

Variables aléatoires et espérance

I. Espaces probabilisés et variables aléatoires

La théorie de la mesure fournit un cadre unique pour les probabilités. En effet, un espace probabilisé n'est rien d'autre qu'un espace mesuré de masse totale égale à 1. Ainsi, tous les théorèmes d'intégration s'appliquent sans modification : convergence monotone, convergence dominée, Fubini.

1. Espace probabilisé

Définition :

Un **espace probabilisé** est un triplet $(\Omega, \mathcal{A}, P)$ où Ω est un ensemble, $\mathcal{A}$ une tribu sur Ω et P une mesure sur $\mathcal{A}$ telle que $P(\Omega) = 1$. Les éléments de $\mathcal{A}$ sont les **événements**. Une propriété vraie hors d'un événement de probabilité nulle est dite vraie **presque sûrement** (p.s.).

Toutes les propriétés d'une mesure restent donc valables. Par exemple, P est croissante, sous-additive et continue le long des suites monotones d'événements. De plus, $P(A^c) = 1 - P(A)$ pour tout événement A .

Exemple :

L'espace $([0,1], \mathcal{B}([0,1]), \lambda)$, où λ est la mesure de Lebesgue, est un espace probabilisé. Il modélise le tirage d'un point « au hasard » dans $[0,1]$. D'autre part, si $(p_n)_n \in \mathbb{N}$ est une suite positive de somme 1, alors $P = \sum_n p_n \delta_n$ est une probabilité sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$.

2. Variables aléatoires et lois

Définition :

Une **variable aléatoire** à valeurs dans un espace mesurable $(E, \mathcal{E})$ est une application mesurable $X : \Omega \rightarrow E$. Sa **loi** est la mesure image $P_X = X_*P$, définie par $P_X(B) = P(X \in B) = P(X^{-1}(B))$ pour tout $B \in \mathcal{E}$.

La loi P_X est une probabilité sur E . Autrement dit, une variable aléatoire transporte la probabilité de Ω vers l'espace d'arrivée. Pour $E = \mathbb{R}$, la tribu est celle des boréliens. Il suffit alors de vérifier que $\{X \leq x\} \in \mathcal{A}$ pour tout réel x , car les demi-droites $] -\infty, x]$ engendrent $\mathcal{B}(\mathbb{R})$.

Proposition :

La **fonction de répartition** $F_X(x) = P(X \leq x)$ caractérise la loi d'une variable réelle. Autrement dit, deux variables de même fonction de répartition ont la même loi.

Démonstration :

Les demi-droites $] -\infty, x]$ forment une classe stable par intersection finie qui engendre $\mathcal{B}(\mathbb{R})$. Or deux probabilités qui coïncident sur une telle classe coïncident sur la tribu engendrée : c'est le lemme d'unicité issu du théorème de classe monotone. Par conséquent, $P_X = P_Y$ dès que $F_X = F_Y$.

Deux cas particuliers dominent la pratique. D'abord, la loi est **discrète** si X prend ses valeurs dans un ensemble dénombrable. Ensuite, elle est **à densité** s'il existe $f \geq 0$ intégrable avec

$P_X(B) = \int_B f \, d\lambda$. Dans les deux cas, la loi est une mesure à densité, respectivement par rapport à la mesure de comptage et à la mesure de Lebesgue.

II. Espérance et méthode de la fonction muette

1. L'espérance est une intégrale

Définition :

Si $X \geq 0$, on pose $E[X] = \int_{\Omega} X dP \in [0, +\infty]$. Si $E[|X|] < +\infty$, on dit que X est **intégrable** et on pose $E[X] = \int_{\Omega} X dP$. La **variance** d'une variable de carré intégrable est $\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$.

L'espérance hérite ainsi de toutes les propriétés de l'intégrale. Elle est linéaire, croissante et vérifie $|E[X]| \leq E[|X|]$. De plus, les théorèmes de convergence monotone et dominée s'écrivent directement avec E . Enfin, l'inégalité de Markov $P(|X| \geq a) \leq E[|X|]/a$ découle de $a 1_{\{|X| \geq a\}} \leq |X|$.

2. Théorème de transfert

Théorème :

Soit X à valeurs dans $(E, \mathcal{E})$ et $h : E \rightarrow \mathbb{R}$ mesurable. Si $h \geq 0$ ou si $h(X)$ est intégrable, alors

$$E[h(X)] = \int_E h dP_X.$$

En particulier, si X a une densité f sur $\mathbb{R}^d$, on a $E[h(X)] = \int_{\mathbb{R}^d} h(x) f(x) dx$. Si X

est discrète, on a $E[h(X)] = \sum_x h(x) P(X = x)$.

Démonstration :

Pour $h = 1_B$, l'égalité s'écrit $P(X \in B) = P_X(B)$: c'est la définition de la loi. Ensuite, par linéarité, elle vaut pour les fonctions étagées positives. Toute fonction mesurable positive est limite croissante de fonctions étagées positives. Donc le théorème de convergence monotone donne l'égalité pour $h \geq 0$. Enfin, on écrit $h = h^+ - h^-$ dans le cas intégrable.

3. La fonction muette

Le théorème de transfert se lit aussi à l'envers. Si l'on sait écrire $E[h(X)]$ sous la forme $\int h f d\lambda$ pour toute fonction test h , alors on connaît la loi de X . C'est la **méthode de la fonction muette**.

Proposition :

Soit X un vecteur aléatoire de $\mathbb{R}^d$ et $f \geq 0$ d'intégrale 1. Si

$E[h(X)] = \int h(x) f(x) dx$ pour toute fonction h mesurable bornée, ou seulement pour toute h continue à support compact, alors X a pour densité f .

Méthode :

Pour trouver la loi de $Y = \varphi(X)$: on fixe h mesurable bornée ; on écrit

$E[h(\varphi(X))] = \int h(\varphi(x)) f_X(x) dx$; on effectue le changement de variable $y = \varphi(x)$ pour obtenir $\int h(y) g(y) dy$; on conclut que g est la densité de Y . Si φ n'est pas injective, on découpe l'intégrale en morceaux où elle l'est.

Exemple :

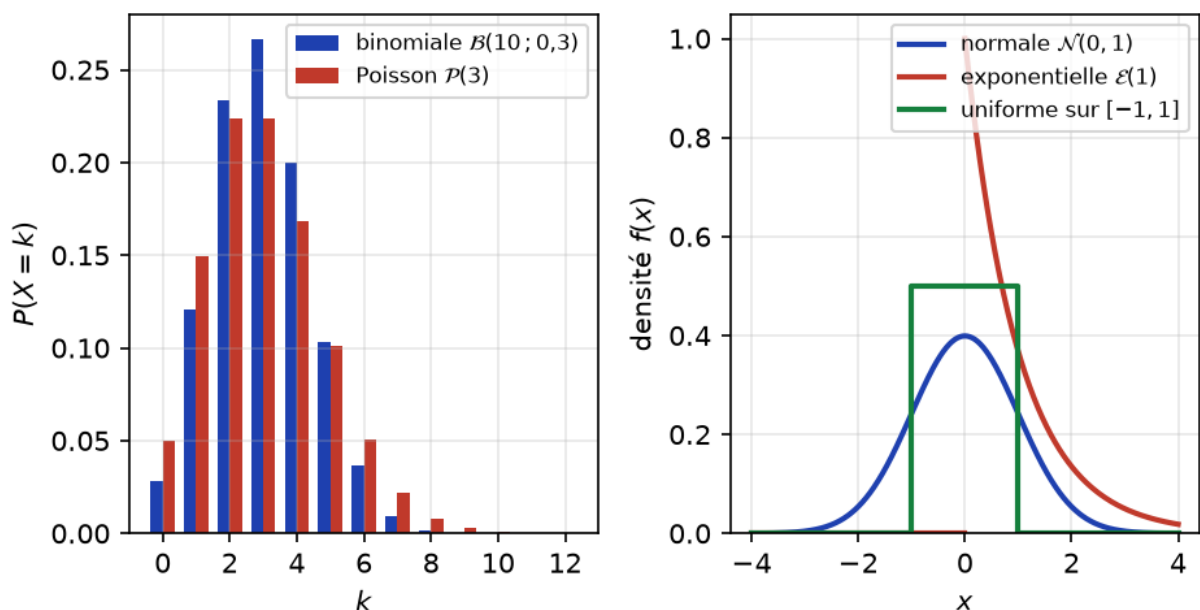
Soit X de loi $\mathcal{N}(0,1)$ et $Y = X^2$. Par parité, on a

$$E[h(X^2)] = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} h(x^2) e^{-x^2/2} dx = \int_0^{+\infty} h(y) \frac{e^{-y/2}}{\sqrt{2\pi y}} dy,$$

avec $y = x^2$ et $dx = dy/(2\sqrt{y})$. Ainsi, Y a pour densité $y \rightarrow e^{-y/2}/\sqrt{2\pi y}$ sur $]0, +\infty[$. C'est la loi du khi-deux à un degré de liberté.

III. Lois usuelles discrètes et à densité

Voici les lois à connaître, avec leur espérance et leur variance. La figure ci-dessous compare d'abord deux lois discrètes, puis trois densités classiques. On remarque que le même langage, celui des mesures, décrit les deux familles.



Définition :

Lois discrètes :

- **Bernoulli** $\mathcal{B}(p)$: $P(X=1) = p$, $P(X=0) = 1-p$; $E = p$,
 $\text{Var} = p(1-p)$.
- **binomiale** $\mathcal{B}(n,p)$: $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$; $E = np$,
 $\text{Var} = np(1-p)$.
- **géométrique** $\mathcal{G}(p)$: $P(X=k) = (1-p)^{k-1} p$ pour $k \geq 1$; $E = 1/p$,
 $\text{Var} = (1-p)/p^2$.
- **Poisson** $\mathcal{P}(\lambda)$: $P(X=k) = e^{-\lambda} \lambda^k / k!$; $E = \text{Var} = \lambda$.

Définition :

Lois à densité :

- **uniforme** sur $[a,b]$: $f = \frac{1}{b-a} 1_{[a,b]}$; $E = \frac{a+b}{2}$, $\text{Var} = \frac{(b-a)^2}{12}$.
- **exponentielle** $\mathcal{E}(\lambda)$: $f(x) = \lambda e^{-\lambda x} 1_{x>0}$; $E = 1/\lambda$, $\text{Var} = 1/\lambda^2$.
- **gamma** $\Gamma(a, \lambda)$: $f(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} 1_{x>0}$; $E = a/\lambda$, $\text{Var} = a/\lambda^2$.
- **normale** $\mathcal{N}(m, \sigma^2)$: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/(2\sigma^2)}$; $E = m$, $\text{Var} = \sigma^2$.
- **Cauchy** : $f(x) = \frac{1}{\pi(1+x^2)}$; elle n'a pas d'espérance.

Remarque :

La loi exponentielle est la seule loi à densité **sans mémoire** :

$P(X > s+t | X > s) = P(X > t)$. Par ailleurs, $\mathcal{E}(\lambda) = \Gamma(1, \lambda)$. Enfin, si Z suit $\mathcal{N}(0,1)$, alors $m + \sigma Z$ suit $\mathcal{N}(m, \sigma^2)$.

IV. Indépendance de tribus et de variables aléatoires

1. Définitions

Définition :

Des sous-tribus $\mathcal{A}_1, \dots, \mathcal{A}_n$ de $\mathcal{A}$ sont **indépendantes** si

$P(A_1 \cap \dots \cap A_n) = P(A_1) \dots P(A_n)$ pour tout choix de $A_i \in \mathcal{A}_i$. Des variables $X_1, \dots, X_n$ sont indépendantes si les tribus $\sigma(X_i) = X_i^{-1}(\mathcal{E}_i)$ le sont. Une famille infinie est indépendante si toute sous-famille finie l'est.

Attention :

L'indépendance deux à deux n'entraîne pas l'indépendance mutuelle. Par exemple, avec deux dés, les événements « premier dé pair », « second dé pair » et « somme paire » sont indépendants deux à deux, mais pas mutuellement.

Lemme :

Si, pour chaque i , la tribu $\mathcal{A}_i$ est engendrée par une classe $\mathcal{C}_i$ stable par intersection finie et contenant Ω , alors il suffit de vérifier la formule produit pour $A_i \in \mathcal{C}_i$.

Ce lemme résulte encore du théorème de classe monotone. Ainsi, des variables réelles $X_1, \dots, X_n$ sont indépendantes si et seulement si $P(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_i P(X_i \leq x_i)$ pour tous réels x_i .

2. Critères d'indépendance

Théorème :

Les assertions suivantes sont équivalentes :

- $X_1, \dots, X_n$ sont indépendantes ;
- la loi du vecteur $(X_1, \dots, X_n)$ est la mesure produit $P_{X_1} \otimes \dots \otimes P_{X_n}$;
- $E[h_1(X_1) \dots h_n(X_n)] = \prod_i E[h_i(X_i)]$ pour toutes fonctions h_i mesurables bornées.

Dans le cas à densité, X et Y sont indépendantes si et seulement si le couple a une densité de la forme $f(x, y) = g(x) k(y)$ presque partout.

Démonstration :

Supposons les variables indépendantes. Les deux mesures $P_{(X_1, \dots, X_n)}$ et $\bigotimes_i P_{X_i}$ coïncident sur les pavés $B_1 \times \dots \times B_n$. Or les pavés forment une classe stable par intersection qui engendre la tribu produit. Donc les deux mesures sont égales. Ensuite, le théorème de Fubini appliqué à la mesure produit donne la formule avec les h_i . Enfin, cette formule avec $h_i = 1_{B_i}$ redonne la définition.

Corollaire :

Si X et Y sont indépendantes et intégrables, alors XY est intégrable et $E[XY] = E[X] E[Y]$. De plus, si X et Y ont des densités f et g , la somme $X + Y$ a pour densité le produit de convolution $f * g(s) = \int f(x) g(s - x) dx$.

Méthode :

Pour démontrer l'indépendance de X et Y : calculer la loi du couple, par exemple sa densité par la fonction muette ; puis vérifier qu'elle se factorise en un produit d'une fonction de x et d'une fonction de y . Pour démontrer la non-indépendance, il suffit d'exhiber deux événements $\{X \in A\}$ et $\{Y \in B\}$ qui ne vérifient pas la formule produit.

Exemple :

Soient X et Y indépendantes de lois $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$. Alors, par la formule du binôme,

$$P(X+Y = n) = \sum_{k=0}^n e^{-\lambda} \frac{\lambda^k}{k!} e^{-\mu} \frac{\mu^{n-k}}{(n-k)!} = e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^n}{n!}.$$

Par conséquent, $X + Y$ suit la loi $\mathcal{P}(\lambda + \mu)$.

V. Vecteurs aléatoires et changement de variables

Un vecteur aléatoire est une variable aléatoire à valeurs dans $\mathbb{R}^d$. Ses coordonnées sont des variables réelles, dont les lois sont les **lois marginales**. Si le vecteur (X, Y) a une densité f , alors X a pour densité $x \rightarrow \int f(x, y) dy$, grâce au théorème de Fubini. En revanche, les marginales ne déterminent pas la loi du couple.

Théorème :

Soit X un vecteur aléatoire de densité f , nulle hors d'un ouvert U de $\mathbb{R}^d$. Soit $\varphi : U \rightarrow V$ un C^1 -difféomorphisme entre ouverts. Alors $Y = \varphi(X)$ a pour densité

$$g(y) = f(\varphi^{-1}(y)) |\det J_{\varphi^{-1}}(y)| 1_V(y).$$

Démonstration :

Soit h mesurable bornée. Par transfert, $E[h(Y)] = \int_U h(\varphi(x)) f(x) dx$. La formule de changement de variables, avec $x = \varphi^{-1}(y)$, donne

$$E[h(Y)] = \int_V h(y) f(\varphi^{-1}(y)) |\det J_{\varphi^{-1}}(y)| dy.$$

La méthode de la fonction muette permet alors de conclure.

Méthode :

Pour calculer la loi d'un vecteur image : compléter si besoin Y en un vecteur de même dimension que X ; vérifier que l'application est un difféomorphisme entre les bons ouverts ; exprimer l'application réciproque ; calculer la valeur absolue de son jacobien ; enfin, obtenir la marginale voulue en intégrant les variables auxiliaires.

Exemple :

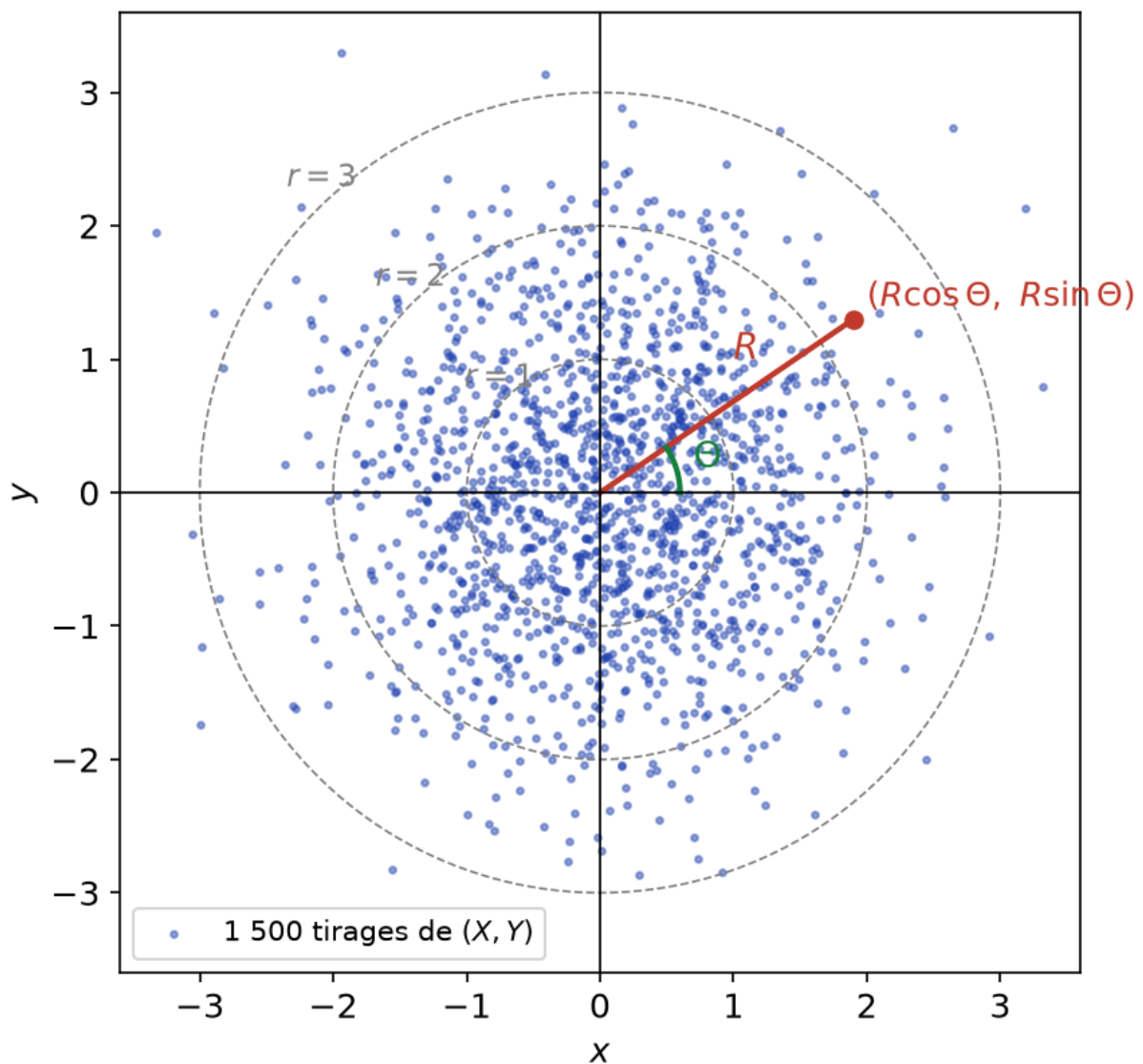
Soient X, Y indépendantes de loi $\mathcal{N}(0,1)$. Le couple a pour densité $\frac{1}{2\pi} e^{-(x^2+y^2)/2}$. On pose $(X, Y) = (R \cos \Theta, R \sin \Theta)$ avec $R > 0$ et $\Theta \in]0, 2\pi[$. Le jacobien du passage

en polaires vaut r . Donc (R, Θ) a pour densité

$$\frac{1}{2\pi} 1_{]0, 2\pi[}(\theta) \times r e^{-r^2/2} 1_{r > 0}.$$

Cette densité se factorise. Par conséquent, R et Θ sont indépendantes et Θ est uniforme sur $]0, 2\pi[$.

La figure ci-dessous illustre ce résultat. Le nuage de points est invariant par rotation, et la distance à l'origine ne renseigne pas sur l'angle.



VI. Covariance et vecteurs gaussiens

1. Covariance

Définition :

Pour X, Y de carré intégrable, la **covariance** est

$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$. Pour un vecteur $X = (X_1, \dots, X_d)$, la **matrice de covariance** est $\Gamma_X = (\text{Cov}(X_i, X_j))_{i,j}$.

Propriété :

La matrice Γ_X est symétrique et positive, car $a^T \Gamma_X a = \text{Var}(\langle a, X \rangle) \geq 0$. De plus, pour toute matrice A de taille $k \times d$, on a $\Gamma_{AX} = A \Gamma_X A^T$. Enfin, des variables indépendantes ont une covariance nulle, mais la réciproque est fausse.

Exemple :

Soit X uniforme sur $[-1, 1]$ et $Y = X^2$. Alors

$\text{Cov}(X, Y) = E[X^3] - E[X]E[X^2] = 0$ par imparité. Pourtant, Y est une fonction de X , donc les deux variables ne sont pas indépendantes.

2. Vecteurs gaussiens

Définition :

Un vecteur X de $\mathbb{R}^d$ est **gaussien** si toute combinaison linéaire $\langle a, X \rangle$ suit une loi normale,

éventuellement dégénérée (une constante est une loi normale de variance nulle).

Théorème :

Soit $\mathbf{X}$ un vecteur gaussien de moyenne $\mathbf{m}$ et de covariance $\mathbf{\Gamma}$. Alors :

- sa fonction caractéristique vaut $\varphi_{\mathbf{X}}(t) = \exp(i \langle t, \mathbf{m} \rangle - \frac{1}{2} t^T \mathbf{\Gamma} t)$; sa loi, notée $\mathcal{N}_d(\mathbf{m}, \mathbf{\Gamma})$, ne dépend donc que de $\mathbf{m}$ et $\mathbf{\Gamma}$;
- pour toute matrice $\mathbf{A}$ et tout vecteur $\mathbf{b}$, $\mathbf{AX} + \mathbf{b}$ est gaussien de loi $\mathcal{N}(\mathbf{Am} + \mathbf{b}, \mathbf{A}\mathbf{\Gamma}\mathbf{A}^T)$;
- si $\mathbf{\Gamma}$ est inversible, $\mathbf{X}$ a pour densité
$$\frac{1}{(2\pi)^{d/2} \sqrt{\det \mathbf{\Gamma}}} \exp\left(-\frac{1}{2} (\mathbf{x}-\mathbf{m})^T \mathbf{\Gamma}^{-1} (\mathbf{x}-\mathbf{m})\right) ;$$
- les coordonnées de $\mathbf{X}$ sont indépendantes si et seulement si $\mathbf{\Gamma}$ est diagonale. Plus généralement, deux sous-vecteurs de $\mathbf{X}$ sont indépendants si et seulement si leurs covariances croisées sont nulles.

Démonstration :

Pour le dernier point, supposons $\mathbf{\Gamma}$ diagonale de coefficients σ_j^2 . Alors

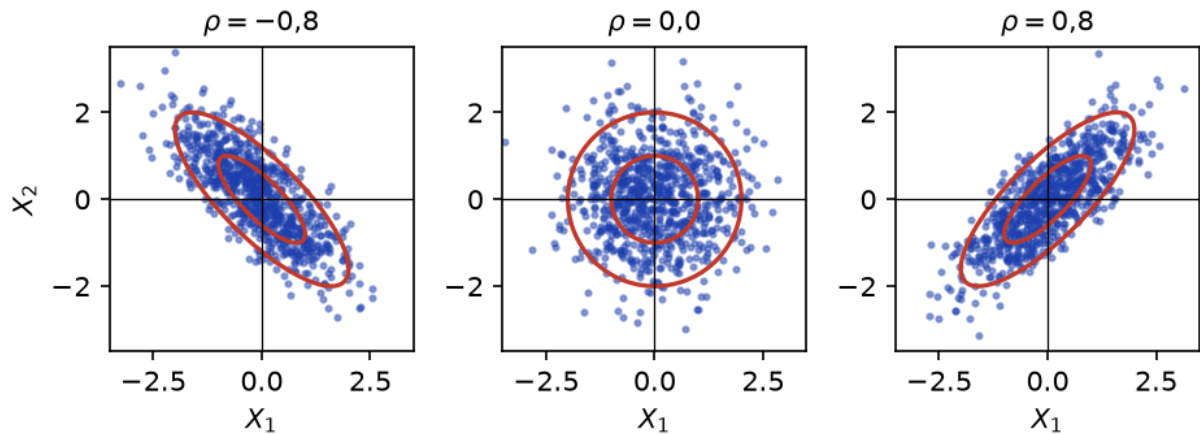
$$\varphi_{\mathbf{X}}(t) = \prod_j \exp(i t_j m_j - \frac{1}{2} \sigma_j^2 t_j^2) = \prod_j \varphi_{X_j}(t_j).$$
 Or la fonction caractéristique caractérise la loi (partie VII). Donc la loi de $\mathbf{X}$ est le produit des lois marginales. Autrement dit, les coordonnées sont indépendantes. La réciproque est vraie pour toutes variables indépendantes.

Attention :

Des variables gaussiennes non corrélées ne sont pas forcément indépendantes. L'hypothèse

essentielle est que le **vecteur** soit gaussien, et pas seulement chacune de ses coordonnées.

Les lignes de niveau de la densité gaussienne sont des ellipses centrées en m . Comme le montre la figure ci-dessous, leurs axes sont dirigés par les vecteurs propres de Γ . Ainsi, le coefficient de corrélation ρ règle l'inclinaison et l'aplatissement du nuage.



Exemple :

Pour simuler $\mathcal{N}_d(m, \Gamma)$, on choisit A telle que $AA^T = \Gamma$, par exemple par la factorisation de Cholesky. Ensuite, on pose $X = m + AZ$, où Z a des coordonnées indépendantes de loi $\mathcal{N}(0,1)$. En effet, Z est gaussien de covariance I_d , donc X est gaussien de covariance $A I_d A^T = \Gamma$.

VII. Fonction caractéristique

1. Définition et propriétés

Définition :

La **fonction caractéristique** d'un vecteur aléatoire X de $\mathbb{R}^d$ est $\varphi_X(t) = E[e^{i\langle t, X \rangle}]$, pour $t \in \mathbb{R}^d$. C'est la transformée de Fourier de la loi P_X .

Elle est toujours définie, car $|e^{i\langle t, X \rangle}| = 1$. De plus, $\varphi_X(0) = 1$, $|\varphi_X| \leq 1$ et φ_X est continue par convergence dominée. Enfin, $\varphi_{aX+b}(t) = e^{ibt} \varphi_X(at)$ en dimension 1.

Théorème :

La fonction caractéristique **caractérise la loi** : si $\varphi_X = \varphi_Y$, alors $P_X = P_Y$. De plus, si φ_X est intégrable sur $\mathbb{R}$, alors X a une densité continue donnée par la formule d'inversion

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi_X(t) dt.$$

Proposition :

Si X et Y sont indépendantes, alors $\varphi_{X+Y} = \varphi_X \varphi_Y$. De plus, X et Y sont indépendantes si et seulement si $\varphi_{(X,Y)}(s,t) = \varphi_X(s) \varphi_Y(t)$ pour tous s, t . Enfin, si $E[|X|^n] < +\infty$, alors φ_X est de classe C^n et $\varphi_X^{(k)}(0) = i^k E[X^k]$ pour $k \leq n$.

2. Calculs classiques

Démonstration :

Calculons φ pour X de loi $\mathcal{N}(0,1)$. Comme $|x e^{itx}| e^{-x^2/2}$ est intégrable indépendamment de t , on peut dériver sous l'intégrale :

$$\varphi'(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} i x e^{itx} e^{-x^2/2} dx.$$

Une intégration par parties, avec $x e^{-x^2/2}$ comme dérivée de $-e^{-x^2/2}$, donne $\varphi'(t) = -t \varphi(t)$. Or $\varphi(0) = 1$, donc $\varphi(t) = e^{-t^2/2}$. Ensuite, $\varphi_{m+\sigma Z}(t) = e^{imt - \sigma^2 t^2/2}$.

Méthode :

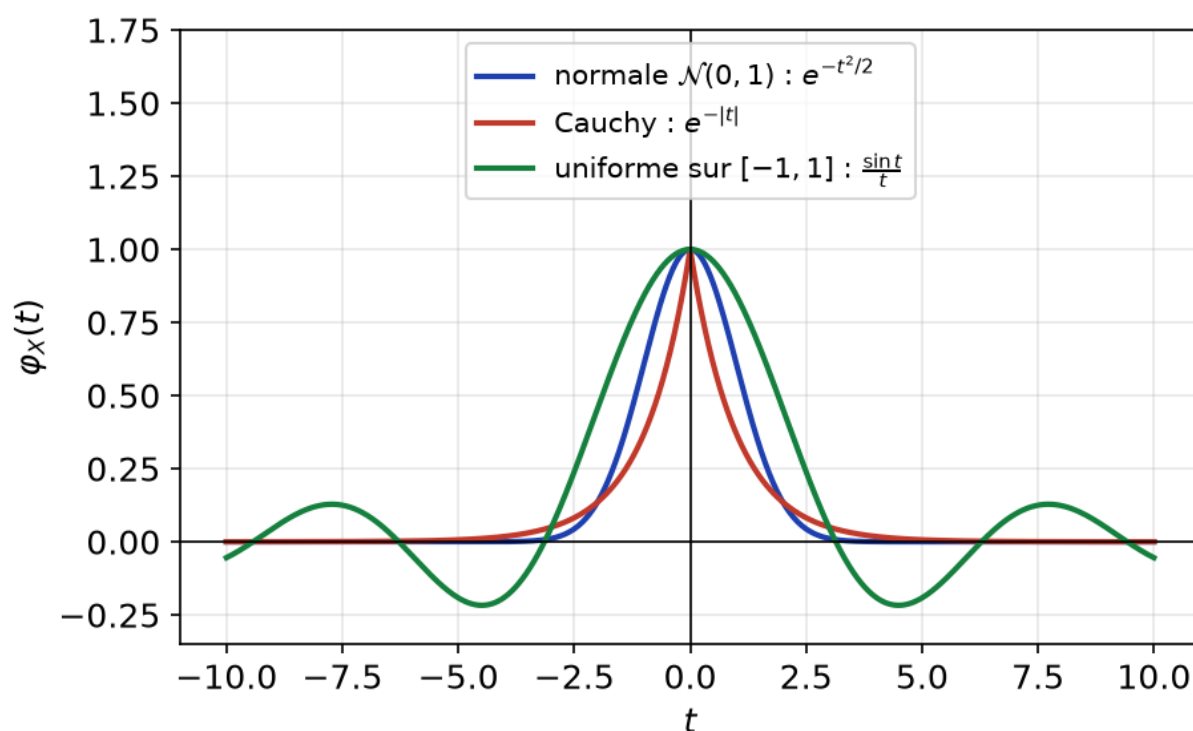
Pour calculer une fonction caractéristique : pour une loi discrète, sommer une série (souvent géométrique ou exponentielle) ; pour une densité, intégrer $e^{itx} f(x)$, en reconnaissant une exponentielle complexe ; si l'intégrale ne se calcule pas, établir une équation différentielle vérifiée par φ ; enfin, pour une somme de variables indépendantes, faire le produit des fonctions caractéristiques.

Exemple :

On obtient ainsi les résultats suivants :

- $\mathcal{B}(n, p) : (1 - p + p e^{it})^n$; $\mathcal{P}(\lambda) : \exp(\lambda(e^{it} - 1))$;
- uniforme sur $[-a, a] : \frac{\sin(at)}{at}$; $\mathcal{E}(\lambda) : \frac{\lambda}{\lambda - it}$;
- Cauchy : $e^{-|t|}$, qui n'est pas dérivable en 0, ce qui est cohérent avec l'absence d'espérance.

La figure ci-dessous compare trois fonctions caractéristiques. Une loi très étalée donne une fonction caractéristique concentrée près de 0. À l'inverse, une densité régulière donne une fonction caractéristique qui décroît vite.



VIII. Espérance conditionnelle

1. Cas discret

Définition :

Soit Y une variable discrète et X intégrable. Pour y tel que $P(Y = y) > 0$, on pose

$$E[X | Y = y] = \frac{E[X 1_{\{Y = y\}}]}{P(Y = y)}.$$

L'**espérance conditionnelle** de X sachant Y est la variable aléatoire $E[X | Y] = g(Y)$, où $g(y) = E[X | Y = y]$.

2. Cas à densité

Définition :

Si (X, Y) a une densité f et si $f_Y(y) > 0$, la **densité conditionnelle** de X sachant $Y = y$ est $f_{X|Y=y}(x) = f(x, y)/f_Y(y)$. On pose alors $g(y) = \int x f_{X|Y=y}(x) dx$ et $E[X|Y] = g(Y)$.

Théorème :

Dans les deux cas, $E[X|Y]$ est l'unique variable de la forme $g(Y)$, à égalité presque sûre près, telle que $E[X h(Y)] = E[g(Y) h(Y)]$ pour toute h mesurable bornée. En conséquence :

- $E[E[X|Y]] = E[X]$;
- $E[k(Y) X|Y] = k(Y) E[X|Y]$ pour k bornée ;
- si X est indépendante de Y , alors $E[X|Y] = E[X]$;
- si X est de carré intégrable, $E[X|Y]$ est la projection orthogonale de X sur les variables de la forme $k(Y)$ dans L^2 .

Démonstration :

Traisons le cas à densité. Par Fubini, on a

$$E[X h(Y)] = \int h(y) \left(\int x f(x, y) dx \right) dy = \int h(y) g(y) f_Y(y) dy = E[g(Y) h(Y)].$$

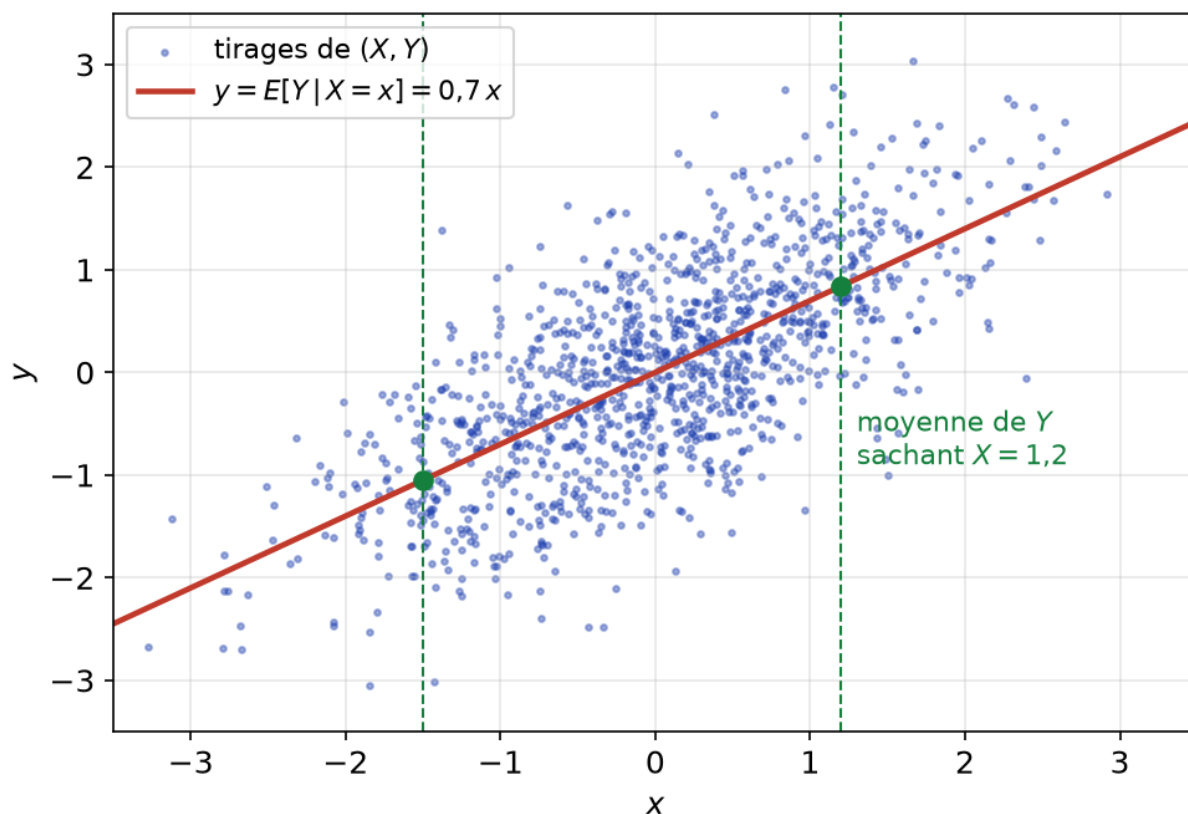
Pour l'unicité, si deux fonctions $g_1(Y)$ et $g_2(Y)$ conviennent, on choisit $h = 1_{\{g_1 > g_2\}}$.

On obtient alors $E[(g_1 - g_2)(Y) 1_{\{g_1(Y) > g_2(Y)\}}] = 0$, donc $g_1(Y) \leq g_2(Y)$ p.s. Par symétrie, on conclut. Enfin, on prend $h = 1$ pour obtenir le premier point.

Exemple :

Soit (X, Y) un vecteur gaussien centré, avec $\text{Var}(X) = \text{Var}(Y) = 1$ et $\text{Cov}(X, Y) = \rho$. La variable $Y - \rho X$ est non corrélée à X . Or le vecteur $(X, Y - \rho X)$ est gaussien, donc ces deux variables sont indépendantes. Par conséquent, $E[Y | X] = \rho X + E[Y - \rho X] = \rho X$.

La figure ci-dessous illustre cet exemple pour $\rho = 0,7$. Pour chaque abscisse x , la moyenne des ordonnées des points proches de la verticale est $0,7x$. Ainsi, l'espérance conditionnelle d'un vecteur gaussien est une fonction affine : c'est la droite de régression.



Ce qu'il faut retenir

- Un espace probabilisé est un espace mesuré de masse 1 ; une variable aléatoire est une fonction mesurable ; sa loi est la mesure image.
- L'espérance est une intégrale : linéarité, convergence monotone et dominée, Fubini s'appliquent directement.
- Le théorème de transfert calcule $E[h(X)]$ avec la loi seule ; lu à l'envers, c'est la méthode de la fonction muette.
- Des variables sont indépendantes si la loi du vecteur est le produit des lois ; pour des densités, il suffit que la densité jointe se factorise.
- La loi de $\varphi(X)$ s'obtient par changement de variables, avec la valeur absolue du jacobien de φ^{-1} .
- La covariance est bilinéaire ; $\Gamma_{AX} = A \Gamma_X A^T$; covariance nulle n'implique pas indépendance.
- Un vecteur gaussien est déterminé par m et Γ ; ses coordonnées sont indépendantes si et seulement si Γ est diagonale.
- La fonction caractéristique caractérise la loi et transforme les sommes indépendantes en produits.
- L'espérance conditionnelle $E[X | Y]$ est une fonction de Y , caractérisée par $E[X h(Y)] = E[E[X | Y] h(Y)]$.

Questions fréquentes sur variables aléatoires et espérance

Pourquoi définir l'espérance comme une intégrale de Lebesgue ?

Parce qu'une seule définition, $E[X] = \int X dP$, couvre à la fois les lois discrètes, les lois à densité et les lois mixtes. De plus, tous les théorèmes d'intégration deviennent disponibles : convergence monotone, convergence dominée, Fubini. On évite ainsi de refaire chaque démonstration deux fois.

Comment utiliser la méthode de la fonction muette ?

On fixe une fonction h mesurable bornée et on calcule $E[h(Y)]$ par le théorème de transfert. Ensuite, par un changement de variable, on met cette intégrale sous la forme $\int h(y)g(y)dy$. On conclut alors que g est la densité de Y . La méthode évite de calculer une fonction de répartition.

Deux variables gaussiennes non corrélées sont-elles indépendantes ?

Pas toujours. C'est vrai si le couple (X, Y) est un vecteur gaussien : sa loi ne dépend que de la moyenne et de la matrice de covariance, qui est alors diagonale. En revanche, avec X de loi $N(0,1)$ et $Y = \varepsilon X$, où ε vaut ± 1 indépendamment, X et Y sont gaussiennes, non corrélées, mais pas indépendantes.

À quoi sert la fonction caractéristique ?

Elle caractérise la loi et transforme une somme de variables indépendantes en produit. On l'utilise donc pour identifier la loi d'une somme, par exemple de variables normales ou de Poisson. Elle donne aussi les moments par dérivation en 0, et elle sera l'outil central du théorème central limite.

Pour aller plus loin en L3

- Les énoncés : [exercices de maths en L3 sur variables aléatoires et espérance](#)
- À maîtriser avant : [Intégrale de Lebesgue et théorèmes de convergence](#), [Mesure produit, Fubini et changement de variables](#)
- Chapitre précédent : [Espaces de Hilbert et analyse de Fourier](#)
- Chapitre suivant : [Convergences, grands nombres et central limite](#)
- Tester vos connaissances : [QCM de maths en L3 par chapitre](#)
- Le sommaire : [tous les chapitres de maths de L3](#) et la [licence de maths de L1 à L3](#)