum_{j=0}^k P(X = j, Y = k - j)$$

Par indépendance, $P(X = j, Y = k - j) = P(X = j)P(Y = k - j)$.

$$\begin{aligned} P(S = k) &= \sum_{j=0}^k \left(\frac{e^{-\lambda_1} \lambda_1^j}{j!} \right) \left(\frac{e^{-\lambda_2} \lambda_2^{k-j}}{(k-j)!} \right) \\ &= e^{-(\lambda_1 + \lambda_2)} \sum_{j=0}^k \frac{\lambda_1^j \lambda_2^{k-j}}{j!(k-j)!} \end{aligned}$$

On multiplie et on divise par $k!$ pour faire apparaître le coefficient binomial :

$$\begin{aligned} P(S = k) &= \frac{e^{-(\lambda_1 + \lambda_2)}}{k!} \sum_{j=0}^k \frac{k!}{j!(k-j)!} \lambda_1^j \lambda_2^{k-j} \\ &= \frac{e^{-(\lambda_1 + \lambda_2)}}{k!} \sum_{j=0}^k \binom{k}{j} \lambda_1^j \lambda_2^{k-j} \end{aligned}$$

La somme est l'expansion du binôme de Newton pour $(\lambda_1 + \lambda_2)^k$.

$$P(S = k) = \frac{e^{-(\lambda_1 + \lambda_2)} (\lambda_1 + \lambda_2)^k}{k!}$$

C'est la PMF d'une loi Poisson($\lambda_1 + \lambda_2$).

Ce résultat est très intuitif :

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 $\lambda_1 = 5$ /heure et les appels à $\lambda_2 = 10$ /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 :

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

7.9 Exercices

Exercice 1 : Loi Jointe et Marginales

Soit le tableau suivant représentant la loi de probabilité jointe $P(X = x, Y = y)$ d'un couple de variables aléatoires (X, Y) .

X \ Y	0	1	2
0	0.1	0.2	0.1
1	0.3	0.1	0.2

1. Vérifiez qu'il s'agit bien d'une loi de probabilité.

2. Calculez la loi marginale de X , $P(X = x)$.
3. Calculez la loi marginale de Y , $P(Y = y)$.

Exercice 2 : Calcul de Probabilité Jointe

En utilisant la loi jointe de l'exercice 1 :

1. Calculez $P(X = 0, Y \leq 1)$.
2. Calculez $P(X = Y)$.
3. Calculez $P(X > Y)$.

Exercice 3 : Indépendance (Loi Jointe)

En utilisant la loi jointe de l'exercice 1 :

1. Calculez $P(X = 0) \times P(Y = 0)$.
2. Comparez ce résultat à $P(X = 0, Y = 0)$.
3. Les variables X et Y sont-elles indépendantes ? Justifiez.

Exercice 4 : Espérances Marginales

En utilisant les lois marginales calculées à l'exercice 1 :

1. Calculez l'espérance $E[X]$.
2. Calculez l'espérance $E[Y]$.

Exercice 5 : Espérance d'une Fonction (LOTUS)

En utilisant la loi jointe de l'exercice 1, calculez $E[XY]$. (Indice : $E[XY] = \sum_x \sum_y (xy)P(X = x, Y = y)$).

Exercice 6 : Covariance (Calcul)

En utilisant les résultats des exercices 4 et 5, calculez la covariance $\text{Cov}(X, Y)$.

Exercice 7 : Variances Marginales

En utilisant les lois marginales de l'exercice 1 et les espérances de l'exercice 4 :

1. Calculez $E[X^2]$ et $\text{Var}(X)$.
2. Calculez $E[Y^2]$ et $\text{Var}(Y)$.

Exercice 8 : Corrélation (Calcul)

En utilisant les résultats des exercices 6 et 7, calculez le coefficient de corrélation $\text{Corr}(X, Y)$.

Exercice 9 : Variance d'une Somme (Non Indépendant)

Soient X et Y deux variables aléatoires telles que $\text{Var}(X) = 10$, $\text{Var}(Y) = 5$ et $\text{Cov}(X, Y) = 2$. Calculez $\text{Var}(X + Y)$.

Exercice 10 : Variance d'une Différence (Indépendant)

Soient X et Y deux variables aléatoires **indépendantes** telles que $\text{Var}(X) = 16$ et $\text{Var}(Y) = 9$.

1. Que vaut $\text{Cov}(X, Y)$?
2. Calculez $\text{Var}(X - Y)$. (Rappel : $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y)$).

Exercice 11 : Bilinéarité de la Covariance

Soient X, Y, Z trois variables aléatoires. Exprimez $\text{Cov}(X + Y, Z)$ en fonction des covariances des variables individuelles.

Exercice 12 : Variance d'une Combinaison Linéaire

Soient X et Y deux variables aléatoires indépendantes avec $\text{Var}(X) = 4$ et $\text{Var}(Y) = 2$. Calculez $\text{Var}(3X - 5Y + 1)$.

Exercice 13 : Variance d'une Somme (Dés)

On lance deux dés équilibrés D_1 et D_2 . Soit $S = D_1 + D_2$. On rappelle que pour un dé, $\text{Var}(D_i) = 35/12$.

1. Les variables D_1 et D_2 sont-elles indépendantes ?
2. Calculez $\text{Var}(S)$.

Exercice 14 : Covariance et Variance

Soit X une variable aléatoire. En utilisant la bilinéarité de la covariance, montrez que $\text{Cov}(X, X) = \text{Var}(X)$.

Exercice 15 : Covariance avec une Constante

Soit X une variable aléatoire et c une constante. Montrez que $\text{Cov}(X, c) = 0$. (Indice : $E[c] = c$ et $E[Xc] = cE[X]$).

Exercice 16 : Standardisation (Centrer-Réduire)

Soit X une variable aléatoire avec $E[X] = 10$ et $\text{Var}(X) = 4$. Soit $Z = \frac{X - E[X]}{\sqrt{\text{Var}(X)}} = \frac{X - 10}{2}$ la variable standardisée.

1. Calculez $E[Z]$.
2. Calculez $\text{Var}(Z)$.

Exercice 17 : Corrélation et Standardisation

Soit $\text{Corr}(X, Y) = 0.5$. Soient Z_X et Z_Y les versions standardisées de X et Y . Que vaut $\text{Cov}(Z_X, Z_Y)$? (Indice : regardez l'intuition de la corrélation).

Exercice 18 : Somme de Lois de Poisson

Un magasin reçoit des clients au comptoir A selon $X \sim \text{Poisson}(\lambda_1 = 5 \text{ clients/heure})$ et au comptoir B selon $Y \sim \text{Poisson}(\lambda_2 = 3 \text{ clients/heure})$. On suppose que X et Y sont indépendantes. Soit $S = X + Y$ le nombre total de clients arrivant au magasin en une heure.

1. Quelle est la loi de S ? Donnez son nom et son paramètre.
2. Quelle est la probabilité qu'exactly 6 clients au total arrivent en une heure, $P(S = 6)$?

Exercice 19 : Corrélation Nulle mais Dépendance

Soit X une variable aléatoire $X \in \{-1, 0, 1\}$, avec $P(X = -1) = 1/3$, $P(X = 0) = 1/3$, $P(X = 1) = 1/3$. Soit $Y = X^2$.

1. Calculez $E[X]$.
2. Calculez $E[XY]$. (Indice : $E[XY] = E[X^3]$).
3. Calculez $\text{Cov}(X, Y)$.
4. Les variables X et Y sont-elles indépendantes ?

Exercice 20 : Bornes de la Corrélation

Soit X une variable aléatoire et $Y = -3X + 5$. Sans faire de calcul, que vaut $\text{Corr}(X, Y)$? Justifiez.

7.10 Corrections des Exercices

Correction Exercice 1 : Loi Jointe et Marginales

1. On somme toutes les probabilités du tableau : $0.1 + 0.2 + 0.1 + 0.3 + 0.1 + 0.2 = 1.0$. Puisque la somme est 1 et toutes les probabilités sont non négatives, c'est une loi valide.
2. Loi marginale de X (somme des lignes) : $P(X = 0) = P(X = 0, Y = 0) + P(X = 0, Y = 1) + P(X = 0, Y = 2) = 0.1 + 0.2 + 0.1 = 0.4$. $P(X = 1) = P(X = 1, Y = 0) + P(X = 1, Y = 1) + P(X = 1, Y = 2) = 0.3 + 0.1 + 0.2 = 0.6$.
3. Loi marginale de Y (somme des colonnes) : $P(Y = 0) = P(X = 0, Y = 0) + P(X = 1, Y = 0) = 0.1 + 0.3 = 0.4$. $P(Y = 1) = P(X = 0, Y = 1) + P(X = 1, Y = 1) = 0.2 + 0.1 = 0.3$. $P(Y = 2) = P(X = 0, Y = 2) + P(X = 1, Y = 2) = 0.1 + 0.2 = 0.3$.

Correction Exercice 2 : Calcul de Probabilité Jointe

1. $P(X = 0, Y \leq 1) = P(X = 0, Y = 0) + P(X = 0, Y = 1) = 0.1 + 0.2 = 0.3$. 2. $P(X = Y) = P(X = 0, Y = 0) + P(X = 1, Y = 1) = 0.1 + 0.1 = 0.2$. 3. $P(X > Y) = P(X = 1, Y = 0) = 0.3$. (C'est la seule case où $x > y$).

Correction Exercice 3 : Indépendance (Loi Jointe)

On utilise les lois marginales de l'exercice 1 : $P(X = 0) = 0.4$ et $P(Y = 0) = 0.4$. 1. $P(X = 0) \times P(Y = 0) = 0.4 \times 0.4 = 0.16$. 2. Dans le tableau joint, $P(X = 0, Y = 0) = 0.1$. 3. Puisque $P(X = 0, Y = 0) \neq P(X = 0) \times P(Y = 0)$ (car $0.1 \neq 0.16$), les variables X et Y **ne sont pas indépendantes**. (Un seul contre-exemple suffit).

Correction Exercice 4 : Espérances Marginales

1. $E[X] = \sum_x xP(X = x) = (0)(P(X = 0)) + (1)(P(X = 1))$ $E[X] = (0)(0.4) + (1)(0.6) = 0.6$.
2. $E[Y] = \sum_y yP(Y = y) = (0)(P(Y = 0)) + (1)(P(Y = 1)) + (2)(P(Y = 2))$ $E[Y] = (0)(0.4) + (1)(0.3) + (2)(0.3) = 0 + 0.3 + 0.6 = 0.9$.

Correction Exercice 5 : Espérance d'une Fonction (LOTUS)

On somme $(xy)P(X = x, Y = y)$ sur les 6 cases. Les termes où $x = 0$ ou $y = 0$ sont nuls.
 $E[XY] = (0 \cdot 0)(0.1) + (0 \cdot 1)(0.2) + (0 \cdot 2)(0.1) + (1 \cdot 0)(0.3) + (1 \cdot 1)(0.1) + (1 \cdot 2)(0.2)$ $E[XY] = 0 + 0 + 0 + 0 + (1)(0.1) + (2)(0.2) = 0.1 + 0.4 = 0.5$.

Correction Exercice 6 : Covariance (Calcul)

On utilise la formule $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$.

$$\text{Cov}(X, Y) = 0.5 - (0.6)(0.9) = 0.5 - 0.54 = -0.04$$

Correction Exercice 7 : Variances Marginales

1. Pour X : $E[X^2] = (0^2)(0.4) + (1^2)(0.6) = 0.6$. $\text{Var}(X) = E[X^2] - (E[X])^2 = 0.6 - (0.6)^2 = 0.6 - 0.36 = 0.24$. 2. Pour Y : $E[Y^2] = (0^2)(0.4) + (1^2)(0.3) + (2^2)(0.3) = 0 + 0.3 + (4)(0.3) = 0.3 + 1.2 = 1.5$. $\text{Var}(Y) = E[Y^2] - (E[Y])^2 = 1.5 - (0.9)^2 = 1.5 - 0.81 = 0.69$.

Correction Exercice 8 : Corrélation (Calcul)

On utilise la formule $\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$.

$$\text{Corr}(X, Y) = \frac{-0.04}{\sqrt{0.24 \times 0.69}} = \frac{-0.04}{\sqrt{0.1656}} \approx \frac{-0.04}{0.4069} \approx -0.098$$

La corrélation est très faible et négative.

Correction Exercice 9 : Variance d'une Somme (Non Indépendant)

On utilise la formule générale :

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

$$\text{Var}(X + Y) = 10 + 5 + 2(2) = 15 + 4 = 19$$

Correction Exercice 10 : Variance d'une Différence (Indépendant)

1. Puisque X et Y sont indépendantes, leur covariance est nulle : $\text{Cov}(X, Y) = 0$. 2. On utilise la formule générale :

$$\text{Var}(X - Y) = \text{Var}(X + (-1)Y) = \text{Var}(X) + \text{Var}(-1 \cdot Y) + 2\text{Cov}(X, -Y)$$

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

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2(0)$$

$$\text{Var}(X - Y) = 16 + 9 = 25$$

Correction Exercice 11 : Bilinearité de la Covariance

La covariance est linéaire sur son premier argument :

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

Correction Exercice 12 : Variance d'une Combinaison Linéaire

On utilise $\text{Var}(aX + bY + c) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$. Ici $a = 3$, $b = -5$, $c = 1$. X et Y sont indépendantes, donc $\text{Cov}(X, Y) = 0$.

$$\begin{aligned}\text{Var}(3X - 5Y + 1) &= (3)^2\text{Var}(X) + (-5)^2\text{Var}(Y) + 0 \\ &= 9 \times (4) + 25 \times (2) = 36 + 50 = 86\end{aligned}$$

(Note : la constante additive $c = 1$ ne change pas la variance).

Correction Exercice 13 : Variance d'une Somme (Dés)

1. Oui, les lancers de deux dés standards sont des événements physiquement indépendants. 2. Puisqu'ils sont indépendants, $\text{Cov}(D_1, D_2) = 0$.

$$\begin{aligned}\text{Var}(S) &= \text{Var}(D_1 + D_2) = \text{Var}(D_1) + \text{Var}(D_2) \\ \text{Var}(S) &= \frac{35}{12} + \frac{35}{12} = \frac{70}{12} = \frac{35}{6}\end{aligned}$$

Correction Exercice 14 : Covariance et Variance

Par définition, $\text{Cov}(A, B) = E[(A - \mu_A)(B - \mu_B)]$. Posons $A = X$ et $B = X$. Alors $\mu_A = \mu_X$ et $\mu_B = \mu_X$.

$$\text{Cov}(X, X) = E[(X - \mu_X)(X - \mu_X)] = E[(X - \mu_X)^2]$$

C'est la définition de $\text{Var}(X)$.

Correction Exercice 15 : Covariance avec une Constante

On utilise la formule de calcul $\text{Cov}(X, c) = E[Xc] - E[X]E[c]$. Par linéarité, $E[Xc] = cE[X]$. L'espérance d'une constante est la constante elle-même : $E[c] = c$.

$$\text{Cov}(X, c) = cE[X] - E[X]c = 0$$

Correction Exercice 16 : Standardisation (Centrer-Réduire)

$Z = \frac{X-10}{2} = \frac{1}{2}X - 5$. 1. Calcul de $E[Z]$ par linéarité :

$$E[Z] = E\left[\frac{1}{2}X - 5\right] = \frac{1}{2}E[X] - 5 = \frac{1}{2}(10) - 5 = 5 - 5 = 0$$

2. Calcul de $\text{Var}(Z)$ par les propriétés de la variance :

$$\begin{aligned}\text{Var}(Z) &= \text{Var}\left(\frac{1}{2}X - 5\right) = \left(\frac{1}{2}\right)^2 \text{Var}(X) = \frac{1}{4}\text{Var}(X) \\ \text{Var}(Z) &= \frac{1}{4}(4) = 1\end{aligned}$$

Par définition, une variable standardisée a une moyenne de 0 et une variance de 1.

Correction Exercice 17 : Corrélation et Standardisation

La corrélation $\text{Corr}(X, Y)$ EST, par définition, la covariance des versions standardisées Z_X et Z_Y .

$$\text{Cov}(Z_X, Z_Y) = \text{Corr}(X, Y) = 0.5$$

Correction Exercice 18 : Somme de Lois de Poisson

1. Puisque X et Y sont des v.a. de Poisson **indépendantes**, leur somme $S = X + Y$ suit aussi une **loi de Poisson**. Le nouveau paramètre est la somme des paramètres : $\lambda_S = \lambda_1 + \lambda_2 =$

$5 + 3 = 8$. Donc, $S \sim \text{Poisson}(\lambda = 8)$. 2. On cherche $P(S = 6)$ pour $S \sim \text{Poisson}(8)$.

$$P(S = 6) = \frac{e^{-8} 8^6}{6!} = \frac{e^{-8} \times 262144}{720} = 364.08 \times e^{-8} \approx 0.122$$

Correction Exercice 19 : Corrélation Nulle mais Dépendance

1. $E[X] = (-1)(1/3) + (0)(1/3) + (1)(1/3) = -1/3 + 0 + 1/3 = 0$. 2. $E[XY] = E[X(X^2)] = E[X^3]$. $E[X^3] = (-1)^3(1/3) + (0)^3(1/3) + (1)^3(1/3) = (-1)(1/3) + 0 + (1)(1/3) = 0$. 3. $\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0 - (0)E[Y] = 0$. Les variables sont **non corrélées**. 4. Les variables X et Y sont-elles indépendantes ? Non. Test : $P(X = 1, Y = 0) \stackrel{?}{=} P(X = 1)P(Y = 0)$.
- $P(X = 1, Y = 0) = P(X = 1, X^2 = 0) = 0$. - $P(X = 1) = 1/3$. - $P(Y = 0) = P(X^2 = 0) = P(X = 0) = 1/3$. - $P(X = 1)P(Y = 0) = (1/3)(1/3) = 1/9$. Puisque $0 \neq 1/9$, elles **ne sont pas indépendantes**. C'est un exemple classique de dépendance non linéaire avec covariance nulle.

Correction Exercice 20 : Bornes de la Corrélation

Y est une fonction linéaire parfaite de X : $Y = aX + b$ avec $a = -3$ et $b = 5$. La corrélation $\text{Corr}(X, Y)$ mesure la force de la relation *linéaire*. Puisqu'elle est parfaite, la corrélation doit être ± 1 . Le coefficient $a = -3$ est négatif, donc la relation est décroissante. Par conséquent, $\text{Corr}(X, Y) = -1$.

7.11 Exercices Python

Ces exercices appliquent les concepts de distributions multivariées (covariance, corrélation, variance d'une somme) à des données financières réelles. Nous allons analyser la relation entre les rendements boursiers de deux entreprises technologiques : Google (GOOG) et Microsoft (MSFT).

Nous travaillerons avec les **rendements journaliers** (variation en pourcentage), qui sont des variables aléatoires continues. L'espérance $E[X]$ sera estimée par la moyenne empirique (`.mean()`) et la variance $\text{Var}(X)$ par la variance empirique (`.var()`).

```
!pip install yfinance
import yfinance as yf
import pandas as pd
import numpy as np

# Définir les tickers et la période
tickers = ["GOOG", "MSFT"]
start_date = "2020-01-01"
end_date = "2024-12-31"

# Télécharger les prix de clôture ajustés
data = yf.download(tickers, start=start_date, end=end_date)["Adj Close"]

# Calculer les rendements journaliers en pourcentage
returns = data.pct_change().dropna()

# Renommer les colonnes pour plus de clarté
returns.columns = ["GOOG_Return", "MSFT_Return"]

# "returns" est notre DataFrame principal.
# X = returns["GOOG_Return"]
# Y = returns["MSFT_Return"]
```

Exercice 1 : Espérances et Variances Marginales

Soit X la v.a. "Rendement journalier de GOOG" et Y la v.a. "Rendement journalier de MSFT".

Votre tâche :

1. Calculer l'espérance empirique $E[X]$ et $E[Y]$. (Que remarquez-vous sur leur ordre de grandeur ?)
2. Calculer la variance empirique $\text{Var}(X)$ et $\text{Var}(Y)$.
3. Calculer l'écart-type empirique σ_X et σ_Y . Laquelle des deux actions est la plus "volatile" ?

Exercice 2 : Standardisation (Centrer-Réduire)

Le concept de variable centrée réduite $Z = \frac{X - \mu_X}{\sigma_X}$ est très utilisé en finance (par ex : "Z-score").

Votre tâche :

1. Récupérer $E[X]$ (la moyenne) et σ_X (l'écart-type) des rendements de GOOG de l'exercice 1.
2. Prendre le **dernier** rendement journalier de GOOG dans le jeu de données.
3. Calculer le "Z-score" de ce dernier rendement.
4. Interpréter ce score (par ex : "Le dernier jour, GOOG a performé à X écarts-types de sa moyenne...").

Exercice 3 : Covariance (Calcul via LOTUS)

Calculez la covariance entre les rendements de GOOG (X) et de MSFT (Y) en utilisant la formule $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$.

Votre tâche :

1. Récupérer $E[X]$ et $E[Y]$ de l'exercice 1.
2. Calculer $E[XY]$ en utilisant le "Théorème de Transfert" (LOTUS) sur les données empiriques (Indice : calculez la moyenne de la série $X \times Y$).
3. Appliquer la formule pour trouver $\text{Cov}(X, Y)$.
4. Le signe est-il positif ou négatif? Qu'est-ce que cela implique intuitivement?

Exercice 4 : Corrélation (Calcul)

La covariance de l'exercice 3 dépend des unités (rendement au carré). Nous allons la normaliser pour obtenir la corrélation $r \in [-1, 1]$.

Votre tâche :

1. Récupérer $\text{Cov}(X, Y)$ (Exercice 3) et σ_X, σ_Y (Exercice 1).
2. Appliquer la formule : $\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$.
3. Interpréter ce coefficient. La relation linéaire entre GOOG et MSFT est-elle forte ou faible?

Exercice 5 : Linéarité de l'Espérance (Portefeuille)

Soit un portefeuille P composé à 60% de GOOG (X) et 40% de MSFT (Y). Le rendement du portefeuille est $P = 0.6X + 0.4Y$. La théorie dit : $E[P] = E[0.6X + 0.4Y] = 0.6E[X] + 0.4E[Y]$.

Votre tâche :

1. En utilisant $E[X]$ et $E[Y]$ de l'exercice 1, calculer l'espérance **théorique** $E[P]$.
2. **Vérification empirique :**
 - Créer la série de données $P_{series} = 0.6 \times X + 0.4 \times Y$.
 - Calculer l'espérance empirique de P_{series} (`.mean()`).
3. Comparer votre résultat théorique (1) et empirique (2).

Exercice 6 : Variance d'un Portefeuille (Variance d'une Somme)

Continuons avec le portefeuille $P = 0.6X + 0.4Y$. La variance **théorique** est : $\text{Var}(P) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$.

Votre tâche :

1. En utilisant $\text{Var}(X)$, $\text{Var}(Y)$ (Ex 1) et $\text{Cov}(X, Y)$ (Ex 3), calculer $\text{Var}(P)$ en appliquant la formule ci-dessus.
2. **Vérification empirique :**
 - Utiliser la série P_{series} de l'exercice 5.
 - Calculer la variance empirique de P_{series} (`.var()`).
3. Comparer votre résultat théorique (1) et empirique (2).

Exercice 7 : Le Bénéfice de la Diversification

La diversification (le fait que $\text{Corr}(X, Y) \neq 1$) réduit le risque. Nous allons le prouver. Le risque (l'écart-type) d'un portefeuille *n'est pas* la moyenne pondérée des risques.

Votre tâche :

1. Calculer le risque du portefeuille σ_P (l'écart-type) en prenant la racine carrée de $\text{Var}(P)$ (calculée à l'exercice 6).
2. Calculer la "moyenne pondérée des risques" : $\sigma_{\text{moy}} = 0.6\sigma_X + 0.4\sigma_Y$ (en utilisant σ_X, σ_Y de l'Ex 1).
3. Comparer σ_P et σ_{moy} . Lequel est le plus petit ?
4. (Conclusion) Pourquoi $\sigma_P < \sigma_{\text{moy}}$? (Indice : $\text{Corr}(X, Y)$).

Exercice 8 : Vérification des Bornes de Corrélation

Le théorème stipule que si $Y = aX + b$, alors $\text{Corr}(X, Y) = \pm 1$. Vérifions cela.

Votre tâche :

1. Soit X la série des rendements de GOOG.
2. Créer une nouvelle variable $Z = -3X + 0.005$ (une relation linéaire négative parfaite).
3. Calculer la corrélation empirique entre X et Z . (Vous pouvez utiliser `X.corr(Z)`).
4. Le résultat est-il conforme au théorème ?

Exercice 9 : Loi Jointe (Discrétisation)

Transformons nos variables continues X et Y en variables de Bernoulli discrètes. Soit $X_{\text{bern}} = 1$ si le rendement de GOOG est positif (> 0), et 0 sinon. Soit $Y_{\text{bern}} = 1$ si le rendement de MSFT est positif (> 0), et 0 sinon. Nous voulons trouver la PMF jointe $P(X_{\text{bern}} = x, Y_{\text{bern}} = y)$.

Votre tâche :

1. Créer les deux séries discrètes X_{bern} et Y_{bern} .
2. Utiliser `pandas.crosstab` pour créer un tableau de contingence (les effectifs).
3. Normaliser ce tableau par l'effectif total pour obtenir la **loi jointe** (PMF jointe).
4. Quelle est la probabilité que les deux actions aient un rendement positif le même jour, $P(X_{\text{bern}} = 1, Y_{\text{bern}} = 1)$?

Exercice 10 : Lois Marginales et Indépendance (Discret)

En utilisant la loi jointe $P(X_{\text{bern}}, Y_{\text{bern}})$ de l'exercice 9.

Votre tâche :

1. Calculer la loi marginale $P(X_{\text{bern}} = x)$ (somme des lignes).
2. Calculer la loi marginale $P(Y_{\text{bern}} = y)$ (somme des colonnes).
3. Les variables X_{bern} et Y_{bern} sont-elles indépendantes ?
4. Justifiez en comparant $P(X_{\text{bern}} = 1, Y_{\text{bern}} = 1)$ au produit $P(X_{\text{bern}} = 1) \times P(Y_{\text{bern}} = 1)$.

8 Moments d'une distribution

8.1 Définitions fondamentales des moments

Après avoir défini l'espérance (μ) et la variance (σ^2), qui sont les moments d'ordre 1 et 2, nous pouvons généraliser cette idée pour capturer des informations plus subtiles sur la forme d'une distribution.

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).

8.2 Asymétrie (Skewness)

Le premier moment nous donne la tendance centrale. Le deuxième moment (la variance) nous donne la dispersion. Le troisième moment, lui, va nous renseigner sur la *symétrie* de la distribution.

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 (Symétrique)** : La distribution est symétrique. Les écarts positifs et négatifs s'annulent. Typiquement : Moyenne = Médiane = Mode.
- **Skewness > 0 (Queue à droite)** : La distribution présente une queue longue à droite. Les grandes valeurs positives sont amplifiées par le cube. Les valeurs extrêmes tirent la moyenne vers la droite.
- **Skewness < 0 (Queue à gauche)** : La distribution présente une queue longue à gauche. Les écarts négatifs dominent. Les valeurs extrêmes tirent la moyenne vers la gauche.

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

Remarque : Pourquoi Standardiser ?

En standardisant d'abord $\left(\frac{X-\mu}{\sigma}\right)$, 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).

8.3 Propriétés de symétrie

Le skewness est une mesure numérique de l'asymétrie. Mais nous pouvons aussi définir la symétrie de manière formelle.

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.

8.4 Aplatissement (Kurtosis)

Après l'asymétrie (ordre 3), le moment d'ordre 4 nous informe sur "l'épaisseur" des queues de la distribution, c'est-à-dire la probabilité d'obtenir des valeurs très éloignées de la moyenne.

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 (basée sur l'Excess Kurtosis)

- **Leptokurtique (Excess Kurtosis > 0)** : Kurtosis total > 3. Distribution pointue avec des queues épaisses. Les événements extrêmes sont plus probables que pour une loi normale.
- **Mésokurtique (Excess Kurtosis = 0)** : Kurtosis total = 3. C'est la référence (loi normale).
- **Platykurtique (Excess Kurtosis < 0)** : Kurtosis total < 3. Distribution aplatie avec des queues légères et un centre large. Les événements extrêmes sont moins probables.

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

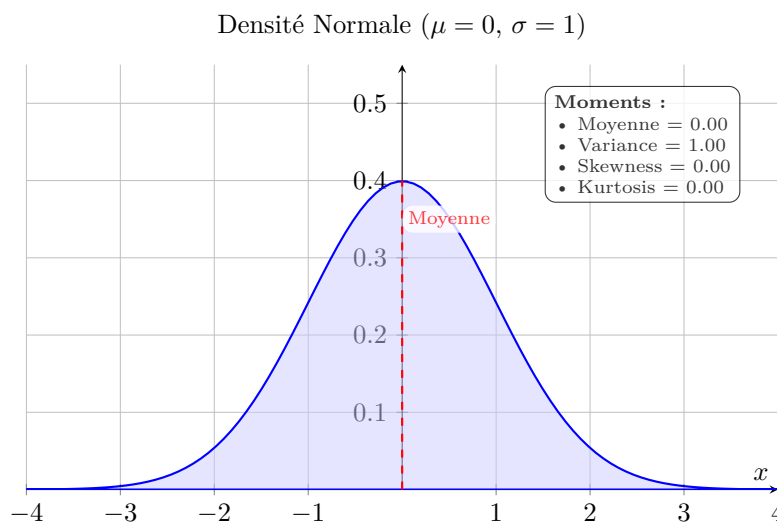
Pourquoi σ^4 au dénominateur ?

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

8.5 Exemples de distributions

Pour bien fixer les idées, comparons le skewness et le kurtosis de plusieurs distributions classiques. Notez que dans les graphiques suivants, le "Kurtosis" affiché est l'*excess kurtosis* (centré à 0).

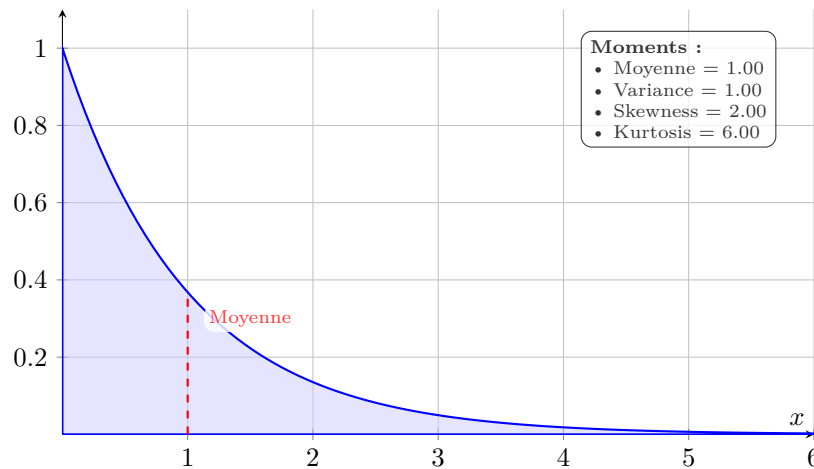
Exemple : La Distribution Normale (Mésokurtique)



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)

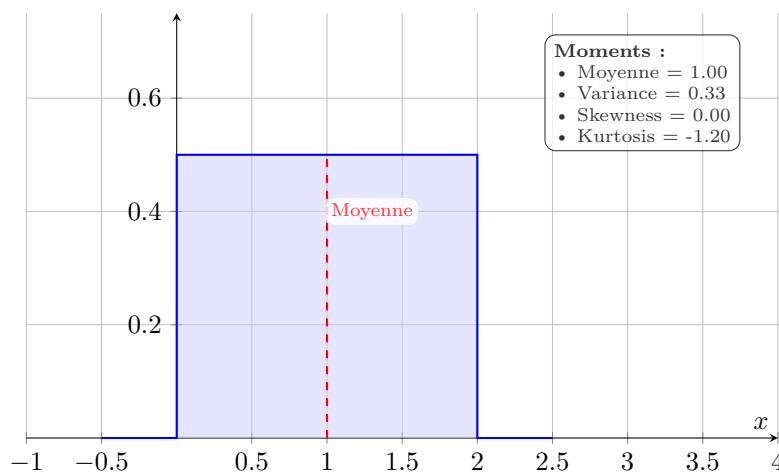
Densité Exponentielle ($\lambda = 1$)



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.

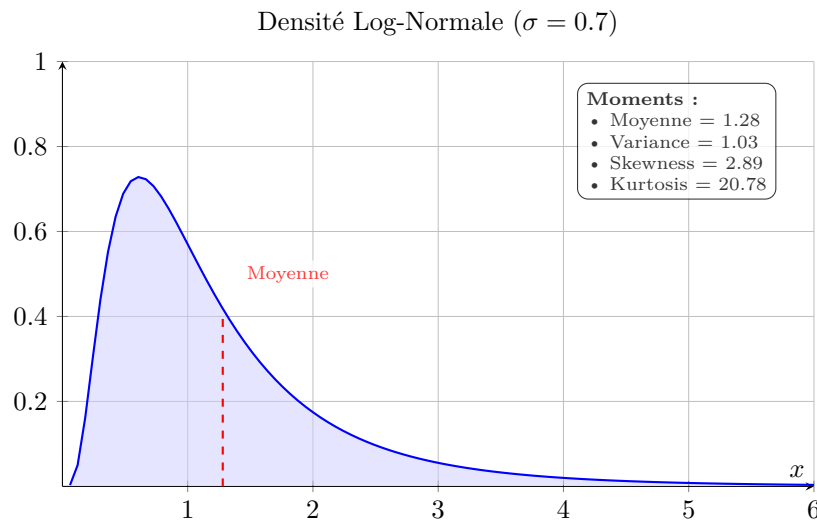
Exemple : La Distribution Uniforme (Platykurtique)

Densité Uniforme ($a = 0, b = 2$)



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.

Nous avons défini les moments d'une *distribution* (moments de population), tels que $\mu = E[X]$ ou $\sigma^2 = E[(X - \mu)^2]$. Ce sont des valeurs théoriques, la "vérité" sous-jacente.

En pratique, nous ne connaissons presque jamais cette "vérité". Nous ne disposons que de données. Notre but est d'utiliser ces données pour *estimer* les moments de la population.

8.6 Moments d'échantillon (Sample Moments)

Définition : Moments d'Échantillon

Soit X_1, X_2, \dots, X_n un échantillon de n observations.

- La **moyenne d'échantillon** (notre "meilleure estimation" de μ) est :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

- La **variance d'échantillon (non biaisée)** (notre "meilleure estimation" de σ^2) est :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

De même, on peut calculer un *skewness d'échantillon* et un *kurtosis d'échantillon* en utilisant \bar{X} et s , qui seront nos estimations du vrai skewness et du vrai kurtosis de la population.

Exemple : Application : Contrôle Qualité

Imaginez une usine qui produit des sacs de sucre de 1kg.

- **Population** : L'infinité de tous les sacs de sucre que la machine produira.
- **Moment de population (inconnu)** : Le poids moyen *réel* μ que la machine verse, et la variance *réelle* σ^2 (sa constance).
- **Problème** : Nous ne pouvons pas peser tous les sacs !
- **Solution** : Nous prélevons un **échantillon** de $n = 10$ sacs.
Nous les pesons : $\{1002g, 998g, 1001g, 995g, 1003g, 1000g, 997g, 1005g, 999g, 1000g\}$.

· **Calcul des moments d'échantillon :**

- $\bar{X} = (1002 + 998 + \dots + 1000)/10 = 1000g$.
- $s^2 = \frac{1}{10-1} ((1002 - 1000)^2 + (998 - 1000)^2 + \dots) = 7.33g^2$.

· **Conclusion :** Notre meilleure estimation est que la machine est bien réglée sur $\mu = 1000g$. L'écart-type de notre échantillon est $s = \sqrt{7.33} \approx 2.7g$. Nous pouvons utiliser cela pour affirmer, par exemple, que 95% des sacs se situent probablement entre $1000 \pm 2s$ (si la distribution est normale).

Remarque : L'Intuition du " $n - 1$ "

Pourquoi diviser par $n - 1$ pour la variance ? C'est la **correction de Bessel**.

Imaginez un échantillon de 1 seule personne ($n = 1$). Sa taille est 170cm.

- Quelle est la moyenne de l'échantillon ? $\bar{X} = 170$ cm.
- Quelle est la variance de l'échantillon ? $\sum (X_i - \bar{X})^2 = (170 - 170)^2 = 0$.
- Si on divisait par $n = 1$, on estimerait que la variance de la population est 0. C'est absurde ! Cela voudrait dire que tout le monde mesure 170cm.

En divisant par $n - 1$ (donc $1 - 1 = 0$), la formule devient $0/0$ (indéfinie), ce qui nous dit à juste titre : "Je ne peux pas estimer la dispersion avec une seule personne."

Intuition plus générale : Nous "perdons un degré de liberté". Pour calculer la variance, nous avons besoin de connaître la moyenne. Mais nous ne connaissons pas la vraie moyenne μ . Nous devons donc utiliser \bar{X} , une *estimation*. Le fait d'utiliser une estimation calculée à partir de ce même échantillon introduit un léger biais (nos données sont, par définition, centrées sur \bar{X}). Diviser par $n - 1$ au lieu de n "gonfle" légèrement le résultat pour compenser ce biais.

8.7 Fonctions génératrices des moments (MGF)

Définition : Fonction Génératrice des Moments (MGF)

La **fonction génératrice des moments** (MGF) d'une variable aléatoire X , notée $M_X(t)$, est définie comme :

$$M_X(t) = E[e^{tX}]$$

Intuition : L'ADN, le Code-Barres, ou le Fichier .zip

Ce concept est abstrait, alors utilisons des analogies :

Analogie 1 : L'ADN ou l'Empreinte Digitale

- La MGF est l'**empreinte digitale unique** d'une distribution.
- Elle "compresse" *toutes* les informations sur votre distribution (moyenne, variance, skewness, kurtosis, etc.) en une seule, unique fonction.
- Si deux distributions ont la même MGF, elles sont identiques. C'est la **propriété d'unicité**.

Analogie 2 : Le Code-Barres

- Pensez à une distribution (ex : Loi Normale) comme à un produit au supermarché.
- La MGF, $M_X(t)$, est son **code-barres unique**.
- Le processus de "génération de moments" (que nous verrons ci-dessous) est le **scanner**.
- En scannant le code-barres ($M_X(t)$), vous pouvez obtenir n'importe quelle information :
- Scan 1 ($M'_X(0)$) → vous donne le prix ($E[X]$).
- Scan 2 ($M''_X(0)$) → vous donne le poids ($E[X^2]$).
- Scan 3 ($M'''_X(0)$) → vous donne le pays d'origine ($E[X^3]$).

Pourquoi e^{tX} ? La "magie" vient du développement en série de Taylor de e^x :

$$e^{tX} = 1 + (tX) + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \dots$$

Quand on prend l'espérance, $E[\cdot]$, les puissances de X (c'est-à-dire $X, X^2, X^3 \dots$) apparaissent. Ce sont les moments ! La MGF "stocke" tous ces moments en les organisant comme coefficients d'un polynôme infini en t .

8.8 Génération des moments via les MGF

Théorème : Moments par Dérivation

Si la MGF $M_X(t)$ existe, alors le m -ième moment non centré $E[X^m]$ est la m -ième dérivée de $M_X(t)$, évaluée en $t = 0$:

$$E[X^m] = \frac{d^m}{dt^m} M_X(t) \Big|_{t=0} = M_X^{(m)}(0)$$

Exemple : Application : La Loi de Poisson

Une loi de Poisson modélise le nombre d'événements (ex : appels à un centre d'appels) par heure. Soit $X \sim \text{Poisson}(\lambda)$, où λ est le nombre moyen d'appels.

La MGF (l'ADN) d'une loi de Poisson est (on l'admet) :

$$M_X(t) = e^{\lambda(e^t - 1)}$$

Utilisons notre "scanner" (les dérivées) pour trouver les moments.

1. Trouver la Moyenne $E[X]$: On dérive une fois (règle de la chaîne) :

$$M'_X(t) = \frac{d}{dt} \left(e^{\lambda(e^t - 1)} \right) = \underbrace{e^{\lambda(e^t - 1)}}_{\text{répète}} \cdot \underbrace{(\lambda e^t)}_{\text{dérivée interne}}$$

Maintenant, on évalue en $t = 0$:

$$E[X] = M'_X(0) = e^{\lambda(e^0 - 1)} \cdot (\lambda e^0) = e^{\lambda(1-1)} \cdot (\lambda \cdot 1) = e^0 \cdot \lambda = 1 \cdot \lambda = \lambda$$

Résultat : La moyenne est λ , ce qui est la définition même du paramètre de la loi de Poisson. Parfait.

2. Trouver $E[X^2]$ (pour la variance) : On dérive une seconde fois (règle du produit sur $M'_X(t) = (\lambda e^t) \cdot (e^{\lambda(e^t - 1)})$) :

$$M''_X(t) = \underbrace{(\lambda e^t)}_{\text{dérivée de u}} \cdot \underbrace{(e^{\lambda(e^t - 1)})}_v + \underbrace{(\lambda e^t)}_u \cdot \underbrace{(e^{\lambda(e^t - 1)} \cdot \lambda e^t)}_{\text{dérivée de v}}$$

Maintenant, on évalue en $t = 0$ (tous les e^0 deviennent 1) :

$$E[X^2] = M''_X(0) = (\lambda \cdot 1) \cdot (e^{\lambda(1-1)}) + (\lambda \cdot 1) \cdot (e^{\lambda(1-1)} \cdot \lambda \cdot 1)$$

$$E[X^2] = (\lambda) \cdot (e^0) + (\lambda) \cdot (e^0 \cdot \lambda) = \lambda \cdot 1 + \lambda \cdot (1 \cdot \lambda) = \lambda + \lambda^2$$

3. Trouver la Variance $\text{Var}(X)$: $\text{Var}(X) = E[X^2] - (E[X])^2 = (\lambda + \lambda^2) - (\lambda)^2 = \lambda$ **Résultat :** Nous avons prouvé par les MGF que pour une loi de Poisson, Moyenne = Variance = λ . C'est une propriété fondamentale de cette loi.

8.9 Sommes de variables aléatoires indépendantes via les MGF

C'est la super-puissance des MGF.

Théorème : MGF d'une Somme

Soient X et Y deux variables aléatoires **indépendantes**. Soit $S = X + Y$. Alors la MGF de S est le produit des MGF individuelles :

$$M_S(t) = M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$$

Intuition : La Magie de l'Exponentielle

Pourquoi est-ce vrai ? $M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX} \cdot e^{tY}]$. Parce que X et Y sont indépendantes, $E[f(X)g(Y)] = E[f(X)]E[g(Y)]$. Donc, $E[e^{tX} \cdot e^{tY}] = E[e^{tX}] \cdot E[e^{tY}] = M_X(t) \cdot M_Y(t)$. Les MGF transforment une opération analytiquement horrible (la "convolution" de densités) en une simple multiplication algébrique.

Exemple : Application : Portefeuille d'Actifs ou Tailles Humaines

C'est l'un des théorèmes les plus importants des statistiques. **Problème :** Soit X la taille d'un homme, $X \sim N(\mu_X, \sigma_X^2)$. Soit Y la taille d'une femme, $Y \sim N(\mu_Y, \sigma_Y^2)$. Si on les choisit au hasard, quelle est la loi de la somme de leurs tailles $S = X + Y$?

1. **ADN de X :** La MGF d'une loi Normale $N(\mu, \sigma^2)$ est $M(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$.
2. **ADN de X et Y :** $M_X(t) = \exp(\mu_X t + \frac{1}{2}\sigma_X^2 t^2)$ $M_Y(t) = \exp(\mu_Y t + \frac{1}{2}\sigma_Y^2 t^2)$
3. **ADN de $S = X + Y$ (on multiplie) :** $M_S(t) = M_X(t) \cdot M_Y(t) = \exp(\mu_X t + \frac{1}{2}\sigma_X^2 t^2) \cdot \exp(\mu_Y t + \frac{1}{2}\sigma_Y^2 t^2)$
4. **Simplification** (en additionnant les exposants) : $M_S(t) = \exp((\mu_X t + \mu_Y t) + (\frac{1}{2}\sigma_X^2 t^2 + \frac{1}{2}\sigma_Y^2 t^2))$ $M_S(t) = \exp((\mu_X + \mu_Y)t + \frac{1}{2}(\sigma_X^2 + \sigma_Y^2)t^2)$
5. **Conclusion (par unicité) :** Regardez cet ADN ! C'est l'ADN d'une loi Normale ! Le nouveau μ est $(\mu_X + \mu_Y)$. La nouvelle σ^2 est $(\sigma_X^2 + \sigma_Y^2)$.

Résultat : Nous avons prouvé que **la somme de deux Normales indépendantes est une nouvelle Normale**. Si $X \sim N(175cm, 7^2)$ et $Y \sim N(165cm, 6^2)$, alors $S \sim N(340cm, 7^2 + 6^2 = 85)$. Notez que les écarts-types *ne s'additionnent pas* ($\sqrt{85} \approx 9.2 \neq 7 + 6$). Ce sont les variances qui s'additionnent.

8.10 Exercices

Exercice 1 : Concepts (Moments)

1. À quoi correspond le 1er moment non centré, $E[X^1]$?
2. À quoi correspond le 2ème moment centré, $E[(X - \mu)^2]$?

Exercice 2 : Interprétation (Skewness)

Une distribution des salaires dans une entreprise a un skewness de +2.5. Qu'est-ce que cela signifie concrètement sur la répartition des salaires ?

Exercice 3 : Interprétation (Skewness)

Si une distribution est parfaitement symétrique, que vaut son skewness ?

Exercice 4 : Interprétation (Kurtosis)

Une distribution des rendements d'un actif financier a un "excess kurtosis" de +5.0.

1. Comment appelle-t-on ce type de distribution ?
2. Qu'est-ce que cela implique sur la probabilité des "krachs" (événements extrêmes) par rapport à une loi normale ?

Exercice 5 : Interprétation (Kurtosis)

Une distribution a un "excess kurtosis" de -1.0.

1. Comment appelle-t-on ce type de distribution ?
2. Comment décririez-vous visuellement son "pic" et ses "queues" par rapport à une loi normale ?

Exercice 6 : Comparaison (Skewness)

La distribution exponentielle est-elle symétrique, asymétrique à gauche ou asymétrique à droite ? Quel est le signe de son skewness ?

Exercice 7 : Comparaison (Kurtosis)

La distribution uniforme est-elle mésokurtique, leptokurtique ou platykurtique ? Son "excess kurtosis" est-il positif, négatif ou nul ?

Exercice 8 : Propriété de Symétrie

Soit X une variable aléatoire symétrique autour de sa moyenne μ . Montrez que son 3ème moment centré, $E[(X - \mu)^3]$, est nul (et donc que son skewness est nul).

Exercice 9 : Vérification de Symétrie (PDF)

La PDF d'une loi X est $f(x) = \frac{1}{2}e^{-|x-3|}$. Cette distribution est-elle symétrique ? Si oui, autour de quel point μ ?

Exercice 10 : Symétrie et Moments

Si une distribution a un skewness de 0, est-elle forcément symétrique ? (Indice : Pensez à une distribution qui aurait des queues asymétriques mais qui s'annuleraient pour le moment d'ordre 3).

Exercice 11 : Définitions (Échantillon vs Pop.)

Quelle est la différence conceptuelle entre σ^2 et s^2 ?

Exercice 12 : Correction de Bessel ($n - 1$)

Pourquoi utilise-t-on $n - 1$ au dénominateur pour s^2 ? Que se passerait-il si nous avions un échantillon de $n = 1$ et que nous divisons par n ?

Exercice 13 : Calcul (Moments d'Échantillon)

On observe l'échantillon de données suivant : $\{2, 3, 10\}$.

1. Calculez la moyenne d'échantillon \bar{X} .
2. Calculez la variance d'échantillon non biaisée s^2 .

Exercice 14 : MGF (Propriété de base)

Soit $M_X(t)$ la MGF d'une variable X . Que vaut $M_X(0)$?

Exercice 15 : MGF (Série de Taylor)

Le développement en série de Taylor d'une MGF est $M_X(t) = 1 + 5t + 14t^2 + O(t^3)$. (Rappel : $M_X(t) = 1 + E[X]t + E[X^2]\frac{t^2}{2!} + \dots$)

1. Trouvez $E[X]$.
2. Trouvez $E[X^2]$.
3. Calculez $\text{Var}(X)$.

Exercice 16 : MGF (Calcul de moments)

La MGF de la loi Exponentielle de paramètre λ est $M_X(t) = \frac{\lambda}{\lambda - t}$.

1. Calculez $M'_X(t)$ et trouvez $E[X] = M'_X(0)$.
2. Calculez $M''_X(t)$ et trouvez $E[X^2] = M''_X(0)$.

Exercice 17 : MGF (Variance de l'Exponentielle)

En utilisant les résultats de l'exercice 16, déduisez $\text{Var}(X)$ pour une loi Exponentielle(λ).

Exercice 18 : MGF (Loi Normale)

La MGF de $X \sim \mathcal{N}(\mu, \sigma^2)$ est $M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$. Calculez $M'_X(t)$ et vérifiez que $M'_X(0) = \mu$.

Exercice 19 : MGF (Loi Normale)

En utilisant la MGF de l'exercice 18, calculez $M''_X(t)$ et montrez que $E[X^2] = \mu^2 + \sigma^2$.

Exercice 20 : MGF (Loi Normale)

En utilisant les résultats des exercices 18 et 19, retrouvez la formule de la variance $\text{Var}(X)$.

Exercice 21 : Propriété des Sommes

Soient X et Y deux variables aléatoires **indépendantes**. Soit $S = X + Y$. Comment la MGF de S , $M_S(t)$, est-elle liée à $M_X(t)$ et $M_Y(t)$?

Exercice 22 : Application (Somme de Poissons)

$X \sim \text{Poisson}(\lambda_1)$ a pour MGF $M_X(t) = e^{\lambda_1(e^t-1)}$. $Y \sim \text{Poisson}(\lambda_2)$ a pour MGF $M_Y(t) = e^{\lambda_2(e^t-1)}$. X et Y sont indépendantes.

1. Trouvez la MGF de $S = X + Y$.
2. En reconnaissant la forme de la MGF de S , quelle est la loi de S ?

Exercice 23 : Application (Somme de Binomiales)

La MGF de $X \sim \text{Bin}(n, p)$ est $M_X(t) = (pe^t + (1-p))^n$. Soit $Y \sim \text{Bin}(m, p)$ (même p), indépendante de X . Quelle est la loi de $S = X + Y$? Justifiez avec les MGF.

Exercice 24 : Application (Transformation Linéaire)

Soit $M_X(t)$ la MGF de X . Soit $Y = aX + b$. Exprimez $M_Y(t)$ en fonction de $M_X(t)$. (Indice : $M_Y(t) = E[e^{t(aX+b)}]$).

Exercice 25 : Application (Moyenne d'Échantillon)

Soient X_1, \dots, X_n des v.a. i.i.d. (indépendantes et identiquement distribuées) avec la MGF $M_X(t)$. Soit $\bar{X} = \frac{1}{n} \sum X_i$ la moyenne d'échantillon. Exprimez la MGF de \bar{X} en fonction de $M_X(t)$. (Indice : utilisez les résultats des exercices 21 et 24).

8.11 Corrections des Exercices

Correction Exercice 1 : Concepts (Moments)

1. Le 1er moment non centré $E[X^1]$ est l'espérance μ . 2. Le 2ème moment centré $E[(X - \mu)^2]$ est la variance σ^2 .

Correction Exercice 2 : Interprétation (Skewness)

Un skewness de +2.5 est fortement positif. Cela signifie que la distribution des salaires est **asymétrique à droite**. Concrètement : la grande majorité des employés a un salaire regroupé (autour de la médiane), mais il existe une "longue queue" de quelques individus avec des salaires très élevés. Ces valeurs extrêmes "tirent" la moyenne vers la droite (Moyenne > Médiane).

Correction Exercice 3 : Interprétation (Skewness)

Si une distribution est parfaitement symétrique (comme la loi normale ou la loi uniforme), les écarts positifs et négatifs à la moyenne s'annulent parfaitement. Le skewness est **nul (0)**.

Correction Exercice 4 : Interprétation (Kurtosis)

1. Un excess kurtosis > 0 est dit **leptokurtique**. 2. Cela signifie que la distribution a des "queues plus épaisses" que la loi normale. La probabilité d'événements extrêmes (très grands gains ou très grandes pertes, comme un "krach") est **plus élevée** que ce qu'un modèle normal ne prédirait.

Correction Exercice 5 : Interprétation (Kurtosis)

1. Un excess kurtosis < 0 est dit **platykurtique**. 2. Visuellement, par rapport à une loi normale, la distribution est plus "aplatie" ou "carrée". Elle a un pic central moins prononcé et des queues plus "légères" (les événements extrêmes sont moins probables). La distribution uniforme est un exemple classique.

Correction Exercice 6 : Comparaison (Skewness)

La distribution exponentielle (ex : temps d'attente) a un pic à 0 et une longue queue vers les grandes valeurs. Elle est **asymétrique à droite**, et son skewness est **positif** (Skew = 2).

Correction Exercice 7 : Comparaison (Kurtosis)

La distribution uniforme est "plate" et n'a pas de pic central ni de queues épaisses. Elle est **platykurtique**. Son "excess kurtosis" est **négatif** (Kurtosis Excessif = -1.2).

Correction Exercice 8 : Propriété de Symétrie

Soit $Y = X - \mu$. La symétrie implique que Y et $-Y$ (qui est $\mu - X$) ont la même distribution. $E[(X - \mu)^3] = E[Y^3]$. Puisque Y et $-Y$ ont la même distribution, ils ont les mêmes moments : $E[Y^k] = E[(-Y)^k]$ pour tout k . Pour $k = 3$: $E[Y^3] = E[(-Y)^3] = E[(-1)^3 Y^3] = E[-Y^3] = -E[Y^3]$. La seule valeur qui est égale à son opposé est 0. $E[Y^3] = -E[Y^3] \implies 2E[Y^3] = 0 \implies E[Y^3] = 0$. Donc, $E[(X - \mu)^3] = 0$.

Correction Exercice 9 : Vérification de Symétrie (PDF)

On teste la condition $f(x) = f(2\mu - x)$. Le candidat pour μ est 3. $f(2\mu - x) = f(2(3) - x) = f(6 - x)$. $f(6 - x) = \frac{1}{2}e^{-|(6-x)-3|} = \frac{1}{2}e^{-|3-x|}$. Puisque $|3 - x| = |-(x - 3)| = |x - 3|$, on a $f(6 - x) = \frac{1}{2}e^{-|x-3|} = f(x)$. Oui, la distribution est symétrique autour de $\mu = 3$.

Correction Exercice 10 : Symétrie et Moments

Non. Le Skewness nul est une condition nécessaire pour la symétrie, but pas suffisante. On peut construire des distributions (assez exotiques) qui sont non symétriques, mais où les asymétries se compensent d'une manière qui annule le 3ème moment, résultant en un skewness de 0. Cependant, dans la plupart des cas pratiques, Skewness = 0 est un très bon indicateur de symétrie.

Correction Exercice 11 : Définitions (Échantillon vs Pop.)

σ^2 (variance de population) est un **paramètre** théorique. C'est la "vraie" variance de l'ensemble de la population (souvent inconnue). s^2 (variance d'échantillon) est une **statistique**. C'est une *estimation* de σ^2 calculée à partir d'un sous-ensemble de données (l'échantillon).

Correction Exercice 12 : Correction de Bessel ($n - 1$)

Si $n = 1$ (un seul échantillon X_1), la moyenne d'échantillon est $\bar{X} = X_1$. La somme des carrés des écarts est $\sum (X_i - \bar{X})^2 = (X_1 - X_1)^2 = 0$. Si on divisait par $n = 1$, on obtiendrait $s^2 = 0/1 = 0$. Cela estimerait faussement que la population n'a pas de variance, ce qui est absurde. Diviser par $n - 1$ (donc $1 - 1 = 0$) donne $0/0$, une forme indéfinie, ce qui nous dit correctement qu'on ne peut pas estimer une dispersion à partir d'un seul point.

Correction Exercice 13 : Calcul (Moments d'Échantillon)

Échantillon : $\{2, 3, 10\}$. $n = 3$. 1. $\bar{X} = \frac{1}{3}(2 + 3 + 10) = \frac{15}{3} = 5$. 2. $s^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2 = \frac{1}{2} [(2 - 5)^2 + (3 - 5)^2 + (10 - 5)^2] = \frac{1}{2} [(-3)^2 + (-2)^2 + (5)^2] = \frac{1}{2} [9 + 4 + 25] = \frac{38}{2} = 19$.

Correction Exercice 14 : MGF (Propriété de base)

$M_X(t) = E[e^{tX}]$. En $t = 0$, $M_X(0) = E[e^{0 \cdot X}] = E[e^0] = E[1] = 1$. Toute MGF valide doit valoir 1 en $t = 0$.

Correction Exercice 15 : MGF (Série de Taylor)

On identifie les coefficients de la série $M_X(t) = 1 + \frac{E[X]}{1!}t + \frac{E[X^2]}{2!}t^2 + \dots$. 1. Le coefficient de t est $E[X]$. On lit $5t$. Donc $E[X] = 5$. 2. Le coefficient de t^2 est $E[X^2]/2!$. On lit $14t^2$. $E[X^2]/2 = 14 \implies E[X^2] = 28$. 3. $\text{Var}(X) = E[X^2] - (E[X])^2 = 28 - (5)^2 = 28 - 25 = 3$.

Correction Exercice 16 : MGF (Calcul de moments)

$M_X(t) = \lambda(\lambda - t)^{-1}$. 1. $M'_X(t) = \lambda \cdot (-1)(\lambda - t)^{-2} \cdot (-1) = \lambda(\lambda - t)^{-2}$. $E[X] = M'_X(0) = \lambda(\lambda - 0)^{-2} = \lambda(\lambda^{-2}) = 1/\lambda$. 2. $M''_X(t) = \lambda \cdot (-2)(\lambda - t)^{-3} \cdot (-1) = 2\lambda(\lambda - t)^{-3}$. $E[X^2] = M''_X(0) = 2\lambda(\lambda - 0)^{-3} = 2\lambda(\lambda^{-3}) = 2/\lambda^2$.

Correction Exercice 17 : MGF (Variance de l'Exponentielle)

$\text{Var}(X) = E[X^2] - (E[X])^2$. $\text{Var}(X) = (2/\lambda^2) - (1/\lambda)^2 = 2/\lambda^2 - 1/\lambda^2 = 1/\lambda^2$.

Correction Exercice 18 : MGF (Loi Normale)

$M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$. On utilise la règle de la chaîne : $(\exp(u))' = \exp(u) \cdot u'$. $M'_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2) \cdot \frac{d}{dt}(\mu t + \frac{1}{2}\sigma^2 t^2)$. $M'_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2) \cdot (\mu + \sigma^2 t)$. $E[X] = M'_X(0) =$

$$\exp(0) \cdot (\mu + 0) = 1 \cdot \mu = \mu.$$

Correction Exercice 19 : MGF (Loi Normale)

On dérive $M'_X(t)$ (règle du produit $u \cdot v$) : $u = \exp(\mu t + \frac{1}{2}\sigma^2 t^2) \implies u' = u \cdot (\mu + \sigma^2 t)$
 $v = (\mu + \sigma^2 t) \implies v' = \sigma^2$ $M''_X(t) = u'v + uv' = [\exp(\dots)(\mu + \sigma^2 t)] \cdot (\mu + \sigma^2 t) + [\exp(\dots)] \cdot (\sigma^2)$ On évalue en $t = 0$: $E[X^2] = M''_X(0) = [\exp(0)(\mu)] \cdot (\mu) + [\exp(0)] \cdot (\sigma^2)$
 $E[X^2] = (1 \cdot \mu) \cdot \mu + 1 \cdot \sigma^2 = \mu^2 + \sigma^2.$

Correction Exercice 20 : MGF (Loi Normale)

$\text{Var}(X) = E[X^2] - (E[X])^2$ $\text{Var}(X) = (\mu^2 + \sigma^2) - (\mu)^2 = \sigma^2.$ La MGF confirme bien que le paramètre σ^2 est la variance.

Correction Exercice 21 : Propriété des Sommes

Si X et Y sont indépendantes, la MGF de la somme est le **produit** des MGF : $M_{X+Y}(t) = M_X(t) \cdot M_Y(t).$

Correction Exercice 22 : Application (Somme de Poissons)

- $M_S(t) = M_X(t) \cdot M_Y(t) = e^{\lambda_1(e^t-1)} \cdot e^{\lambda_2(e^t-1)}$ $M_S(t) = e^{\lambda_1(e^t-1)+\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)}.$
- On reconnaît la MGF d'une loi de Poisson. Le paramètre est ce qui multiplie $(e^t - 1)$. Donc S suit une loi de Poisson de paramètre $(\lambda_1 + \lambda_2)$. $S \sim \text{Poisson}(\lambda_1 + \lambda_2).$

Correction Exercice 23 : Application (Somme de Binomiales)

$M_S(t) = M_X(t) \cdot M_Y(t) = (pe^t + (1-p))^n \cdot (pe^t + (1-p))^m$ $M_S(t) = (pe^t + (1-p))^{n+m}.$
 On reconnaît la MGF d'une loi Binomiale avec $n + m$ essais et une probabilité de succès p .
 $S \sim \text{Bin}(n + m, p).$

Correction Exercice 24 : Application (Transformation Linéaire)

$M_Y(t) = E[e^{tY}] = E[e^{t(aX+b)}] = E[e^{taX+tb}]$ $M_Y(t) = E[e^{taX} \cdot e^{tb}].$ Puisque e^{tb} est une constante : $M_Y(t) = e^{tb}E[e^{(ta)X}].$ Par définition, $E[e^{(ta)X}]$ est la MGF de X évaluée au point (ta) . $M_Y(t) = e^{tb}M_X(ta).$

Correction Exercice 25 : Application (Moyenne d'Échantillon)

Soit $S = \sum X_i$. $\bar{X} = S/n = (1/n)S$. 1. D'abord, la MGF de la somme S (Exercice 21) : $M_S(t) = M_{X_1}(t) \cdot \dots \cdot M_{X_n}(t) = [M_X(t)]^n$ (car i.i.d.) 2. Ensuite, on utilise la transformation linéaire $\bar{X} = aS + b$ avec $a = 1/n$ et $b = 0$ (Exercice 24) : $M_{\bar{X}}(t) = e^{bt}M_S(at) = e^0 M_S(t/n)$
 $M_{\bar{X}}(t) = M_S(t/n).$ 3. On combine les deux : $M_{\bar{X}}(t) = [M_X(t/n)]^n.$

8.12 Exercices Python

La loi log-normale est fondamentale en finance. Elle repose sur l'idée que si les **log-rendements** d'une action $X_i = \ln(P_i/P_{i-1})$ sont (approximativement) normaux, alors le prix futur P_t , qui est un **produit** de ces rendements ($P_t = P_0 \times e^{X_1} \times \dots \times e^{X_t}$), suivra une loi log-normale.

Nous allons estimer les paramètres μ et σ^2 de la loi normale sous-jacente à partir des log-rendements journaliers de Microsoft (MSFT) et Google (GOOG), puis utiliser la théorie log-normale pour modéliser les prix.

```
!pip install yfinance
import yfinance as yf
import pandas as pd
import numpy as np
from scipy.stats import norm # Moteur pour les calculs de CDF/PDF
import matplotlib.pyplot as plt

# Définir les tickers et la période
tickers = ["MSFT", "GOOG"]
start_date = "2020-01-01"
end_date = "2024-12-31"
```

```

# Telecharger les prix de cloture ajustés
data = yf.download(tickers, start=start_date, end=end_date)["Adj Close"]

# Calculer les LOG-RENDEMENTS journaliers
log_returns = np.log(data / data.shift(1)).dropna()

# Renommer les colonnes
log_returns.columns = ["MSFT_LogReturn", "GOOG_LogReturn"]

# 'log_returns' est notre DataFrame.
# X_msft = log_returns["MSFT_LogReturn"]
# X_goog = log_returns["GOOG_LogReturn"]

```

Exercice 1 : Estimer les Paramètres μ et σ^2

Soit P_t le prix de MSFT. Le modèle suppose que $X = \ln(P_t/P_{t-1}) \sim \mathcal{N}(\mu, \sigma^2)$. Les paramètres μ et σ^2 sont les paramètres "log-normaux".

Votre tâche :

1. Estimer μ (l'espérance du log-rendement journalier) pour MSFT.
2. Estimer σ^2 (la variance du log-rendement journalier) pour MSFT.
3. Estimer σ (l'écart-type du log-rendement journalier) pour MSFT.

Exercice 2 : Test de Normalité (Graphique)

La théorie log-normale repose sur la normalité des log-rendements X . Vérifions-le visuellement.

Votre tâche :

1. Utiliser μ et σ (pour MSFT) de l'Exercice 1.
2. **(Plot)** Tracer l'histogramme des log-rendements **empiriques** de MSFT (Indice : `plt.hist(..., density=True, bins=50)`).
3. **(Plot)** Superposer la PDF **théorique** de la loi normale $\mathcal{N}(\mu, \sigma^2)$ sur cet histogramme.
4. (Indice : Créez un `np.linspace`, calculez la PDF avec `norm.pdf(x, loc=mu, scale=sigma)`, puis `plt.plot()`).
5. (Conclusion) La cloche théorique s'ajuste-t-elle bien aux données réelles ?

Exercice 3 : Asymétrie (Prix vs Log-Rendements)

La théorie dit que les log-rendements X sont symétriques (Normaux), mais que les prix P_t sont asymétriques à droite (Log-Normaux).

Votre tâche :

1. Calculer la moyenne et la médiane de la série des **log-rendements** de MSFT.
2. Calculer la moyenne et la médiane de la série des **prix** de MSFT (la colonne `data['MSFT']`).
3. Comparer les deux paires. Les log-rendements sont-ils symétriques (moyenne \approx médiane) ? Les prix sont-ils asymétriques (moyenne $>$ médiane) ?
4. **(Plot)** Créer deux histogrammes côte à côte (`plt.subplot`) pour visualiser la distribution des log-rendements et celle des prix.

Exercice 4 : Espérance vs Médiane (Théorique)

Soit $Y = P_t/P_{t-1} = e^X$ la variable "ratio de prix journalier". $Y \sim \text{Log-}\mathcal{N}(\mu, \sigma^2)$. Théorie : $\text{Med}(Y) = e^\mu$ et $E[Y] = e^{\mu + \sigma^2/2}$.

Votre tâche :

1. Utiliser μ et σ^2 (pour MSFT) de l'Exercice 1.
2. Calculer la médiane **théorique** $\text{Med}(Y)$.
3. Calculer l'espérance **théorique** $E[Y]$.
4. Vérifier que $E[Y] > \text{Med}(Y)$, confirmant l'asymétrie.

Exercice 5 : Espérance Théorique vs Empirique

Vérifions le calcul de $E[Y]$ de l'exercice 4 de manière empirique.

Votre tâche :

1. Créer la série Y (ratio de prix journalier) : $Y = \exp(X_{\text{msft}})$.
2. Calculer l'espérance **empirique** de Y (la moyenne de cette série Y).
3. Comparer cette valeur empirique à l'espérance **théorique** $e^{\mu+\sigma^2/2}$ calculée à l'exercice 4.

Exercice 6 : Variance Théorique vs Empirique

Théorie : $\text{Var}(Y) = (e^{\sigma^2} - 1) \cdot e^{2\mu+\sigma^2}$.

Votre tâche :

1. Utiliser μ et σ^2 (pour MSFT) de l'Exercice 1.
2. Calculer la variance **théorique** $\text{Var}(Y)$ en utilisant la formule ci-dessus.
3. Calculer la variance **empirique** de la série Y (créée à l'Ex 5).
4. Comparer les deux résultats.

Exercice 7 : Modélisation du Prix Futur (Paramètres)

Modélisons le prix de GOOG dans $t = 20$ jours ouvrés (environ 1 mois). Le prix P_{20} est log-normal si l'on suppose $P_{20} = P_0 \cdot e^{X_{20}}$, où P_0 est le prix actuel. Le log-rendement total $X_{20} = \ln(P_{20}/P_0)$ suit $X_{20} \sim \mathcal{N}(t\mu, t\sigma^2)$.

Votre tâche :

1. Estimer μ_G et σ_G^2 (journaliers) pour GOOG (similaire à l'Ex 1).
2. Définir $t = 20$.
3. Calculer $\mu_{20} = t\mu_G$ (l'espérance du log-rendement sur 20 jours).
4. Calculer $\sigma_{20}^2 = t\sigma_G^2$ (la variance du log-rendement sur 20 jours).

Exercice 8 : Calcul de Probabilité (Prix Futur)

En utilisant les paramètres μ_{20} et $\sigma_{20} = \sqrt{\sigma_{20}^2}$ de l'exercice 7 pour GOOG :

Votre tâche :

1. Calculer la probabilité que GOOG ait un rendement positif sur 20 jours.
2. On cherche $P(P_{20} > P_0) \implies P(P_{20}/P_0 > 1) \implies P(\ln(P_{20}/P_0) > \ln(1))$.
3. Calculer $P(X_{20} > 0)$.
4. (Indice : Standardiser 0 avec μ_{20} et σ_{20} , puis utiliser $1 - \Phi(z)$).
5. **(Plot)** Tracer la PDF de $X_{20} \sim \mathcal{N}(\mu_{20}, \sigma_{20}^2)$ et hachurer la zone $x > 0$.

Exercice 9 : Calcul de Probabilité (Perte > 5%)

En utilisant les paramètres μ_{20} et σ_{20} de l'exercice 7 pour GOOG :

Votre tâche :

1. Calculer la probabilité que GOOG perde plus de 5% sur 20 jours.
2. On cherche $P(P_{20} < 0.95 \times P_0) \implies P(P_{20}/P_0 < 0.95)$.
3. Calculer $P(X_{20} < \ln(0.95))$.
4. (Indice : Standardiser $\ln(0.95)$ avec μ_{20} et σ_{20} , puis utiliser $\Phi(z)$).
5. **(Plot)** Tracer la PDF de X_{20} et hachurer la zone $x < \ln(0.95)$.

Exercice 10 : Problème Inverse (Intervalle de Confiance)

Trouvons l'intervalle de 95% pour le prix de GOOG dans 20 jours. Nous cherchons les bornes y_1, y_2 telles que $P(y_1 \leq P_{20} \leq y_2) = 0.95$. On suppose un intervalle centré sur la loi normale sous-jacente (entre $z = -1.96$ et $z = +1.96$).

Votre tâche :

1. Trouver $z_{\text{inf}} = -1.96$ et $z_{\text{sup}} = +1.96$.
2. "Dé-standardiser" ces Z-scores pour trouver les log-rendements x_1 et x_2 : $x = \mu_t + z\sigma_t$

(en utilisant μ_{20} et σ_{20} de l'Ex 7).

3. Convertir ces log-rendements en ratios de prix $y = e^x$.
4. (Conclusion) L'intervalle de 95% pour le ratio de prix est $[y_1, y_2]$.

Exercice 11 : Calcul de la Médiane vs Espérance (Prix Futur)

Pour le prix de GOOG dans 20 jours, $P_{20} = P_0 \cdot Y_{20}$, où $Y_{20} \sim \text{Log-}\mathcal{N}(\mu_{20}, \sigma_{20}^2)$.

Votre tâche :

1. Calculer le ratio de prix **médian** attendu : $\text{Med}(Y_{20}) = e^{\mu_{20}}$.
2. Calculer le ratio de prix **moyen** (espérance) attendu : $E[Y_{20}] = e^{\mu_{20} + \sigma_{20}^2/2}$.
3. (Conclusion) Lequel est le plus élevé? Pourquoi est-ce important pour un investisseur?

9 Les Loïs des Grands Nombres (LLN)

Dans la section précédente, nous avons fait une distinction cruciale entre les **moments de population** (les "vraies" valeurs théoriques, inconnues, comme μ et σ^2) et les **moments d'échantillon** (nos estimations calculées à partir des données, comme \bar{X} et s^2).

Par exemple, nous avons défini la moyenne d'échantillon $\bar{X} = \frac{1}{n} \sum X_i$ comme notre "meilleure estimation" de la moyenne de population μ . Mais qu'est-ce qui nous garantit que cette estimation est "bonne"? Qu'est-ce qui nous assure que si nous collectons plus de données (en augmentant n), notre \bar{X} se rapprocherait de μ ?

La réponse à cette question fondamentale est fournie par les **Lois des Grands Nombres (LLN)**. Elles forment le pont théorique entre les probabilités (la théorie) et les statistiques (la pratique).

Intuition : L'Idée Fondamentale : L'Exemple du Dé

Supposons que nous voulons connaître la valeur moyenne d'un lancer de dé équilibré.

- **Moment de Population** : Nous savons par la théorie que $\mu = E[X] = \frac{1+2+3+4+5+6}{6} = 3.5$.
- **Moments d'Échantillon** : Nous n'avons pas cette information, alors nous lançons le dé.
 - $n = 2$ lancers : On obtient (2, 6). $\bar{X}_2 = (2+6)/2 = 4.0$. (Assez loin de 3.5)
 - $n = 10$ lancers : On obtient (1, 6, 3, 3, 5, 2, 4, 1, 6, 4). $\bar{X}_{10} = 3.5$. (Pile dessus!)
 - $n = 100$ lancers : On obtiendra $\bar{X}_{100} \approx 3.48$ (par exemple).
 - $n = 1,000,000$ lancers : On obtiendra $\bar{X}_{1,000,000} \approx 3.5001$ (par exemple).

La Loi des Grands Nombres formalise cette intuition : à mesure que $n \rightarrow \infty$, la moyenne de notre échantillon \bar{X}_n **converge** vers la vraie moyenne μ .

La distinction entre les lois "Faible" et "Forte" réside dans la *manière* dont nous définissons cette convergence.

9.1 L'Inégalité de Chebyshev

Avant de prouver la Loi Faible, nous avons besoin d'un outil fondamental qui relie la variance d'une variable à la probabilité qu'elle s'éloigne de sa moyenne. C'est l'Inégalité de Chebyshev.

Sa puissance réside dans son universalité : elle s'applique à *n'importe quelle* distribution, à condition qu'elle ait une moyenne et une variance finies.

Théorème : Inégalité de Chebyshev

Soit Y une variable aléatoire avec une espérance finie $\mu = E[Y]$ et une variance finie $\sigma^2 = \text{Var}(Y)$.

Alors, pour tout nombre réel $k > 0$:

$$P(|Y - \mu| \geq k) \leq \frac{\text{Var}(Y)}{k^2} = \frac{\sigma^2}{k^2}$$

Preuve : Preuve de l'Inégalité de Chebyshev

Nous présentons la preuve pour une variable aléatoire continue Y de densité $f(y)$. La preuve pour le cas discret est similaire en remplaçant les intégrales par des sommes.

1. Par définition, la variance σ^2 est $E[(Y - \mu)^2]$:

$$\sigma^2 = E[(Y - \mu)^2] = \int_{-\infty}^{\infty} (y - \mu)^2 f(y) dy$$

2. Nous pouvons scinder cette intégrale en deux parties : la région où Y est proche de μ ($|y - \mu| < k$) et la région où Y est loin de μ ($|y - \mu| \geq k$) :

$$\sigma^2 = \int_{|y-\mu|<k} (y - \mu)^2 f(y) dy + \int_{|y-\mu|\geq k} (y - \mu)^2 f(y) dy$$

3. L'intégrande $(y - \mu)^2 f(y)$ est toujours non-négative (un carré fois une densité). Par conséquent, la première intégrale est ≥ 0 . En la supprimant, nous ne pouvons que diminuer la valeur totale :

$$\sigma^2 \geq \int_{|y - \mu| \geq k} (y - \mu)^2 f(y) dy$$

4. Maintenant, concentrons-nous sur la région d'intégration : $|y - \mu| \geq k$. Dans cette région, par définition, nous avons $(y - \mu)^2 \geq k^2$.
5. Nous pouvons remplacer $(y - \mu)^2$ par k^2 dans l'intégrale. Puisque nous remplaçons un terme par quelque chose de plus petit ou égal, la valeur de l'intégrale diminue (ou reste égale) :

$$\sigma^2 \geq \int_{|y - \mu| \geq k} k^2 f(y) dy$$

6. k^2 est une constante, nous pouvons la sortir de l'intégrale :

$$\sigma^2 \geq k^2 \int_{|y - \mu| \geq k} f(y) dy$$

7. Par définition, l'intégrale de la densité $f(y)$ sur la région $|y - \mu| \geq k$ n'est autre que la probabilité $P(|Y - \mu| \geq k)$.

$$\sigma^2 \geq k^2 \cdot P(|Y - \mu| \geq k)$$

8. En réarrangeant les termes (puisque $k > 0$, $k^2 > 0$), nous obtenons l'inégalité désirée :

$$P(|Y - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

Cette preuve est un cas particulier de l'Inégalité de Markov (appliquée à la variable aléatoire non-négative $X = (Y - \mu)^2$ et à la constante $a = k^2$).

Intuition : Comprendre l'Inégalité de Chebyshev

Cette formule peut être lue comme suit :

"La probabilité de s'écarter de la moyenne (μ) d'au moins k est bornée par la variance divisée par k^2 ."

- **Le rôle de la variance (σ^2) :** Si la variance est grande, la borne supérieure est élevée. L'inégalité nous dit "il est possible que la variable s'éloigne", ce qui est logique pour une grande dispersion. Si la variance est faible, la borne est basse, ce qui force la probabilité d'être loin à être faible.
- **Le rôle de l'écart (k) :** Le terme k^2 au dénominateur est crucial. Il signifie que la probabilité de s'écarter de la moyenne diminue *quadratiquement* avec la distance k . Être très loin est (relativement) très improbable.

Exemple : Une Borne Universelle

Exprimons l'inégalité en termes d'écarts-types (en posant $k = c \cdot \sigma$) :

$$P(|Y - \mu| \geq c\sigma) \leq \frac{\sigma^2}{(c\sigma)^2} = \frac{1}{c^2}$$

- **Pour $c = 2$:** $P(|Y - \mu| \geq 2\sigma) \leq \frac{1}{4} = 25\%$. Peu importe la distribution (symétrique, asymétrique, bizarre...), la probabilité d'être à 2 écarts-types ou plus de la moyenne est **au maximum** de 25%. (Pour une loi normale, cette probabilité est bien plus faible, $\approx 4.55\%$).
- **Pour $c = 3$:** $P(|Y - \mu| \geq 3\sigma) \leq \frac{1}{9} \approx 11.1\%$. La probabilité d'être à 3 écarts-types ou plus est au maximum de 11.1%. (Pour une loi normale, c'est $\approx 0.27\%$).

Chebyshev fournit une borne "garantie", bien que souvent non optimale. Elle est l'outil parfait pour la preuve qui suit.

9.2 La Loi Faible des Grands Nombres (LFGN / WLLN)

La loi faible stipule que la probabilité que notre moyenne d'échantillon s'écarte de la vraie moyenne de plus qu'une petite quantité ϵ tend vers zéro. C'est une **convergence en probabilité**.

Définition : Convergence en Probabilité

On dit qu'une suite de variables aléatoires Y_n converge en probabilité vers une constante c , noté $Y_n \xrightarrow{P} c$, si pour tout $\epsilon > 0$ (aussi petit soit-il) :

$$\lim_{n \rightarrow \infty} P(|Y_n - c| > \epsilon) = 0$$

Intuition : Comprendre la Convergence en Probabilité

La définition $P(|\bar{X}_n - \mu| > \epsilon) \rightarrow 0$ signifie :

- ϵ est votre **marge d'erreur** acceptable (ex : 0.01).
- $|\bar{X}_n - \mu|$ est l'erreur réelle de votre estimation.
- $P(\dots)$ est la probabilité que votre erreur **dépasse** votre marge.
- $\lim_{n \rightarrow \infty}(\dots) = 0$ signifie : "Si vous prenez un échantillon n suffisamment grand, la probabilité de faire une erreur plus grande que ϵ devient négligeable."

C'est une affirmation sur ce qui se passe pour un n fixe et très grand.

Théorème : Loi Faible des Grands Nombres (Khinchine)

Soit X_1, X_2, \dots, X_n une suite de variables aléatoires **i.i.d.** (indépendantes et identiquement distribuées) avec une espérance finie $E[X_i] = \mu$. Soit $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ la moyenne d'échantillon. Alors, \bar{X}_n converge en probabilité vers μ :

$$\bar{X}_n \xrightarrow{P} \mu$$

Preuve : Preuve (simplifiée) via l'Inégalité de Chebyshev

La loi faible de Khinchine ne nécessite qu'une moyenne finie. Cependant, si nous ajoutons la condition que la **variance** σ^2 est **aussi finie**, la preuve devient très simple.

Elle repose directement sur l'Inégalité de Chebyshev, que nous venons de voir. Nous l'appliquons à la variable aléatoire $Y = \bar{X}_n$.

1. Identifions les termes pour l'inégalité $P(|Y - E[Y]| \geq k) \leq \frac{\text{Var}(Y)}{k^2}$:
 - Notre variable est $Y = \bar{X}_n$.
 - Son espérance est $E[Y] = E[\bar{X}_n] = \mu$.
 - Sa variance est $\text{Var}(Y) = \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$.
 - Notre écart k est la marge d'erreur ϵ .
2. (Rappel du calcul de la variance de \bar{X}_n) : Puisque les X_i sont i.i.d., $\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n} \sum X_i\right) = \frac{1}{n^2} \sum \text{Var}(X_i) = \frac{1}{n^2} (n\sigma^2) = \frac{\sigma^2}{n}$.
3. Appliquons l'inégalité de Chebyshev avec ces termes :

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2/n}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

4. Prenons maintenant la limite quand $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \epsilon) \leq \lim_{n \rightarrow \infty} \frac{\sigma^2}{n\epsilon^2}$$

5. Puisque σ^2 et ϵ^2 sont des constantes finies, le terme de droite $\frac{\text{constante}}{n}$ tend vers 0.
6. Comme une probabilité ne peut pas être négative, nous avons :

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > \epsilon) = 0$$

C'est exactement la définition de la convergence en probabilité.

9.3 La Loi Forte des Grands Nombres (LFGN / SLLN)

La loi forte est une affirmation beaucoup plus puissante. Elle ne dit pas seulement qu'un "gros" écart est improbable pour un n "grand" ; elle dit que la probabilité que la suite \bar{X}_n ne converge pas vers μ est nulle. C'est une **convergence presque sûre**.

Définition : Convergence Presque Sûre

On dit qu'une suite de variables aléatoires Y_n converge presque sûrement vers une constante c , noté $Y_n \xrightarrow{p.s.} c$, si :

$$P\left(\lim_{n \rightarrow \infty} Y_n = c\right) = 1$$

Théorème : Loi Forte des Grands Nombres (Kolmogorov)

Soit X_1, X_2, \dots, X_n une suite de variables aléatoires **i.i.d.** avec une espérance finie $E[X_i] = \mu$. Alors, \bar{X}_n converge presque sûrement vers μ :

$$\bar{X}_n \xrightarrow{p.s.} \mu$$

Remarque : Forte implique Faible

La convergence "presque sûre" (SLLN) est une condition plus stricte que la convergence "en probabilité" (WLLN). Si une suite converge presque sûrement, elle converge aussi en probabilité. L'inverse n'est pas toujours vrai.

9.4 Différence : Faible vs. Forte

Intuition : Faible vs. Forte : L'Analogie du Casino

Soit \bar{X}_n votre gain moyen par partie après avoir joué n fois à la roulette. La vraie moyenne (l'avantage de la maison) est $\mu = -0.053$ (pour une roulette américaine).

- **Loi Faible (WLLN)** : "Si vous prévoyez de jouer $n = 1$ million de parties ce soir. La probabilité qu'à la fin de votre millionième partie, votre moyenne $\bar{X}_{1,000,000}$ soit loin de -0.053 (par exemple, que vous soyez gagnant, $\bar{X}_n > 0$) est infinitésimale."
- C'est une affirmation sur la distribution de \bar{X}_n à un point fixe n (très grand). Elle n'exclut pas la possibilité théorique (mais improbable) que si vous continuez à jouer, votre moyenne \bar{X}_n puisse à nouveau diverger follement avant de reconverger plus tard.
- **Loi Forte (SLLN)** : "Si vous jouez à la roulette pour l'éternité, en regardant la séquence de vos moyennes $\bar{X}_1, \bar{X}_2, \bar{X}_3, \dots, \bar{X}_n, \dots$ "
- "La probabilité que cette **séquence entière** ne converge pas exactement vers $\mu = -0.053$ est de 0."
- C'est une affirmation sur la **trajectoire complète**. Elle dit que, avec une probabilité de 1, la trajectoire de \bar{X}_n va s'approcher de μ et **ne plus s'en écarter** de manière significative.

En résumé :

- **Faible** : Pour n assez grand, un écart est **improbable**.
- **Forte** : La **trajectoire** converge vers μ (avec une probabilité de 1).

9.5 Application : La Méthode de Monte-Carlo

La Loi Forte des Grands Nombres est le moteur de l'une des techniques de calcul les plus puissantes : la simulation de Monte-Carlo. Elle nous permet d'estimer des quantités complexes (comme des intégrales) en utilisant le hasard.

Exemple : Estimer la valeur de π

Problème : Comment calculer π sans formule géométrique ?

Méthode (Statistique) :

1. Imaginez un carré de côté 1 (de $(0, 0)$ à $(1, 1)$). Son aire est $A_{\text{carré}} = 1$.

2. Imaginez un quart de cercle de rayon $r = 1$ inscrit dans ce carré. Son aire est $A_{\text{cercle}} = \frac{1}{4}\pi r^2 = \frac{\pi}{4}$.
3. Le *ratio* des aires est $\frac{A_{\text{cercle}}}{A_{\text{carré}}} = \frac{\pi/4}{1} = \frac{\pi}{4}$.

Simulation :

1. Nous allons "lancer des fléchettes" au hasard sur ce carré n fois.
2. Pour ce faire, nous générons n paires de nombres aléatoires (X_i, Y_i) , où $X_i \sim U(0, 1)$ et $Y_i \sim U(0, 1)$.
3. Pour chaque point i , nous vérifions s'il a atterri **dans le cercle**. La condition est $X_i^2 + Y_i^2 \leq 1$.
4. Nous définissons une nouvelle variable aléatoire Z_i (de Bernoulli) :

$$Z_i = \begin{cases} 1 & \text{si } X_i^2 + Y_i^2 \leq 1 \quad (\text{le point est dans le cercle}) \\ 0 & \text{sinon} \end{cases}$$

Application de la LLN :

- Quelle est la "vraie moyenne" μ de cette variable Z_i ?
- $\mu = E[Z_i] = 1 \cdot P(Z_i = 1) + 0 \cdot P(Z_i = 0) = P(Z_i = 1)$.
- $P(Z_i = 1)$ est la probabilité qu'un point aléatoire tombe dans le cercle. Puisque les points sont uniformes, cette probabilité est simplement le ratio des aires !
- Donc, la vraie moyenne (inconnue) est $\mu = \frac{A_{\text{cercle}}}{A_{\text{carré}}} = \frac{\pi}{4}$.
- Comment estimer μ ? Nous utilisons la moyenne d'échantillon \bar{Z}_n :

$$\bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_i = \frac{\text{Nombre de points dans le cercle}}{n}$$

- Par la **Loi Forte des Grands Nombres**, nous avons la garantie que :

$$\bar{Z}_n \xrightarrow{p.s.} \mu = \frac{\pi}{4}$$

Conclusion : Pour estimer π , il suffit de calculer \bar{Z}_n (une simple proportion) et de la multiplier par 4.

$$\pi \approx 4 \cdot \bar{Z}_n$$

Plus notre nombre de simulations n est grand, plus la SLLN nous garantit que notre estimation sera proche de la vraie valeur de π .

9.6 Exercices

Pour tous les exercices de calcul, vous pouvez utiliser les valeurs (arrondies) suivantes pour la fonction de répartition de la loi normale standard $\Phi(z) = P(Z \leq z)$:

- $\Phi(0) = 0.5$
- $\Phi(0.08) \approx 0.5319$
- $\Phi(0.1) \approx 0.5398$
- $\Phi(1.0) \approx 0.8413$
- $\Phi(1.5) \approx 0.9332$
- $\Phi(1.58) \approx 0.9429$
- $\Phi(1.645) \approx 0.95$
- $\Phi(1.75) \approx 0.9599$
- $\Phi(1.96) \approx 0.975$
- $\Phi(2.0) \approx 0.9772$
- $\Phi(2.33) \approx 0.99$
- $\Phi(2.5) \approx 0.9938$

- $\Phi(3.0) \approx 0.9987$
- $\Phi(10.0) \approx 1.0$

Et rappelez-vous la propriété de symétrie : $\Phi(-z) = 1 - \Phi(z)$.

Exercice 1 : Chebyshev (Calcul de base)

Une variable aléatoire Y a une moyenne $\mu = 50$ et une variance $\sigma^2 = 16$. Utilisez l'inégalité de Chebyshev pour trouver une borne supérieure à $P(|Y - 50| \geq 10)$.

Exercice 2 : Chebyshev (En termes d'écarts-types)

Quelle est la borne *universelle* (valable pour toute distribution) pour la probabilité qu'une variable aléatoire s'écarte de sa moyenne de plus de 5 écarts-types ?

Exercice 3 : Chebyshev (Application à \bar{X}_n)

Soit X_i une suite de v.a. i.i.d. avec $\mu = 100$ et $\sigma^2 = 400$. Soit $n = 100$.

1. Calculez $E[\bar{X}_{100}]$ et $\text{Var}(\bar{X}_{100})$.
2. En utilisant Chebyshev, bornez $P(|\bar{X}_{100} - 100| \geq 4)$.

Exercice 4 : CLT (Distribution de \bar{X}_n)

On prélève un échantillon de $n = 64$ observations d'une population de moyenne $\mu = 20$ et de variance $\sigma^2 = 16$. Quelle est la distribution approximative de la moyenne d'échantillon \bar{X}_{64} selon le TCL ?

Exercice 5 : CLT (Calcul de probabilité pour \bar{X}_n)

En utilisant les informations de l'exercice 4, calculez $P(\bar{X}_{64} \leq 21)$.

Exercice 6 : CLT (Calcul de probabilité pour \bar{X}_n)

On prélève $n = 100$ observations d'une population de moyenne $\mu = 50$ et d'écart-type $\sigma = 5$. Calculez $P(49 \leq \bar{X}_{100} \leq 51)$.

Exercice 7 : CLT (Inverse pour \bar{X}_n)

Soit \bar{X}_n la moyenne de $n = 36$ v.a. i.i.d. de moyenne $\mu = 10$ et de variance $\sigma^2 = 81$. Trouvez la valeur c telle que $P(\bar{X}_{36} \leq c) \approx 0.9772$.

Exercice 8 : CLT (Distribution de S_n)

On prélève un échantillon de $n = 100$ observations d'une population de moyenne $\mu = 10$ et de variance $\sigma^2 = 4$. Quelle est la distribution approximative de la somme $S_{100} = \sum X_i$ selon le TCL ?

Exercice 9 : CLT (Calcul de probabilité pour S_n)

En utilisant les informations de l'exercice 8, calculez $P(S_{100} > 1020)$.

Exercice 10 : CLT (Application S_n)

Un ascenseur doit transporter $n = 49$ personnes. Le poids de chaque personne est une v.a. i.i.d. de moyenne $\mu = 70\text{kg}$ et d'écart-type $\sigma = 14\text{kg}$. Calculez la probabilité que le poids total S_{49} dépasse 3500 kg.

Exercice 11 : CLT (Inverse pour S_n)

Soit S_n la somme de $n = 64$ v.a. i.i.d. de moyenne $\mu = 5$ et de variance $\sigma^2 = 1$. Trouvez la valeur c telle que $P(S_{64} > c) \approx 0.05$.

Exercice 12 : Règle d'Approximation

Peut-on approximer une loi $X \sim \text{Bin}(n = 40, p = 0.1)$ par une loi normale en utilisant la règle $np \geq 10$ et $n(1 - p) \geq 10$?

Exercice 13 : Règle d'Approximation

Peut-on approximer une loi $X \sim \text{Bin}(n = 500, p = 0.04)$ par une loi normale en utilisant la règle $np \geq 10$ et $n(1 - p) \geq 10$?

Exercice 14 : Paramètres d'Approximation

Soit $X \sim \text{Bin}(100, 0.3)$. On souhaite l'approximer par $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. Calculez μ_Y et σ_Y^2 .

Exercice 15 : Correction de Continuité (Règles)

Soit X une variable binomiale (discrète) et Y son approximation normale (continue). Traduisez les probabilités discrètes suivantes en probabilités continues :

1. $P(X = 50)$
2. $P(X \geq 50)$
3. $P(X \leq 49)$
4. $P(X < 49)$

Exercice 16 : Correction de Continuité (Règles)

Traduisez les probabilités discrètes suivantes en probabilités continues :

1. $P(X > 30)$
2. $P(30 < X < 40)$
3. $P(30 \leq X \leq 40)$

Exercice 17 : Binomiale (Calcul $P(X$

On lance une pièce équilibrée 100 fois. Soit X le nombre de "Pile". Utilisez l'approximation normale (avec correction de continuité) pour estimer $P(X = 50)$. (C'est l'exemple du texte).

Exercice 18 : Binomiale (Calcul $P(X \geq k)$)

On lance une pièce équilibrée 100 fois ($X \sim \text{Bin}(100, 0.5)$). Estimez $P(X \geq 60)$.

Exercice 19 : Binomiale (Calcul $P(X \leq k)$)

Un traitement a un taux de succès de $p = 0.2$. On l'administre à $n = 400$ patients. Soit X le nombre de succès. Estimez $P(X \leq 70)$.

Exercice 20 : Binomiale (Calcul $P(X > k)$)

En utilisant la situation de l'exercice 19 ($X \sim \text{Bin}(400, 0.2)$), estimez $P(X > 92)$.

Exercice 21 : Binomiale (Calcul $P(k_1 \leq X \leq k_2)$)

Un sondage est mené auprès de $n = 100$ personnes. On suppose que $p = 0.6$ est la probabilité qu'une personne soutienne une mesure. Soit X le nombre de supporters. Estimez $P(50 \leq X \leq 65)$.

Exercice 22 : TCL vs Chebyshev

Soit \bar{X}_{100} la moyenne de $n = 100$ v.a. i.i.d. de moyenne $\mu = 10$ et $\sigma^2 = 25$. Calculez $P(|\bar{X}_{100} - 10| \geq 1)$ en utilisant :

1. L'inégalité de Chebyshev.
2. Le Théorème Central Limite.

Exercice 23 : Taille d'Échantillon (CLT)

Soient X_i des v.a. i.i.d. avec $\mu = 0$ et $\sigma^2 = 1$. Combien d'échantillons n faut-il pour garantir que $P(|\bar{X}_n| \leq 0.1) \geq 0.95$? (Indice : $P(-1.96 \leq Z \leq 1.96) = 0.95$).

Exercice 24 : LLN vs CLT

Considérons \bar{X}_n pour des X_i i.i.d. avec $\mu = 10, \sigma^2 = 100$.

1. Calculez $P(9 \leq \bar{X}_n \leq 11)$ pour $n = 100$.
2. Calculez $P(9 \leq \bar{X}_n \leq 11)$ pour $n = 10000$.

3. Comment ce résultat illustre-t-il la LLN ?

Exercice 25 : Binomiale (Sans Correction)

Soit $X \sim \text{Bin}(400, 0.1)$. Estimez $P(X \leq 30.5)$ sans utiliser la correction de continuité (c'est-à-dire en approximant $P(X \leq 30.5) \approx P(Y \leq 30.5)$). (Ceci permet de comparer avec l'exercice 19).

9.7 Corrections des Exercices

Correction Exercice 1 : Chebyshev (Calcul de base)

$P(|Y - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$. $P(|Y - 50| \geq 10) \leq \frac{16}{10^2} = \frac{16}{100} = 0.16$. La probabilité est d'au maximum 16%.

Correction Exercice 2 : Chebyshev (En termes d'écart-types)

On pose $k = 5\sigma$. $P(|Y - \mu| \geq 5\sigma) \leq \frac{\sigma^2}{(5\sigma)^2} = \frac{\sigma^2}{25\sigma^2} = \frac{1}{25} = 0.04$. La probabilité est d'au maximum 4%.

Correction Exercice 3 : Chebyshev (Application à \bar{X}_n)

1. $E[\bar{X}_{100}] = \mu = 100$. $\text{Var}(\bar{X}_{100}) = \frac{\sigma^2}{n} = \frac{400}{100} = 4$. 2. On applique Chebyshev à \bar{X}_{100} (avec $\mu = 100, \sigma^2 = 4$) et $k = 4$. $P(|\bar{X}_{100} - 100| \geq 4) \leq \frac{\text{Var}(\bar{X}_{100})}{k^2} = \frac{4}{4^2} = \frac{4}{16} = 0.25$.

Correction Exercice 4 : CLT (Distribution de \bar{X}_n)

$E[\bar{X}_{64}] = \mu = 20$. $\text{Var}(\bar{X}_{64}) = \frac{\sigma^2}{n} = \frac{16}{64} = 0.25$. Selon le TCL, $\bar{X}_{64} \approx \mathcal{N}(20, 0.25)$.

Correction Exercice 5 : CLT (Calcul de probabilité pour \bar{X}_n)

On standardise $\bar{X}_{64} \approx \mathcal{N}(20, 0.25)$, donc $\sigma_{\bar{X}} = \sqrt{0.25} = 0.5$. On cherche $P(\bar{X}_{64} \leq 21)$. $Z = \frac{\bar{X}_n - \mu}{\sigma_{\bar{X}}} = \frac{21 - 20}{0.5} = \frac{1}{0.5} = 2$. $P(\bar{X}_{64} \leq 21) = P(Z \leq 2) = \Phi(2) \approx 0.9772$.

Correction Exercice 6 : CLT (Calcul de probabilité pour \bar{X}_n)

$\mu = 50, \sigma = 5, n = 100$. $\bar{X}_{100} \approx \mathcal{N}(\mu, \sigma^2/n)$. $\mu_{\bar{X}} = 50$. $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 5/\sqrt{100} = 0.5$. On cherche $P(49 \leq \bar{X}_{100} \leq 51)$. $Z_1 = \frac{49 - 50}{0.5} = -2$. $Z_2 = \frac{51 - 50}{0.5} = 2$. $P(-2 \leq Z \leq 2) = \Phi(2) - \Phi(-2) = \Phi(2) - (1 - \Phi(2)) = 2\Phi(2) - 1$. $P \approx 2(0.9772) - 1 = 1.9544 - 1 = 0.9544$. (La règle des 95%).

Correction Exercice 7 : CLT (Inverse pour \bar{X}_n)

$\mu = 10, \sigma^2 = 81, n = 36$. $\bar{X}_{36} \approx \mathcal{N}(\mu, \sigma^2/n)$. $\mu_{\bar{X}} = 10$. $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 9/\sqrt{36} = 9/6 = 1.5$. On cherche c tel que $P(\bar{X}_{36} \leq c) \approx 0.9772$. $P(Z \leq \frac{c - 10}{1.5}) = 0.9772$. D'après la table, le z correspondant est 2.0. $\frac{c - 10}{1.5} = 2 \implies c - 10 = 3 \implies c = 13$.

Correction Exercice 8 : CLT (Distribution de S_n)

$E[S_{100}] = n\mu = 100 \times 10 = 1000$. $\text{Var}(S_{100}) = n\sigma^2 = 100 \times 4 = 400$. Selon le TCL, $S_{100} \approx \mathcal{N}(1000, 400)$.

Correction Exercice 9 : CLT (Calcul de probabilité pour S_n)

On standardise $S_{100} \approx \mathcal{N}(1000, 400)$, donc $\sigma_{S_n} = \sqrt{400} = 20$. On cherche $P(S_{100} > 1020)$. $Z = \frac{S_n - n\mu}{\sigma_{S_n}} = \frac{1020 - 1000}{20} = \frac{20}{20} = 1$. $P(S_{100} > 1020) = P(Z > 1) = 1 - \Phi(1) \approx 1 - 0.8413 = 0.1587$.

Correction Exercice 10 : CLT (Application S_n)

$n = 49, \mu = 70, \sigma = 14$. $S_{49} \approx \mathcal{N}(n\mu, n\sigma^2)$. $E[S_{49}] = n\mu = 49 \times 70 = 3430$. $\sigma_{S_n} = \sigma\sqrt{n} = 14 \times \sqrt{49} = 14 \times 7 = 98$. On cherche $P(S_{49} > 3500)$. $Z = \frac{3500 - 3430}{98} = \frac{70}{98} \approx 0.714$. $P(Z > 0.714)$. (Valeur non fournie, mais le calcul est le Z-score).

Correction Exercice 11 : CLT (Inverse pour S_n)

$n = 64, \mu = 5, \sigma^2 = 1$. $S_{64} \approx \mathcal{N}(n\mu, n\sigma^2)$. $E[S_{64}] = 64 \times 5 = 320$. $\sigma_{S_n} = \sigma\sqrt{n} = 1 \times \sqrt{64} = 8$. On cherche c tel que $P(S_{64} > c) \approx 0.05$. $P(Z > \frac{c-320}{8}) = 0.05$. $P(Z \leq z) = 0.95$. D'après la table, $z \approx 1.645$. $\frac{c-320}{8} = 1.645 \implies c = 320 + 8(1.645) = 320 + 13.16 = 333.16$.

Correction Exercice 12 : Règle d'Approximation

$n = 40, p = 0.1$. $np = 40 \times 0.1 = 4$. $n(1-p) = 40 \times 0.9 = 36$. Puisque $np = 4$ est < 10 , la règle n'est pas satisfaite. L'approximation normale n'est pas recommandée.

Correction Exercice 13 : Règle d'Approximation

$n = 500, p = 0.04$. $np = 500 \times 0.04 = 20$. $n(1-p) = 500 \times 0.96 = 480$. Les deux conditions ($20 \geq 10$ et $480 \geq 10$) sont satisfaites. L'approximation est valide.

Correction Exercice 14 : Paramètres d'Approximation

$X \sim \text{Bin}(100, 0.3)$. $\mu_Y = np = 100 \times 0.3 = 30$. $\sigma_Y^2 = np(1-p) = 100 \times 0.3 \times 0.7 = 21$. L'approximation est $Y \sim \mathcal{N}(30, 21)$.

Correction Exercice 15 : Correction de Continuité (Règles)

1. $P(X = 50) \approx P(49.5 \leq Y \leq 50.5)$ 2. $P(X \geq 50) \approx P(Y \geq 49.5)$ 3. $P(X \leq 49) \approx P(Y \leq 49.5)$ 4. $P(X < 49) = P(X \leq 48) \approx P(Y \leq 48.5)$

Correction Exercice 16 : Correction de Continuité (Règles)

1. $P(X > 30) = P(X \geq 31) \approx P(Y \geq 30.5)$ 2. $P(30 < X < 40) = P(31 \leq X \leq 39) \approx P(30.5 \leq Y \leq 39.5)$ 3. $P(30 \leq X \leq 40) \approx P(29.5 \leq Y \leq 40.5)$

Correction Exercice 17 : Binomiale (Calcul $P(X)$)

$X \sim \text{Bin}(100, 0.5)$. $\mu = 50, \sigma^2 = 25, \sigma = 5$. On cherche $P(X = 50) \approx P(49.5 \leq Y \leq 50.5)$. $Z_1 = \frac{49.5-50}{5} = -0.1$. $Z_2 = \frac{50.5-50}{5} = 0.1$. $P(-0.1 \leq Z \leq 0.1) = \Phi(0.1) - \Phi(-0.1) = \Phi(0.1) - (1 - \Phi(0.1)) = 2\Phi(0.1) - 1$. $P \approx 2(0.5398) - 1 = 1.0796 - 1 = 0.0796$.

Correction Exercice 18 : Binomiale (Calcul $P(X \geq k)$)

$X \sim \text{Bin}(100, 0.5)$. $\mu = 50, \sigma = 5$. On cherche $P(X \geq 60) \approx P(Y \geq 59.5)$. $Z = \frac{59.5-50}{5} = \frac{9.5}{5} = 1.9$. $P(Z \geq 1.9) = 1 - \Phi(1.9)$. (Valeur $\Phi(1.9)$ non fournie, mais $1 - \Phi(1.96) \approx 0.025$).

Correction Exercice 19 : Binomiale (Calcul $P(X \leq k)$)

$X \sim \text{Bin}(400, 0.2)$. $np = 80, n(1-p) = 320$. (Règle OK). $\mu = 80, \sigma^2 = 80 \times 0.8 = 64, \sigma = 8$. On cherche $P(X \leq 70) \approx P(Y \leq 70.5)$. $Z = \frac{70.5-80}{8} = \frac{-9.5}{8} \approx -1.19$. $P(Z \leq -1.19) = 1 - \Phi(1.19)$. (Valeur non fournie).

Correction Exercice 20 : Binomiale (Calcul $P(X > k)$)

$X \sim \text{Bin}(400, 0.2)$. $\mu = 80, \sigma = 8$. On cherche $P(X > 92) = P(X \geq 93) \approx P(Y \geq 92.5)$. $Z = \frac{92.5-80}{8} = \frac{12.5}{8} = 1.5625$. $P(Z \geq 1.5625) \approx P(Z \geq 1.58) = 1 - \Phi(1.58) \approx 1 - 0.9429 = 0.0571$.

Correction Exercice 21 : Binomiale (Calcul $P(k_1 \leq X \leq k_2)$)

$X \sim \text{Bin}(100, 0.6)$. $np = 60, n(1-p) = 40$. (Règle OK). $\mu = 60, \sigma^2 = 60 \times 0.4 = 24, \sigma = \sqrt{24} \approx 4.9$. On cherche $P(50 \leq X \leq 65) \approx P(49.5 \leq Y \leq 65.5)$. $Z_1 = \frac{49.5-60}{4.9} = \frac{-10.5}{4.9} \approx -2.14$. $Z_2 = \frac{65.5-60}{4.9} = \frac{5.5}{4.9} \approx 1.12$. $P(-2.14 \leq Z \leq 1.12) = \Phi(1.12) - \Phi(-2.14) = \Phi(1.12) - (1 - \Phi(2.14))$. (Valeurs non fournies).

Correction Exercice 22 : TCL vs Chebyshev

$n = 100, \mu = 10, \sigma^2 = 25$. $\bar{X}_{100} \approx \mathcal{N}(10, 25/100 = 0.25)$. $\sigma_{\bar{X}} = 0.5$. On cherche $P(|\bar{X}_{100} - 10| \geq 1)$. 1. **Chebyshev** : $k = 1$, $\text{Var}(\bar{X}_{100}) = 0.25$. $P \leq \frac{\text{Var}(\bar{X}_{100})}{k^2} = \frac{0.25}{1^2} = 0.25$. (Borne $\leq 25\%$). 2. **CLT** : $P(|\bar{X}_{100} - 10| \geq 1) = P(Z \geq \frac{1}{0.5}) + P(Z \leq \frac{-1}{0.5}) = P(Z \geq 2) + P(Z \leq -2)$. $P = (1 - \Phi(2)) + \Phi(-2) = (1 - \Phi(2)) + (1 - \Phi(2)) = 2(1 - \Phi(2))$. $P \approx 2(1 - 0.9772) = 2(0.0228) = 0.0456$. (Probabilité $\approx 4.56\%$).

Correction Exercice 23 : Taille d'Échantillon (CLT)

$\bar{X}_n \approx \mathcal{N}(\mu, \sigma^2/n) = \mathcal{N}(0, 1/n)$. $\sigma_{\bar{X}} = 1/\sqrt{n}$. On veut $P(|\bar{X}_n| \leq 0.1) \geq 0.95$, soit $P(-0.1 \leq \bar{X}_n \leq 0.1) \geq 0.95$. On standardise : $P\left(\frac{-0.1-0}{1/\sqrt{n}} \leq Z \leq \frac{0.1-0}{1/\sqrt{n}}\right) \geq 0.95$. $P(-0.1\sqrt{n} \leq Z \leq 0.1\sqrt{n}) \geq 0.95$. On sait que $P(-1.96 \leq Z \leq 1.96) = 0.95$. On doit donc avoir $0.1\sqrt{n} \geq 1.96$. $\sqrt{n} \geq 19.6 \implies n \geq (19.6)^2 = 384.16$. Il faut $n = 385$ échantillons au minimum.

Correction Exercice 24 : LLN vs CLT

$\mu = 10, \sigma^2 = 100$. 1. $n = 100$: $\bar{X}_{100} \approx \mathcal{N}(10, 100/100 = 1)$. $\sigma_{\bar{X}} = 1$. $P(9 \leq \bar{X}_{100} \leq 11) = P\left(\frac{9-10}{1} \leq Z \leq \frac{11-10}{1}\right) = P(-1 \leq Z \leq 1)$. $P = 2\Phi(1) - 1 \approx 2(0.8413) - 1 = 0.6826$. 2. $n = 10000$: $\bar{X}_{10000} \approx \mathcal{N}(10, 100/10000 = 0.01)$. $\sigma_{\bar{X}} = 0.1$. $P(9 \leq \bar{X}_{10000} \leq 11) = P\left(\frac{9-10}{0.1} \leq Z \leq \frac{11-10}{0.1}\right) = P(-10 \leq Z \leq 10)$. $P \approx \Phi(10) - \Phi(-10) \approx 1 - (1 - 1) = 1$. 3. **Illustration** : Le CLT montre *comment* la LLN fonctionne. En augmentant n , la variance de \bar{X}_n s'effondre (de 1 à 0.01), forçant la distribution de \bar{X}_n à se concentrer massivement autour de $\mu = 10$. La probabilité que \bar{X}_n soit proche de μ tend vers 1.

Correction Exercice 25 : Binomiale (Sans Correction)

$X \sim \text{Bin}(400, 0.1)$. On approxime $P(X \leq 30.5)$. $\mu = np = 40$. $\sigma^2 = np(1-p) = 36$. $\sigma = 6$. $Y \sim \mathcal{N}(40, 36)$. On cherche $P(Y \leq 30.5)$. $Z = \frac{30.5-40}{6} = \frac{-9.5}{6} \approx -1.583$. $P(Z \leq -1.58) = \Phi(-1.58) = 1 - \Phi(1.58) \approx 1 - 0.9429 = 0.0571$. (Note : C'est le même calcul que l'exercice 19, $P(X \leq 30)$, car $P(X \leq 30) \approx P(Y \leq 30.5)$).

9.8 Exercices Python

Ces exercices utilisent les **Lois des Grands Nombres (LLN)** pour estimer des paramètres de population à partir de simulations. La LLN garantit que la moyenne d'échantillon (\bar{X}_n) converge vers la vraie espérance (μ) lorsque n devient grand.

Nous utiliserons les données de **yfinance** pour établir des "vraies" valeurs (μ, σ^2), puis nous simulerons des échantillons plus petits pour voir comment la moyenne de l'échantillon (\bar{X}_n) s'approche de μ .

```
!pip install yfinance
import yfinance as yf
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

# Définir les tickers et la période
tickers = ["GOOG"]
start_date = "2010-01-01"
end_date = "2024-12-31"

# Télécharger les prix de clôture ajustés
data = yf.download(tickers, start=start_date, end=end_date)["Adj Close"]

# Calculer les rendements journaliers en pourcentage
returns = data.pct_change().dropna()

# 'returns' est notre DataFrame principal.
# Considerons cette grande série de données comme notre "Population"
# pour les besoins de ces exercices.
population_mean = returns.mean()
population_var = returns.var()
population_std = returns.std()

print(f"--- Population (GOOG 2010-2024) ---")
print(f"Vraie Moyenne (mu) = {population_mean:.6f}")
print(f"Vraie Variance (sigma^2) = {population_var:.6f}")
```

Exercice 1 : Vérification de la Loi Faible (WLLN)

La WLLN dit que $P(|\bar{X}_n - \mu| > \epsilon) \rightarrow 0$ lorsque $n \rightarrow \infty$. Nous allons vérifier que la variance de \bar{X}_n diminue avec n , ce qui est la clé de la preuve de Chebyshev.

La théorie dit : $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$.

Votre tâche :

1. Utiliser σ^2 (la variance de la population) calculée ci-dessus.
2. Calculer la variance **théorique** de la moyenne d'échantillon $\text{Var}(\bar{X}_n)$ pour $n = 10$, $n = 100$, et $n = 1000$.
3. (Conclusion) Comment la variance de notre estimateur \bar{X}_n évolue-t-elle lorsque n augmente ? Qu'est-ce que cela implique sur la précision de notre estimation ?

Exercice 2 : Inégalité de Chebyshev

Chebyshev donne une borne universelle : $P(|\bar{X}_n - \mu| \geq k) \leq \frac{\text{Var}(\bar{X}_n)}{k^2}$. Utilisons $n = 100$ et $\epsilon = 0.01$ (soit un écart de 1% du rendement journalier).

Votre tâche :

1. Utiliser $\text{Var}(\bar{X}_{100})$ calculée à l'exercice 1.
2. Fixer $k = \epsilon = 0.01$.
3. Calculer la borne supérieure de probabilité (le côté droit de l'inégalité).
4. (Conclusion) Interpréter cette borne : "Pour un échantillon de 100 jours, la probabilité que notre moyenne d'échantillon soit erronée de plus de 1% est, au maximum, de..."

Exercice 3 : Simulation de la Loi Forte (SLLN) - Trajectoire

La SLLN dit que $P(\lim_{n \rightarrow \infty} \bar{X}_n = \mu) = 1$. Nous allons simuler une "trajectoire" de \bar{X}_n pour visualiser cette convergence.

Votre tâche :

1. Prendre les 1000 premiers rendements de la série `returns`.
2. Calculer la moyenne d'échantillon cumulative \bar{X}_n pour $n = 1, 2, 3, \dots, 1000$.
3. (Indice : utiliser `.expanding().mean()` de pandas).
4. (**Plot**) Tracer \bar{X}_n en fonction de n (de 1 à 1000).
5. (**Plot**) Tracer une ligne horizontale constante à la "vraie" moyenne μ (la `population_mean`).
6. (Conclusion) La trajectoire de \bar{X}_n converge-t-elle vers μ ?

Exercice 4 : Méthode de Monte-Carlo (Estimation de Probabilité)

Nous voulons estimer $p = P(X > 0.02)$, la probabilité d'un "gros jour positif" (rendement > 2%). La vraie valeur p est la proportion empirique sur toute la "population". L'estimation \bar{Z}_n est la proportion sur un échantillon de n jours.

Votre tâche :

1. Calculer la "vraie" probabilité p (notre μ) en comptant la proportion de `returns > 0.02` sur tout le dataset.
2. Simuler 500 expériences. Dans **chaque** expérience :
 - Tirer un échantillon de $n = 50$ jours (avec remise) de `returns`.
 - Estimer \bar{Z}_{50} (la proportion de jours > 0.02 dans cet échantillon).
3. (**Plot**) Tracer l'histogramme de vos 500 estimations \bar{Z}_{50} .
4. (**Plot**) Ajouter une ligne verticale à la "vraie" moyenne p .
5. (Conclusion) Les estimations sont-elles centrées autour de la vraie valeur, comme prédit par la LLN ?

Exercice 5 : Estimation de π (Monte-Carlo Pur)

Appliquons l'exemple du cours pour estimer π en utilisant la LLN. Nous cherchons $\mu = \pi/4$. Nous estimons μ par $\bar{Z}_n = \frac{\text{Points dans le cercle}}{n}$.

Votre tâche (avec NumPy) :

1. Définir $n = 1,000,000$.
2. Générer n coordonnées $X \sim U(0,1)$ et n coordonnées $Y \sim U(0,1)$.
3. Calculer $Z_i = 1$ si $X_i^2 + Y_i^2 \leq 1$, et 0 sinon. (Indice : `np.where` ou une comparaison booléenne).
4. Calculer \bar{Z}_n (la moyenne de Z).
5. Calculer votre estimation de $\pi \approx 4 \cdot \bar{Z}_n$.
6. (Conclusion) Votre estimation est-elle proche de `math.pi` ?

10 Le Théorème Central Limite (TCL)

10.1 Introduction : L'omniprésence de la loi normale

Dans la section précédente, la Loi des Grands Nombres (LLN) nous a donné une garantie fondamentale : la moyenne d'échantillon \bar{X}_n converge vers la vraie moyenne μ lorsque n devient grand.

$$\bar{X}_n \xrightarrow{p.s.} \mu$$

La LLN nous dit **où** la moyenne d'échantillon converge (vers la constante μ), mais elle ne nous dit rien sur la *forme* de la distribution de \bar{X}_n autour de μ pour un n grand, mais fini.

Le **Théorème Central Limite (TCL)** comble cette lacune. Il décrit la *manière* dont \bar{X}_n converge, en nous donnant la forme de sa distribution. C'est sans doute le théorème le plus important des statistiques.

Intuition : L'Idée Fondamentale

Intuitivement, ce résultat affirme qu'une **somme** d'un grand nombre de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) tend, le plus souvent, à suivre une **loi normale** (aussi appelée loi de Laplace-Gauss ou "courbe en cloche").

Ce théorème et ses généralisations offrent une explication à l'omniprésence de la loi normale dans la nature. De nombreux phénomènes (la taille d'un individu, l'erreur de mesure d'un instrument, le bruit de fond d'un signal) sont le résultat de l'addition d'un très grand nombre de petites perturbations aléatoires. Le TCL nous dit que le résultat de cette somme sera, inévitablement, distribué selon une loi normale.

10.2 L'illustration : la somme des "Pile ou Face"

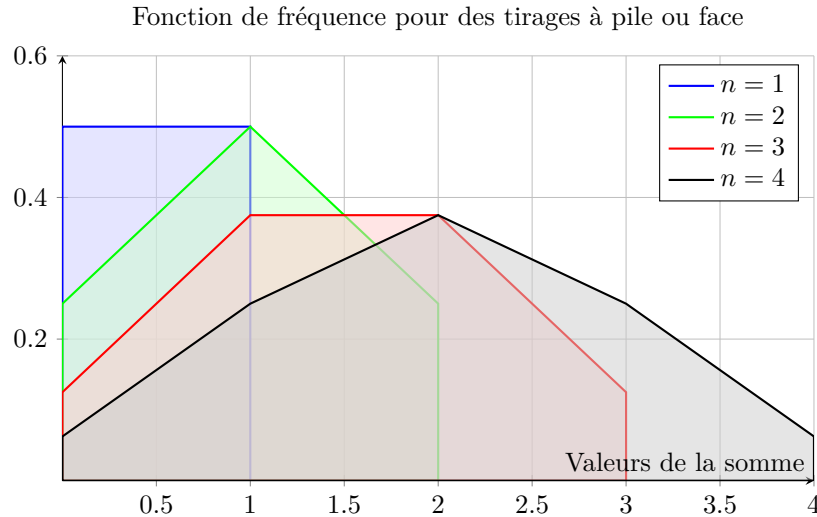
Prenons l'exemple le plus simple pour illustrer ce phénomène : le jeu de "pile ou face".

Exemple : Distribution de la Somme de n Lancers

Soit X_i le résultat du i -ème lancer, avec $X_i = 1$ pour "Face" (probabilité 0,5) et $X_i = 0$ pour "Pile" (probabilité 0,5). La distribution d'origine (pour $n = 1$) n'est pas du tout une courbe en cloche : c'est une distribution discrète avec deux bâtons de même hauteur.

Considérons la **somme** $S_n = X_1 + X_2 + \dots + X_n$, qui représente le nombre total de "Face" obtenus en n lancers.

- **Pour $n = 1$** : La distribution de S_1 est :
 - Valeurs de la somme : $\{0, 1\}$
 - Fréquences : $\{0.5, 0.5\}$
- **Pour $n = 2$** : Les sommes possibles sont $\{0, 1, 2\}$. La distribution de S_2 est :
 - Valeurs de la somme : $\{0, 1, 2\}$
 - Fréquences : $\{0.25, 0.5, 0.25\}$ (elle forme un triangle).
- **Pour $n = 3$** : Les sommes possibles sont $\{0, 1, 2, 3\}$. La distribution de S_3 est :
 - Valeurs de la somme : $\{0, 1, 2, 3\}$
 - Fréquences : $\{0.125, 0.375, 0.375, 0.125\}$



Graphiquement, on constate que plus le nombre de tirages n augmente (par exemple, jusqu'à $n = 12$), plus la courbe de fréquence (qui reste discrète) se rapproche d'une courbe en cloche symétrique, caractéristique de la loi normale.

10.3 Distribution de la population vs. Distribution d'échantillonnage

Le point le plus remarquable du TCL est qu'il fonctionne *quelle que soit* la distribution de départ.

Intuition : Population vs. Échantillonnage

Imaginez deux univers de distributions :

- **1. La Distribution de la Population (X_i) :** C'est la loi de nos variables X_i individuelles. Elle peut avoir **n'importe quelle forme** (par exemple, une distribution bimodale, asymétrique, ou uniforme). Cette distribution a une "vraie" moyenne μ et un "vrai" écart-type σ .
- **2. La Distribution d'Échantillonnage (\bar{X}_n) :** C'est la distribution de la *moyenne* $\bar{X}_n = (X_1 + \dots + X_n)/n$, calculée sur des échantillons de taille n . C'est la distribution de "toutes les moyennes d'échantillon possibles".

Le TCL énonce la relation magique entre les deux :

Quelle que soit la forme de la distribution de la population, plus la taille de l'échantillon n croît, plus la distribution d'échantillonnage de la moyenne \bar{X}_n est proche d'une loi normale (gaussienne).

De plus, les paramètres de cette loi normale sont :

- **Moyenne :** La distribution de \bar{X}_n est centrée sur la même moyenne μ que la population.
- **Écart-type :** La distribution de \bar{X}_n est beaucoup plus resserrée. Son écart-type (appelé "erreur standard") est $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

Cette dispersion σ/\sqrt{n} qui tend vers 0 est la manifestation de la Loi des Grands Nombres. Le TCL précise que la *forme* de cette convergence est gaussienne.

10.4 Énoncé formel du Théorème Central Limite

Pour énoncer le théorème formellement, nous devons d'abord définir les propriétés de la somme S_n et de la moyenne \bar{X}_n .

Soit X_1, \dots, X_n des variables aléatoires i.i.d. avec $E[X_i] = \mu$ et $\text{Var}(X_i) = \sigma^2$.

- **La Somme $S_n = \sum X_i$:**
 - Espérance : $E[S_n] = E[\sum X_i] = \sum E[X_i] = n\mu$
 - Variance : $\text{Var}(S_n) = \text{Var}(\sum X_i) = \sum \text{Var}(X_i) = n\sigma^2$
 - Écart-type : $\sigma_{S_n} = \sqrt{n\sigma^2} = \sigma\sqrt{n}$

- **La Moyenne** $\bar{X}_n = S_n/n$:
 - Espérance : $E[\bar{X}_n] = E[S_n/n] = \frac{1}{n}E[S_n] = \frac{1}{n}(n\mu) = \mu$
 - Variance : $\text{Var}(\bar{X}_n) = \text{Var}(S_n/n) = \frac{1}{n^2}\text{Var}(S_n) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$
 - Écart-type : $\sigma_{\bar{X}_n} = \sqrt{\sigma^2/n} = \frac{\sigma}{\sqrt{n}}$

Nous voyons que la distribution de S_n s'étale (variance $\rightarrow \infty$) tandis que celle de \bar{X}_n se contracte (variance $\rightarrow 0$). Pour étudier la *forme* de la convergence, nous créons une variable "stable" en la centrant (soustrayant la moyenne) et en la réduisant (divisant par l'écart-type). C'est la variable Z_n .

Théorème : Théorème Central Limite (Lindeberg-Lévy)

Soit X_1, X_2, \dots, X_n une suite de variables aléatoires **i.i.d.** (indépendantes et identiquement distribuées) suivant la même loi D . Supposons que l'**espérance** μ et l'**écart-type** σ de cette loi D existent, sont finis, et $\sigma \neq 0$.

Considérons la variable aléatoire standardisée Z_n :

$$Z_n = \frac{S_n - E[S_n]}{\sigma_{S_n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

Cette variable est équivalente à la moyenne standardisée :

$$Z_n = \frac{\bar{X}_n - E[\bar{X}_n]}{\sigma_{\bar{X}_n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

(Pour tout n , Z_n est une variable centrée-réduite : $E[Z_n] = 0$ et $\text{Var}(Z_n) = 1$).

Alors, la suite de variables aléatoires $Z_1, Z_2, \dots, Z_n, \dots$ **converge en loi** vers une variable aléatoire Z qui suit la **loi normale centrée réduite** $N(0, 1)$, lorsque n tend vers l'infini.

Cela signifie que si Φ est la fonction de répartition de la loi $N(0, 1)$, alors pour tout réel z :

$$\lim_{n \rightarrow \infty} P(Z_n \leq z) = \lim_{n \rightarrow \infty} P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = \Phi(z)$$

10.5 Applications Pratiques du TCL

Le TCL n'est pas seulement une curiosité mathématique ; c'est le fondement de l'inférence statistique. Voici comment l'appliquer concrètement pour résoudre des problèmes.

Exemple : La taille des individus

Contexte : La taille des individus dans une population suit une courbe en cloche. Pourquoi ? Car elle est la **somme** de milliers de petites influences (gènes, nutrition, etc.). Le TCL s'applique.

Données : Supposons que dans une population, la taille X des individus ait une espérance $\mu = 175$ cm et un écart-type $\sigma = 8$ cm. (Note : la loi de X n'est pas forcément normale, même si en pratique elle l'est).

Problème : On prélève un échantillon aléatoire de $n = 64$ individus. Quelle est la probabilité que la **moyenne de cet échantillon** (\bar{X}_{64}) soit supérieure à 177 cm ?

Solution :

1. Identifier les paramètres :

- Moyenne de la population : $\mu = 175$ cm
- Écart-type de la population : $\sigma = 8$ cm
- Taille de l'échantillon : $n = 64$

2. Appliquer le TCL : Puisque $n = 64$ est grand (généralement $n \geq 30$ est suffisant), le TCL s'applique. La distribution d'échantillonnage de la moyenne \bar{X}_n suit approximativement une loi normale.

$$\bar{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

3. Calculer les paramètres de la loi normale de \bar{X}_n :

- Espérance de \bar{X}_n : $E[\bar{X}_n] = \mu = 175$ cm.
- Écart-type de \bar{X}_n (appelé "Erreur Standard") :

$$\sigma_{\bar{X}_n} = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{64}} = \frac{8}{8} = 1 \text{ cm}$$

Donc, $\bar{X}_{64} \approx N(175, 1^2)$.

4. **Standardiser (Calculer le Z-score)** : Nous cherchons $P(\bar{X}_{64} > 177)$. Nous transformons cette valeur en un score Z pour utiliser la loi normale centrée réduite $N(0, 1)$.

$$Z = \frac{\bar{X}_n - \mu}{\sigma_{\bar{X}_n}} = \frac{177 - 175}{1} = 2$$

5. **Trouver la probabilité** : Chercher $P(\bar{X}_{64} > 177)$ revient à chercher $P(Z > 2)$. En utilisant la table de la loi normale (ou une calculatrice) :

$$P(Z > 2) = 1 - P(Z \leq 2) = 1 - \Phi(2)$$

Sachant que $\Phi(2) \approx 0.9772$,

$$P(Z > 2) = 1 - 0.9772 = 0.0228$$

Conclusion : Il y a environ 2.28% de chances qu'un échantillon de 64 personnes ait une taille moyenne supérieure à 177 cm.

Exemple : Remplissage de bouteilles

Contexte : Une machine remplit des bouteilles de soda. Le volume versé X_i fluctue légèrement. La loi de X_i est inconnue.

Données : La machine est réglée pour verser en moyenne $\mu = 500$ ml. L'écart-type du processus est connu et vaut $\sigma = 6$ ml. Pour un contrôle, on prélève un échantillon de $n = 36$ bouteilles.

Problème : On considère que la machine est dérégulée si la moyenne de l'échantillon \bar{X}_{36} est inférieure à 498 ml. Quelle est la probabilité d'une "fausse alarme" (c'est-à-dire, la machine fonctionne bien à $\mu = 500$, mais l'échantillon a une moyenne $\bar{X}_{36} < 498$) ?

Solution :

1. **Identifier les paramètres** : $\mu = 500$ ml, $\sigma = 6$ ml, $n = 36$.
2. **Appliquer le TCL** : $n = 36 \geq 30$, donc le TCL s'applique.

$$\bar{X}_{36} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

3. **Calculer les paramètres de \bar{X}_{36}** :

- Espérance : $E[\bar{X}_{36}] = \mu = 500$ ml.
- Erreur Standard : $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{36}} = \frac{6}{6} = 1$ ml.

Donc, $\bar{X}_{36} \approx N(500, 1^2)$.

4. **Standardiser (Calculer le Z-score)** : Nous cherchons la probabilité $P(\bar{X}_{36} < 498)$.

$$Z = \frac{\bar{X}_n - \mu}{\sigma_{\bar{X}_n}} = \frac{498 - 500}{1} = -2$$

5. **Trouver la probabilité** : Chercher $P(\bar{X}_{36} < 498)$ revient à chercher $P(Z < -2)$.

$$P(Z < -2) = \Phi(-2)$$

Par symétrie de la loi normale, $\Phi(-z) = 1 - \Phi(z)$.

$$P(Z < -2) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228$$

Conclusion : Il y a 2.28% de chances d'avoir une fausse alarme, c'est-à-dire de croire à tort que la machine est dérégulée alors qu'elle fonctionne normalement.

Exemple : Rendement d'un portefeuille (sur la Somme)

Contexte : Le rendement quotidien X_i d'un actif est très volatile. On s'intéresse au rendement annuel **total**, qui est la **somme** des rendements quotidiens.

Données : Supposons que le rendement quotidien X_i ait une espérance $\mu = 0.04\%$ et un écart-type $\sigma = 1\%$. (La loi de X_i est inconnue, mais μ et σ existent). Il y a $n = 252$ jours de trading dans l'année.

Problème : Quelle est la probabilité que le rendement annuel total $S_{252} = X_1 + \dots + X_{252}$ soit négatif (inférieur à 0) ?

Solution :

1. **Identifier les paramètres (pour une seule v.a. X_i) :** $\mu = 0.0004$, $\sigma = 0.01$, $n = 252$.
2. **Appliquer le TCL (pour la somme S_n) :** $n = 252$ est grand. Le TCL s'applique à la somme S_n .

$$S_n \approx N(n\mu, n\sigma^2)$$

3. **Calculer les paramètres de la loi normale de S_{252} :**

- Espérance de S_{252} : $E[S_n] = n\mu = 252 \times 0.0004 = 0.1008$ (soit 10.08%).
- Variance de S_{252} : $\text{Var}(S_n) = n\sigma^2 = 252 \times (0.01)^2 = 252 \times 0.0001 = 0.0252$.
- Écart-type de S_{252} : $\sigma_{S_n} = \sqrt{n\sigma^2} = \sqrt{0.0252} \approx 0.1587$ (soit 15.87%).

Donc, $S_{252} \approx N(0.1008, 0.1587^2)$.

4. **Standardiser (Calculer le Z-score) :** Nous cherchons $P(S_{252} < 0)$.

$$Z = \frac{S_n - E[S_n]}{\sigma_{S_n}} = \frac{0 - 0.1008}{0.1587} \approx -0.635$$

5. **Trouver la probabilité :** Chercher $P(S_{252} < 0)$ revient à chercher $P(Z < -0.635)$.

$$P(Z < -0.635) = \Phi(-0.635) = 1 - \Phi(0.635)$$

En interpolant dans la table, $\Phi(0.635) \approx 0.7373$.

$$P(Z < -0.635) \approx 1 - 0.7373 = 0.2627$$

Conclusion : Malgré une espérance de rendement quotidien positive, il y a environ 26.3% de chances que le rendement annuel total soit négatif.

Exemple : Estimation d'une proportion (Marge d'erreur)

Contexte : On veut estimer la proportion p de votants qui approuvent un candidat. On modélise chaque personne i par une variable de Bernoulli X_i (1 si "oui", 0 si "non"). L'espérance de la population est $\mu = E[X_i] = p$. La variance de la population est $\sigma^2 = \text{Var}(X_i) = p(1-p)$. Le résultat du sondage est la moyenne d'échantillon $\bar{X}_n = \hat{p}$ (la proportion observée).

Données : On sonde $n = 1000$ personnes. Le résultat est que 540 personnes disent "oui". Donc $\hat{p} = 540/1000 = 0.54$.

Problème : Calculer l'intervalle de confiance à 95% pour la vraie proportion p (la fameuse "marge d'erreur").

Solution :

1. **Appliquer le TCL :** $n = 1000$ est grand. Le TCL nous dit que la proportion d'échantillon $\hat{p} = \bar{X}_n$ suit une loi normale :

$$\hat{p} \approx N\left(p, \frac{p(1-p)}{n}\right)$$

2. **Formule de l'Intervalle de Confiance :** Un intervalle de confiance à 95% est centré sur notre estimation \hat{p} et s'étend de ± 1.96 erreurs standard (car $P(-1.96 \leq Z \leq 1.96) = 0.95$).

$$I.C_{.95\%} = [\hat{p} - 1.96 \cdot \sigma_{\hat{p}} ; \hat{p} + 1.96 \cdot \sigma_{\hat{p}}]$$

où $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$.

3. **Estimer l'Erreur Standard** : Problème : nous ne connaissons pas p (c'est ce que nous cherchons!). Nous ne pouvons donc pas calculer $\sigma_{\hat{p}}$. **Solution** : Nous l'estimons en utilisant notre meilleur estimateur pour p , qui est $\hat{p} = 0.54$.

$$\text{Erreur Standard Estimée (SE)} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

$$SE = \sqrt{\frac{0.54 \times (1 - 0.54)}{1000}} = \sqrt{\frac{0.54 \times 0.46}{1000}} = \sqrt{\frac{0.2484}{1000}} \approx \sqrt{0.0002484} \approx 0.01576$$

4. **Calculer la Marge d'Erreur** : La marge d'erreur (ME) est la demi-largeur de l'intervalle.

$$ME = 1.96 \times SE = 1.96 \times 0.01576 \approx 0.0309$$

5. **Construire l'Intervalle** :

$$I.C._{95\%} = [0.54 - 0.0309 ; 0.54 + 0.0309] = [0.5091 ; 0.5709]$$

Conclusion : Avec 54% d'intentions de vote sur un échantillon de 1000 personnes, nous sommes confiants à 95% que la vraie proportion p dans la population se situe entre 50.9% et 57.1%. La marge d'erreur du sondage est de $\pm 3.1\%$.

10.6 Exercices

Pour tous les exercices de calcul, vous pouvez utiliser les valeurs (arrondies) suivantes pour la fonction de répartition de la loi normale standard $\Phi(z) = P(Z \leq z)$:

- $\Phi(0) = 0.5$
- $\Phi(1.0) \approx 0.8413$
- $\Phi(1.5) \approx 0.9332$
- $\Phi(1.96) \approx 0.975$
- $\Phi(2.0) \approx 0.9772$
- $\Phi(2.5) \approx 0.9938$
- $\Phi(3.0) \approx 0.9987$

Et rappelez-vous la propriété de symétrie : $\Phi(-z) = 1 - \Phi(z)$.

Exercice 1 : Paramètres de \bar{X}_n

Soit X_i une suite de v.a. i.i.d. avec $\mu = 50$ et $\sigma^2 = 100$. Soit $n = 25$. Calculez $E[\bar{X}_{25}]$ et $\text{Var}(\bar{X}_{25})$.

Exercice 2 : Paramètres de S_n

Soit X_i une suite de v.a. i.i.d. avec $\mu = 10$ et $\sigma = 3$. Soit $n = 16$. Calculez $E[S_{16}]$ et l'écart-type $\sigma_{S_{16}}$.

Exercice 3 : Paramètres de \bar{X}_n (bis)

Soit X_i une suite de v.a. i.i.d. avec $\mu = 70$ et $\sigma = 10$. Soit $n = 400$. Calculez $E[\bar{X}_{400}]$ et l'écart-type $\sigma_{\bar{X}_{400}}$.

Exercice 4 : Paramètres de S_n (bis)

Soit X_i une suite de v.a. i.i.d. avec $\mu = 0.5$ et $\sigma^2 = 0.01$. Soit $n = 64$. Calculez $E[S_{64}]$ et $\text{Var}(S_{64})$.

Exercice 5 : Retrouver σ^2 (via \bar{X}_n)

La moyenne d'échantillon \bar{X}_n de $n = 49$ observations a une variance $\text{Var}(\bar{X}_{49}) = 2$. Quelle est la variance σ^2 de la population d'origine ?

Exercice 6 : Retrouver n (via S_n)

La somme S_n d'observations i.i.d. a une variance $\text{Var}(S_n) = 300$. La variance de la population est $\sigma^2 = 12$. Quelle est la taille de l'échantillon n ?

Exercice 7 : CLT pour \bar{X}_n (Queue Droite)

Soit X_i i.i.d. avec $\mu = 100$ et $\sigma = 15$. Soit $n = 36$. Calculez $P(\bar{X}_{36} > 105)$.

Exercice 8 : CLT pour \bar{X}_n (Queue Gauche)

Soit X_i i.i.d. avec $\mu = 50$ et $\sigma = 8$. Soit $n = 64$. Calculez $P(\bar{X}_{64} < 48)$.

Exercice 9 : CLT pour \bar{X}_n (Intervalle)

Soit X_i i.i.d. avec $\mu = 20$ et $\sigma = 5$. Soit $n = 100$. Calculez $P(19 \leq \bar{X}_{100} \leq 21.25)$.

Exercice 10 : Application \bar{X}_n (Bouteilles)

Une machine remplit des bouteilles avec $\mu = 500$ ml et $\sigma = 6$ ml. On prend $n = 36$ bouteilles. Calculez $P(\bar{X}_{36} > 501.5)$.

Exercice 11 : Application \bar{X}_n (Tailles)

La taille des individus a $\mu = 175$ cm et $\sigma = 8$ cm. On prend $n = 64$ individus. Calculez $P(\bar{X}_{64} < 173)$.

Exercice 12 : Application \bar{X}_n (Notes)

Les notes à un examen ont $\mu = 70$ et $\sigma = 12$. Une classe de $n = 36$ étudiants est un échantillon. Calculez $P(\bar{X}_{36} < 67)$.

Exercice 13 : CLT pour S_n (Queue Droite)

Soit X_i i.i.d. avec $\mu = 10$ et $\sigma = 2$. Soit $n = 100$. Calculez $P(S_{100} > 1020)$.

Exercice 14 : CLT pour S_n (Queue Gauche)

Soit X_i i.i.d. avec $\mu = 5$ et $\sigma = 4$. Soit $n = 64$. Calculez $P(S_{64} \leq 304)$.

Exercice 15 : CLT pour S_n (Intervalle)

Soit X_i i.i.d. avec $\mu = 2$ et $\sigma = 3$. Soit $n = 36$. Calculez $P(S_{36} > 90)$.

Exercice 16 : Application S_n (Ascenseur)

Un ascenseur transporte $n = 49$ personnes. Poids : $\mu = 70$ kg, $\sigma = 14$ kg. Calculez $P(S_{49} > 3528)$.

Exercice 17 : Application S_n (Rendement)

Le rendement quotidien X_i a $\mu = 0.001$ et $\sigma = 0.01$. Soit $n = 100$. Calculez $P(S_{100} > 0.2)$.

Exercice 18 : Application S_n (Rendement)

En utilisant les données de l'exercice 17, calculez $P(S_{100} < 0)$.

Exercice 19 : Inverse (Trouver c pour \bar{X}_n)

Soit X_i i.i.d. avec $\mu = 50$ et $\sigma = 10$. Soit $n = 100$. Trouvez la valeur c telle que $P(\bar{X}_{100} \leq c) \approx 0.8413$.

Exercice 20 : Inverse (Trouver c pour S_n)

Soit X_i i.i.d. avec $\mu = 10$ et $\sigma = 3$. Soit $n = 36$. Trouvez la valeur c telle que $P(S_{36} \leq c) \approx 0.0013$.

Exercice 21 : Inverse (Taille d'échantillon n)

Une population a $\mu = 0$ et $\sigma = 10$. Quelle taille d'échantillon n faut-il pour que $P(|\bar{X}_n| \leq 1) \geq 0.95$? (Indice : $P(-1.96 \leq Z \leq 1.96) = 0.95$).

Exercice 22 : Inverse (Taille d'échantillon n)

Une population a $\mu = 100$ et $\sigma = 20$. Quelle taille d'échantillon n faut-il pour que $P(\bar{X}_n \geq 102) \leq 0.0228$?

Exercice 23 : Paramètres (Proportion)

On sonde $n = 400$ personnes. La vraie proportion p est 0.25. On modélise $X_i \sim \text{Bern}(p)$.

1. Calculez $\mu = E[X_i]$ et $\sigma^2 = \text{Var}(X_i)$.
2. Calculez $E[\hat{p}]$ et $\text{Var}(\hat{p})$ (où $\hat{p} = \bar{X}_n$).

Exercice 24 : Calcul (Proportion)

On sonde $n = 100$ personnes. La vraie proportion est $p = 0.5$. Calculez la probabilité que la proportion observée \hat{p} soit supérieure à 0.6, $P(\hat{p} > 0.6)$.

Exercice 25 : Marge d'Erreur (Proportion)

Un sondage sur $n = 1000$ personnes donne un résultat $\hat{p} = 0.54$. Calculez la marge d'erreur à 95% (c'est-à-dire $1.96 \times SE_{\hat{p}}$).

10.7 Corrections des Exercices**Correction Exercice 1 : Paramètres de \bar{X}_n**

$$E[\bar{X}_{25}] = \mu = 50. \text{Var}(\bar{X}_{25}) = \frac{\sigma^2}{n} = \frac{100}{25} = 4.$$

Correction Exercice 2 : Paramètres de S_n

$$E[S_{16}] = n\mu = 16 \times 10 = 160. \text{Var}(S_{16}) = n\sigma^2 = 16 \times 3^2 = 16 \times 9 = 144. \sigma_{S_{16}} = \sqrt{\text{Var}(S_{16})} = \sqrt{144} = 12.$$

Correction Exercice 3 : Paramètres de \bar{X}_n (bis)

$$E[\bar{X}_{400}] = \mu = 70. \sigma_{\bar{X}_{400}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{400}} = \frac{10}{20} = 0.5.$$

Correction Exercice 4 : Paramètres de S_n (bis)

$$E[S_{64}] = n\mu = 64 \times 0.5 = 32. \text{Var}(S_{64}) = n\sigma^2 = 64 \times 0.01 = 0.64.$$

Correction Exercice 5 : Retrouver σ^2 (via \bar{X}_n)

$$\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n} \implies \sigma^2 = n \cdot \text{Var}(\bar{X}_n). \sigma^2 = 49 \times 2 = 98.$$

Correction Exercice 6 : Retrouver n (via S_n)

$$\text{Var}(S_n) = n\sigma^2 \implies n = \frac{\text{Var}(S_n)}{\sigma^2}. n = \frac{300}{12} = 25.$$

Correction Exercice 7 : CLT pour \bar{X}_n (Queue Droite)

$$\mu = 100, \sigma = 15, n = 36. \bar{X}_{36} \approx \mathcal{N}(\mu, \sigma^2/n). \mu_{\bar{X}} = 100. \sigma_{\bar{X}} = \frac{15}{\sqrt{36}} = \frac{15}{6} = 2.5. \text{ On cherche } P(\bar{X}_{36} > 105). Z = \frac{105-100}{2.5} = \frac{5}{2.5} = 2. P(Z > 2) = 1 - \Phi(2) \approx 1 - 0.9772 = 0.0228.$$

Correction Exercice 8 : CLT pour \bar{X}_n (Queue Gauche)

$$\mu = 50, \sigma = 8, n = 64. \bar{X}_{64} \approx \mathcal{N}(\mu, \sigma^2/n). \mu_{\bar{X}} = 50. \sigma_{\bar{X}} = \frac{8}{\sqrt{64}} = \frac{8}{8} = 1. \text{ On cherche } P(\bar{X}_{64} < 48). Z = \frac{48-50}{1} = -2. P(Z < -2) = \Phi(-2) = 1 - \Phi(2) \approx 1 - 0.9772 = 0.0228.$$

Correction Exercice 9 : CLT pour \bar{X}_n (Intervalle)

$$\mu = 20, \sigma = 5, n = 100. \bar{X}_{100} \approx \mathcal{N}(\mu, \sigma^2/n). \mu_{\bar{X}} = 20. \sigma_{\bar{X}} = \frac{5}{\sqrt{100}} = \frac{5}{10} = 0.5. \text{ On cherche } P(19 \leq \bar{X}_{100} \leq 21.25). Z_1 = \frac{19-20}{0.5} = -2. Z_2 = \frac{21.25-20}{0.5} = \frac{1.25}{0.5} = 2.5. P(-2 \leq Z \leq 2.5) = \Phi(2.5) - \Phi(-2) = \Phi(2.5) - (1 - \Phi(2)). P \approx 0.9938 - (1 - 0.9772) = 0.9938 - 0.0228 = 0.9710.$$

Correction Exercice 10 : Application \bar{X}_n (Bouteilles)

$\mu = 500, \sigma = 6, n = 36$. $\bar{X}_{36} \approx \mathcal{N}(\mu, \sigma^2/n)$. $\mu_{\bar{X}} = 500$. $\sigma_{\bar{X}} = \frac{6}{\sqrt{36}} = \frac{6}{6} = 1$. On cherche $P(\bar{X}_{36} > 501.5)$. $Z = \frac{501.5-500}{1} = 1.5$. $P(Z > 1.5) = 1 - \Phi(1.5) \approx 1 - 0.9332 = 0.0668$.

Correction Exercice 11 : Application \bar{X}_n (Tailles)

$\mu = 175, \sigma = 8, n = 64$. $\bar{X}_{64} \approx \mathcal{N}(\mu, \sigma^2/n)$. $\mu_{\bar{X}} = 175$. $\sigma_{\bar{X}} = \frac{8}{\sqrt{64}} = \frac{8}{8} = 1$. On cherche $P(\bar{X}_{64} < 173)$. $Z = \frac{173-175}{1} = -2$. $P(Z < -2) = \Phi(-2) = 1 - \Phi(2) \approx 1 - 0.9772 = 0.0228$.

Correction Exercice 12 : Application \bar{X}_n (Notes)

$\mu = 70, \sigma = 12, n = 36$. $\bar{X}_{36} \approx \mathcal{N}(\mu, \sigma^2/n)$. $\mu_{\bar{X}} = 70$. $\sigma_{\bar{X}} = \frac{12}{\sqrt{36}} = \frac{12}{6} = 2$. On cherche $P(\bar{X}_{36} < 67)$. $Z = \frac{67-70}{2} = -1.5$. $P(Z < -1.5) = \Phi(-1.5) = 1 - \Phi(1.5) \approx 1 - 0.9332 = 0.0668$.

Correction Exercice 13 : CLT pour S_n (Queue Droite)

$\mu = 10, \sigma = 2, n = 100$. $S_{100} \approx \mathcal{N}(n\mu, n\sigma^2)$. $E[S_{100}] = 100 \times 10 = 1000$. $\sigma_{S_n} = \sigma\sqrt{n} = 2 \times \sqrt{100} = 2 \times 10 = 20$. On cherche $P(S_{100} > 1020)$. $Z = \frac{1020-1000}{20} = \frac{20}{20} = 1$. $P(Z > 1) = 1 - \Phi(1) \approx 1 - 0.8413 = 0.1587$.

Correction Exercice 14 : CLT pour S_n (Queue Gauche)

$\mu = 5, \sigma = 4, n = 64$. $S_{64} \approx \mathcal{N}(n\mu, n\sigma^2)$. $E[S_{64}] = 64 \times 5 = 320$. $\sigma_{S_n} = \sigma\sqrt{n} = 4 \times \sqrt{64} = 4 \times 8 = 32$. On cherche $P(S_{64} \leq 304)$. $Z = \frac{304-320}{32} = \frac{-16}{32} = -0.5$. $P(Z \leq -0.5) = \Phi(-0.5) = 1 - \Phi(0.5)$. (Valeur non fournie).

Correction Exercice 15 : CLT pour S_n (Intervalle)

$\mu = 2, \sigma = 3, n = 36$. $S_{36} \approx \mathcal{N}(n\mu, n\sigma^2)$. $E[S_{36}] = 36 \times 2 = 72$. $\sigma_{S_n} = \sigma\sqrt{n} = 3 \times \sqrt{36} = 3 \times 6 = 18$. On cherche $P(S_{36} > 90)$. $Z = \frac{90-72}{18} = \frac{18}{18} = 1$. $P(Z > 1) = 1 - \Phi(1) \approx 1 - 0.8413 = 0.1587$.

Correction Exercice 16 : Application S_n (Ascenseur)

$\mu = 70, \sigma = 14, n = 49$. $S_{49} \approx \mathcal{N}(n\mu, n\sigma^2)$. $E[S_{49}] = 49 \times 70 = 3430$. $\sigma_{S_n} = \sigma\sqrt{n} = 14 \times \sqrt{49} = 14 \times 7 = 98$. On cherche $P(S_{49} > 3528)$. $Z = \frac{3528-3430}{98} = \frac{98}{98} = 1$. $P(Z > 1) = 1 - \Phi(1) \approx 1 - 0.8413 = 0.1587$.

Correction Exercice 17 : Application S_n (Rendement)

$\mu = 0.001, \sigma = 0.01, n = 100$. $S_{100} \approx \mathcal{N}(n\mu, n\sigma^2)$. $E[S_{100}] = 100 \times 0.001 = 0.1$. $\sigma_{S_n} = \sigma\sqrt{n} = 0.01 \times \sqrt{100} = 0.01 \times 10 = 0.1$. On cherche $P(S_{100} > 0.2)$. $Z = \frac{0.2-0.1}{0.1} = \frac{0.1}{0.1} = 1$. $P(Z > 1) = 1 - \Phi(1) \approx 1 - 0.8413 = 0.1587$.

Correction Exercice 18 : Application S_n (Rendement)

On utilise les mêmes paramètres : $S_{100} \approx \mathcal{N}(0.1, 0.1^2)$. On cherche $P(S_{100} < 0)$. $Z = \frac{0-0.1}{0.1} = -1$. $P(Z < -1) = \Phi(-1) = 1 - \Phi(1) \approx 1 - 0.8413 = 0.1587$.

Correction Exercice 19 : Inverse (Trouver c pour \bar{X}_n)

$\mu = 50, \sigma = 10, n = 100$. $\bar{X}_{100} \approx \mathcal{N}(\mu, \sigma^2/n)$. $\mu_{\bar{X}} = 50$. $\sigma_{\bar{X}} = \frac{10}{\sqrt{100}} = 1$. On cherche c tel que $P(\bar{X}_{100} \leq c) \approx 0.8413$. $P(Z \leq \frac{c-50}{1}) = 0.8413$. D'après la table, le z correspondant est 1.0. $\frac{c-50}{1} = 1 \implies c = 51$.

Correction Exercice 20 : Inverse (Trouver c pour S_n)

$\mu = 10, \sigma = 3, n = 36$. $S_{36} \approx \mathcal{N}(n\mu, n\sigma^2)$. $E[S_{36}] = 36 \times 10 = 360$. $\sigma_{S_n} = \sigma\sqrt{n} = 3 \times \sqrt{36} = 18$. On cherche c tel que $P(S_{36} \leq c) \approx 0.0013$. $P(Z \leq \frac{c-360}{18}) = 0.0013$. On sait $\Phi(3) \approx 0.9987$, donc $\Phi(-3) = 1 - 0.9987 = 0.0013$. Le z correspondant est -3.0 . $\frac{c-360}{18} = -3 \implies c = 360 - 3(18) = 360 - 54 = 306$.

Correction Exercice 21 : Inverse (Taille d'échantillon n)

$\mu = 0, \sigma = 10$. On veut $P(|\bar{X}_n| \leq 1) \geq 0.95$. $P(-1 \leq \bar{X}_n \leq 1) \geq 0.95$. Standardisation : $\sigma_{\bar{X}} = \frac{10}{\sqrt{n}}$. $P\left(\frac{-1-0}{10/\sqrt{n}} \leq Z \leq \frac{1-0}{10/\sqrt{n}}\right) \geq 0.95$. $P\left(\frac{-\sqrt{n}}{10} \leq Z \leq \frac{\sqrt{n}}{10}\right) \geq 0.95$. On sait $P(-1.96 \leq Z \leq 1.96) = 0.95$. On doit donc avoir $\frac{\sqrt{n}}{10} \geq 1.96$. $\sqrt{n} \geq 19.6 \implies n \geq (19.6)^2 = 384.16$. Il faut $n = 385$ au minimum.

Correction Exercice 22 : Inverse (Taille d'échantillon n)

$\mu = 100, \sigma = 20$. On veut $P(\bar{X}_n \geq 102) \leq 0.0228$. $P(Z \geq z) \leq 0.0228$. On sait $P(Z \geq 2) = 1 - \Phi(2) \approx 0.0228$. Donc on a besoin que notre Z-score soit ≥ 2 . $Z = \frac{102-100}{20/\sqrt{n}} = \frac{2}{20/\sqrt{n}} = \frac{2\sqrt{n}}{20} = \frac{\sqrt{n}}{10}$. On pose $Z \geq 2 \implies \frac{\sqrt{n}}{10} \geq 2$. $\sqrt{n} \geq 20 \implies n \geq 400$. Il faut $n = 400$ au minimum.

Correction Exercice 23 : Paramètres (Proportion)

$X_i \sim \text{Bern}(p)$ avec $p = 0.25$. 1. $\mu = E[X_i] = p = 0.25$. $\sigma^2 = \text{Var}(X_i) = p(1-p) = 0.25 \times 0.75 = 0.1875$. 2. $\hat{p} = \bar{X}_n$. $n = 400$. $E[\hat{p}] = \mu = 0.25$. $\text{Var}(\hat{p}) = \frac{\sigma^2}{n} = \frac{0.1875}{400} \approx 0.00046875$.

Correction Exercice 24 : Calcul (Proportion)

$p = 0.5, n = 100$. $\hat{p} \approx \mathcal{N}(p, \frac{p(1-p)}{n})$. $E[\hat{p}] = 0.5$. $\text{Var}(\hat{p}) = \frac{0.5 \times 0.5}{100} = \frac{0.25}{100} = 0.0025$. $\sigma_{\hat{p}} = \sqrt{0.0025} = 0.05$. On cherche $P(\hat{p} > 0.6)$. $Z = \frac{0.6-0.5}{0.05} = \frac{0.1}{0.05} = 2$. $P(Z > 2) = 1 - \Phi(2) \approx 1 - 0.9772 = 0.0228$.

Correction Exercice 25 : Marge d'Erreur (Proportion)

$n = 1000, \hat{p} = 0.54$. On estime l'erreur standard $SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. $SE = \sqrt{\frac{0.54 \times (1-0.54)}{1000}} = \sqrt{\frac{0.2484}{1000}} \approx 0.01576$. La marge d'erreur à 95% est $ME = 1.96 \times SE$. $ME = 1.96 \times 0.01576 \approx 0.0309$. (Soit $\pm 3.09\%$).

10.8 Exercices Python

Ces exercices appliquent le Théorème Central Limite (TCL) aux données financières. Nous considérerons l'ensemble des rendements journaliers de Google (GOOG) sur une longue période comme notre **population** (dont nous connaissons le "vrai" μ et σ^2). Nous simulerons ensuite un **échantillonnage** (tirer n jours au hasard) pour voir comment la **distribution d'échantillonnage de la moyenne** (\bar{X}_n) se comporte.

```
!pip install yfinance
import yfinance as yf
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

# Définir le ticker et une longue période pour notre "population"
ticker = "GOOG"
start_date = "2010-01-01"
end_date = "2024-12-31"

# Télécharger les prix de clôture ajustés
data = yf.download(ticker, start=start_date, end=end_date)["Adj Close"]

# Calculer les rendements journaliers (notre population X)
returns = data.pct_change().dropna()

# Calculer les "vrais" paramètres de la population
population_mean = returns.mean()
population_var = returns.var()
population_std = returns.std()

print(f"--- Paramètres de la Population (GOOG 2010-2024) ---")
print(f"Mu (moyenne) = {population_mean:.6f}")
```

```
print(f"Sigma^2 (variance) = {population_var:.6f}")
print(f"Sigma (ecart-type) = {population_std:.6f}")
print(f"N total (population) = {len(returns)}")
```

Exercice 1 : Distribution de la Population vs. TCL

La population des rendements journaliers X_i n'est pas parfaitement normale. La distribution de \bar{X}_n , en revanche, devrait l'être.

Votre tâche :

1. **(Plot)** Tracer l'histogramme de la **population** (la série **returns** complète).
2. (Conclusion) La distribution de la population ressemble-t-elle à une loi normale parfaite ? (Regardez les queues).

Exercice 2 : Paramètres de la Distribution d'Échantillonnage (Théorie)

Le TCL prédit les paramètres de la distribution de \bar{X}_n . Supposons que nous prenions des échantillons de taille $n = 30$.

Votre tâche :

1. Calculer l'espérance **théorique** de la moyenne d'échantillon, $E[\bar{X}_{30}]$.
2. Calculer la variance **théorique** de la moyenne d'échantillon, $\text{Var}(\bar{X}_{30}) = \sigma^2/n$.
3. Calculer l'erreur standard **théorique** (l'écart-type) de la moyenne d'échantillon, $\sigma_{\bar{X}_{30}} = \sigma/\sqrt{n}$.
4. (Utilisez les μ et σ^2 de la population calculés dans la cellule de setup).

Exercice 3 : Simulation de la Distribution d'Échantillonnage

Vérifions le TCL par simulation. Nous allons générer $k = 1000$ échantillons de taille $n = 30$ et calculer la moyenne de chacun.

Votre tâche :

1. Créer une liste (ou un array) vide **sample_means**.
2. Boucler $k = 1000$ fois :
 - Tirer un échantillon de $n = 30$ rendements de la **returns** (avec `np.random.choice(..., size=30, replace=True)`).
 - Calculer la moyenne de cet échantillon.
 - Ajouter cette moyenne à votre liste **sample_means**.
3. Vous avez maintenant 1000 valeurs de \bar{X}_{30} .

Exercice 4 : Vérification des Paramètres (Empirique vs. Théorie)

Utilisons les 1000 moyennes d'échantillon (**sample_means**) de l'exercice 3.

Votre tâche :

1. Calculer la moyenne **empirique** des 1000 moyennes d'échantillon.
2. Comparer ce résultat à l'espérance **théorique** $E[\bar{X}_{30}]$ de l'exercice 2.
3. Calculer la variance **empirique** des 1000 moyennes d'échantillon.
4. Comparer ce résultat à la variance **théorique** $\text{Var}(\bar{X}_{30})$ de l'exercice 2.

Exercice 5 : Visualisation du TCL (Plot)

C'est la visualisation la plus importante. Nous allons superposer la distribution empirique des moyennes (de l'Ex 3) et la distribution normale théorique (de l'Ex 2).

Votre tâche :

1. **(Plot)** Tracer l'histogramme des 1000 **sample_means** (de l'Ex 3). Assurez-vous d'utiliser `density=True`.
2. **(Plot)** Sur le **même** graphique, tracer la PDF de la loi normale **théorique** prédite par le TCL.
3. (Indice : $X \sim \mathcal{N}(\mu, \sigma^2/n)$). Utilisez `scipy.stats.norm.pdf()` avec `loc = population_mean` et `scale = (population_std / np.sqrt(30))`).

4. (Conclusion) La prédiction du TCL correspond-elle à la simulation ?

Exercice 6 : L'Effet de la Taille n (Plot)

La convergence est plus rapide lorsque n est grand.

Votre tâche :

1. Répéter l'exercice 3, mais pour $n = 5$ (créer 1000 `sample_means_5`).
2. Répéter l'exercice 3, mais pour $n = 100$ (créer 1000 `sample_means_100`).
3. **(Plot)** Créer trois histogrammes côte à côte (`plt.subplot`) pour $n = 5$, $n = 30$ (de l'Ex 3), et $n = 100$.
4. (Conclusion) Que constatez-vous à propos de la variance (largeur de la cloche) de la distribution de \bar{X}_n lorsque n augmente ?

Exercice 7 : Application du TCL (Calcul de Probabilité pour \bar{X}_n)

Utilisons le TCL pour répondre à une question pratique. Quelle est la probabilité que le rendement **moyen** sur un mois de trading ($n = 21$ jours) soit positif ?

Votre tâche :

1. On cherche $P(\bar{X}_{21} > 0)$.
2. Identifier μ et σ (de la population).
3. Calculer l'erreur standard $\sigma_{\bar{X}_{21}} = \sigma/\sqrt{21}$.
4. Standardiser la valeur 0 : $Z = (0 - \mu)/\sigma_{\bar{X}_{21}}$.
5. Calculer la probabilité $P(Z > z) = 1 - \Phi(z)$ en utilisant `scipy.stats.norm.cdf()`.

Exercice 8 : Application du TCL (Calcul de Probabilité pour S_n)

Quelle est la probabilité que le rendement **total** (la somme) sur une année ($n = 252$ jours) soit supérieur à 10% ?

Votre tâche :

1. On cherche $P(S_{252} > 0.10)$.
2. Le TCL dit $S_n \approx \mathcal{N}(n\mu, n\sigma^2)$.
3. Calculer l'espérance de la somme : $E[S_{252}] = n\mu = 252 \times \mu$.
4. Calculer l'écart-type de la somme : $\sigma_{S_{252}} = \sigma\sqrt{n} = \sigma \times \sqrt{252}$.
5. Standardiser la valeur 0.10 : $Z = (0.10 - E[S_{252}])/\sigma_{S_{252}}$.
6. Calculer la probabilité $P(Z > z) = 1 - \Phi(z)$.

Exercice 9 : TCL pour les Proportions (Binomiale)

Le TCL s'applique aussi aux proportions (qui sont des moyennes de Bernoulli). Soit p la probabilité qu'un jour soit un "jour de hausse" (rendement > 0).

Votre tâche :

1. Estimer la "vraie" proportion p (notre μ) en calculant la proportion de jours de hausse dans toute la population `returns`.
2. On sonde $n = 100$ jours. Quelle est la probabilité que notre sondage ($\hat{p} = \bar{X}_{100}$) montre une majorité de jours de baisse ($\hat{p} < 0.5$) ?
3. L'erreur standard pour une proportion est $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$.
4. Standardiser 0.5 : $Z = (0.5 - p)/\sigma_{\hat{p}}$.
5. Calculer la probabilité $P(Z < z) = \Phi(z)$.

Exercice 10 : Marge d'Erreur (Intervalle de Confiance)

C'est l'application la plus courante du TCL dans les médias. Quelle est la "marge d'erreur" à 95% pour notre estimation de p (proportion de jours de hausse) si l'on utilise un échantillon de $n = 1000$ jours ?

Votre tâche :

1. Utiliser le p (proportion de la population) estimé à l'exercice 9.

2. Calculer l'erreur standard pour $n = 1000$: $SE = \sqrt{p(1-p)/1000}$.
3. La marge d'erreur à 95% est $ME = 1.96 \times SE$. (Car $P(-1.96 \leq Z \leq 1.96) \approx 0.95$).
4. (Conclusion) Interpréter le résultat : "Notre estimation de la proportion de jours de hausse sera correcte à $\pm [ME]$ près, 95% du temps."