

Variables aléatoires discrètes

I. Variables aléatoires discrètes et loi

On fixe un espace probabilisé $(\Omega, \mathcal{A}, P)$. Une variable aléatoire décrit une quantité numérique qui dépend du hasard. Dans ce chapitre, elle ne prend qu'un nombre fini ou dénombrable de valeurs. Ainsi, toutes les probabilités se calculent par des sommes, finies ou non. C'est pourquoi la théorie des séries joue ici un rôle central.

Définition :

Une **variable aléatoire discrète** est une application $X : \Omega \rightarrow \mathbb{R}$ telle que $X(\Omega)$ est fini ou dénombrable et telle que, pour tout $x \in X(\Omega)$, l'ensemble $\{X = x\} = X^{-1}(\{x\})$ appartient à $\mathcal{A}$.

Les événements $\{X = x\}$, pour $x \in X(\Omega)$, sont deux à deux disjoints et leur réunion vaut Ω . Autrement dit, ils forment un système complet d'événements. Par conséquent, la σ -additivité donne

$$\sum_{x \in X(\Omega)} P(X = x) = 1.$$

Définition :

La **loi** de X est la donnée de $X(\Omega)$ et des nombres $p_x = P(X = x)$ pour $x \in X(\Omega)$. Pour toute partie B de $\mathbb{R}$, on a alors

$$P(X \in B) = \sum_{x \in X(\Omega) \cap B} p_x.$$

Réciproquement, toute famille (p_x) de réels positifs indexée par un ensemble dénombrable et de somme 1 est la loi d'une variable aléatoire sur un espace convenable. On admet ce résultat. En pratique, il suffit donc de vérifier la positivité et la somme égale à 1.

Définition :

La **fonction de répartition** de X est $F_X: \mathbb{R} \rightarrow [0, 1]$, $F_X(x) = P(X \leq x)$.

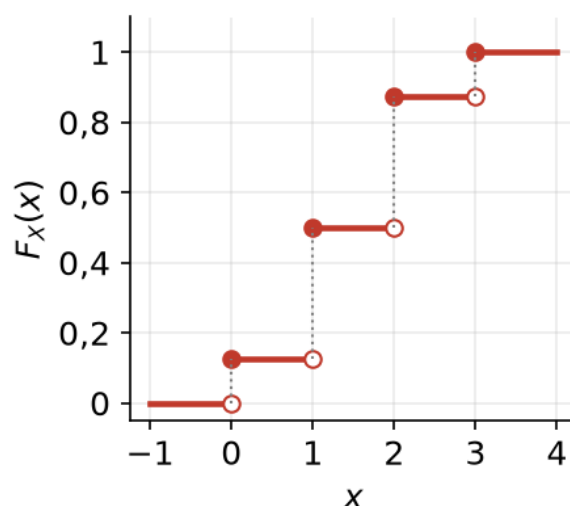
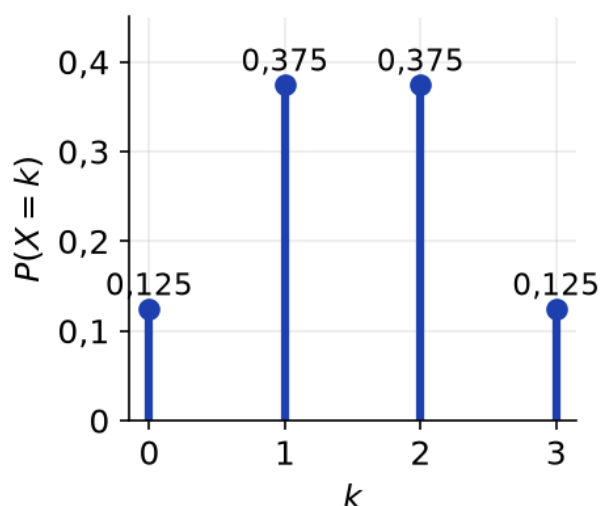
Propriété :

La fonction F_X est croissante, continue à droite, de limite 0 en $-\infty$ et 1 en $+\infty$. De plus, pour tout réel x , $P(X = x) = F_X(x) - \lim_{t \rightarrow x^-} F_X(t)$. Enfin, F_X caractérise la loi de X .

Démonstration :

Si $x \leq y$, alors $\{X \leq x\} \subset \{X \leq y\}$, d'où la croissance. Ensuite, on applique la continuité monotone de P . La suite décroissante $\{X \leq x + 1/n\}$ a pour intersection $\{X \leq x\}$, ce qui donne la continuité à droite. De même, la suite croissante $\{X \leq x - 1/n\}$ a pour réunion $\{X < x\}$. Donc la limite à gauche vaut $P(X < x)$, et le saut vaut $P(X = x)$. Les limites en $\pm\infty$ se traitent de la même façon.

Pour une variable discrète, la fonction de répartition est donc une fonction en escalier. Elle saute de p_x en chaque valeur x . La figure ci-dessous montre le diagramme en bâtons et la fonction de répartition du nombre de piles obtenus en trois lancers d'une pièce équilibrée.



Exemple :

Soit X le nombre de piles en trois lancers. Alors $X(\Omega) = \{0, 1, 2, 3\}$ et $P(X = k) = \binom{3}{k} / 8$. Ainsi, $F_X(x) = 0$ pour $x < 0$, $F_X(x) = 1/8$ sur $[0, 1[$, $5/8$ sur $[1, 2[$, $7/8$ sur $[2, 3[$ et 1 pour $x \geq 3$.

II. Loys discrètes usuelles

1. Loys finies

Les loys suivantes modélisent des expériences répétées un nombre fini de fois. D'abord, la loi uniforme traduit l'équiprobabilité. Ensuite, les lois de Bernoulli et binomiale comptent des succès. Enfin, la loi hypergéométrique décrit des tirages sans remise.

Définition :

- **Loi uniforme** sur $\{1, \dots, n\}$: $P(X = k) = \frac{1}{n}$ pour $1 \leq k \leq n$.

- **Loi de Bernoulli** $\mathcal{B}(p)$, avec $p \in [0, 1]$: $P(X = 1) = p$ et $P(X = 0) = 1 - p$.
- **Loi binomiale** $\mathcal{B}(n, p)$: $P(X = k) = \binom{n}{k} p^k (1 - p)^{n - k}$ pour $0 \leq k \leq n$.
- **Loi hypergéométrique** $\mathcal{H}(N, K, n)$: on tire sans remise n boules dans une urne de N boules dont K sont gagnantes, et X compte les gagnantes. Alors

$$P(X = k) = \binom{K}{k} \binom{N - K}{n - k} / \binom{N}{n}.$$

La somme des probabilités binomiales vaut $(p + 1 - p)^n = 1$ par la formule du binôme. De même, pour la loi hypergéométrique, la somme vaut 1 grâce à la formule de Vandermonde. En effet,

$$\sum_k \binom{K}{k} \binom{N - K}{n - k} = \binom{N}{n}.$$

Propriété :

Si n épreuves de Bernoulli indépendantes ont la même probabilité de succès p , le nombre de succès suit la loi $\mathcal{B}(n, p)$.

2. Loix infinies : géométrique et de Poisson

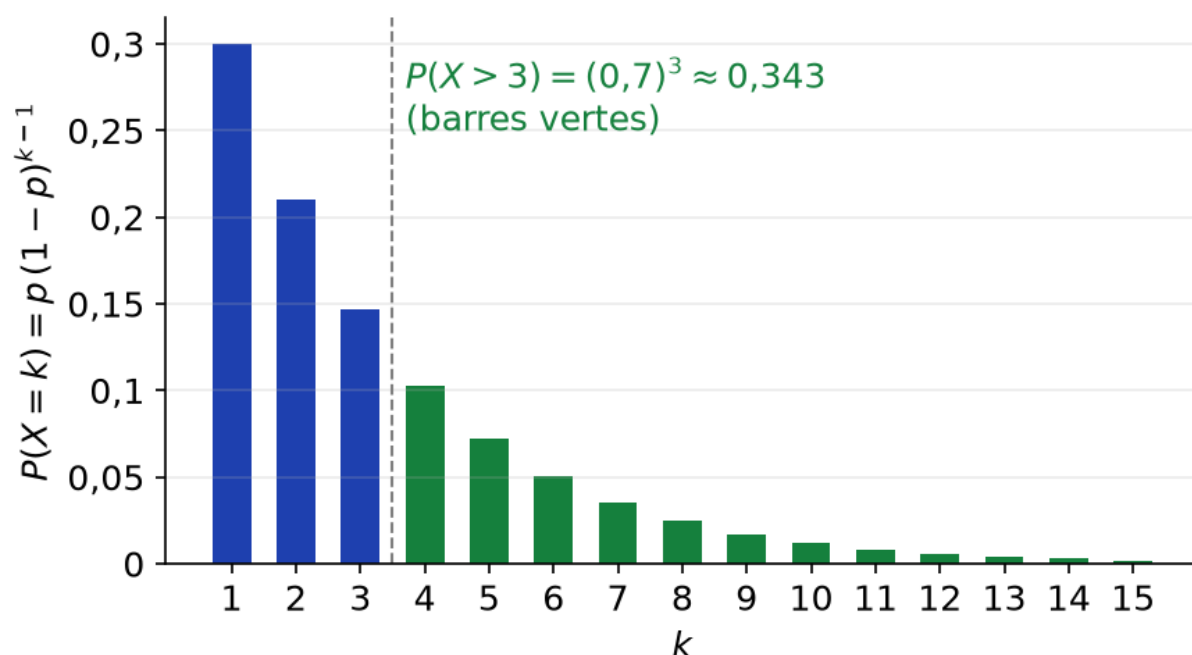
Définition :

- **Loi géométrique** $\mathcal{G}(p)$, $p \in]0, 1]$: $X(\mathcal{O}) = \mathbb{N}^*$ et
 $P(X = k) = p (1 - p)^{k - 1}$. C'est le rang du premier succès dans une suite d'épreuves de Bernoulli indépendantes.
- **Loi de Poisson** $\mathcal{P}(\lambda)$, $\lambda > 0$: $X(\mathcal{O}) = \mathbb{N}$ et $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$.

Pour la loi géométrique, on reconnaît une série géométrique de raison $q = 1 - p$. Ainsi,

$$\sum_{k \geq 1} p q^{k-1} = \frac{p}{1-q} = 1. \text{ Pour la loi de Poisson, la série exponentielle donne } \sum_k \frac{\lambda^k}{k!} = e^\lambda.$$

De plus, la queue de la loi géométrique est simple : $P(X > n) = q^n$, car les n premières épreuves sont des échecs. La figure suivante l'illustre pour $p = 0,3$.



Propriété :

La loi géométrique est **sans mémoire** : pour tous entiers $m, n \geq 0$,

$P(X > m + n \mid X > m) = P(X > n)$. Réciproquement, une variable à valeurs dans $\mathbb{N}^*$ sans mémoire, avec $P(X > 1) < 1$, suit une loi géométrique.

Démonstration :

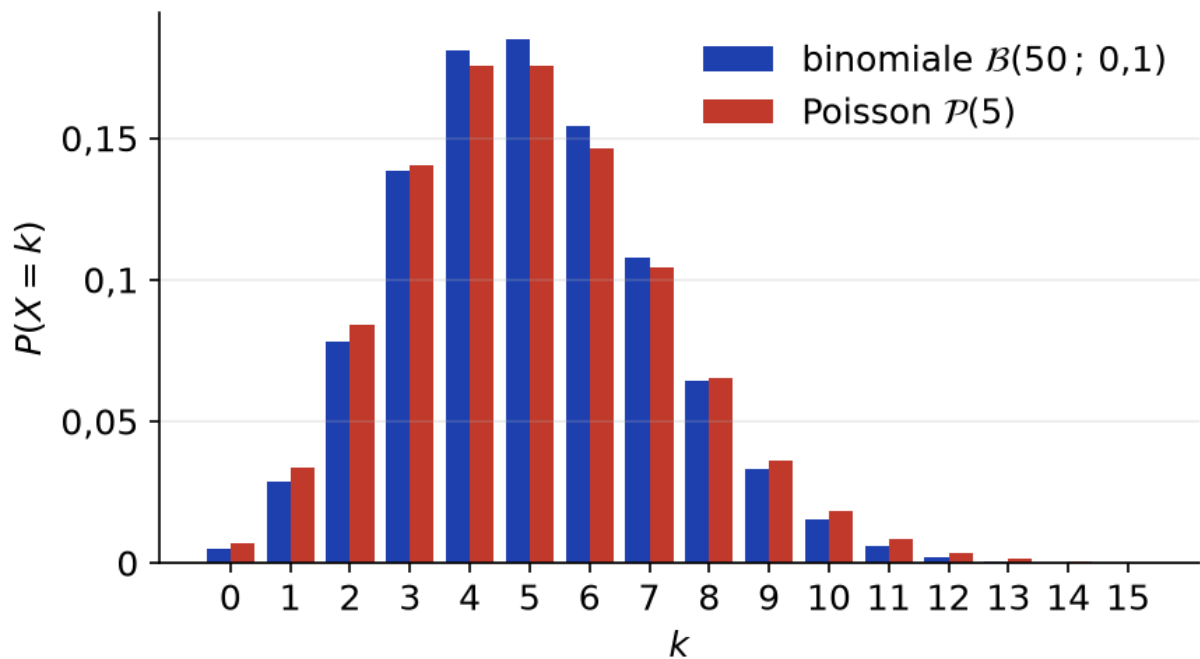
Comme $\{X > m + n\} \subset \{X > m\}$, on obtient

$P(X > m + n \mid X > m) = \frac{q^{m+n}}{q^m} = q^n$. Pour la réciproque, posons $u_n = P(X > n)$. L'hypothèse s'écrit $u_{m+n} = u_m u_n$, donc $u_n = u_1^n$. On pose alors $p = 1 - u_1 > 0$, et $P(X = n) = u_{n-1} - u_n = p(1-p)^{n-1}$.

Proposition :

Si $p_n \in [0, 1]$ et $n p_n \rightarrow \lambda > 0$, alors pour tout entier k ,
$$\binom{n}{k} p_n^k (1 - p_n)^{n-k} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!} \text{ quand } n \rightarrow +\infty.$$

En effet, $\binom{n}{k} p_n^k \sim \frac{(n p_n)^k}{k!}$, et $(1 - p_n)^{n-k} = \exp((n-k) \ln(1 - p_n)) \rightarrow e^{-\lambda}$. Par conséquent, la loi de Poisson approche une binomiale de grand n et de petit p . Comme le montre la figure ci-dessous, $\mathcal{B}(50, 0,1)$ et $\mathcal{P}(5)$ sont déjà très proches.



III. Espérance d'une variable aléatoire discrète

L'espérance est la moyenne des valeurs pondérée par leurs probabilités. Cependant, lorsque $X(\omega)$ est infini, cette moyenne est une série. Or l'ordre d'énumération de $X(\omega)$ n'est pas canonique. On exige donc la convergence absolue, qui rend la somme indépendante de l'ordre.

Définition :

On dit que X **admet une espérance** si la famille $(x P(X = x))_{x \in X(\Omega)}$ est sommable, c'est-à-dire si $\sum_{x \in X(\Omega)} |x| P(X = x) < +\infty$. On pose alors

$$E(X) = \sum_{x \in X(\Omega)} x P(X = x).$$

Remarque :

Si X est positive, la somme $\sum x P(X = x)$ a toujours un sens dans $[0, +\infty]$. On écrit alors $E(X) = +\infty$ quand elle diverge. De plus, toute variable bornée, en particulier toute variable finie, admet une espérance.

Exemple :

Si $X \sim \mathcal{G}(p)$, la série entière $\sum k x^{k-1}$ a pour somme $\frac{1}{(1-x)^2}$ sur $] -1, 1[$. Les termes sont positifs, donc la convergence absolue est acquise, et

$$E(X) = p \sum_{k \geq 1} k q^{k-1} = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

1. Théorème de transfert

On veut souvent l'espérance de $f(X)$ sans calculer la loi de $f(X)$. Le théorème suivant l'autorise. Sa preuve repose sur la sommation par paquets des familles sommables.

Théorème :

Théorème de transfert. Soit $f : X(\Omega) \rightarrow \mathbb{R}$. La variable $f(X)$ admet une espérance si et

seulement si $\sum_{x \in X(\Omega)} |f(x)| P(X = x) < +\infty$. Dans ce cas,

$$E(f(X)) = \sum_{x \in X(\Omega)} f(x) P(X = x).$$

Démonstration :

Posons $Y = f(X)$. Pour $y \in Y(\Omega)$, l'ensemble $I_y = f^{-1}(\{y\}) \cap X(\Omega)$ est dénombrable, et $P(Y = y) = \sum_{x \in I_y} P(X = x)$. Les I_y partitionnent $X(\Omega)$. D'après le théorème de sommation par paquets pour les familles positives,

$$\sum_x |f(x)| P(X = x) = \sum_y \sum_{x \in I_y} |y| P(X = x) = \sum_y |y| P(Y = y).$$

Ainsi, les deux sommabilités sont équivalentes. Ensuite, on reprend le même calcul sans valeurs absolues, ce qui est licite pour une famille sommable.

2. Propriétés de l'espérance

Théorème :

Soient X et Y deux variables discrètes admettant une espérance, et a, b deux réels.

- **Linéarité** : $aX + bY$ admet une espérance et $E(aX + bY) = a E(X) + b E(Y)$.
- **Positivité** : si $X \geq 0$, alors $E(X) \geq 0$, avec égalité si et seulement si $P(X = 0) = 1$.
- **Croissance** : si $X \leq Y$, alors $E(X) \leq E(Y)$.
- **Domination** : si $|X| \leq Y$ et si Y admet une espérance, alors X en admet une.

Pour la linéarité, on applique le transfert au couple (X, Y) , qui est une variable discrète à valeurs dans $\mathbb{R}^2$. En effet, $\sum_{(x, y)} |ax + by| P(X = x, Y = y)$ est majorée par $|a| E(|X|) + |b| E(|Y|)$ grâce aux lois marginales. Les autres points s'en déduisent.

Méthode :

Pour calculer une espérance :

1. si X est une somme de variables simples (par exemple d'indicateurs), utilisez la linéarité, sans chercher la loi de X ;
2. sinon, écrivez la série $\sum x P(X = x)$ et justifiez sa convergence absolue ;
3. reconnaissez une série géométrique dérivée, une série exponentielle ou une somme télescopique ;
4. pour $E(f(X))$, appliquez le théorème de transfert, sans déterminer la loi de $f(X)$.

Exemple :

Pour $X \sim \mathcal{B}(n, p)$, écrivez $X = X_1 + \dots + X_n$, où X_i est l'indicateur du succès i . Chaque X_i suit $\mathcal{B}(p)$, donc $E(X_i) = p$. Par linéarité, $E(X) = np$. De même, pour la loi $\mathcal{H}(N, K, n)$, chaque tirage donne une boule gagnante avec probabilité K/N . Ainsi, $E(X) = nK/N$.

IV. Variance, écart type et covariance

L'espérance ne mesure pas la dispersion. Pour cela, on étudie l'écart quadratique moyen à l'espérance. Il faut d'abord que X^2 admette une espérance.

Proposition :

Si X^2 admet une espérance, alors X admet une espérance. De plus, si X^2 et Y^2 admettent une espérance, XY en admet une.

Démonstration :

On utilise les inégalités $|x| \leq \frac{1+x^2}{2}$ et $|xy| \leq \frac{x^2+y^2}{2}$. Il suffit ensuite d'appliquer le critère de domination.

Définition :

Si $E(X^2)$ existe, la **variance** de X est $V(X) = E((X - E(X))^2)$. L'**écart type** est $\sigma(X) = \sqrt{V(X)}$.

Propriété :

Formule de König-Huygens : $V(X) = E(X^2) - E(X)^2$. De plus, $V(aX + b) = a^2 V(X)$. Enfin, $V(X) = 0$ si et seulement si X est presque sûrement constante.

La formule de König-Huygens se prouve en développant le carré, puis par linéarité. D'autre part, pour les lois à valeurs entières, on calcule souvent $E(X(X-1))$ plutôt que $E(X^2)$. En effet, le facteur $k(k-1)$ se simplifie bien avec $k!$ ou avec une dérivée seconde.

Exemple :

Soit $X \sim \mathcal{P}(\lambda)$. Par transfert,

$$E(X(X-1)) = \sum_{k \geq 2} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} = \lambda^2 e^{-\lambda} \sum_{j \geq 0} \frac{\lambda^j}{j!} = \lambda^2. \text{ De même, } E(X) = \lambda. \text{ Donc } V(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Définition :

Si X et Y admettent un moment d'ordre 2, leur **covariance** est

$$\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y). \text{ Si } \sigma(X)\sigma(Y) \neq 0, \text{ le coefficient de corrélation est } \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

Théorème :

La covariance est bilinéaire et symétrique, et $\text{Cov}(X, X) = V(X)$. De plus :

- $|\text{Cov}(X, Y)| \leq \sigma(X)\sigma(Y)$, donc $|\rho(X, Y)| \leq 1$;
- $V(X_1 + \dots + X_n) = \sum_{i=1}^n V(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j).$

Démonstration :

La première inégalité est l'inégalité de Cauchy-Schwarz pour la forme bilinéaire symétrique positive $(U, V) \rightarrow E(UV)$, appliquée à $U = X - E(X)$ et $V = Y - E(Y)$. Pour la variance d'une somme, on développe $\text{Cov}(\sum_i X_i, \sum_j X_j)$ par bilinéarité.

Le tableau ci-dessous rassemble les moments des lois usuelles, avec $q = 1 - p$.

Propriété :

- Uniforme sur $\{1, \dots, n\}$: $E(X) = \frac{n+1}{2}$, $V(X) = \frac{n^2-1}{12}$.
- Bernoulli $\mathcal{B}(p)$: $E(X) = p$, $V(X) = pq$.
- Binomiale $\mathcal{B}(n, p)$: $E(X) = np$, $V(X) = npq$.
- Géométrique $\mathcal{G}(p)$: $E(X) = \frac{1}{p}$, $V(X) = \frac{q}{p^2}$.
- Poisson $\mathcal{P}(\lambda)$: $E(X) = V(X) = \lambda$.

V. Couples de variables aléatoires et indépendance

1. Loi conjointe, lois marginales et lois conditionnelles

Un couple (X, Y) de variables discrètes est une variable discrète à valeurs dans $\mathbb{R}^2$. Sa loi, dite **loi conjointe**, est la famille des $P(X = x, Y = y)$. On en déduit les lois de X et de Y , appelées lois marginales.

Propriété :

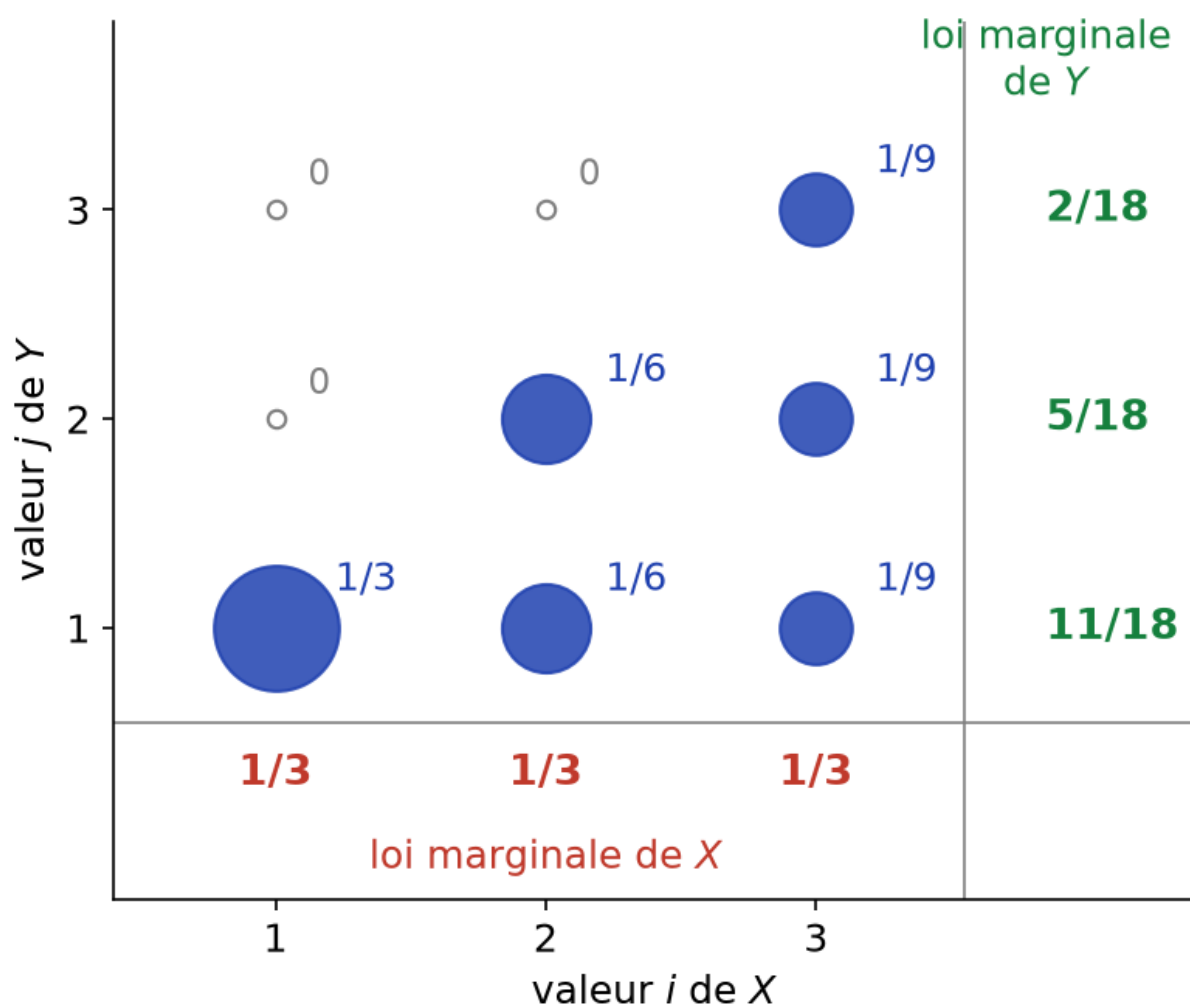
Pour tout $x \in X(\Omega)$, $P(X = x) = \sum_{y \in Y(\Omega)} P(X = x, Y = y)$. Si $P(Y = y) > 0$, la **loi conditionnelle** de X sachant $\{Y = y\}$ est $x \rightarrow P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$.

La première formule découle de la formule des probabilités totales, appliquée au système complet $(\{Y = y\})_y$. Attention : les lois marginales ne déterminent pas la loi conjointe en général.

Exemple :

On choisit X uniformément dans $\{1, 2, 3\}$. Ensuite, sachant $X = i$, on choisit Y uniformément dans $\{1, \dots, i\}$. Alors $P(X = i, Y = j) = \frac{1}{3i}$ pour $j \leq i$. La loi marginale de Y s'obtient en sommant par ligne : $P(Y = 1) = \frac{1}{3} + \frac{1}{6} + \frac{1}{9} = \frac{11}{18}$, $P(Y = 2) = \frac{5}{18}$ et $P(Y = 3) = \frac{2}{18}$.

Comme le montre la figure ci-dessous, chaque disque a une aire proportionnelle à $P(X = i, Y = j)$. Les lois marginales se lisent en marge, d'où leur nom.



2. Indépendance

Définition :

Les variables discrètes $X_1, \dots, X_n$ sont **indépendantes** si, pour tous $x_1, \dots, x_n$,
$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i).$$
 Une suite de variables est indépendante si toute sous-famille finie l'est.

Théorème :

Soient X et Y indépendantes.

- Pour toutes fonctions f et g , les variables $f(X)$ et $g(Y)$ sont indépendantes.
- Si X et Y admettent une espérance, XY aussi, et $E(XY) = E(X)E(Y)$.
- Par conséquent, $\text{Cov}(X, Y) = 0$ et $V(X + Y) = V(X) + V(Y)$ dès que les variances existent.

Démonstration :

Traitons le deuxième point. Par transfert appliqué au couple, puis par indépendance,

$$\sum_{x,y} |xy| P(X = x, Y = y) = \sum_x |x| P(X = x) \sum_y |y| P(Y = y).$$
 Ce produit est fini, donc la famille est sommable. On refait alors le calcul sans valeurs absolues, grâce au théorème de Fubini pour les familles sommables.

Attention :

La réciproque est fautive. Si X est uniforme sur $\{-1, 0, 1\}$ et $Y = X^2$, alors $\text{Cov}(X, Y) = E(X^3) = 0$. Pourtant, $P(X = 0, Y = 1) = 0 \neq P(X = 0)P(Y = 1) = \frac{2}{9}$.

3. Loi d'une somme de variables indépendantes

Propriété :

Si X et Y sont indépendantes et à valeurs dans $\mathbb{N}$, alors, pour tout $n \in \mathbb{N}$,
$$P(X + Y = n) = \sum_{k=0}^n P(X = k) P(Y = n - k).$$
C'est le **produit de convolution** des deux lois.

En effet, $\{X + Y = n\}$ est la réunion disjointe des $\{X = k, Y = n - k\}$ pour $0 \leq k \leq n$. Il reste à utiliser l'indépendance.

Méthode :

Pour déterminer la loi d'une somme $S = X + Y$ de variables indépendantes :

1. déterminez d'abord $S(\Omega)$;
2. décomposez $\{S = n\}$ selon les valeurs de X , en respectant les contraintes de $Y(\Omega)$;
3. factorisez $P(X = k, Y = n - k)$ grâce à l'indépendance, puis reconnaissez une somme connue (formule du binôme, Vandermonde) ;
4. à défaut, passez par les fonctions génératrices (partie VI).

Exemple :

Si $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ sont indépendantes, alors

$P(X + Y = n) = e^{-(\lambda + \mu)} \sum_{k=0}^n \frac{\lambda^k \mu^{n-k}}{k! (n-k)!} = e^{-(\lambda + \mu)} \frac{(\lambda + \mu)^n}{n!}$ par la formule du binôme. Ainsi, $X + Y \sim \mathcal{P}(\lambda + \mu)$.

VI. Fonction génératrice

Pour une variable à valeurs entières, on range les probabilités comme coefficients d'une série entière. On transforme alors les convolutions en produits. De plus, les moments se lisent sur les dérivées en 1.

Définition :

Soit X à valeurs dans $\mathbb{N}$. Sa **fonction génératrice** est

$$G_X(t) = E(t^X) = \sum_{n=0}^{+\infty} P(X = n) t^n.$$

Propriété :

Le rayon de convergence de cette série entière est au moins 1. Elle converge normalement sur $[-1, 1]$, et $G_X(1) = 1$. De plus, G_X caractérise la loi : $P(X = n) = \frac{G_X^{(n)}(0)}{n!}$.

En effet, pour $|t| \leq 1$, on a $|P(X = n) t^n| \leq P(X = n)$, terme général d'une série convergente. Ensuite, l'unicité des coefficients d'une série entière donne la caractérisation.

Théorème :

La variable X admet une espérance si et seulement si G_X est dérivable à gauche en 1 , et alors $E(X) = G'_X(1)$. De même, X admet un moment d'ordre 2 si et seulement si G_X est deux fois dérivable à gauche en 1 , et alors $E(X(X-1)) = G''_X(1)$ et

$$V(X) = G''_X(1) + G'_X(1) - G'_X(1)^2.$$

Remarque :

Si le rayon de convergence est strictement supérieur à 1 , ce théorème résulte simplement de la dérivation terme à terme d'une série entière à l'intérieur de son disque. Le cas d'un rayon égal à 1 demande un argument de convergence monotone, que l'on admet.

Théorème :

Si X et Y sont indépendantes et à valeurs dans $\mathbb{N}$, alors $G_{X+Y} = G_X G_Y$ sur $[-1, 1]$.

Démonstration :

Les variables t^X et t^Y sont indépendantes, comme fonctions de X et de Y . Elles sont bornées pour $|t| \leq 1$. Donc $E(t^{X+Y}) = E(t^X t^Y) = E(t^X) E(t^Y)$.

Voici les fonctions génératrices usuelles, avec $q = 1 - p$:

Propriété :

- $\mathcal{B}(p) : G(t) = q + pt ; \mathcal{B}(n, p) : G(t) = (q + pt)^n .$
- $\mathcal{G}(p) : G(t) = \frac{pt}{1 - qt}$ pour $|t| < \frac{1}{q} .$
- $\mathcal{P}(\lambda) : G(t) = e^{\lambda(t - 1)}$ pour tout réel $t .$

Méthode :

Pour trouver la loi d'une somme de variables indépendantes à valeurs entières, calculez le produit des fonctions génératrices. Ensuite, reconnaissez la fonction génératrice d'une loi connue. Enfin, concluez par unicité. Par exemple, si $X \sim \mathcal{B}(n, p)$ et $Y \sim \mathcal{B}(m, p)$ sont indépendantes, $G_{X+Y}(t) = (q + pt)^{n+m}$, donc $X + Y \sim \mathcal{B}(n + m, p)$.

Exemple :

Pour $X \sim \mathcal{G}(p)$, on a $G_X(t) = \frac{pt}{1 - qt}$, dont le rayon vaut $\frac{1}{q} > 1$. On calcule $G'_X(t) = \frac{p}{(1 - qt)^2}$ et $G''_X(t) = \frac{2pq}{(1 - qt)^3}$. Ainsi, $E(X) = \frac{1}{p}$ et $V(X) = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2}$.

VII. Inégalités de concentration et loi faible des grands nombres

1. Inégalités de Markov et de Bienaymé-Tchebychev

Théorème :

Inégalité de Markov. Si $X \geq 0$ admet une espérance, alors pour tout $a > 0$,
$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Démonstration :

On a l'inégalité $a 1_{\{X \geq a\}} \leq X$, car $X \geq 0$. Par croissance de l'espérance,
$$a P(X \geq a) \leq E(X).$$

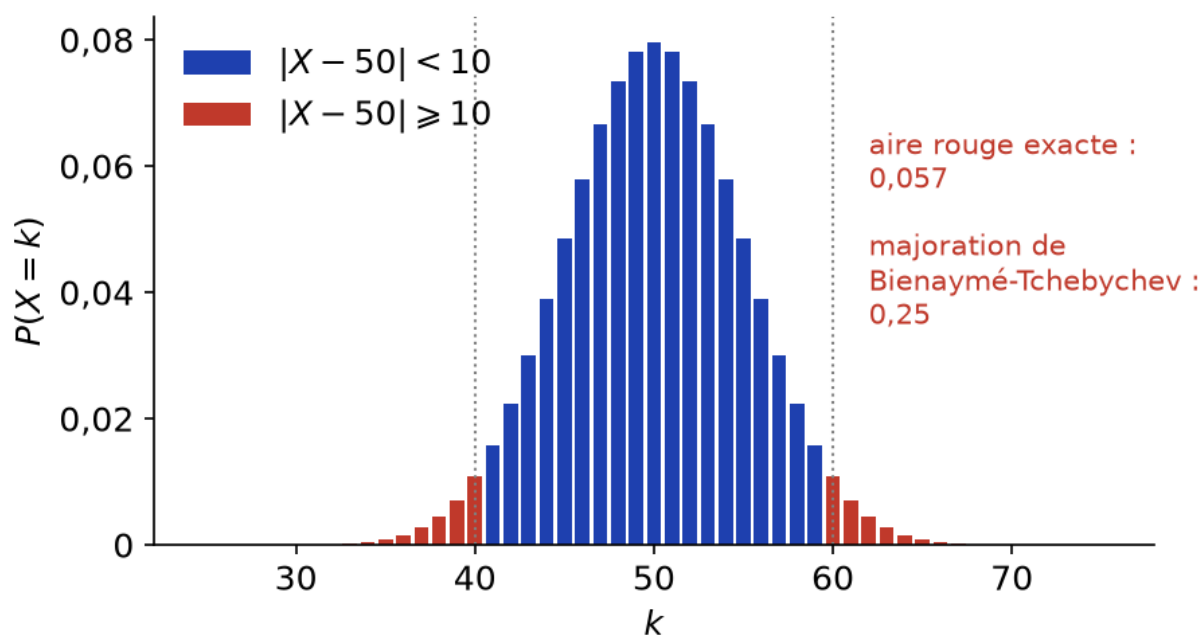
Corollaire :

Inégalité de Bienaymé-Tchebychev. Si X admet une variance, alors pour tout $a > 0$,
$$P(|X - E(X)| \geq a) \leq \frac{V(X)}{a^2}.$$

Démonstration :

On applique l'inégalité de Markov à la variable positive $(X - E(X))^2$ et au réel a^2 . En effet, les événements $\{|X - E(X)| \geq a\}$ et $\{(X - E(X))^2 \geq a^2\}$ sont égaux.

Ces majorations sont universelles, donc souvent grossières. Par exemple, pour $X \sim \mathcal{B}(100, 1/2)$, on a $V(X) = 25$. Ainsi, $P(|X - 50| \geq 10) \leq 0,25$. Or la valeur exacte est proche de 0,057, comme le montre la figure ci-dessous.



2. Loi faible des grands nombres

Théorème :

Loi faible des grands nombres. Soit $(X_n)_{n \geq 1}$ une suite de variables indépendantes, de même loi, admettant une variance σ^2 . On note m leur espérance et $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$. Alors, pour tout $\varepsilon > 0$,

$$P(|\bar{X}_n - m| \geq \varepsilon) \leq \frac{\sigma^2}{n \varepsilon^2} \xrightarrow[n \rightarrow +\infty]{} 0.$$

Démonstration :

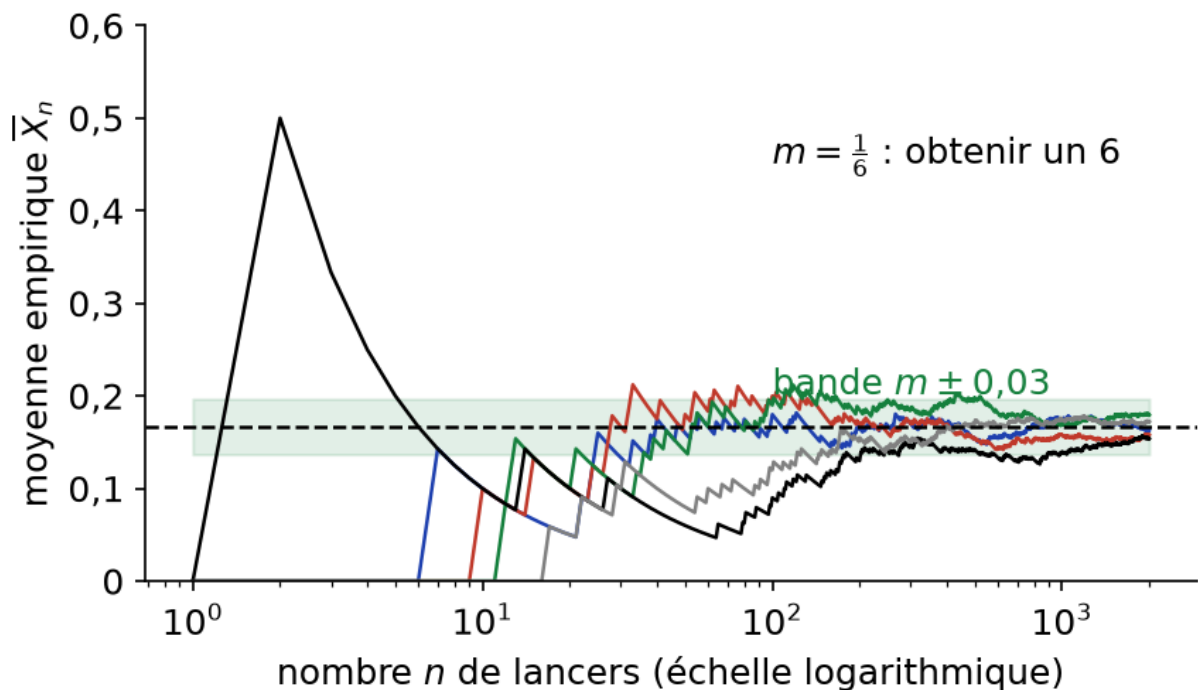
Par linéarité, $E(\bar{X}_n) = m$. Ensuite, les X_i sont indépendantes, donc leurs covariances sont nulles. Ainsi, $V(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{\sigma^2}{n}$. Il reste à appliquer l'inégalité de Bienaymé-Tchebychev.

Remarque :

La preuve n'utilise que la nullité des covariances et une majoration uniforme des variances. Elle s'étend donc à des variables deux à deux non corrélées. Plus généralement, il suffit que

$$V(X_1 + \dots + X_n) = o(n^2).$$

On dit que $\bar{X}_n$ converge en probabilité vers m . Concrètement, la fréquence d'un événement se stabilise autour de sa probabilité. La figure suivante montre cinq suites de moyennes empiriques pour l'événement « obtenir un 6 » avec un dé équilibré.



Ce qu'il faut retenir

- Une variable discrète est déterminée en loi par les $P(X = x)$, positifs et de somme 1 ; sa fonction de répartition est en escalier.
- Lois usuelles : uniforme, Bernoulli, binomiale, hypergéométrique, géométrique (sans mémoire,

$P(X > n) = q^n$) et Poisson.

- L'espérance exige la **convergence absolue** de $\sum x P(X = x)$; le théorème de transfert calcule $E(f(X))$ sans la loi de $f(X)$.
- Linéarité : décomposez une variable en somme d'indicateurs pour obtenir son espérance.
- $V(X) = E(X^2) - E(X)^2$, et $V(\sum X_i) = \sum V(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j)$.
- Indépendance : la loi conjointe est le produit des marginales ; elle entraîne une covariance nulle, mais la réciproque est fausse.
- La loi d'une somme indépendante est une convolution ; sa fonction génératrice est le produit $G_X G_Y$.
- $E(X) = G'_X(1)$ et $V(X) = G''_X(1) + G'_X(1) - G'_X(1)^2$.
- Markov : $P(X \geq a) \leq E(X)/a$; Bienaymé-Tchebychev : $P(|X - E(X)| \geq a) \leq V(X)/a^2$.
- Loi faible des grands nombres : $P(|\bar{X}_n - m| \geq \epsilon) \leq \sigma^2 / (n \epsilon^2)$.

Questions fréquentes sur variables aléatoires discrètes

Pourquoi exige-t-on la convergence absolue pour définir l'espérance ?

Lorsque $X(\omega)$ est infini, l'espérance est une somme infinie, et l'ensemble des valeurs n'a pas d'ordre naturel. La convergence absolue garantit que la somme ne dépend pas de l'ordre choisi. Elle autorise aussi la sommation par paquets, qui démontre le théorème de transfert et la linéarité.

Une covariance nulle prouve-t-elle l'indépendance ?

Non. L'indépendance entraîne une covariance nulle, mais la réciproque est fausse en général. Par exemple, si X est uniforme sur $\{-1, 0, 1\}$ et $Y = X^2$, la covariance est nulle alors que Y est une fonction de X . La réciproque est vraie pour deux variables de Bernoulli.

Comment obtenir l'espérance et la variance avec la fonction génératrice ?

Pour une variable à valeurs entières, on a $E(X) = G'_X(1)$ et $E(X(X-1)) = G''_X(1)$, dérivées à gauche si le rayon vaut 1. On en déduit $V(X) = G''_X(1) + G'_X(1) - G'_X(1)^2$. Cette méthode est très efficace pour une somme de variables indépendantes.

À quoi sert l'inégalité de Bienaymé-Tchebychev si elle est peu précise ?

Elle est universelle : elle ne demande que la variance, sans connaître la loi. C'est pourquoi elle démontre la loi faible des grands nombres et donne des tailles d'échantillon garanties. Pour une loi précise, la probabilité exacte est souvent bien plus petite que la borne.

Pour aller plus loin en L2

- Les énoncés : [exercices de maths en L2 sur variables aléatoires discrètes](#)
- À maîtriser avant : [Espaces probabilisés et conditionnement](#), [Séries numériques à termes positifs](#)
- Chapitre précédent : [Espaces probabilisés et conditionnement](#)
- Chapitre suivant : [Suites et séries de fonctions](#)
- Tester vos connaissances : [QCM de maths en L2 par chapitre](#)
- Le sommaire : [tous les chapitres de maths de L2](#) et la [licence de maths de L1 à L3](#)