I. Espace de proba, variable aléatoire

1. Modéliser les probabilités

espace de probabilité : $(\Omega, \mathcal{F}, P)$
Ω : ensemble
P : mesure de probabilité (c'est-à-dire $P(\Omega) = 1$)
$\mathcal{F}$: tribu des ensembles mesurables (donnée "technique")

événement = ensemble mesurable

On se place toujours dans un espace $(\Omega, \mathcal{F}, P)$, même si ce n'est pas précisé.

Exemple. On lance un dé.
Soit A l'événement "le résultat du dé est pair".

$\Omega = \{1, \dots, 6\}$,
P = probabilité uniforme sur Ω,
$\mathcal{F} = \mathcal{P}(\Omega)$ (ensemble des parties de Ω).
$A = \{2, 4, 6\}$.
$P(A) = 3/6 = 1/2$.

Variable aléatoire réelle (v.a.) :
$X : \Omega \to \mathbb{R}$ fonction mesurable.

Loi d'une v.a. X (notée P_X) :
mesure de probabilité sur $\mathbb{R}$ définie par :

$\forall B \subset \mathbb{R}$ borélien,

$$\begin{aligned} P_X(B) &\stackrel{\text{def}}{=} P(X \in B) \\ &= P(\{\omega \in \Omega \mid X(\omega) \in B\}) \end{aligned}$$

Ω est souvent une "boîte noire".
Ce qu'on connaît, c'est la loi de X.

Exemple. On lance un dé et une pièce.
Soit X le résultat du dé et Y le résultat de la pièce.

$\forall\, 1 \le i \le 6$, $P_X(i) = P(X = i) = 1/6$,
$P_Y(pile) = P_Y(face) = 1/2$.

Que valent Ω et P ? On peut prendre
$\Omega = \{1, \dots, 6\} \times \{pile, face\}$,
P = proba uniforme ($P(i, \varepsilon) = 1/12$)
mais en général, ce n'est pas nécessaire de le savoir pour faire des calculs de probabilité.

Exemple. La cantine est ouverte entre midi et 13h, représenté par l'intervalle $[0, 1]$. L'heure d'arrivée de Bob à la cantine est une v.a. $X : \Omega \to [0, 1]$.

Bob arrive entre midi et 12h30 de façon uniforme :
– la loi P_X est concentrée sur $[0, 1/2]$.
– si $a, b \in [0, 1/2]$, $P_X([a, b])$ est proportionnelle à la longueur de $[a, b]$.

D'où $P_X([a, b]) = 2(b - a)$.

le coefficient 2 vient du fait que $P_X([0, 1/2]) = 1$.

Par exemple,
P(*Bob arrive avant 12h15*) $= P_X([0, 1/4]) = 1/2$.

Alice arrive à la cantine entre 12h et 13h, mais plus fréquemment autour de 12h15. C'est une v.a. $Y : \Omega \to [0, 1]$. Comment donner la loi de Y ?

La densité de P_Y est une fonction $f : \mathbb{R} \to \mathbb{R}^+$ telle que

$$\forall A \in \mathcal{B}(\mathbb{R}), \quad P_Y(A) = \int_A f(x)\, dx.$$

2. Espérance, formule de transfert

L'**espérance** est synonyme d'intégrale :

$$\text{si } X\colon \Omega \to \mathbb{R}, \quad E(X) \overset{\text{def}}{=} \int_\Omega X(\omega)\, dP(\omega).$$

Plus généralement :
Soit $X\colon \Omega \to E$ une v.a. et $\varphi\colon E \to \mathbb{R}^n$ une fonction mesurable bornée.

$$E(\varphi(X)) \overset{\text{def}}{=} \int_\Omega \varphi(X(\omega))\, dP(\omega).$$

peut être défini pour des φ non bornées, il faut $\varphi(X) \in L^1$.

formule de transfert (ou de transport) :

$$E(\varphi(X)) = \int_{\mathbb{R}^n} \varphi(x)\, dP_X(x).$$

Ceci découle du fait que c'est vrai pour $\varphi = \mathbb{1}_A$.

$$\begin{aligned} E(\mathbb{1}_A(X)) &= \int \mathbb{1}_A(X(\omega)) dP(\omega) = \int \mathbb{1}_{X^{-1}(A)}(\omega) dP(\omega) \\ &= P(X^{-1}(A)) = P_X(A) = \int \mathbb{1}_A(x) dP_X(x) \end{aligned}$$

(puis utiliser la densité des fonctions étagées)

Généralisation aux v.a. vectorielles : si $X\colon \Omega \to \mathbb{R}^d$ et $\varphi\colon \mathbb{R}^d \to \mathbb{R}$, alors $E(\varphi(X)) = \int_{\mathbb{R}^d} \varphi(x)\, dP_X(x)$.

Espérance, moments, variance.

Soit X une v.a. réelle. Son **espérance** est :

$$E(X) = \int_\Omega X(\omega) dP(\omega) = \int_\mathbb{R} x\, dP_X(x).$$

On dit que X est **centrée** si $E(X) = 0$.

On peut se ramener à des v.a. centrées en soustrayant leur espérance (souvent utile pour étudier la variation autour de la moyenne).

Pour tout $n \in \mathbb{N}^*$, le **moment d'ordre n** de X est

$$E(X^n) = \int_\mathbb{R} x^n\, dP_X(x)$$

moment d'ordre 1 = espérance

La **variance** est le moment d'ordre 2 centré :
si $m = E(X)$,

$$\mathrm{Var}(X) \overset{\text{def}}{=} E\left[(X-m)^2\right] = E(X^2) - m^2.$$

Les moments d'ordre n ne peuvent pas être définis pour toutes les v.a. : pour cela, il faut que X^n soit une fonction intégrable, autrement dit $X \in L^n$.
$\mathrm{Var}(X)$ *existe ssi* $X \in L^2$

Loi continue (= loi à densité).

Si X est une v.a. de densité $f(x)$, alors $E(\varphi(X)) = \int_\mathbb{R} \varphi(x) f(x)\, dx$.
En particulier, $E(X) = \int_\mathbb{R} x f(x)\, dx$.

Découle de $E(\mathbb{1}_A(X)) = P_X(A) = \int_A f(x)dx = \int \mathbb{1}_A(x) f(x) dx$.

Ceci explique la notation $dP_X = f(x)dx$.

Exemple. Soit X une v.a. de loi exponentielle $\mathcal{E}(\lambda)$, c'est-à-dire $dP_X = \mathbb{1}_{\mathbb{R}^+}(x)\lambda e^{-\lambda x}dx$.

$$E(X) = \int_\mathbb{R} x dP_X(x) = \int_0^{+\infty} x\lambda e^{-\lambda x} dx$$

Intégration par parties ($u' = \lambda e^{-\lambda x}, v = x,\ u = -e^{-\lambda x}, v' = 1$) :

$$\begin{aligned} \int_0^M x\lambda e^{-\lambda x} dx &= \left[-xe^{-\lambda x}\right]_0^M + \int_0^M e^{-\lambda x} dx \to 1/\lambda \\ &= \left[-xe^{-\lambda x} - \frac{1}{\lambda} e^{-\lambda x}\right]_0^M \\ &= -\left(M + \frac{1}{\lambda}\right) e^{-\lambda M} + \frac{1}{\lambda} \\ &\to \frac{1}{\lambda} \text{ quand } M \to +\infty. \end{aligned}$$

D'où $E(X) = \frac{1}{\lambda}$.

Par un calcul similaire : $\mathrm{Var}(X) = E(X^2) - E(X) = 1/\lambda^2$.

Exemple : loi normale

Soit X de loi $\mathcal{N}(0,1)$, c'est-à-dire $dP_X = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ (on admet que c'est une loi de probabilité, c'est-à-dire que $\frac{1}{\sqrt{2\pi}} \int_\mathbb{R} e^{-x^2/2}\, dx = 1$).

Montrons que $E(X) = 0$ et $\mathrm{Var}(X) = 1$.

Espérance.
$E(X) = \int_\mathbb{R} \frac{1}{\sqrt{2\pi}}\ x e^{-x^2/2}\, dx = 0$ car la fonction $xe^{-x^2/2}$ est impaire.

Remarque : pour que l'espérance soit bien définie, il faut aussi montrer que X est L^1, ce qu'on voit facilement en mettant des valeurs absolues dans l'intégrale.

Variance.
$\mathrm{Var}(X) = \frac{1}{\sqrt{2\pi}} \int_\mathbb{R} x^2 e^{-x^2/2}\, dx$.
Intégration par parties [avec $u = x, v' = xe^{-x^2/2}$] :

$$\mathrm{Var}(X) = \left[-\frac{xe^{-x^2/2}}{\sqrt{2\pi}}\right] + \frac{1}{\sqrt{2\pi}} \int_\mathbb{R} e^{-x^2/2}\, dx$$

$$\mathrm{Var}(X) = 0 + P_X(\mathbb{R}) = 1.$$

Pour montrer que, pour la loi $\mathcal{N}(m,\sigma^2)$, l'espérance est m et la variance σ^2 : prendre Y de loi $\mathcal{N}(m,\sigma^2)$ et $X = \frac{Y-m}{\sigma}$.

$P(X \in B) = P(Y \in \sigma B + m) =$
$\int_{\sigma B+m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-m)^2}{2\sigma^2}\right) dy$

Par changement de variable $x = \frac{y-m}{\sigma}$, on trouve que la densité de X est $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, autrement dit X suit une loi $\mathcal{N}(0,1)$. Comme $Y = \sigma X + m$, on a $E(Y) = E(X) + m = m$ et
$\mathrm{Var}(Y) = \sigma^2 \mathrm{Var}(X) = \sigma^2$.

Loi discrète.

X est une v.a. discrète si elle prend (p.s.) ses valeurs dans un ensemble fini ou dénombrable $\{x_i ; i \in I\}$.

La loi de X est donnée par $\forall i \in I,\ P(X = x_i) = p_i$.

On peut écrire $P_X = \sum_{i \in I} p_i \, \delta_{x_i}$,
où δ_a est la mesure de Dirac en a.

On a : $E(\varphi(X)) = \sum_{i \in I} p_i \, \varphi(x_i)$.

En particulier, $E(X) = \sum_{i \in I} p_i \, x_i$.

car $\int \varphi(t)\delta_x(t) = \varphi(x)$.

Exemple trivial.

On a 100 notes entre 0 et 20, on veut calculer leur moyenne. Si on veut l'exprimer en terme de v.a. :

$\Omega = \{1, \dots, 100\}$, P est la proba uniforme sur Ω,
$X : \Omega \to \{0, \dots, 20\}$, où $X(i) = n_i$ est la i-ème note.

$$E(X) = \sum_{i=1}^{n} \frac{1}{100} X(i) = \frac{1}{100} \sum_{i=1}^{n} n_i.$$

(c'est le calcul direct de $\int_\Omega X(\omega) \, dP(\omega)$)

Soit k_n le nombre de notes égales à n. Alors la loi de X est donnée par $P(X = n) = \dfrac{k_n}{100}$,
autrement dit $P_X = \sum_{i=0}^{20} k_i \, \delta_i$.

Avec la formule de transfert, on trouve

$$E(X) = \sum_{i=0}^{20} \frac{k_n}{100} n = \frac{1}{100} \sum_{i=0}^{20} k_n \, n.$$

Exemples.

Loi de Bernoulli.
Si $P_X = b(p)$ alors

$$E(X) = (1-p) \times 0 + p \times 1 = p.$$

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p).$$

Loi de Poisson.
Si $P_X = \mathcal{P}(\lambda)$ alors

$$\begin{aligned} E(X) &= \sum_{k \geq 0} e^{-\lambda} \frac{\lambda^k}{k!} k \\ &= \lambda e^{-\lambda} \sum_{k \geq 1} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \cdot e^{\lambda} \end{aligned}$$

d'où $E(X) = \lambda$.

(voir l'exercice 2 pour le calcul de la variance)

II. Probabilités conditionnelles

1. Définition

Définition.
Soit A et B deux événements, avec $P(B) > 0$. La **probabilité conditionnelle** de A sachant B est

$$P(A|B) \stackrel{\text{def}}{=} \frac{P(A \cap B)}{P(B)}.$$

Exemple. Bill arrive en retard en cours la moitié du temps, et ne vient pas une fois sur 4 en moyenne. Le cours commence sans lui. Quelle est la probabilité qu'il vienne aujourd'hui ?

A : Bill vient en cours
B : Bill n'est pas là au début du cours.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/2}{1/2 + 1/4} = 2/3.$$

Si X est une variable aléatoire, on peut considérer des événements de la forme $\{X \in A\}$.

Exemple.
On lance un dé. X est le résultat obtenu. Y vaut "pair" ou "impair" selon la parité du résultat.

$$P(X = 6 \mid Y = \text{pair}) = \frac{P(X = 6, Y = \text{pair})}{P(Y = \text{pair})} = \frac{1}{3}$$

$$P(X = 6 \mid Y = \text{impair}) = 0$$

Exemple : la loi exponentielle est sans mémoire

Soit X une v.a. de loi exponentielle $\mathcal{E}(\lambda)$ ($\lambda > 0$).

Montrons que, pour tous $s, t \geq 0$,

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

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

Or $P(X > y) = \int_y^{+\infty} \lambda e^{-\lambda x} dx = e^{-\lambda y}$ *si* $y \geq 0$.

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

Ce résultat montre que la loi exponentielle est une loi sans mémoire. Selon l'exercice 6, c'est la seule sous l'hypothèse du b). En fait, cette hypothèse n'est pas nécessaire mais le résultat est alors plus difficile à montrer (voir [Billingsley, A20, p568]).

2. Formule des probabilités totales

Soit $(B_i)_{i \in I}$ une famille finie ou dénombrable d'événements formant une partition de Ω, avec $P(B_i) > 0$ pour tout i. Alors

$$P(A) = \sum_{i \in I} P(A|B_i) P(B_i).$$

Formule surtout utilisée quand Ω est fini ou dénombrable.

Exemple : test pour détecter une maladie.
– personne malade : test positif dans 95% des cas.
– personne saine : test négatif dans 95% des cas.

La maladie touche 1 personne sur 1000. On teste quelqu'un. Probabilité d'avoir un test positif ?

$M = \{\text{personnes malades}\}, S = \{\text{personnes saines}\}$,
$A = \{\text{personnes testées positif}\}$.

$$P(A) = P(A|M)P(M) + P(A|S)P(S)$$

$P(A) = 0,95 \times 0,001 + 0,05 \times 0,999 = 0,0509.$

3. Formule de Bayes

Formule de Bayes "de base" :

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

Il suffit d'écrire $P(A|B)P(B) = P(A \cap B) = P(B|A)P(A)$.

$P(A)$: probabilité a priori de A.
$P(A|B)$: probabilité a posteriori de A sachant B.

Formule de Bayes "améliorée" :

Soit $(A_j)_{j \in I}$ une partition de Ω, et B un événement. On suppose $P(B) > 0$ et $P(A_j) > 0$ pour tout $j \in I$. Alors

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j \in I} P(B|A_j)P(A_j)}.$$

Formule précédente avec $P(B) = \sum_{j \in I} P(B|A_j)P(A_j)$.

Exemple : "faux positifs" en médecine.
On a un test pour savoir si quelqu'un a une maladie donnée.
– S'il est malade : test positif dans 99,9% des cas.
– S'il est sain : test négatif dans 99,9% des cas.

On suppose que la maladie touche une personne sur 1 000. On teste quelqu'un, le test est positif. Quelle est la probabilité qu'il soit malade ?

M : la personne est malade.
A : le test est positif.

$$P(M|A) = \underbrace{\frac{P(A|M)P(M)}{P(A)}}_{\frac{0{,}999 \times 0{,}001}{0{,}999 \times 0{,}001 + 0{,}001 \times 0{,}999}} = 0{,}5$$

III. Indépendance

1. Définitions de l'indépendance

Intuitivement, A et B sont indépendants si la connaissance de B ne donne pas d'information sur A, c'est-à-dire $P(A|B) = P(A)$. Ce qui équivaut à $P(A \cap B) = P(A)P(B)$.

Ex : si on lance un dé et une pièce, les événements "obtenir 6" et "obtenir face" sont indépendants.

Définition. Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé.
– Deux événements A, B sont dits **indépendants** si $P(A \cap B) = P(A)P(B)$. On note $A \perp B$.

– Si $A_i \in \mathcal{F}$, la famille d'événements $(A_i)_{i \in I}$ est dite **indépendante** si pour tout sous-ensemble fini $\{i_1, \ldots, i_n\}$ de I,

$$P\left(\bigcap_{k=1}^{n} A_{i_k}\right) = \prod_{k=1}^{n} P(A_{i_k}).$$

Contre-exemple.
Soit $\Omega = \{1, \ldots, 8\}$ *et* P *la probabilité uniforme.*

Soit $A = \{1,2,3,4\}$, $B = \{4,5,6,7\}$, $C = \{2,4,6,8\}$.
On a $P(A) = P(B) = P(C) = \frac{1}{2}$.

$P(A \cap B \cap C) = P(\{4\}) = \frac{1}{8} = P(A)P(B)P(C)$.
Mais $P(A \cap B) = \frac{1}{8} \neq P(A)P(B) = \frac{1}{4}$.
Donc (A, B, C) *ne sont pas indépendants.*

Contre-exemple.
Soit $\Omega = \{1, \ldots, 4\}$ *et* P *la probabilité uniforme.*
Soit $A = \{1,2\}$, $B = \{1,3\}$, $C = \{1,4\}$.
On a $P(A) = P(B) = P(C) = 1/2$.

$P(A \cap B) = P(\{1\}) = 1/4 = P(A)P(B)$ *donc* $A \perp B$. *De même,* $A \perp C$ *et* $B \perp C$, *autrement dit* A, B, C *sont 2 à 2 indépendants.*

$P(A \cap B \cap C) = P(\{1\}) = 1/4 \neq P(A)P(B)P(C)$
donc A, B, C *ne sont pas indépendants.*

Définition.
Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé et X_i des v.a. définies sur Ω à valeurs dans E_i (E_i étant muni de la tribu $\mathcal{G}_i$).

La famille de v.a. $(X_i)_{i\in I}$ est **indépendante** si pour tous $A_i \in \mathcal{G}_i$, $i \in I$, la famille d'événements $(X_i \in A_i)_{i\in I}$ est indépendante, c'est-à-dire pour tout sous-ensemble fini $\{i_1, \ldots, i_n\}$ de I

$$P\left(X_{i_k} \in A_{i_k}, 1 \leq k \leq n\right) = \prod_{k=1}^{n} P(X_{i_k} \in A_{i_k}).$$

Remarque. Les X_i doivent être définis sur le même espace Ω sinon l'expression de gauche n'a pas de sens.

$$P\left(X_{i_k} \in A_{i_k}, 1 \leq k \leq n\right) = P\left(\{\omega \in \Omega; X_{i_k}(\omega) \in A_{i_k}, 1 \leq k \leq n\}\right)$$

Exemple. On lance un dé et une pièce.
X = résultat du dé, Y = résultat de la pièce.
Ce sont des v.a. indépendantes.

Quel Ω prendre pour pouvoir définir X, Y sur le même ensemble ?

$\Omega = \{1, \ldots, 6\} \times \{pile, face\}$.

X est la projection sur la première coordonnée : $X\colon \Omega \to \{1, \ldots, 6\}$.

Y est la projection sur la deuxième coordonnée : $Y\colon \Omega \to \{pile, face\}$.

Pour montrer l'indépendance d'une famille finie de v.a. $X_1, \ldots, X_n$, il est inutile de prendre un sous-ensemble d'indices, il suffit de tester que :

$$P\left(X_k \in A_k, 1 \leq k \leq n\right) = \prod_{k=1}^{n} P(X_k \in A_k)$$

pour tous $A_k \in \mathcal{G}_k, 1 \leq k \leq n$.

Remarque. *Cette propriété est fausse pour des événements (voir le 1er contre-exemple donné précédemment).*

Preuve.
Supposons qu'on a l'égalité ci-dessus et montrons l'indépendance des X_i. Soit J un sous-ensemble fini de $\{1, \ldots, n\}$. On pose $A_i = E_i$ si $i \notin J$.

$$\begin{aligned} P(X_i \in A_i, i \in J) &= P(X_i \in A_i, 1 \leq i \leq n) \\ &= \prod_{1\leq i\leq n} P(X_i \in A_i) \quad \textit{(hypothèse)} \end{aligned}$$

Or

$$\prod_{1\leq i\leq n} P(X_i \in A_i) = \prod_{i\in J} P(X_i \in A_i) \times 1.$$

Donc $P(X_i \in A_i, i \in J\} = \prod_{i\in J} P(X_i \in A_i)$. $\square$

Pour montrer l'indépendance d'une famille infinie $(X_k)_{k\geq 1}$, il suffit de tester que pour tout $n \geq 1$:

$$P\left(X_k \in A_k, 1 \leq k \leq n\right) = \prod_{k=1}^{n} P(X_k \in A_k)$$

pour tous $A_k \in \mathcal{G}_k, 1 \leq k \leq n$.

Preuve.
Supposons qu'on a l'égalité ci-dessus et montrons l'indépendance des $(X_k)_{k\geq 1}$. Si J est un sous-ensemble fini de $\mathbb{N}^*$, il existe n tel que $J \subset \{1, \ldots, n\}$. On pose $A_i = E_i$ si $i \in \{1, \ldots, n\} \setminus J$. On montre ensuite, de façon similaire à la preuve précédente, que

$$\begin{aligned} P(X_i \in A_i, i \in J) &= P(X_i \in A_i, 1 \leq i \leq n) \\ &= \prod_{1\leq i\leq n} P(X_i \in A_i) \quad \textit{(hypothèse)} \\ &= \prod_{i\in J} P(X_i \in A_i) \end{aligned}$$

ce qui montre l'indépendance. $\square$

Si X_i est une v.a. discrète, on peut se contenter de regarder les événements de la forme $\{X_i = a_i\}$, car tout événement $\{X_i \in A_i\}$ est une union disjointe de tels événements.

Exemple. *On lance un dé équilibré. $X = 0$ si le résultat est pair, $X = 1$ sinon. $Y = 0$ si le résultat est multiple de 3, $Y = 1$ sinon. Pour montrer que $X \perp Y$, il suffit de considérer $P(X = 0, Y = 0)$, $P(X = 0, Y = 1)$, $P(X = 1, Y = 0)$ et $P(X = 1, Y = 1)$.*

Définition.
Une famille de sous-tribus $(\mathcal{F}_i)_{i \in I}$ (avec $\mathcal{F}_i \subset \mathcal{F}$) est dite **indépendante** si toute famille d'événements $(A_i \in \mathcal{F}_i)_{i \in I}$, est indépendante.

Proposition.
La famille d'événements $(A_i)_{i \in I}$ est indépendante si et seulement si la famille de sous-tribus $(\sigma(A_i))_{i \in I}$ est indépendante, où $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$ est la tribu engendrée par A.

La famille de v.a. $(X_i)_{i \in I}$ est indépendante si et seulement si la famille de tribus $(X_i^{-1}(\mathcal{G}_i))_{i \in I}$ est indépendante.

2. Indépendance et loi produit

Mesure produit (définition/théorème).
Soit μ une mesure sur $(\Omega, \mathcal{F})$ et ν une mesure sur $(\Omega', \mathcal{F}')$. $\mathcal{F} \otimes \mathcal{F}'$ est la tribu de $\Omega \times \Omega'$ engendrée par les ensembles $A \times B$ avec $A \in \mathcal{F}, B \in \mathcal{F}'$.

Pour tous $A \in \mathcal{F}, B \in \mathcal{F}'$, on définit

$$\mu \otimes \nu(A \times B) = \mu(A)\nu(B).$$

Alors $\mu \otimes \nu$ s'étend de façon unique en une mesure sur $(\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}')$.
De plus, si μ et ν sont des probabilités alors $\mu \otimes \nu$ est une probabilité.

Formulation de l'indépendance en terme de lois
Soit $(X_i)_{1 \leq i \leq n}$ des v.a. définies sur Ω. Les v.a. $X_1, \ldots, X_n$ sont indépendantes si et seulement si la loi de la v.a. $Y = (X_1, \ldots, X_n)$ est égale à $P_{X_1} \otimes P_{X_2} \otimes \cdots \otimes P_{X_n}$.

Si les X_i sont des v.a. réelles alors Y est une v.a. vectorielle à valeurs dans $\mathbb{R}^n$ et la loi produit est une loi sur $\mathbb{R}^n$.

Découle directement de la définition de l'indépendance et des mesures produits.

Si X, Y sont des v.a. réelles indépendantes, la loi de (X, Y) est $P_X \otimes P_Y$. Pour intégrer, appliquer Fubini :

$$\begin{aligned} E(\varphi(X,Y)) &= \int_{\mathbb{R}^2} \varphi(x,y) d(P_X \otimes P_Y)(x,y) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \varphi(x,y) dP_X(x) \right) dP_Y(y) \end{aligned}$$

soit en résumé :

$$d(P_X \otimes P_Y)(x,y) = dP_X(x) dP_Y(y).$$

Proposition.
Soit X, Y des v.a. réelles intégrables indépendantes. Alors XY est intégrable et $E(XY) = E(X)E(Y)$.
Par Fubini-Tonelli :

$$\underbrace{\int_{\mathbb{R}^2} |xy| \, dP_X(x) dP_Y(y)}_{E(|XY|)} = \underbrace{\int_{\mathbb{R}} |x| \, dP_X(x)}_{E(|X|)} \underbrace{\int_{\mathbb{R}} |y| \, dP_Y(y)}_{E(|Y|)} < +\infty$$

puis par Fubini, $E(XY) = E(X)E(Y)$.

Si X et Y sont des v.a. indépendantes, de lois continues de densités $dP_X = f(x)dx$ et $dP_Y = g(y)dy$, alors la v.a. (X, Y) a une loi continue de densité $dP_{(X,Y)} = f(x)g(y)dxdy$, et

$$E(\varphi(X,Y)) = \int_{\mathbb{R}^2} \varphi(x,y) f(x) g(y) dx dy.$$

3. Premières propriétés de l'indépendance

Théorème.
Soit $X_1, X_2, \ldots, X_n$ des v.a. indépendantes.
– Soit $\varphi_1, \varphi_2, \ldots, \varphi_n$ des fonctions mesurables. Alors les v.a. $\varphi_1(X_1), \ldots, \varphi_n(X_n)$ sont indépendantes.
Si X_i est à valeurs dans E_i, alors φ_i doit être définie sur E_i (à valeurs dans un certain F_i).
– Si $1 \leq k < n$, alors les v.a. $Y = (X_1, \ldots, X_k)$ et $Y' = (X_{k+1}, \ldots, X_n)$ sont indépendantes (théorème des coalitions).

Exemples.
– Si R, θ sont deux v.a. réelles indépendantes, alors R^2 et $\cos(\theta)$ sont indépendantes.

– Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. indépendantes et $M_n = \dfrac{X_1 + \cdots + X_n}{n}$. Alors $M_n \perp X_{n+1}$.
En effet, $Y = (X_1, \ldots, X_n) \perp X_{n+1}$ et $M_n = \varphi(Y)$.

4. Lemme de Borel-Cantelli

Notation.
Soit $(A_n)_{n\geq 1}$ une suite d'événements. On note

$$\limsup_{n\to+\infty} A_n \stackrel{\text{def}}{=} \bigcap_{k=1}^{+\infty} \bigcup_{n\geq k} A_n.$$

$\limsup_{n\to+\infty} A_n = \{\omega \in \Omega \mid \omega \in A_n \text{ pour une infinité de } n\}$

$$\liminf_{n\to+\infty} A_n \stackrel{\text{def}}{=} \bigcup_{k=1}^{+\infty} \bigcap_{n\geq k} A_n$$
$$= \{\omega \in \Omega \mid \omega \in A_n \text{ à partir d'un certain rang}\}.$$

On a $(\limsup A_n)^c = \liminf(A_n^c)$.

Théorème (lemme de Borel-Cantelli).
Soit $(A_n)_{n\geq 1}$ une suite d'événements.

1) Si $\sum_{n\geq 1} P(A_n) < +\infty$ alors $P(\limsup_{n\to+\infty} A_n) = 0$.

2) Si $\sum_{n\geq 1} P(A_n) = +\infty$ et si la famille $(A_n)_{n\geq 1}$ est indépendante alors $P(\limsup_{n\to+\infty} A_n) = 1$.

Exemple : le singe dactylographe de Borel.
Soit m une suite de lettres de longueur L. Soit $(X_n)_{n\geq 1}$ une suite de v.a. indépendantes de loi uniforme sur les caractères $\{c_1, \ldots, c_N\}$. Existe-t-il $n \geq 0$ tel que $X_{n+1} \cdots X_{n+L}$ forment le mot m ?

Soit $A_n = \{(X_{n+1} \cdots X_{n+L}) = m\}$.
Les $(A_n)_{n\geq 0}$ ne sont pas indépendants, par contre les $(A_{nL})_{n\geq 0}$ sont indépendants. $P(A_{nL}) = \dfrac{1}{N^L} > 0$, donc $\sum_{n\geq 0} P(A_{nL}) = +\infty$. On applique le lemme de Borel-Cantelli (2) : $P(\limsup_n A_{nL}) = 1$, autrement dit on écrit le mot m une infinité de fois p.s.

Remarque. le temps d'attente moyen est de l'ordre de N^L.

Preuve Borel-Cantelli. *[Barbe-Ledoux p99]*
1) Soit $B_k = \bigcup_{n\geq k} A_n$.
On a $P(B_k) \leq \sum_{n=k}^{+\infty} P(A_n) \to 0$ (reste d'une série convergente).
Or $\limsup_n A_n \subset B_k$ pour tout k donc $P(\limsup_n A_n) = 0$.
2) Les (A_n^c) sont indépendants, donc

$$P\left(\bigcap_{k=n}^{p} A_k^c\right) = \prod_{k=n}^{p} P(A_k^c) = \prod_{k=n}^{p} \underbrace{(1 - P(A_k))}_{\leq e^{-P(A_k)}}.$$

(on utilise que $1 - x \leq e^{-x}$ pour tout x), donc

$$P\left(\bigcap_{k=n}^{+\infty} A_k^c\right) \leq P\left(\bigcap_{k=n}^{p} A_k^c\right) \leq \exp\left(-\sum_{k=n}^{p} P(A_k)\right).$$

Or pour n fixé $\sum_{k=n}^{+\infty} P(A_k) \to +\infty$ par hypothèse, donc $\exp\left(-\sum_{k=n}^{p} P(A_k)\right) \to 0$ quand $p \to +\infty$, d'où $P\left(\bigcap_{k=n}^{+\infty} A_k^c\right) = 0$.
Donc

$$P\left(\left(\bigcap_{k=n}^{+\infty} A_k^c\right)^c\right) = P\left(\bigcup_{k=n}^{+\infty} A_k\right) = 1,$$

et $P\left(\bigcap_{n=1}^{+\infty} \bigcup_{k=n}^{+\infty} A_k\right) = 1$
(intersection dénombrable d'ensembles de mesures 1).

5. Loi du 0-1

Définition.
Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé et $(\mathcal{F}_n)_{n\geq 1}$ des sous-tribus de $\mathcal{F}$.
On note $\sigma\Big(\mathcal{F}_n, \mathcal{F}_{n+1}, \dots\Big)$ la tribu engendrée par les tribus $(\mathcal{F}_k)_{k\geq n}$, et on pose

$$\mathcal{F}^\infty = \bigcap_{n\geq 1} \sigma\Big(\mathcal{F}_n, \mathcal{F}_{n+1}, \dots\Big).$$

$\mathcal{F}^\infty$ est appelée **tribu asymptotique**.

Théorème (loi du 0-1 de Kolmogorov).
Soit $(\mathcal{F}_n)_{n\geq 1}$ des sous-tribus indépendantes, c'est-à-dire $\forall A_n \in \mathcal{F}_n$, la famille $(A_n)_{n\geq 1}$ est indépendante. Alors pour tout événement $A \in \mathcal{F}^\infty$, $P(A) = 0$ ou 1.

Exemple.
Soit $(X_n)_{n\geq 1}$ des v.a. réelles sur $(\Omega, \mathcal{F}, P)$, et

$$\mathcal{F}_n = \sigma(X_n) = \{X^{-1}(B) \in \mathcal{F} \mid B \in \mathcal{B}(\mathbb{R})\}.$$

Si les v.a. $(X_n)_{n\geq 1}$ sont indépendantes, les tribus $(\mathcal{F}_n)_{n\geq 1}$ sont indépendantes.

Si $A = \{\omega \in \Omega \mid X_n(\omega) = 0$ pour une infinité de $n\}$, Alors $A \in \mathcal{F}^\infty$, car cette propriété est déterminée uniquement par $(X_n)_{n\geq k}$ pour k arbitrairement grand. Donc $P(A) = 0$ ou 1.

$\omega \in A \Leftrightarrow \forall N,\ \exists n \geq N, X_n(\omega) = 0$ *donc*

$$A = \underbrace{\bigcap_{n\geq N} \underbrace{\bigcup_{n\geq N} X_n^{-1}(\{0\})}_{\in \sigma(\mathcal{F}_N, \mathcal{F}_{N+1}, \dots)}}_{\in \mathcal{F}^\infty}$$

Soit

$$B = \left\{\omega \in \Omega \mid \lim_{n\to+\infty} \frac{X_1(\omega) + \cdots + X_n(\omega)}{n} \text{ existe}\right\}.$$

Alors $B \in \mathcal{F}^\infty$ car, pour ω fixé :

$$\frac{X_1(\omega) + \cdots + X_n(\omega)}{n} \text{ converge } \iff \frac{X_k(\omega) + \cdots + X_n(\omega)}{n} \text{ converge.}$$

Donc si on pose $S_n = X_1 + \cdots + X_n$, alors $\frac{S_n}{n}$ converge p.s. ou diverge p.s.

On peut pousser le raisonnement plus loin si $\frac{S_n}{n}$ *converge p.s.*

Notons Z_∞ *la v.a. réelle égale à la limite* $\lim_{n\to+\infty} \frac{S_n}{n}$. *Posons*

$$A_t = \{\omega \in \Omega \mid Z_\infty(\omega) \leq t\} \text{ pour tout } t \in \mathbb{R}.$$

$A_t \in \mathcal{F}^\infty$, *donc* $P(A_t) = 0$ *ou* 1. *De plus, si* $t \leq t'$, $A_t \subset A_{t'}$, *donc la fonction* $t \mapsto P(A_t)$ *est croissante. On a également*

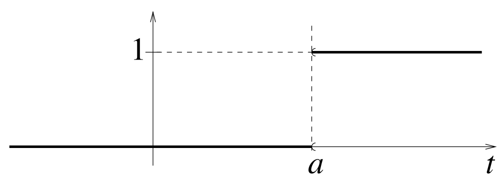
$$\lim_{t\to-\infty} P(A_t) = P(Z_\infty \leq -\infty) = 0$$

et

$$\lim_{t\to+\infty} P(A_t) = P(Z_\infty \leq +\infty) = 1$$

(Z_t est finie p.s. puisque la limite existe par hypothèse).

Ceci implique qu'il existe $a \in \mathbb{R}$ *tel que* $P(A_t) = 0$ *pour tout* $t < a$ *et* $P(A_t) = 1$ *pour tout* $t > a$. *Autrement dit,* $t \mapsto (A_t)$ *est de la forme :*



Par conséquent, $P(Z_t \geq a + \varepsilon) = 1 - P(A_{a+\varepsilon}) = 0$ *pour tout* $\varepsilon > 0$, *d'où* $P(Z_t > a) = 0$. *De plus,* $P(Z_t < a) = 0$, *donc* $Z_t = a$ *p.s.*
Conclusion : Si $\frac{S_n}{n}$ *converge p.s, alors elle converge p.s. vers une constante.*

IV. Caractériser une loi

1. Fonction de répartition

Définition.
Soit X une v.a. réelle. La **fonction de répartition** de X est la fonction $F_X : \mathbb{R} \to [0,1]$ définie par

$$\forall t \in \mathbb{R},\ F_X(t) \stackrel{\text{def}}{=} P(X \le t).$$

Proposition.
a) F_X est une fonction croissante à valeurs dans $[0,1]$.
b) $\lim\limits_{t\to-\infty} F_X(t) = 0$ et $\lim\limits_{t\to+\infty} F_X(t) = 1$.
c) $\forall a \in \mathbb{R}$, F_X est continue à droite de a et

$$F_X(a-) \stackrel{\text{def}}{=} \lim_{t\uparrow a} F_X(t) \text{ existe et vaut } P(X < a).$$

Remarque. *Une fonction F vérifiant les propriétés a), b), c) est la fonction de répartition d'une certaine v.a.*

Théorème . Soit X et Y des v.a. réelles. Alors $P_X = P_Y$ si et seulement si $F_X = F_Y$.

Preuve proposition.
a) F_X est croissante car si $t \le t'$ alors $]-\infty, t] \subset]-\infty, t']$.

b) $\emptyset = \bigcap_{n\ge0}]-\infty, -n]$ (intersection décroissante) donc $0 = \lim\limits_{n\to+\infty} F_X(-n)$. De plus, $F_X(-n-1) \le F_X(t) \le F_X(-n)$ si $-n-1 \le t \le -n$ donc $0 = \lim\limits_{t\to-\infty} F_X(t)$.

$\mathbb{R} = \bigcup_{n\ge0}]-\infty, n]$ (union croissante) donc $1 = \lim\limits_{t\to+\infty} F_X(t)$.

c) $\{X \le t\} = \bigcap_{n\ge1} \{X \le t + 1/n\}$, c'est une intersection décroissante donc $F_X(t) = \lim F_X(t + 1/n)$. Comme F_X est croissante, on a $F_X(t) = \lim_{x\downarrow t} F_X(x)$, c'est-à-dire que F est continue à droite de t.

$\{X < t\} = \bigcup_{n\ge1} \{X \le t - 1/n\}$, c'est une union croissante donc $P(X < t) = F_X(t-)$. □

Le théorème se montre à l'aide du lemme suivant :

Lemme. *Soit P_1 et P_2 deux mesures de probabilité sur $(\Omega, \mathcal{F})$ telles que $\forall C \in \mathcal{C},\ P_1(C) = P_2(C)$, où $\mathcal{C}$ est un π-système (c.à.d. une famille stable par intersection finie) qui engendre $\mathcal{F}$. Alors $P_1 = P_2$.*

Propriétés.
- $P(X = a) = F_X(a) - F_X(a-)$.
- $P(a < X \le b) = F_X(b) - F_X(a)$ *si* $a < b$.
- $P(X > a) = 1 - F_X(a)$.

Exemple. *Fonction de répartition de* $\min(X, Y)$, *où X, Y sont des v.a. réelles indépendantes : voir l'exercice 10.*

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

$$F_X(t) = \int_{-\infty}^{t} f(x)\, dx.$$

La fonction F_X est continue sur $\mathbb{R}$.

Si f est continue en t, alors F_X est dérivable en t et $F_X'(t) = f(t)$.

Exemple. *Si $P_X = \mathcal{E}(\lambda)$, alors $F_X(t) = 0$ si $t \le 0$ et $F_X(t) = \int_0^t \lambda e^{-\lambda x}\, dx = 1 - e^{-\lambda t}$ si $t \ge 0$.*
$\rightsquigarrow$ dessin.

Réciproquement :

Proposition.
Soit X une v.a. réelle. S'il existe une fonction f telle que $F_X(t) = \int_{-\infty}^{t} f(x)\, dx$, alors X a une loi continue de densité f.

En particulier, si F_X est <u>continue sur $\mathbb{R}$</u> et C^1 par morceaux, alors la densité de X est $f(x) = F_X'(x)$ (on peut donner une valeur arbitraire à $f(x)$ aux points où F_X' n'est pas définie).

Exemple. *Soit X une v.a. de loi normale $\mathcal{N}(0,1)$. Quelle est la loi de $Z = e^X$? Voir l'exercice 3.*

Exemple. *Soit X une v.a. de loi $\mathcal{N}(0,1)$. Loi de $Y = X^2$?*
$P(Y \leq t) = 0$ si $t < 0$ et $P(Y \leq t) = P(-\sqrt{t} \leq X \leq \sqrt{t})$ si $t \geq 0$.
Donc, pour $t \geq 0$, $F_Y(t) = F_X(\sqrt{t}) - F_X(-\sqrt{t})$.

$F_X(t)$ dérivable de dérivée $\frac{1}{\sqrt{2\pi}}e^{-t^2/2}$, donc F_Y dérivable sur $]0,+\infty[$ et $F_Y'(t)$ est la densité de Y.

$F_Y'(t) = \frac{1}{2\sqrt{t}}(F_X'(\sqrt{t}) + F_X'(-\sqrt{t}))$. Donc la densité de Y est $f_Y(t) = \frac{1}{2\sqrt{t}}(f_X(\sqrt{t}) + f_X(-\sqrt{t}))$ si $t > 0$.
$f_Y(t) = \mathbb{1}_{\mathbb{R}_+^*}(t)\frac{1}{\sqrt{2\pi t}}e^{-t/2}$.

(voir aussi l'exercice 4)

Exemple. *Soit X, Y 2 v.a. réelles indépendantes.*
Loi de $M = \max(X,Y)$ et $m = \min(X,Y)$?

$P(M \leq t) = P(X \leq t$ et $Y \leq t) = P(X \leq t)P(Y \leq t)$.
Donc $F_M(t) = F_X(t)F_Y(t)$.

Si X, Y ont des densités continues alors M aussi (de densité $f_M(t) = f_X(t)F_Y(t) + F_X(t)f_Y(t)$).

De même, $P(m > t) = P(X > t)P(Y > t)$,
càd $1 - F_m(t) = (1 - F_X(t))(1 - F_Y(t))$.

Si X est une v.a. discrète, la fonction de répartition est en escaliers, les marches sont aux valeurs prises par X, la hauteur de la marche en x_i est $P(x_i)$.

Remarque. *Les fonctions de répartition ne sont pas très utiles pour les v.a. discrètes.*

Exemples.
Dessiner F_X pour $\mathcal{B}(2,1/2)$ (valeurs $1/4, 1/2, 1/4$).

2. Fonction caractéristique

Définition.
Soit X une v.a. réelle. La **fonction caractéristique** de X est la fonction $\varphi_X : \mathbb{R} \to \mathbb{C}$ telle que

$$\varphi_X(t) \stackrel{\text{def}}{=} E(e^{itX}) = \int_\Omega e^{itX(\omega)}\,dP(\omega) = \int_\mathbb{R} e^{itx}\,dP_X(x).$$

Remarque. *$|e^{itX}| = 1$ donc c'est intégrable et φ_X est définie sur $\mathbb{R}$ quelle que soit la loi de X.*

Si $dP_X = f(x)dx$, alors $\varphi_X(t) = \int_\mathbb{R} e^{itx}f(x)\,dx$.

C'est la **transformée de Fourier** de f.
(le coefficient de la transformée de Fourier "proba" est différent du coefficient de la transformée "classique".)

Propriétés.
- $\forall t \in \mathbb{R}$, $|\varphi_X(t)| \leq 1$.
- $\varphi_X(0) = 1$.

Théorème de Lévy.
Soit X et Y des v.a. réelles. Alors $P_X = P_Y$ si et seulement si $\varphi_X = \varphi_Y$.

$\Rightarrow$: trivial. $\Leftarrow$: admis (difficile).

Théorème (formule d'inversion de Fourier).
Soit X une v.a. réelle. Si sa fonction caractéristique φ_X est intégrable alors la loi de X est continue et sa densité est donnée par

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

(peu utilisé en proba)

Théorème.
Soit X une v.a. réelle. Si $X \in L^n$ alors φ_X est n fois dérivable et

$$\forall 1 \leq k \leq n,\ \varphi_X^{(k)}(0) = i^k E(X^k).$$

Réciproquement, si φ_X est k fois dérivable en 0 et $2n \leq k$ alors $X \in L^{2n}$.

Exemple . *Montrons que la fonction caractéristique de $\mathcal{N}(0,1)$ est $\varphi(t) = e^{-t^2/2}$. [Foata-Fuchs p165]*

$\varphi(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{itx} e^{-x^2/2} dx$, *donc φ est dérivable et*
$\varphi'(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} ixe^{itx} e^{-x^2/2} dx.$

Faisons une intégration par parties
$(u = ie^{itx}, v' = xe^{-x^2/2}, u' = ite^{itx}, v = -e^{x^2/2})$

$$\varphi'(t) = \underbrace{\frac{1}{\sqrt{2\pi}} \left[ie^{itx} e^{x^2/2} \right]_{-\infty}^{+\infty}}_{=0} - \underbrace{\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} te^{itx} e^{-x^2/2} dx}_{=t\varphi(t)}$$

On obtient une équation différentielle :

$$\varphi'(t) = -t\varphi(t).$$

Donc $\varphi(t)$ est de la forme $Ce^{-t^2/2}$.
De plus, $\varphi(0) = 1$ donc $C = 1$.

Conclusion : $\varphi(t) = e^{-t^2/2}$.

Remarque. *φ est C^∞ donc une v.a. de loi $\mathcal{N}(0,1)$ est dans L^n pour tout n, et on peut calculer ses moments à l'aide des dérivées de φ.*

Calcul de l'espérance et de la variance de $\mathcal{N}(0,1)$ à partir de la fonction caractéristique

Soit X une v.a. de loi $\mathcal{N}(0,1)$. On a $\varphi_X(t) = e^{-t^2/2}$.

φ_X est C^∞ donc $X \in L^n$ pour tout n.

$\varphi'_X(t) = -te^{-t^2/2}$, $\varphi'_X(0) = 0$ *donc $E(X) = 0$.*

$\varphi''_X(t) = (t^2 - 1)e^{-t^2/2}$, $\varphi''_X(0) = -1 = -E(X^2)$

donc $E(X^2) = 1$ et $\mathrm{Var}(X^2) = 1$.

Si $Y = \sigma X + m$, alors Y est de loi $\mathcal{N}(m, \sigma^2)$. En effet :

$$\begin{aligned} P(Y \in B) &= P\left(X \in \frac{B-m}{\sigma}\right) \\ &= \int_{\frac{B-m}{\sigma}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\ &= \int_B \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-m)^2}{2\sigma^2}\right) dy \end{aligned}$$

(changement de variable $y = \sigma x + m$)

On a $E(Y) = E(X) + m = m$ et $\mathrm{Var}(Y) = \sigma^2 \mathrm{Var}(X) = \sigma^2$.
Autrement dit, m et σ^2 sont l'espérance et la variance de la loi $\mathcal{N}(m, \sigma^2)$.

Fonction caractéristique de Y :

$$\varphi_Y(t) = E\left(e^{it(\sigma X + m)}\right) = e^{itm} E(e^{it\sigma X}) = e^{itm} \varphi_X(\sigma t)$$

donc $\varphi_Y(t) = e^{itm - \sigma^2 t^2/2}$.

3. Fonction génératrice

[Ouvrard 1 p138, Foata-Fuchs p99]

Définition.
Soit X une v.a. à valeurs dans $\mathbb{N}$ et $p_n = P(X = n)$.
La **fonction génératrice** de X est définie par

$$G_X(x) \overset{\text{def}}{=} \sum_{n=0}^{+\infty} p_n x^n.$$

Cette série entière est définie au moins sur $[-1, 1]$.

Dans $\mathbb{C}$, cette série entière est définie au moins pour tous les $z \in \mathbb{C}$ tels que $|z| \leq 1$.

Propriétés.
- $G_X(1) = 1$ *(car $G_X(1) = \sum P(X = n) = P(X \in \mathbb{N}) = 1$).*
- $|G_X(z)| \leq 1$ si $|z| \leq 1$.
- $p_n = \frac{G_X^{(n)}(0)}{n!}$ *(propriété d'une série entière).*

Proposition.
Soit X et Y deux v.a. à valeurs dans $\mathbb{N}$. Alors $G_X = G_Y \iff P_X = P_Y$.

Preuve.
Si $G_X = G_Y$, alors $\forall n \in \mathbb{N},\ P(X = n) = P(Y = n)$. (deux séries entières sont égales ssi tous les coefficients sont égaux).

Proposition.
Soit X et Y deux v.a. indépendantes à valeurs dans $\mathbb{N}$. Alors $G_{X+Y} = G_X G_Y$.

Preuve. *Il suffit de développer les séries G_{X+Y} et $G_X G_Y$.*

Exemple. *[Ouvrard 1 p139]*
Soit X une v.a. de loi de Poisson $\mathcal{P}(\lambda)$.

$$G_X(x) = \sum_{n=0}^{+\infty} e^{-\lambda} \cdot \frac{\lambda^n}{n!} x^n = e^{-\lambda} \sum_{n=0}^{+\infty} \frac{(\lambda x)^n}{n!}$$

donc $G_X(x) = e^{-\lambda} e^{\lambda x} = e^{\lambda(x-1)}$.

Si $P_X = \mathcal{P}(\lambda)$ alors $G_X(x) = e^{\lambda(x-1)}$.

Soit Y une v.a. de loi $\mathcal{P}(\lambda')$.
On a $G_Y(x) = e^{(\lambda'-1)x}$.

Si $X \perp Y$ alors

$$G_{X+Y} = e^{\lambda(x-1)} e^{\lambda'(x-1)} = e^{(\lambda+\lambda')(x-1)}.$$

Donc $X + Y$ suit une loi de Poisson $\mathcal{P}(\lambda + \lambda')$.

Exercice 12 : résultat similaire pour $\mathcal{B}(n, p)$.

Proposition.
Soit X une v.a. à valeurs dans $\mathbb{N}$.
- $X \in L^1 \iff G_X$ est dérivable en 1. Dans ce cas, $G'_X(1) = E(X)$.
- $X \in L^2 \iff G_X$ est 2 fois dérivable en 1. Dans ce cas, $G''_X(1) = E(X^2) - E(X)$.

Remarque. G_X est définie au moins sur $[-1, 1]$. Si G_X n'est pas définie au delà de 1, alors $G'_X(1)$ est en fait une dérivée à gauche de 1.
Si G_X est définie sur un intervalle ouvert contenant 1, alors G_X est C^∞ en 1 (série entière).

Exemple.
Soit X une v.a. de loi de Poisson $\mathcal{P}(\lambda)$.
On a vu que $G_X(x) = e^{\lambda(x-1)}$.

$G'_X(x) = \lambda e^{\lambda(x-1)}$ et $G'_X(1) = \lambda$. Donc $E(X) = \lambda$.

$G''_X(x) = \lambda^2 e^{\lambda(x-1)}$ et $G''_X(1) = \lambda^2$. Donc
$E(X^2) = \lambda^2 + \lambda$,
$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \lambda$.

Si $P_X = \mathcal{P}(\lambda)$ alors $E(X) = \lambda$ et $\mathrm{Var}(X) = \lambda$.

4. Transformée de Laplace

[Barbe-Ledoux, p70]

Soit X une v.a. réelle. La **transformée de Laplace** de X est la fonction

$$L_X \colon t \mapsto E(e^{tX}),$$

définie sur l'ensemble des t pour lesquels $e^{tX} \in L^1$ (cet ensemble contient toujours 0).

La transformée de Laplace, si elle est définie sur un voisinage de 0, caractérise la loi.

Proposition. Si L_X est définie sur un voisinage de 0, alors L_X est analytique et

$$\forall n \geq 0,\ (L_X)^{(n)}(0) = E(X^n).$$

L_X est aussi appelée **fonction génératrice des moments**.

Attention ! parfois, la transformée de Laplace est définie par $t \mapsto E(e^{-tX})$, soit $L_X(-t)$ avec les notations précédentes ! c'est par exemple le cas dans [Foata-Franchi-Fuchs]

V. Somme de v.a. indépendantes

On va souvent étudier des sommes de v.a. du type $X_1 + X_2 + \cdots + X_n$, où les X_i sont le plus souvent des v.a. indépendantes de même loi.

1. Variance

Propriété.
Si X et Y sont deux v.a. réelles indépendantes, alors $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

En particulier, si $(X_i)_{1\leq i\leq n}$ sont des v.a. réelles indépendantes de même loi, et si $S_n = X_1 + \cdots + X_n$, alors : $\mathrm{Var}(S_n) = n\mathrm{Var}(X_1)$, $E(S_n) = nE(X_1)$.

D'où : $\mathrm{Var}(\frac{S_n}{n}) = \frac{\mathrm{Var}(X_1)}{n}$ et $E(\frac{S_n}{n}) = E(X_1)$.

La propriété précédente est en fait vraie pour des v.a. non corrélées.

Définition. *Soit X, Y 2 v.a, $m_X = E(X)$ et $m_Y = E(Y)$. La covariance de X et Y est*

$$\mathrm{Cov}(X,Y) \stackrel{\text{def}}{=} E(XY) - m_X m_Y = E\left[(X-m_X)(Y-m_Y)\right].$$

Deux v.a. réelles X, Y sont dites **non corrélées** *si* $\mathrm{Cov}(X,Y) = 0$.

Si X et Y sont deux v.a. réelles non corrélées, alors $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Deux v.a. indépendantes sont non corrélées.
La réciproque n'est pas vraie.

Exemple. *Soit X de loi $\mathcal{N}(0,1)$ et $Y = X^2$.*

$E(X) = 0$ *par symétrie, et de même* $E(XY) = E(X^3) = 0$.

$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = 0$.

X et Y ne sont pas indépendantes vu la définition de Y.

2. Fonction caractéristique d'une somme

Théorème.
Soit X et Y deux v.a. réelles indépendantes. Alors $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$ pour tout $t \in \mathbb{R}$.

Preuve.
$\varphi_{X+Y}(t) = E(e^{it(X+Y)}) = E(e^{itX}e^{itY})$. *Or* $e^{itX} \perp e^{itY}$ *donc* $\varphi_{X+Y}(t) = E(e^{itX})E(e^{itY})$. □

Exemple.
Soit X et Y des v.a. indépendantes de loi $\mathcal{N}(m,\sigma^2)$ et $\mathcal{N}(m',\sigma'^2)$ respectivement. Leurs fonctions caractéristiques sont $\varphi_X(t) = e^{itm-\sigma^2t^2/2}$ et $\varphi_Y(t) = e^{itm'-\sigma'^2t^2/2}$ (voir page 47).

La fonction caractéristique de $X+Y$ est

$$e^{itm-\sigma^2t^2/2}.e^{itm'-\sigma'^2t^2/2} = e^{it(m+m')-(\sigma^2+\sigma'^2)t^2/2}.$$

Comme la fonction caractéristique caractérise la loi d'une v.a., on en déduit que la loi de $X+Y$ est :

$$\mathcal{N}(m+m', \sigma^2+\sigma'^2).$$

Exemple.
Calcul de la fonction caractéristique de $\mathcal{B}(n,p)$:

$$\begin{aligned}\varphi(t) &= \sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} e^{itk} \\ &= \sum_{k=0}^{n} C_n^k (pe^{it})^k (1-p)^{n-k} \\ &= (pe^{it}+1-p)^n\end{aligned}$$

Soit $X_1, \ldots, X_n$ des v.a. indépendantes de même loi $b(p) = \mathcal{B}(1,p)$. On a $\varphi_{X_i} = pe^{it}+1-p$.

Si $S_n = X_1 + \cdots + X_n$, alors $\varphi_{S_n}(t) = (pe^{it}+1-p)^n$. On en déduit que $P_{S_n} = \mathcal{B}(n,p)$. De plus :
$E(S_n) = nE(X_1) = np$,
$\mathrm{Var}(S_n) = n\mathrm{Var}(X_1) = np(1-p)$

$X_1 + \cdots + X_n$ est de loi $\mathcal{B}(n,p)$ et $X_{n+1} + \cdots + X_{n+m}$ est de loi $\mathcal{B}(m,p)$. Ces v.a. sont indépendantes, et leur somme S_{n+m} est de loi $\mathcal{B}(n+m,p)$.

exercice 15 : résultat similaire pour $\mathcal{P}(\lambda)$
[Barbe-Ledoux p67+91].

3. Loi d'une somme de v.a. indépendantes

Définition.

Soit μ et ν deux mesures finies sur $\mathbb{R}$. Le **produit de convolution** $\mu * \nu$ est la mesure sur $\mathbb{R}$ définie par : $\forall \phi : \mathbb{R} \to \mathbb{R}$ fonction borélienne bornée,

$$\int_{\mathbb{R}} \phi\, d(\mu * \nu) \overset{\text{def}}{=} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \phi(x+y)\, d\nu(y) \right) d\mu(x)$$

Si μ et ν sont des probabilités alors $\mu * \nu$ est une probabilité.

attention, $x+y$ et pas $x-y$. On verra pourquoi.

Propriétés.

L'opération $*$ est
– commutative,
– associative,
– distributive par rapport à l'addition.

Théorème.

Soit X et Y deux v.a. réelles indépendantes. La loi de $X+Y$ est égale à $P_X * P_Y$.

Preuve du théorème.
Pour identifier P_{X+Y}, on va calculer $\int_{\mathbb{R}} \phi\, dP_{X+Y}$ pour toute fonction mesurable bornée ϕ.

$$\int_{\mathbb{R}} \phi\, dP_{X+Y} = E(\phi(X+Y))$$

$(x,y) \mapsto \phi(x+y)$ est une fonction de $\mathbb{R}^2$ dans $\mathbb{R}$. La loi de (X,Y) est $P_X \otimes P_Y$ donc, par la formule de transfert,

$$\begin{aligned} E(\phi(X+Y)) &= \int_{\mathbb{R}^2} \phi(x+y)\, d(P_X \otimes P_Y)(x,y) \\ &= \int_{\mathbb{R}^2} \phi(x+y)\, dP_X(x) dP_Y(y) \\ &= \int_{\mathbb{R}} \phi\, d(P_X * P_Y) \end{aligned}$$

Par conséquent $P_{X+Y} = P_X * P_Y$. $\square$

Preuve des propriétés.
Le 2 premières propriétés sont immédiates si on utilise le théorème sur les v.a. Soit X, Y, Z des v.a. indépendantes de loi P, Q, R. On a $X+Y = Y+X$ et $(X+Y)+Z = X+(Y+Z)$ donc $P*Q = Q*P$ et $(P*Q)*R = P*(Q*R)$.

Soit $S = aQ + bR$.

$$\int \phi\, d(P*S) = \int \phi(x+y)\, dP(x) dS(y)$$
$$= a \int \phi(x+y)\, dP(x) dQ(y) + b \int \phi(x+y)\, dP(x) dR(y)$$
$$= a \int \phi\, d(P*Q) + b \int \phi\, d(P*R)$$

d'où $P*(aQ+bR) = aP*Q + bP*R$.

Lois continues. Soit X et Y des v.a de densité f et g. Si X et Y sont indépendantes, alors $X+Y$ a une loi de densité $f*g$, où

$$f*g(x) = \int_{\mathbb{R}} f(x-y) g(y)\, dy.$$

Cette définition de convolution sur les mesures (avec $x+y$) est compatible avec la convolution classique sur les fonctions.

Preuve.

$$\begin{aligned} \int_{\mathbb{R}} \phi\, d(P_X * P_Y) &= \int_{\mathbb{R}} \int_{\mathbb{R}} \phi(x+y)\, dP_X(x) dP_Y(y) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \phi(x+y) f(x) g(y)\, dx dy \end{aligned}$$

Si on pose $z = x+y$ alors

$$\begin{aligned} \int_{\mathbb{R}} \phi\, d(P_X * P_Y) &= \int_{\mathbb{R}} \int_{\mathbb{R}} \phi(z) f(z-y) g(y)\, dz dy \\ &= \int_{\mathbb{R}} \phi(z) f*g(z)\, dz \end{aligned}$$

On en déduit que $X+Y$ a une loi de densité $f*g$.

exercice 11 : si X, Y sont des v.a. continues et indépendantes, alors $X \neq Y$ p.s.

Lois discrètes.

Soit X et Y des v.a. indépendantes de lois discrètes.

Si X est de loi δ_a (càd $X = a$ p.s.) et Y de loi δ_b, alors $X+Y = a+b$ p.s.
Donc la loi de $X+Y$ est δ_{a+b}.

$$\boldsymbol{\delta_a * \delta_b = \delta_{a+b}}$$

Pour deux lois discrètes quelconques, la loi $P_X * P_Y$ s'obtient en utilisant la distributivité.

Exemple. $P_X = P_Y = b(\frac{1}{2})$.

$$\begin{aligned} P_X * P_Y &= \left(\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1\right) * \left(\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1\right) \\ &= \frac{1}{4}\underbrace{\delta_0 * \delta_0}_{\delta_0} + \frac{1}{4}\underbrace{\delta_0 * \delta_1}_{\delta_1} + \frac{1}{4}\underbrace{\delta_1 * \delta_0}_{\delta_1} + \frac{1}{4}\underbrace{\delta_1 * \delta_1}_{\delta_2} \\ &= \frac{1}{4}\delta_0 + \frac{1}{2}\delta_1 + \frac{1}{4}\delta_2 = \mathcal{B}(2, 1/2) \end{aligned}$$

VI. Processus de Poisson

[Billingsley p307]

On considère une vitre dans un lieu exposé. À chaque fois qu'elle est cassée, on la remplace. Quel est le nombre de vitres cassées observé au bout du temps t ?

Soit X_n la durée de vie de la n-ième vitre. $(X_n)_{n\geq 1}$ sont des v.a. indépendantes de même loi.

$S_n = X_1 + \cdots + X_n$ est le temps au bout duquel la vitre est cassée pour la n-ième fois.
On pose $S_0 = 0$.

Soit N_t le nombre de bris de vitres observés au temps $t \geq 0$. $N_t = \max\{n \geq 0 \mid S_n \leq t\}$. On a $N_t = 0$ si $X_1 > t$ (pas de vitre changée).
(N_t) est un **processus de comptage**.

On suppose que la durée de vie d'une vitre suit une loi exponentielle $\mathcal{E}(\lambda)$ avec $\lambda > 0$. Montrons que N_t suit alors une loi de Poisson $\mathcal{P}(\lambda t)$.

La densité de X_n est $f(x) = \mathbb{1}_{\mathbb{R}_+}(x)\lambda e^{-\lambda x}$.
La densité de S_n est $f_n = \underbrace{f * f * \cdots * f}_{n \text{ fois}}$.

On montre par récurrence que

$$f_n(x) = \mathbb{1}_{\mathbb{R}_+}(x)\frac{\lambda^n x^{n-1}}{(n-1)!}e^{-\lambda x}.$$

$N_t \geq n \iff S_n \leq t$
(au temps t on a changé au moins n vitres). Donc

$$P(N_t \geq n) = \int_0^t \frac{\lambda}{(n-1)!}(\lambda x)^{n-1}e^{-\lambda x}dx.$$

Faisons une intégration par parties pour $n \geq 2$
$(u = \frac{(\lambda x)^{n-1}}{(n-1)!}, v' = \lambda e^{-\lambda x}, u' = \frac{\lambda(\lambda x)^{n-2}}{(n-2)!}, v = -e^{-\lambda x})$

$$P(N_t \geq n) = \left[-e^{-\lambda x}\frac{(\lambda x)^{n-1}}{(n-1)!}\right]_0^t + \int_0^t \lambda\frac{(\lambda x)^{n-2}}{(n-2)!}e^{-\lambda x}dx$$

$$P(N_t \geq n) = -e^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!} + P(N_t \geq n-1).$$

Pour avoir $N_t = n$, il faut $N_t \geq n$ et $N_t < n+1$, c'est-à-dire $\{N_t = n\} = \{N_t \geq n\} \setminus \{N_t \geq n+1\}$.

Donc $P(N_t = n) = P(N_t \geq n) - P(N_t \geq n+1)$ et

$$P(N_t = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!} \text{ si } n \geq 1.$$

De plus, $P(N_t = 0) = P(X_1 > t) = e^{-\lambda t}$.

Conclusion : N_t suit une loi de Poisson $\mathcal{P}(\lambda t)$.

Remarque : $P(N_t \in \mathbb{N}) = 1$ donc on a $N_t < \infty$ p.s.

Le fait que $N_t < \infty$ p.s. peut être vu comme une conséquence de la loi des grands nombres : $S_n/n \to \lambda$ p.s. avec $\lambda > 0$ donc $Sn \to +\infty$ p.s. et donc $N_t < +\infty$ p.s.

Remarque. On modélise par une loi exponentielle des temps d'attente : temps entre 2 clients, entre 2 passages de taxis, entre 2 accidents.... Le nombre de clients/taxis/accidents pendant une durée donnée suit une loi de Poisson.

$(N_t)_t$ est un **processus de Poisson**. Il vérifie :
i) $N_0 = 0$,
ii) $\forall 0 \leq t_1 < \cdots < t_k$, les v.a. $N_{t_2} - N_{t_1}, \cdots, N_{t_k} - N_{t_{k-1}}$ sont indépendantes,
iii) si $0 \leq s < t$, $N_t - N_s$ est de loi $\mathcal{P}(\lambda(t-s))$.

Preuve de ces propriétés pages suivantes.

Toute famille de v.a. vérifiant ces propriétés est appelée un processus de Poisson.

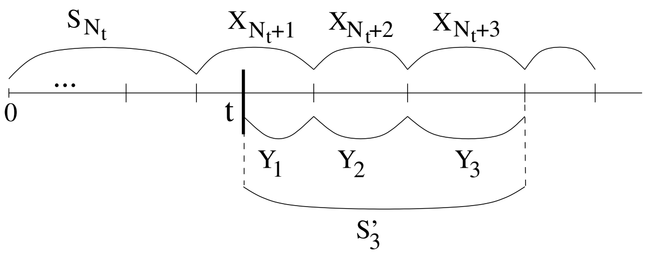
Remarque. Tout processus de comptage vérifiant (i-ii-iii) peut être décrit avec des v.a. exponentielles comme ci-dessus. [Billingsley]

Pour être un processus de comptage, il faut les hypothèses : $\forall \omega, t \mapsto N_t(\omega)$ est continue à droite, les sauts sont égaux à 1 et $\lim_{t\to+\infty} N_t(\omega) = +\infty$.

Processus de Poisson, suite [Billingsley]

On garde les notations précédentes. Montrons que $N_{t+s} - N_t$ est de loi $\mathcal{P}(\lambda s)$ (propriété (iii)) et que cette v.a. est indépendante de N_t (propriété (ii) pour $k = 2$).

On se place au temps t et on observe les pannes suivantes. Comme l'entier N_t vérifie $S_{N_t} \leq t < S_{N_t+1}$, le temps d'attente jusqu'à la prochaine panne est $Y_1 = S_{N_t+1} - t$, les délais entre les pannes suivantes sont $Y_2 = X_{N_t+2}, Y_3 = X_{N_t+3}, \ldots, Y_k = X_{N_t+k}$ pour $k \geq 2$. Soit $S'_k = Y_1 + \cdots + Y_k$.



On pose $S'_0 = 0$, et $N'_s = \max\{k \geq 0 \mid S'_k \leq s\}$ pour $s \geq 0$ (nombre de pannes observées pendant le délai s, à partir du temps t).

On va montrer que si on se place au temps t, on observe la même chose qu'en se plaçant au temps 0. C'est assez naturel puisqu'on considère des lois X_n sans mémoire.

On a $S'_k = S_{N_t+k} - t$ par définition et $N'_s = N_{t+s} - N_t$ car le nombre de pannes observées entre les temps t et $t + s$ est égal au nombre de pannes entre 0 et $t + s$ moins le nombre de pannes entre 0 et t (cela peut également être montré à partir des définitions de N'_s et N_t).

Déterminons la loi de (N_t, Y_1). Il suffit de considérer les événements $\{N_t = n, Y_1 > s\}$ pour déterminer cette loi (parce que N_t est une loi discrète et que les ensembles $]s, +\infty[$ forment un π-système engendrant les boréliens de $\mathbb{R}$).

Soit $n \in \mathbb{N}$ et $s > 0$.

$$\begin{aligned}
\{N_t = n, Y_1 > s\} &= \{S_n \leq t < S_{n+1},\ S_{n+1} - t > s\} \\
&= \{S_n \leq t,\ S_{n+1} > t + s\} \\
&\quad \text{(car si } S_{n+1} > t+s\text{, alors } S_{n+1} > t\text{)} \\
&= \{S_n \leq t,\ X_{n+1} > t + s - S_n\} \\
&\quad \text{(car } S_{n+1} = S_n + X_{n+1}\text{)}
\end{aligned}$$

Donc $P(N_t = n, Y_1 > s) = P(S_n \leq t,\ X_{n+1} > s + t - S_n)$. Si on pose $A = \{(x, y) \in \mathbb{R}^2 \mid x \leq y, y > s + t - x\}$, alors $P(N_t = n, Y_1 > s) = E(\mathbb{1}_A(S_n, X_{n+1}))$.

S_n et X_{n+1} sont indépendants par hypothèse, donc la loi de (S_n, X_{n+1}) est la loi produit $P_{S_n} \otimes P_{X_{n+1}}$, d'où :

$$\begin{aligned}
P(N_t = n, Y_1 > s) &= \int_A dP_{S_n}(x) dP_{X_{n+1}}(y) \\
&= \int_{0 \leq x \leq t} \left(\int_{y > s+t-x} dP_{X_{n+1}}(y) \right) dP_{S_n}(x)
\end{aligned}$$

Or $\displaystyle\int_a^{+\infty} \lambda e^{-\lambda x} dx = e^{-\lambda a}$, donc

$$\begin{aligned}
\int_{y > s+t-x} dP_{X_{n+1}}(y) &= \int_{s+t-x}^{+\infty} \lambda e^{-\lambda y} dy = e^{-\lambda s} e^{-\lambda(t-x)} \\
&= e^{-\lambda s} \int_{y > t-x} dP_{X_{n+1}}(y)
\end{aligned}$$

On a donc

$$\begin{aligned}
P(N_t = n, Y_1 > s) &= e^{-\lambda s} \int_{\{0 \leq x \leq t, y > t-x\}} dP_{X_{n+1}}(y) dP_{S_n}(x) \\
&= e^{-\lambda s} P(S_n \leq t, X_{n+1} > t - S_n) \\
&= e^{-\lambda s} P(S_n \leq t < S_{n+1}) \\
&= e^{-\lambda s} P(N_t = n)
\end{aligned}$$

Si Z est une v.a. de loi $\mathcal{E}(\lambda)$ indépendante avec N_t, alors $P(N_t = n, Z > s) = P(N_t = n)P(Z > s) = P(N_t = n)e^{-\lambda s}$.

On voit alors que la loi de (N_t, Y_1) est la même que la loi de (N_t, Z), qui est égale à $P_{N_t} \otimes \mathcal{E}(\lambda)$. On en déduit que Y_1 est de loi $\mathcal{E}(\lambda)$ et que N_t et Y_1 sont indépendants.

Maintenant, on cherche la loi de $(N_t, Y_1, \ldots, Y_k)$. Soit $s_1, \ldots, s_k > 0$. On a

*$P(N_t = n, Y_1 > s_1, \ldots, Y_k > s_k) =$
$P(S_n \leq t < S_{n+1}, X_{n+1} > s_1 + t - S_n, X_{n+2} > s_2, \ldots, X_{n+k} > s_k\}$
$= P(S_n \leq t < S_{n+1}, X_{n+1} > s_1 + t - S_n) P(X_{n+2} > s_2) \cdots P(X_{n+k} > s_k)$
(l'indépendance des $(X_n)_{n \geq 1}$ implique que (S_n, X_{n+1}) est indépendant de $X_{n+2+}, \ldots, X_{n+k}$)*

$P(S_n \leq t < S_{n+1}, X_{n+1} > s_1 + t - S_n) = P(N_t = n, Y_1 > s_1)$, donc par ce qui précède et en raison des lois des X_i :

$P(N_t = n, Y_1 > s_1, \ldots, Y_k > s_k) = e^{-\lambda s_1} e^{-\lambda s_2} \ldots e^{-\lambda s_k} P(N_t = n)$.

Conclusion : la loi de $(N_t, Y_1, \ldots, Y_k)$ est $P_{N_t} \otimes \mathcal{E}(\lambda) \otimes \cdots \otimes \mathcal{E}(\lambda)$.

La loi de $(N_t, Y_1, \dots, Y_k)$ est une loi produit, ce qui implique que N_t est indépendant des v.a. $(Y_k)_{k\geq 1}$, donc N'_s (qui dépend de $(Y_k)_{k\geq 1}$) et N_t sont indépendants. Or $N'_s = N_{t+s} - N_t$, donc $N_{t+s} - N_t$ et N_t sont des v.a. indépendantes. C'est la propriété (ii) avec $k = 2$, $t_1 = t$ et $t_2 = t + s$.

De plus, ceci montre que les $(Y_k)_{k\geq 1}$ sont des v.a. indépendantes de même loi $\mathcal{E}(\lambda)$. Donc, par analogie entre la définition de N'_s et de N_t, on obtient que $N'_s = N_{t+s} - N_t$ est de loi $\mathcal{P}(\lambda s)$. C'est la propriété (iii).

La propriété (ii) pour k quelconque se montre de la même façon, mais c'est beaucoup plus lourd à écrire.

VII. Convergences d'une suite de v.a.

1. Convergence presque sûre

Définition.
Soit $(X_n)_{n\geq 1}$ et X des v.a. réelles définies sur Ω. On dit que la suite de v.a. $(X_n)_{n\geq 1}$ **converge presque sûrement** vers la v.a. X si $P(\lim_{n\to+\infty} X_n = X) = 1$.

Autrement dit, il existe un sous-ensemble Ω' tel que $P(\Omega') = 1$ et $\forall \omega \in \Omega'$, $\lim_{n\to+\infty} X_n(\omega)$ existe et vaut $X(\omega)$ (convergence simple sur un ensemble de mesure 1).

On note $X_n \xrightarrow{p.s.} X$.

On peut généraliser la définition à des v.a. prenant leurs valeurs dans un même ensemble E, qui doit être un espace topologique (c'est-à-dire qu'on peut définir une limite).

2. Convergence en probabilité

Définition. Soit $(X_n)_{n\geq 1}$ et X des v.a. réelles définies sur Ω. On dit que la suite $(X_n)_{n\geq 1}$ **converge en probabilité** vers la v.a. X si

$$\forall \varepsilon > 0, \ \lim_{n\to+\infty} P(|X_n - X| > \varepsilon) = 0.$$

On note $X_n \xrightarrow{P} X$.

On peut définir la convergence en proba pour les v.a. à valeurs dans $\mathbb{R}^d$ (en prenant la norme au lieu de la valeur absolue).

Propriété. Si $X_n \xrightarrow{p.s.} X$, alors $X_n \xrightarrow{P} X$.

Réciproque partielle :

Théorème.
Si $X_n \xrightarrow{P} X$, alors il existe une sous-suite $(X_{n_k})_{k\geq 1}$ qui converge p.s. vers X.

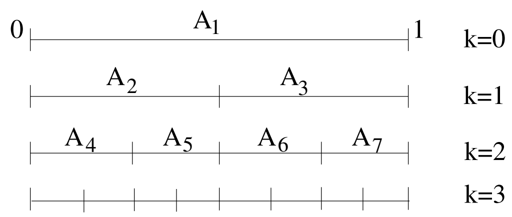
La convergence en proba autorise X_n à s'écarter beaucoup de X sur un ensemble petit qui peut varier avec n, donc pour un ω donné on peut avoir $X_n(\omega)$ loin de $X(\omega)$ avec n aussi grand qu'on veut. Ceci est interdit par la convergence p.s. puisque $\forall \omega \in \Omega'$, $X_n(\omega)$ tend vers $X(\omega)$ (Ω' est un ensemble de mesure 1 dans Ω, il est indépendant de n).

Exemple : convergence en proba mais pas p.s.

On définit $(A_n)_{n\geq 1}$ des sous-intervalles de $[0, 1]$ de la façon suivante : pour $0 \leq j \leq 2^k - 1$,

$$A_{2^k+j} = \left[\frac{j}{2^k}, \frac{j+1}{2^k}\right].$$

Chacun des intervalles $(A_i)_{2^k \leq i < 2^{k+1}}$ est de longueur $1/2^k$, et leur union vaut $[0, 1]$.



On prend P la mesure de Lebesgue sur $\Omega = [0, 1]$ et on pose $X_n = \mathbb{1}_{A_n}$. C'est une v.a. à valeurs dans $\{0, 1\}$.

Soit $0 < \varepsilon < 1$. On a $P(|X_n| \geq \varepsilon) = P(A_n) \leq 1/2^k$ si $n \geq 2^k$. Donc $\lim_{n\to+\infty} P(|X_n| \geq \varepsilon) = 0$, autrement dit $X_n \xrightarrow{P} 0$.

Par contre, $(X_n)_{n\geq 1}$ ne converge pas p.s. vers 0. En effet, fixons $\omega \in [0,1]$. Soit N un entier, et k tel que $2^k \geq N$. On a :

$$\bigcup_{0\leq j<2^k} A_{2^k+j} = [0,1]$$

donc il existe $n \geq 2^k$ tel que $\omega \in A_n$, et donc

$$\forall N, \exists n \geq N, X_n(\omega) = 1.$$

Ceci interdit la convergence de $(X_n(\omega))_{n\geq 1}$ vers 0, et comme ceci est vrai pour tout $\omega \in [0,1]$, on ne peut pas avoir convergence simple sur un ensemble de mesure 1 (càd convergence p.s.).

Propriété. *$(X_n)_{n\geq 1}$ converge p.s. vers X ssi $\forall \varepsilon > 0$, $\lim_{n\to\infty} P\left(\sup_{k\geq n} |X_k - X| \geq \varepsilon\right) = 0$.*

Preuve.
$\lim X_n(\omega) = X(\omega) \Leftrightarrow \forall\varepsilon > 0, \exists n, \forall k \geq n, |X_k(\omega) - X(\omega)| < \varepsilon$. *Soit $\Omega' = \{\omega \in \Omega; \lim X_n(\omega) = X(\omega)\}$. Alors*

$$\Omega' = \bigcap_{\varepsilon>0} \underbrace{\bigcup_{n\geq 1}\bigcap_{k\geq n} \{\omega; |X_n(\omega) - X(\omega)| < \varepsilon\}}_{A(\varepsilon)}.$$

$X_n \xrightarrow{p.s.} X \Leftrightarrow P(\Omega') = 1$. Or $\Omega' \subset A_\varepsilon$ donc si $P(\Omega') = 1$ alors $\forall\varepsilon > 0, P(A_\varepsilon) = 1$. Réciproquement, comme $\Omega' = \bigcap_{p\geq 1} A(\frac{1}{p})$, si $P(A(\varepsilon)) = 1$ pour tout $\varepsilon > 0$ alors $P(\Omega') = 1$.
Donc $X_n \xrightarrow{p.s.} X$ est équivalent à $\forall\varepsilon > 0,\ P(A(\varepsilon)) = 1$.
Autrement dit :

$$(*) \qquad X_n \xrightarrow{p.s.} X \Longleftrightarrow \forall\varepsilon > 0, P(A(\varepsilon)^c) = 0.$$

$$\begin{aligned} A(\varepsilon)^c &= \bigcap_{n\geq 1}\bigcup_{k\geq n} \{\omega \in \Omega; |X_n(\omega) - X(\omega)| \geq \varepsilon\} \\ &= \bigcap_{n\geq 1} \underbrace{\{\omega \in \Omega; \sup_{k\geq n} |X_n(\omega) - X(\omega)| \geq \varepsilon\}}_{B_n(\varepsilon)} \end{aligned}$$

$B_n(\varepsilon)$ est une suite décroissante d'ensembles, donc

$$P((A(\varepsilon)^c) = P\left(\bigcap_{n\geq 1} B_n(\varepsilon)\right) = \lim_{n\to+\infty} P(B_n(\varepsilon)).$$

En remplaçant dans $(*)$, on obtient :
$X_n \overset{p.s.}{\to} X \leftrightarrow \forall\varepsilon > 0, \lim_{n\to+\infty} P(\{\omega; \sup_{k\geq n} |X_n(\omega) - X(\omega)| \geq \varepsilon\}) = 0.$

Cette propriété implique que si $X_n \xrightarrow{p.s.} X$ alors $X_n \xrightarrow{P} X$.

3. Convergence L^p

Définition.
Soit $(X_n)_{n\geq 1}$ et X des v.a. réelles définies sur Ω. La suite de v.a. $(X_n)_{n\geq 1}$ **converge dans** L^p vers la v.a. X si $\lim_{n\to+\infty} \|X_n - X\|_p = 0$, autrement dit

$$\lim_{n\to+\infty} E(|X_n - X|^p) = 0.$$

On note $X_n \xrightarrow{L^p} X$.

On peut considérer des v.a. à valeurs dans $\mathbb{R}^d$.

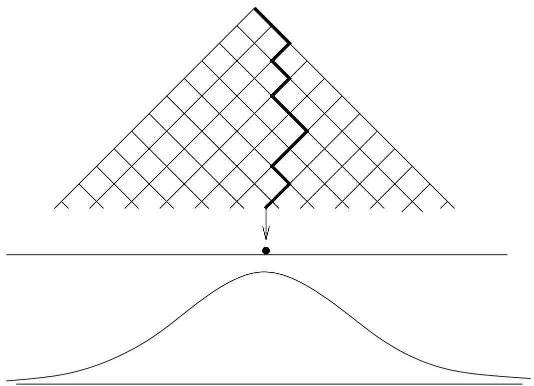
On considère en général la convergence L^1 ou L^2.

Propriété.
Si $X_n \xrightarrow{L^p} X$, alors $X_n \xrightarrow{P} X$.

La convergence L^p n'implique pas la convergence p.s., ni l'inverse.

4. Convergence en loi

a) Sablier de Pascal



On verse du sable par le haut, les lignes représentent des canaux où le sable s'écoule. A chaque bifurcation, un grain de sable choisit gauche ou droite de façon équiprobable. Le tas de sable qui se forme ressemble à une gaussienne. Les grains de sable suivant une loi binomiale ($X = k$ si le sable sort par la sortie k), on a envie de dire que la loi binomiale "ressemble" à la loi normale. La notion de convergence en loi sert à donner un sens à cette ressemblance.

b) Définir la convergence en loi

Soit $(\mu_n)_{n\geq 0}$ et μ des mesures de probabilité sur $\mathbb{R}$. Quand peut-on dire que μ_n converge vers μ ?

On pourrait demander que $\mu_n(B)$ converge vers $\mu(B)$ pour tout borélien B, mais dans certains cas cela ne correspond pas à l'intuition.

Exemple.
Soit $\mu_n = \delta_{1/n}$ et $\mu = \delta_0$ (mesures de Dirac). Soit $B = [a, b]$, $a < b$.
– Si $b \neq 0$, $\mu_n(B) = \mu(B)$ pour tout n assez grand donc $\mu_n(B) \to \mu(B)$.
– Si $b = 0$ alors $\mu(B) = 1$ mais $\forall n \geq 1, \mu_n(B) = 0$.

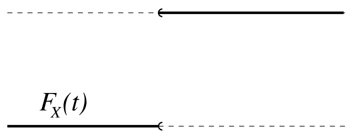
On a pourtant envie de dire que μ_n tend vers μ.

Il y a un problème de "continuité".

On se restreint aux intervalles $]-\infty, t]$ (ensembles qui engendrent les boréliens). Si μ est la loi de X, alors $\mu(]-\infty, t]) = F_X(t)$ (fonction de répartition).

Définition.
Soit $(X_n)_{n\geq 0}$ et X des v.a. réelles. On dit que X_n **converge en loi** vers X si $\lim\limits_{n\to+\infty} F_{X_n}(t) = F_X(t)$ pour tout t qui est point de continuité de F_X. On note $X_n \xrightarrow{loi} X$.

Suite de l'exemple.
Soit X_n de loi $\mu_n = \delta_{1/n}$ et X de loi $\mu = \delta_0$.
$F_{X_n} = \mathbb{1}_{[1/n,+\infty[}$, $F_X = \mathbb{1}_{[0,+\infty[}$.



$\forall t \neq 0$, $F_{X_n}(t) \to F_X(t)$, donc $X_n \xrightarrow{loi} X$.

La fonction F_X n'est pas continue en 0,
$F_{X_n}(0) = 0$ ne converge pas vers $F_X(0) = 1$.

Une mesure μ est également définie par les valeurs des intégrales $\int f d\mu$ pour f dans L^1. On peut se restreindre aux fonction f continues (qui sont denses dans L^1).

Le théorème suivant donne une définition équivalente de la convergence en loi.

Théorème.
$X_n \xrightarrow{loi} X$ si et seulement si, pour toute fonction continue bornée $f : \mathbb{R} \to \mathbb{R}$,

$$\lim_{n\to+\infty} \int_{\mathbb{R}} f(x)\, dP_{X_n}(x) = \int_{\mathbb{R}} f(x)\, dP_X(x)$$

Comme $\int \mathbb{1}_B d\mu = \mu(B)$, le théorème n'est pas vrai pour toutes les fonctions L^1 pour la même raison que précédemment.

Propriété.
Si $X_n \xrightarrow{P} X$, alors $X_n \xrightarrow{loi} X$.

Preuve de la propriété.
On suppose que $X_n \xrightarrow{P} X$. Soit $\varepsilon > 0$.

Si ω est tel que $X_n(\omega) \leq t$, on distingue 2 cas :
– soit $|X_n(\omega) - X(\omega)| < \varepsilon$, et alors $X(\omega) < X_n(\omega) + \varepsilon \leq t + \varepsilon$,
– soit $|X_n(\omega) - X(\omega)| \geq \varepsilon$,
ce qui nous donne :

$\{\omega | X_n(\omega) \leq t\} \subset \{\omega \mid X(\omega) \leq t + \varepsilon\} \cup \{\omega \mid |X_n(\omega) - X(\omega)| \geq \varepsilon\}$.
Donc $P(X_n \leq t) \leq P(X \leq t + \varepsilon) + P(|X_n - X| \geq \varepsilon)$.

Si n tend vers l'infini (avec ε fixé), alors $P(|X_n - X| \geq \varepsilon) \to 0$ (convergence en probabilité). Donc

$$\limsup_{n\to+\infty} F_{X_n}(t) \leq F_X(t + \varepsilon).$$

Si t est un point de continuité de F_X, on fait tendre ε vers 0 et $F_X(t + \varepsilon) \to F_X(t)$. D'où $\limsup\limits_{n\to+\infty} F_{X_n}(t) \leq F_X(t)$.

Inégalité dans l'autre sens :
On écrit $P(X \leq t - \varepsilon) \leq P(X_n \leq t) + P(|X_n - X| \geq \varepsilon)$, d'où

$$F_{X_n}(t) \geq F_X(t - \varepsilon) - P(|X_n - X| \geq \varepsilon).$$

D'où

$$\liminf_{n\to+\infty} F_{X_n}(t) \geq F_X(t - \varepsilon),$$

et si t est un point de continuité de F_X, $\liminf\limits_{n\to+\infty} F_{X_n}(t) \geq F_X(t)$.

Conclusion : $\lim\limits_{n\to+\infty} F_{X_n}(t) = F_X(t)$ pour tout t point de continuité de F_X. □

Propriété. La convergence en loi dépend uniquement des lois des X_n *(si $X_n \xrightarrow{loi} X$ et si les v.a. Y_n et Y ont les mêmes lois que X_n et X alors $Y_n \xrightarrow{loi} Y$).*

Remarque. On ne demande pas que les X_n soient définies sur le même ensemble.

Théorème de Lévy (admis).
Les conditions suivantes sont équivalentes :

i) $X_n \xrightarrow{loi} X$,

ii) $\lim_{n\to+\infty} \varphi_{X_n}(t) = \varphi_X(t)$ pour tout $t \in \mathbb{R}$.

Remarque. Il ne suffit pas d'avoir la convergence des fonctions φ_{X_n}, il faut s'assurer que la limite est la fonction caractéristique d'une certaine loi.

Remarque.
On sait que si $X_n \xrightarrow{loi} X$ alors $\int f dP_{X_n} \to \int f dP_X$ si f est continue bornée. Donc

$$\varphi_{X_n}(t) = \int e^{itx} dP_{X_n}(x) \to \varphi_X(t) = \int e^{itx} dP_X(x)$$

C'est i)⇒ ii). Le théorème de Lévy donne la réciproque. L'idée est que les combinaisons linéaires des fonctions $x \mapsto e^{itx}$ sont suffisamment riches pour approcher les fonctions continues. La preuve, difficile, est souvent admise.

5. Convergence en loi de v.a. discrètes

Proposition.
Soit $(X_n)_{n\geq 0}$ et X des v.a. à valeurs dans $\mathbb{N}$. Alors $X_n \xrightarrow{loi} X$ si et seulement si pour tout entier $k \in \mathbb{N}$,

$$\lim_{n\to+\infty} P(X_n = k) = P(X = k).$$

Preuve. *Supposons que $\forall k \in \mathbb{N}$, $\lim_{n\to\infty} P(X_n = k) = P(X = k)$.*
Tout point $t \notin \mathbb{N}$ est un point de continuité de F_X.
Si $k < t < k+1$, $k \in \mathbb{N}$, alors $F_{X_n}(t) = \sum_{i=0}^{k} P(X_n = i)$.
Si $\forall k \in \mathbb{N}$, $\lim_{n\to+\infty} P(X_n = k) = P(X = k)$ alors si $t < 0$ on a $F_{X_n}(t) = 0 = F_X(t)$ et si $t > 0, t \notin \mathbb{N}$, il existe $k \in \mathbb{N}$ tel que $k < t < k+1$ donc $F_{X_n}(t) = \sum_{i=0}^{k} P(X_n = i)$ converge vers $F_X(t) = \sum_{i=0}^{k} P(X = i)$.

Réciproquement, Si $X_n \xrightarrow{loi} X$, on obtient $\lim_{n\to+\infty} P(X_n = 0) = P(X = 0)$ en prenant $t_0 \in]0, 1[$, puis

$$\lim_{n\to+\infty} P(X_n = 0) + P(X_n = 1) = P(X = 0) + P(X = 1)$$

en prenant $t_1 \in]1, 2[$, d'où $\lim_{n\to+\infty} P(X_n = 1) = P(X = 1)$. Par récurrence $\lim_{n\to+\infty} P(X_n = k) = P(X = k)$.

a) Tirage avec/sans remise

N : population totale.
p : proportion de A dans la population totale.
Le nombre de personnes A est Np

On interroge n personnes, sans remise, et on note Y_N le nombre de A parmi les n.
Y_N suit une loi hypergéométrique :

$$P(Y_N = k) = \frac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}} \text{ si } 0 \leq k \leq \min(n, N_1),$$

et $P(Y_N = k) = 0$ sinon.

Propriété.
$\forall k, \lim_{N\to+\infty} P(Y_N = k) = \binom{n}{k} p^k (1-p)^{n-k}$,
Donc Y_N converge en loi vers $B(n, p)$ quand $N \to \infty$.

Conclusion : les tirages avec et sans remise se ressemblent quand la population est importante.

Ceci justifie la modélisation des sondages par des v.a. indépendantes de loi de Bernoulli.

b) Convergence de lois binomiales vers une loi de Poisson

Théorème.
Soit $(X_n)_{n\geq 1}$ des v.a. de loi binomiale $\mathcal{B}(n, p_n)$ telles que

$$\lim_{n\to+\infty} np_n = \lambda > 0$$

alors X_n converge en loi vers une v.a. de loi de Poisson $\mathcal{P}(\lambda)$ quand $n \to +\infty$.

(cas particulier : $p_n = \lambda/n$.)

approximation d'une binomiale par une loi de Poisson : il est d'usage de remplacer $\mathcal{B}(n, p)$ par $\mathcal{P}(np)$ quand $p < 0,1$ et n est grand (par exemple $n \geq 100$).

Preuve du théorème.

Il suffit de montrer que $\lim_{n\to+\infty} P(X_n=k) = e^{-\lambda}\frac{\lambda^k}{k!}$.

On fixe $k\in\mathbb{N}$.

$$P(X_n=k)=\frac{n!}{k!(n-k)!}p_n^k(1-p_n)^{n-k}$$

On fait apparaître $np_n\to\lambda$, *on garde* $k!$ *:*

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

Calcul de $\lim \ln\left(1-\frac{np_n}{n}\right)^{n-k} = \lim (n-k)\ln\left(1-\frac{np_n}{n}\right)$ *:*

$\lim_{n\to+\infty}(np_n)^k=\lambda^k$.

$\frac{np_n}{n}\sim\frac{\lambda}{n}\to 0$ *et* $\ln(1+x)\sim x$ *quand* $x\to 0$ *donc*

$(n-k)\ln\left(1-\frac{np_n}{n}\right)\sim -(n-k)\frac{np_n}{n}\sim -(n-k)\frac{\lambda}{n}\to -\lambda$

et $\lim_{n\to+\infty}\left(1-\frac{np_n}{n}\right)^{n-k}=e^{-\lambda}$.

Formule de Stirling : $n!\sim n^n e^{-n}\sqrt{2\pi n}$. *Donc*

$$\begin{aligned}\frac{n!}{(n-k)!n^k} &\sim \frac{n^n e^{-n}\sqrt{2\pi n}}{n^k(n-k)^{n-k}e^{-(n-k)}\sqrt{2\pi(n-k)}}\\ &= \left(\frac{n}{n-k}\right)^{n-k}e^{-k}\sqrt{\frac{n}{n-k}}\end{aligned}$$

$$\ln\left(\frac{n}{n-k}\right)^{n-k} = (n-k)\ln\left(1+\frac{k}{n-k}\right)\sim (n-k)\frac{k}{n-k}=k$$

donc $\lim_{n\to+\infty}\left(\frac{n}{n-k}\right)^{n-k}=e^k$ *et* $\lim_{n\to+\infty}\frac{n!}{(n-k)!n^k}=1$.

Conclusion : $\forall k\in\mathbb{N},\ \lim_{n\to+\infty}P(X_n=k)=e^{-\lambda}\frac{\lambda^k}{k!}$. $\square$

c) Exemple modélisé par une loi de Poisson

[Dacunha-Castelle p34]

Une usine produit un tissu avec parfois des défauts. Le nombre moyen de défauts par mètre carré est λ, et les défauts se produisent au hasard, indépendamment les uns des autres.

Soit X le nombre de défauts observés sur un mètre carré.

Proposition. X suit une loi de Poisson $\mathcal{P}(\lambda)$.

Remarque.
Pour la même raison, le nombre d'accidents (ou d'appels à un standard,...) pendant un temps donné est modélisé par une loi de Poisson.

Quand on a des événements indépendants répartis uniformément avec une certaine densité λ, le nombre d'événements pendant le temps 1 suit une loi $\mathcal{P}(\lambda)$. On modélise en fait un nombre infini d'événements répartis uniformément sur $[0,+\infty[$ (ce qui n'a de sens qu'en passant à la limite).

On sait que $E(X)=\lambda$ *(nombre moyen de défauts par mètre carré).*

On découpe le coupon de tissu en n *surfaces égales, et on appelle* $X_1,\dots,X_n$ *le nombre de défauts sur chaque morceau. On a* $X=X_1+\cdots+X_n$.

$X_1,\dots,X_n$ *sont des v.a. indépendantes et de même loi, donc leur espérance vaut* $\frac{E(X)}{n}=\frac{\lambda}{n}$.

Si n *est assez grand, il y a peu de chance qu'on observe 2 défauts sur le même morceau. Il est donc raisonnable d'approcher la loi des* X_i *par une loi de Bernoulli* $b(p)$*, avec* $p=E(X_i)=\frac{\lambda}{n}$.

La somme de n *v.a. indépendantes de loi* $b(p)$ *étant une v.a. de loi binomiale* $\mathcal{B}(n,p)$*, on approche alors la loi de* $X=X_1+\cdots+X_n$ *par* $\mathcal{B}(n,\frac{\lambda}{n})$ *quand* n *tend vers l'infini. D'après le théorème précédent (page 83), des v.a. de loi* $\mathcal{B}(n,\frac{\lambda}{n})$ *convergent en loi vers* $\mathcal{P}(\lambda)$.

Conclusion : X *suit une loi de Poisson* $\mathcal{P}(\lambda)$.

VIII. Loi des grands nombres

On considère $(X_n)_{n\geq 1}$ une suite de v.a. vérifiant certaines hypothèses (portant sur l'indépendance, l'intégrabilité, une loi commune aux X_n,...)

La loi des grands nombres concerne la convergence de la suite

$$\frac{X_1+\cdots+X_n}{n}.$$

Loi faible : convergence en probabilité.
Loi forte : convergence presque sûre.

Selon les hypothèses, plus ou moins fortes, on obtient différentes versions de loi des grands nombres (faible ou forte).

Si $(X_n)_{n\geq 1}$ sont des v.a. de même loi, soit m leur espérance commune et $Y_n = X_n - m$. Alors les Y_n sont centrées (c.-à-d. $E(Y_n) = 0$) et

$$\frac{X_1 + \cdots + X_n}{n} \to m \iff \frac{Y_1 + \cdots + Y_n}{n} \to 0.$$

On peut donc se ramener à des v.a. centrées.

Considérer des v.a. centrées permet des énoncés plus généraux, avec des v.a. qui n'ont pas nécessairement la même loi.

Quand les v.a. $(X_n)_{n\geq 1}$ ont la même loi, on dit aussi qu'elles sont **identiquement distribuées** (iid=indépendantes identiquement distribuées).

Notation.
Dans la suite, on notera $S_n = X_1 + \cdots + X_n$, avec $(X_n)_{n\geq 1}$ des v.a. toutes définies sur le même Ω.

1. Inégalités de Markov et de Tchebychev

Inégalité de Markov.
Soit X une v.a. réelle positive L^1 et $t > 0$.

$$P(X \geq t) \leq \frac{E(X)}{t}.$$

Inégalité de Bienaymé-Tchebychev.
Soit X une v.a. réelle L^2 et $t > 0$.

$$P(|X - E(X)| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}.$$

L'inégalité de Bienaymé-Tchebychev estime la variation de X autour de sa moyenne.

Elle découle de l'inégalité de Markov appliquée à $Y = |X - E(X)|^2$.

L'inégalité de Bienaymé-Tchebychev sert en particulier à démontrer la loi faible des grands nombres.

2. Loi faible des grands nombres

Théorème (loi faible - v.a. L^2 non corrélées).
[Foata-Franchi-Fuchs p226]
Soit $(X_n)_{n\geq 1}$ une suite de v.a. réelles dans L^2, centrées, non corrélées deux à deux (c'est-à-dire $\mathrm{Cov}(X_i, X_j) = 0$ si $i \neq j$).

S'il existe C tel que $\forall n \geq 1,\ \mathrm{Var}(X_n) \leq C$, alors

$$\frac{S_n}{n} \xrightarrow{P} 0.$$

Preuve. Les (X_i) sont non corrélées donc

$$\mathrm{Var}(S_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) \leq nC$$

Les X_i sont centrées donc S_n est centrée aussi.

Inégalité de Bienaymé-Tchebychev appliquée à $\frac{S_n}{n}$:

$$P\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{\mathrm{Var}(S_n)}{\varepsilon^2 n^2} \leq \frac{C}{\varepsilon^2 n}$$

Conclusion :

$$\forall \varepsilon > 0,\ \lim_{n\to+\infty} P\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) = 0,\ \text{càd} \quad \frac{S_n}{n} \xrightarrow{P} 0$$

Remarques.
– La preuve montre également que $P\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right)$ converge en $O(1/n)$ pour ε fixé.
(d'autres résultats donnent une meilleure vitesse de convergence : théorème central limite, grandes déviations)

– On a trivialement la convergence L^2 :

$$E\left(\left|\frac{S_n}{n}\right|^2\right) = \frac{\mathrm{Var}(S_n)}{n^2} \leq \frac{C}{n} \longrightarrow 0.$$

Remarque. *La preuve de la loi faible est trop courte pour faire un développement à elle toute seule.*

La loi faible des grands nombres est utilisée surtout sous les hypothèses suivantes :

Théorème (loi faible L^2 iid).
Soit $(X_n)_{n\geq 1}$ une suite de v.a. réelles L^2, indépendantes et de même loi. Alors

$$\frac{S_n}{n} \xrightarrow{P} E(X_1).$$

C'est une conséquence du théorème précédent pour la suite de v.a. centrées $(X_n - E(X_1))_{n\geq 1}$ car la variance des X_n est constante.

Remarque.
Les lois qu'on a vues sont toutes L^p pour tout $p \geq 1$.
Contre-exemple :
la loi de Cauchy, de densité $\frac{1}{\pi}\frac{1}{1+x^2}$, n'est pas L^1.

3. Loi forte des grands nombres

Théorème (loi forte iid L^2). [Jacod-Protter p179]
Soit $(X_n)_{n\geq 1}$ une suite de v.a. réelles L^2, de même loi, indépendantes. Alors $\dfrac{S_n}{n} \xrightarrow[L^2]{p.s.} E(X_1)$.

On a déjà vu qu'on a convergence dans L^2.

Lemme. $\displaystyle\sum_{i=1}^{+\infty} E(X_i^2) < +\infty \Longrightarrow \lim_{i\to+\infty} X_i = 0 \ p.s.$

Remarque : comme $E_i^2 \geq 0$, $\displaystyle\sum_{i=1}^{+\infty} E(X_i^2)$ est toujours définie, en prenant éventuellement $+\infty$ comme valeur.

Preuve du lemme. *Soit*

$$Y = \sum_{i=1}^{+\infty} E(X_i^2) \quad \text{et} \quad A = \{\omega \in \Omega \mid Y(\omega) = +\infty\}.$$

On a $E(Y) \geq E(Y\mathbb{1}_A)$ donc, si $P(A) > 0$, alors $E(Y) = +\infty$. C'est exclu, donc $P(A) = 0$ et

$$\forall \omega \in \Omega \setminus A,\ Y(\omega) = \sum_{i=1}^{+\infty} X_i^2(\omega) < +\infty.$$

Les termes d'une série convergente tendent vers 0, donc $\forall \omega \in \Omega \setminus A$, $\lim_{i\to+\infty} X_i^2(\omega) = 0$, et donc $\lim_{i\to+\infty} X_i(\omega) = 0$.
Autrement dit, $\lim_{i\to+\infty} X_i = 0$ p.s.□

Preuve du théorème.
On pose $\sigma^2 = \mathrm{Var}(X_1)$ et $Z_n = \frac{S_n}{n} - E(X_1)$.
On a $E(Z_n) = 0$ et $E(Z_n^2) = \mathrm{Var}(Z_n) = \frac{\sigma^2}{n}$, donc

$$E\left(\sum_{n=1}^{+\infty} Z_{n^2}^2\right) = \sum_{n=1}^{+\infty} \frac{\sigma^2}{n^2} < +\infty.$$

En appliquant le lemme, $\lim_{n\to+\infty} Z_{n^2} = 0$ p.s.

C'est la convergence p.s. d'une sous-suite, reste à montrer la convergence de la suite entière.

Soit $n \in \mathbb{N}$. Il existe un unique entier $p = p(n)$ tel que $p^2 \leq n < (p+1)^2$. On va comparer Z_n avec Z_{p^2}.

$$Z_n - \frac{p^2}{n} Z_{p^2} = \frac{1}{n} \sum_{i=p^2+1}^{n} X_i - E(X_1)$$

donc

$$E\left(\left(Z_n - \frac{p^2}{n} Z_{p^2}\right)^2\right) = \frac{n-p^2}{n^2}\sigma^2.$$

De plus, $n - p^2 \leq (p+1)^2 - p^2 = 2p + 1$ et $p \leq \sqrt{n}$, donc $n - p^2 \leq 2\sqrt{n} + 1 \leq 3\sqrt{n}$. D'où

$$\sum_{n=1}^{+\infty} E\left(\left(Z_n - \frac{p(n)^2}{n} Z_{p(n)^2}\right)^2\right) \leq \sum_{n=1}^{+\infty} \frac{3}{n^{3/2}}\sigma^2 < +\infty.$$

En appliquant le lemme, $\lim_{n\to+\infty} Z_n - \dfrac{p(n)^2}{n} Z_{p(n)^2} = 0$ p.s.

Or $\lim_{n\to+\infty} Z_{p(n)^2} = 0$ p.s. par ce qui précède, et $\lim_{n\to+\infty} \dfrac{p(n)^2}{n} = 1$.

Conclusion : $\lim_{n\to+\infty} Z_n = 0$. p.s.

Théorème (loi forte L^2).
Soit $(X_n)_{n\geq 1}$ une suite de v.a. réelles L^2, centrées, indépendantes. Supposons que $\sum_{n\geq 1} \frac{\text{Var}(X_n)}{n^2} < +\infty$.
Alors $\frac{S_n}{n} \xrightarrow[L^2]{p.s.} 0$.

Énoncé dans [Foata-Franchi-Fuchs p229], sans preuve. La preuve de [Jacod-Protter] se généralise en fait à ce cas.

Théorème (loi forte L^1). [Barbe-Ledoux p140]
Soit $(X_n)_{n\geq 1}$ une suite de v.a. réelles L^1, de même loi, indépendantes. Alors $\frac{S_n}{n} \xrightarrow[L^1]{p.s.} E(X_1)$.

Ce qui est intéressant dans le 1er théorème, c'est qu'on ne suppose pas que les X_n ont la même loi.

Dans le 2ème théorème, les X_n sont seulement L^1, hypothèse plus faible que L^2 (mais pas si utile en pratique, la plupart les v.a. utilisées étant L^p pour tout p).

La preuve de la loi forte L^1 utilise la loi forte L^2 puis la densité des fonctions L^2 dans L^1 ; elle est trop longue pour un développement.

Remarque.
La loi forte pour des v.a. bornées (ou L^4) est la plus facile à montrer, cf [Barbe-Ledoux, début preuve théorème 5.2 p140].

Théorème.
Soit $(X_n)_{n\geq 1}$ une suite de v.a. réelles L^4, centrées, indépendantes. On suppose qu'il existe C tel que $\forall n \geq 1,\ E((X_n)^4) \leq C$. Alors $\frac{S_n}{n} \xrightarrow{p.s.} 0$.

Preuve.
$(S_n)^4 = \sum_{1\leq i,j,k,l\leq n} X_i X_j X_k X_l$. On regroupe les termes selon que les indices sont distincts ou non :

$$(S_n)^4 = \sum_i (X_i)^k + 4\sum_{i\neq j}(X_i)^3 X_j + 6\sum_{i\neq j}(X_i)^2(X_j)^2$$
$$+ 12 \sum_{\substack{i,j,k \\ \text{distincts}}} (X_i)^2 X_j X_k + 24 \sum_{\substack{i,j,k,l \\ \text{distincts}}} X_i X_j X_k X_l.$$

($6 = C_4^2$: nombre de façons de choisir la position des i).

*Indépendance : $E(X_i^3 X_j) = E(X_i^3)E(X_j) = 0$ car $E(X_j) = 0$. De même $E(X_i^2 X_j X_k) = 0$ et $E(X_i X_j X_k X_l) = 0$.
Donc $E((S_n)^4) = \sum_i E((X_i)^4) + \sum_{i\neq j} E((X_i)^2)E((X_j)^2)$.*

*Or $E((X_i)^4) \leq C$, et $E((X_i)^2) \leq E((X_i)^4)^{1/2} \times 1 \leq C^{1/2}$ par Cauchy-Schwarz. Donc $E((S_n)^4) \leq nC + 6\frac{n(n-1)}{2}C \leq 4Cn^2$.
On en déduit que $E\left(\left(\frac{S_n}{n}\right)^4\right) \leq \frac{4C}{n^2} \to 0$, donc $\frac{S_n}{n} \xrightarrow{L^4} 0$.*

Soit $Y = \sum_{n\geq 1} \left(\frac{S_n}{n}\right)^4$.

$$E(Y) = \sum_{n\geq 1} \frac{E((S_n)^4)}{n^4} \leq 4C \sum_{n\geq 1} \frac{1}{n^2} < +\infty.$$

*Donc $P(Y = +\infty) = 0$ (car $E(Y) \geq E(Y \mathbb{1}_{Y=+\infty}) = +\infty \times P(Y = +\infty)$). Soit $N = \{Y = +\infty\}$ (ensemble négligeable).
Si $\omega \notin N$ alors $Y(\omega) < +\infty$, or $Y(\omega)$ est une série de termes positifs, donc $\left(\frac{S_n(\omega)}{n}\right)^4 \to 0$, autrement dit $\frac{S_n(\omega)}{n} \to 0$.
Conclusion : $\frac{S_n}{n} \xrightarrow{p.s.} 0$.*

4. Application de la LGN en statistiques : moyenne et variance empiriques

Soit $X_1, \ldots, X_n$ un échantillon de X, c'est-à-dire des v.a. indépendantes de même loi que X.

Estimateur de $m = E(X)$: $M_n = \frac{1}{n}\sum_{k=1}^n X_k$.

Par la loi forte des grands nombres, M_n converge p.s. vers m quand $n \to +\infty$. On dit que M_n est un estimateur consistant de m.

De plus, $E(M_n) = \frac{1}{n}\sum_{k=1}^n E(X_k) = m$.
On dit que M_n est un estimateur sans biais.

Estimation de $\sigma^2 = \text{Var}(X)$: $V_n = \frac{1}{n}\sum_{k=1}^n (X_k - M_n)^2$.

On peut montrer que V_n converge p.s. vers σ^2 et que $E(V_n) = \frac{n-1}{n}\cdot\sigma^2$.
On utilise plutôt $V_n' = \frac{1}{n-1}\sum_{k=1}^n (X_k - M_n)^2$ qui vérifie $E(V_n') = \sigma^2$ (et V_n' converge p.s. vers σ^2).

On pose $X'_k = X_k - m$ *et* $M'_n = M_n - m = \frac{1}{n}\sum_{k=}^{n} X'_k$.

$$\begin{aligned}(X_k - M_n)^2 &= (X'_k - M'_n)^2 = \left(\left(1-\frac{1}{n}\right)X'_k - \frac{1}{n}\sum_{i\neq k} X'_k\right)^2 \\ &= \left(\frac{n-1}{n}\right)^2 X_k'^2 + \frac{1}{n^2}\sum_{i\neq k} X_i'^2 + \textit{ double-produits}\end{aligned}$$

Par indépendance, $E(X'_iX'_k) = E(X'_i)E(X'_k) = 0$ *si* $i \neq k$. *Donc*

$$\begin{aligned}E((X_k - M_n)^2) &= \left(\frac{n-1}{n}\right)^2 E(X_k'^2) + \frac{1}{n^2}\sum_{i\neq k} E(X_i'^2) \\ &= \left(\frac{n-1}{n}\right)^2 \cdot \sigma^2 + \frac{n-1}{n^2}\cdot\sigma^2 \\ &= \frac{n-1}{n}\cdot\sigma^2\end{aligned}$$

On a donc $E\left(\sum_{k=1}^{n}(X_k - M'_n)^2\right) = (n-1)\cdot\sigma^2$.
Ceci explique qu'on préfère diviser par $n-1$ *plutôt que par* n.

Le même calcul, mené jusqu'au bout sans espérance, donne

$$V'_n = \left(1-\frac{2}{n}\left(3-\frac{1}{n-1}\right)\right)\frac{1}{n}\sum_{k=1}^{n} X_k'^2 - \left(6-\frac{2}{n-1}\right)\left(\frac{1}{n}\sum_{k=1}^{n} X'_k\right)^2.$$

La loi des grands nombres donne $\frac{1}{n}\sum_{k=1}^{n} X_k'^2 \to \sigma^2$, *et* $\frac{1}{n}\sum_{k=1}^{n} X'_k \to 0$, *d'où* $V'_n \to \sigma^2$ *p.s.*

5. Grandes déviations

Soit $S_n = X_1 + \cdots + X_n$, avec X_n des v.a. indépendantes de même loi, L^2. La loi faible des grands nombres énonce que $\frac{S_n}{n}$ converge en probabilité vers $E(X_1)$:

$$\forall\varepsilon > 0,\ \lim_{n\to+\infty} P\left(\left|\frac{S_n}{n} - E(X_1)\right| \geq \varepsilon\right) = 0.$$

Les "grandes déviations" consistent à étudier la vitesse de convergence de la quantité ci-dessus.

La preuve de la loi faible des grands nombres fournit une majoration par $\frac{\mathrm{Var}(X_1)}{\varepsilon^2 n}$ (convergence en $O(\frac{1}{n})$).

Sous certaines hypothèses, on peut montrer que la vitesse de convergence est exponentielle. La preuve utilise la transformée de Laplace.

Remarque : la transformée de Laplace vaut
– en proba : $E(e^{sX})$ ($= \int e^{sx}f(x)dx$ si la loi de X a pour densité f).
– en analyse : $\int e^{-sx}f(x)dx$.
Attention à la différence de convention.

Théorème de grandes déviations quand les v.a. X_i suivent une loi de Bernoulli (autrement dit S_n suit une binomiale) :

Théorème. [Lesigne p16]
Soit S_n une v.a. de loi binomiale $B(n,p)$. Pour tout $\varepsilon > 0$. il existe $h^+(\varepsilon) > 0$, $h^-(\varepsilon) > 0$ tels que

$$P\left(\frac{S_n}{n} \geq p + \varepsilon\right) \leq e^{-nh^+(\varepsilon)},$$

$$P\left(\frac{S_n}{n} \leq p - \varepsilon\right) \leq e^{-nh^-(\varepsilon)}.$$

On a également $P\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \leq 2e^{-nh(\varepsilon)}$,
avec $h(\varepsilon) = \min(h^+(\varepsilon), h^-(\varepsilon))$.

Plus précisément, si $\varepsilon < 1 - p$,

$$h^+(\varepsilon) = (p+\varepsilon)\ln\frac{p+\varepsilon}{p} + (1-p-\varepsilon)\ln\frac{1-p-\varepsilon}{1-p}.$$

De façon symétrique, on a, pour $\varepsilon < p$:

$$h^-(\varepsilon) = (p-\varepsilon)\ln\frac{p-\varepsilon}{p} + (1-p+\varepsilon)\ln\frac{1-p+\varepsilon}{1-p}.$$

On écrit $S_n = X_1 + \cdots + X_n$ *avec* X_i *des v.a. indépendantes de loi de Bernoulli* $b(p)$. *Soit* $a > 0$.

$$\begin{aligned}P\left(\frac{S_n}{n} \geq t\right) &= P(e^{aS_n} \geq e^{ant}) \\ &\leq e^{-ant}E(e^{aS_n}) \quad \textit{(inégalité de Markov)} \\ &= e^{-ant}\left(E(e^{aX_1})\right)^n \quad \textit{(indépendance)} \\ &= \exp\left(-n(at - \ln E(e^{aX_1}))\right)\end{aligned}$$

Donc, si on pose $a(t) = \sup_{a>0}(at - \ln E(e^{aX_1}))$, *on a*

$$P\left(\frac{S_n}{n} \geq t\right) \leq e^{-na(t)}.$$

Les t *intéressants sont les* $t \in]p,1[$ *car*
- $P\left(\frac{S_n}{n} \geq t\right) = 0$ *si* $t > 1$,
- $P\left(\frac{S_n}{n} \geq 1\right) = P\left(\frac{S_n}{n} = 1\right) = p^n$,
- $P\left(\frac{S_n}{n} \geq t\right)$ *ne tend pas vers* 0 *si* $t \leq p$.

$E(e^{aX_1}) = (1-p) + pe^a$.
$a(t) = \sup_{x>0} f(x)$ *si on pose* $f(x) = xt - \ln(1 - p + pe^x)$.

Etude de f *quand* $x \in [0, +\infty[$ *(avec* $p < t < 1$*) :*
- $f'(x) = t - \frac{pe^x}{1-p+pe^x}$.
- $f'(x) = 0 \Leftrightarrow t(1-p+pe^x) = pe^x \Leftrightarrow e^x = \frac{t(1-p)}{p(1-t)}$

$\Leftrightarrow x = x_0 = \ln\frac{t(1-p)}{p(1-t)}$ *(*$x_0 > 0$ *car* $\frac{t}{p} > 1$ *et* $\frac{(1-p)}{(1-t)} > 1$*).*
- $f'(0) = t - p > 0$, *et* $f'(x) \to t - 1 < 0$ *quand* $x \to +\infty$ *donc* $f'(x) > 0$ *ssi* $0 \leq x < x_0$.

Conclusion : f *a un unique maximum en* x_0 *et*
$a(t) = f(x_0) = t\ln\frac{t}{p} + (1-t)\ln\frac{1-t}{1-p}$.
$P\left(\frac{S_n}{n} \geq t\right) \leq e^{-nh^+(\varepsilon)}$ *avec* $h^+(\varepsilon) = a(p+\varepsilon)$.

Il existe des résultats de grandes déviations pour des v.a. plus générales.

Soit X une v.a. réelle telle que $E(e^{sX}) < +\infty$ pour tout s dans un voisinage U de 0. Si $s > 0$, on a :

$$\begin{array}{rcl} P(X \geq t) & = & P(e^{sX} \geq e^{st}) \\ & \leq & e^{-st}E(e^{sX}) \quad \text{(inégalité de Markov)} \end{array}$$

Donc : $P(X \geq t) \leq \inf_{s>0} e^{-st}E(e^{sX})$.

Si on écrit $h_X(t) = \sup_{s>0}(st - \ln E(e^{sX}))$, cela donne :

Inégalité de Bernstein, Cramér ou Chernoff.

$$\forall t \in \mathbb{R},\ P(X \geq t) \leq e^{-h_X(t)}.$$

Proposition. *$h_X(t) > 0$ si $t > E(X)$. (admis)*

Considérons $S_n \geq nt$ au lieu de $X \geq t$.
$E(e^{sS_n}) = (E(e^{sX_1}))^n$ par indépendance,
$h_{S_n}(nt) = n \sup_{s>0}(st - \ln E(e^{sX_1})) = nh_{X_1}(t)$. D'où :

Théorème.
Si $t > E(X_1)$, $P\left(\frac{S_n}{n} \geq t\right) \leq e^{-nh_{X_1}(t)}$ avec $h_{X_1}(t) > 0$.

[Dacunha-Castelle p78+p102], [Barbe-Ledoux p62]

Dans certains cas, on peut avoir un équivalent au lieu d'une majoration.

Théorème de Chernoff (grandes déviations).
[Lacroix-Mazliak, "Probabilités", p109]

Soit $(X_n)_{n\geq 1}$ des v.a. réelles indépendantes de même loi, prenant un nombre fini de valeurs, et $S_n = X_1 + \cdots + X_n$. Soit $h(t) = \sup_{s>0}(st - \ln E^{sX_1})$.

Si $t > E(X_1)$ est tel que $P(X_1 > t) > 0$, alors $h(t) > 0$ et

$$\frac{1}{n}\ln\left(P\left(\frac{S_n}{n} \geq t\right)\right) = -h(t) + O\left(\frac{1}{\sqrt{n}}\right).$$

6. Théorème de Glivenko-Cantelli

[Dacunha-Castelle], [Billingsley], [Ouvrard tome 2]

Soit $(X_n)_{n\geq 1}$ des v.a. réelles indépendantes de même loi. La fonction de répartition empirique de $X_1, \ldots, X_n$ est :

$$F_n(x,\omega) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{]-\infty,x]}(X_k(\omega)).$$

Soit F la fonction de répartition commune des X_n.

On a le résultat suivant :
Théorème de Glivenko-Cantelli.

$$\sup_{x\in\mathbb{R}} |F_n(x,\omega) - F(x)| \xrightarrow{p.s.} 0$$

Autrement dit, F_n converge p.s. vers F, uniformément en x.

Ce théorème dit qu'on peut approximer la loi de X (via sa fonction de répartition) en faisant des expériences indépendantes de loi P_X.

Preuve. *On fixe $x \in \mathbb{R}$. On pose $Y_n(\omega) = \mathbb{1}_{]-\infty,x]}(X_k(\omega))$. Ainsi $F_n(x,\omega) = \frac{1}{n}\sum_{k=1}^{n} Y_k(\omega)$.*

Les $(Y_n)_{n\geq 1}$sont des v.a. indépendantes de même loi, bornées, et $E(Y_n) = P(X_n \leq x) = F(x)$. Par la loi forte des grands nombres, $F_n(x,\cdot) \xrightarrow{p.s.} F(x)$ pour x fixé. C'est-à-dire :

$$\forall x, \exists A_x \subset \Omega, P(A_x) = 0, \forall \omega \notin A_x, \lim_{n\to+\infty} F_n(x,\omega) = F(x). \quad (1)$$

De même avec $F_n(x^-,\omega) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{]-\infty,x[}(X_k(\omega))$:

$$\forall x, \exists B_x, P(B_x) = 0, \forall \omega \notin B_x, \lim_{n\to+\infty} F_n(x^-,\omega) = F(x^-). \quad (2)$$

Soit $N \geq 1$ et $1 \leq j \leq N-1$. Soit $x_{jN} = \inf\{x \in \mathbb{R} \mid \frac{j}{N} \leq F(x)\}$.

F est continue à droite en tout point donc $\frac{j}{N} \leq F(x_{kN})$.
De plus, $\frac{j}{N} > F(x)$ si $x < x_{jN}$, donc $\frac{j}{N} \geq \lim_{x\to x_{jN}^-} F(x) = F(x_{jN}^-)$.
Ce qui donne :

$$F(x_{jN}^-) \leq \frac{j}{N} \leq F(x_{jN}).$$

Par convention, on pose
$x_{0N} = -\infty$ et $F(x_{0N}) = 0$ (limite de F en $-\infty$),
$x_{NN} = +\infty$ et $F(x_{NN}^-) = 1$ (limite de F en $+\infty$).

$$\begin{array}{rcl} & \frac{0}{N} & \leq F(x_{0N}) \\ F(x_{1N}^-) \leq & \frac{1}{N} & \leq F(x_{1N}) \\ F(x_{2N}^-) \leq & \frac{2}{N} & \leq F(x_{2N}) \\ & \vdots & \\ F(x_{N-1,N}^-) \leq & \frac{N-1}{N} & \leq F(x_{N-1,N}) \\ F(x_{NN}^-) \leq & \frac{N}{N} & \end{array}$$

En soustrayant les lignes 2 à 2 :

$$\forall 1 \leq j \leq N,\ F(x_{jN}^-) - F(x_{j-1,N}) \leq \frac{1}{N}. \qquad (3)$$

Soit $x \in \mathbb{R}$. Il existe $0 \leq j \leq N-1$ tel que $x_{jN} \leq x < x_{j+1,N}$. Par croissance des fonctions de répartition :

$$F_n(x_{jN},\omega) - F(x_{j+1,N}^-) \leq F_n(x,\omega) - F(x) \leq F_n(x_{j+1,N}^-,\omega) - F(x_{jN}).$$

Par (3) : $F_n(x_{jN},\omega) - F(x_{jN}) - \frac{1}{N} \leq F_n(x,\omega) - F(x) \leq F_n(x_{j+1,N}^-,\omega) - F(x_{j+1,N}^-) + \frac{1}{N}$.

On pose

$$D_n^N(\omega) = \max_{1 \leq j \leq N-1} \{|F_n(x_{jN},\omega) - F(x_{jN})|, |F_n(x_{j+1,N}^-,\omega) - F(x_{j+1,N}^-)|\}.$$

Ce qui précède montre que

$$\forall N \geq 1,\ \forall \omega \in \Omega,\ \sup_{x \in \mathbb{R}} |F_n(x,\omega) - F(x)| \leq D_n^N(\omega) + \frac{1}{N}.$$

On pose

$$C = \bigcup_{N \geq 1} \bigcup_{1 \leq i \leq N-1} (A_{x_{jN}} \cup B_{x_{jN}}).$$

Alors $P(C) = 0$ (union dénombrable de négligeables) et si $\omega \notin C$ alors par (1)+(2), $\lim_{n \to +\infty} D_n^N(\omega) = 0$.

Conclusion :

$$\forall \omega \notin C, \limsup_{n \to +\infty} \sup_{x \in \mathbb{R}} |F_n(x,\omega) - F(x)| \leq \frac{1}{N}$$

Comme ceci est vrai pour tout N,

$$\forall \omega \notin C,\ \lim_{n \to +\infty} \sup_{x \in \mathbb{R}} |F_n(x,\omega) - F(x)| = 0,$$

autrement dit F_n converge p.s. vers F, uniformément en x.

IX. Théorème Central Limite

1. TCL et théorèmes voisins

Théorème central limite (TCL).
Soit $(X_n)_{n \geq 1}$ des v.a. L^2, indépendantes, de même loi, et $S_n = X_1 + \cdots + X_n$. Alors

$$\frac{S_n - nE(X_1)}{\sqrt{n\mathrm{Var}(X_1)}} = \frac{\sqrt{n}}{\sqrt{\mathrm{Var} X_1}} \left(\frac{S_n}{n} - E(X_1) \right)$$

converge en loi vers $\mathcal{N}(0,1)$ quand $n \to +\infty$.

Formulation équivalente :
pour tous a, b (réels ou $\pm\infty$) avec $a \leq b$,

$$\lim_{n \to +\infty} P\left(a \leq \frac{S_n - nE(X_1)}{\sqrt{n\mathrm{Var}(X_1)}} \leq b \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} dt.$$

car la fonction de répartition de $\mathcal{N}(0,1)$ est continue en tout point et $P(a \leq N \leq b) = F_N(b) - F_N(a)$.

Par la loi des grands nombres, $\frac{S_n}{n} \xrightarrow{p.s.} E(X_1)$.
Autrement dit, $\frac{S_n(\omega)}{n} = E(X_1) + o(1)$ p.s.

Le théorème central limite précise le $o(1)$: $\frac{S_n}{n}$ ressemble à

$$E(X_1) + \frac{\sqrt{\mathrm{Var}(X_1)}}{\sqrt{n}} \cdot \mathcal{N}(0,1) + o\left(\frac{1}{\sqrt{n}}\right).$$

attention, c'est un "petit o probabiliste" !
En gros, c'est vrai sur un ensemble de mesure $1-\varepsilon$ (avec la constante dans petit o dépendant de ε).

Preuve du Théorème Central Limite.
Soit $X'_n = \frac{X_n - E(X_n)}{\sqrt{\mathrm{Var}(X_n)}}$. *Alors* $E(X'_n) = 0$, $\mathrm{Var}(X'_n) = 1$ *et les* $(X'_n)_{n\geq 1}$ *sont indépendantes de même loi. Si* $S'_n = X'_1 + \cdots + X'_n$, *il faut montrer que* $\frac{S'_n}{\sqrt{n}}$ *converge en loi vers* $\mathcal{N}(0,1)$.

La fonction caractéristique de $\mathcal{N}(0,1)$ *est* $e^{-t^2/2}$, *donc par le théorème de Lévy, il faut montrer que*

$$\forall t \in \mathbb{R}, \lim_{n\to+\infty} \varphi_{S'_n/\sqrt{n}}(t) = e^{-t^2/2}.$$

On note $\varphi = \varphi_{X'_1}$. *On a* $\varphi_{X'_i} = \varphi$ *pour tout* i.
$\varphi_{S'_n/\sqrt{n}}(t) = E(e^{itS'_n/\sqrt{n}}) = \varphi_{S'_n}(t/\sqrt{n})$,
et $\varphi_{S'_n}(u) = (\varphi(u))^n$ *(indépendance)*,
donc $\varphi_{S'_n/\sqrt{n}}(t) = (\varphi(t/\sqrt{n}))^n$.

Par hypothèse, X'_1 *est* L^2, *donc* φ *est 2 fois dérivable et* $\varphi'(0) = iE(X'_1) = 0$, $\varphi''(0) = -E((X'_1)^2) = -1$.

On fait un développement limité en 0 : $\varphi(u) = 1 - \frac{u^2}{2} + o(u^2)$.
On fixe $t \in \mathbb{R}$ *et on prend* $n \to +\infty$ *:*

$$\begin{aligned}
\ln(\varphi_{S'_n/\sqrt{n}}(t)) &= n \ln(\varphi(t/\sqrt{n})) \\
&= n \ln(1 - \frac{t^2}{2n} + \underbrace{o(t^2/n)}_{o(1/n)}) \\
&= n(-\frac{t^2}{2n} + o(1/n)) = -\frac{t^2}{2} + o(1)
\end{aligned}$$

Donc $\lim_{n\to+\infty} \ln(\varphi_{S'_n/\sqrt{n}}(t)) = -\frac{t^2}{2}$, *autrement dit :*
$\lim_{n\to+\infty} \varphi_{S'_n/\sqrt{n}}(t) = e^{-t^2/2}$ *pour tout* $t \in \mathbb{R}$.

Attention, *il s'agit de logarithme complexe puisque* φ *est à valeurs complexes.*

Quand n *est suffisamment grand,* $u = t/\sqrt{n}$ *est proche de 0 donc* $\varphi(u)$ *est proche de 1. Le logarithme complexe est défini sans problème dans un voisinage de 1. Une justification possible est donnée par le développement de la série entière associée à* $ln(1+z)$, *qui a un rayon de convergence de 1.*

Voici une façon de se débarrasser des log complexes.

$$\varphi_{S'_n/\sqrt{n}}(t) = \left(\varphi(t/\sqrt{n})\right)^n = \left(1 - \frac{t^2}{2n} + \frac{1}{n}\varepsilon(1/n)\right)^n$$

avec $\varepsilon(z)$ *complexe, tendant vers 0 quand* $z \to 0$.

L'idée est que le terme dominant est $\left(1 - \frac{t^2}{2n}\right)^n$, *qui est réel.*

Par la formule du binôme,

$$\varphi_{S'_n/\sqrt{n}}(t) = \left(1 - \frac{t^2}{2n}\right)^n + \sum_{k=1}^{n} \binom{n}{k} \left(1 - \frac{t^2}{2n}\right)^{n-k} \left(\frac{\varepsilon(1/n)}{n}\right)^k$$

$$\begin{aligned}
\left|\varphi_{S'_n/\sqrt{n}}(t) - \left(1 - \frac{t^2}{2n}\right)^n\right| &\leq \sum_{k=1}^{n} \binom{n}{k} \left(1 - \frac{t^2}{2n}\right)^{n-k} \left(\frac{|\varepsilon(1/n)|}{n}\right)^k \\
&\leq \left(1 - \frac{t^2}{2n} + \frac{|\varepsilon(1/n)|}{n}\right)^n - \left(1 - \frac{t^2}{2n}\right)^n
\end{aligned}$$

La quantité $\left(1 - \frac{t^2}{2n} + \frac{|\varepsilon(1/n)|}{n}\right)^n$ *ne fait intervenir que des réels, on peut donc le traiter de façon classique en prenant le logarithme (réel), et montrer que sa limite est* $e^{-t^2/2}$.

Voir [Zuily-Queffelec p540] pour une preuve du TCL du même genre (sans log complexe).

Le théorème de Moivre (historiquement plus ancien) est le TCL pour des lois de Bernoulli.

Théorème de Moivre.
Si S_n suit une loi $\mathcal{B}(n,p)$ alors

$$\lim_{n\to+\infty} P\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} dt.$$

Remarque. *Il est d'usage d'approximer* $\frac{S_n - np}{\sqrt{np(1-p)}}$ *par* $\mathcal{N}(0,1)$ *dès que* $np(1-p) > 18$. *La convergence est plus rapide si* p *est proche de* $1/2$.

Le théorème de Moivre peut se démontrer en approximant les coefficients binomiaux à l'aide de la formule de Stirling, le calcul est lourd mais élémentaire.

Théorème.
Pour tout $\lambda \in \mathbb{R}_+^*$, soit X_λ une v.a. de loi de Poisson $\mathcal{P}(\lambda)$. Alors

$$\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \xrightarrow{loi} \mathcal{N}(0,1) \text{ quand } \lambda \to +\infty.$$

Preuve. *La preuve est similaire à celle du théorème central limite, en utilisant le fait que* $\varphi_{X_\lambda}(t) = \exp(\lambda(e^{it} - 1))$.

Soit $Y_\lambda = \frac{X_\lambda - \lambda}{\sqrt{\lambda}}$.

$$\begin{aligned}
\varphi_{Y_\lambda}(t) &= E\left(e^{i\frac{t}{\sqrt{\lambda}}X_\lambda}\right) e^{-\frac{it\lambda}{\sqrt{\lambda}}} \\
&= \varphi_{X_\lambda}\left(\frac{t}{\sqrt{\lambda}}\right) e^{-\frac{it\lambda}{\sqrt{\lambda}}} \\
&= \exp(\lambda(e^{i\frac{t}{\sqrt{\lambda}}} - 1))e^{-\frac{it\lambda}{\sqrt{\lambda}}} \\
&= \left(\exp\left(e^{i\frac{t}{\sqrt{\lambda}}} - 1 - i\frac{t}{\sqrt{\lambda}}\right)\right)^\lambda
\end{aligned}$$

D'où

$$\begin{aligned}
\log \varphi_{Y_\lambda}(t) &= \lambda\left(e^{i\frac{t}{\sqrt{\lambda}}} - 1 - i\frac{t}{\sqrt{\lambda}}\right) \\
&= \lambda\left(1 + i\frac{t}{\sqrt{\lambda}} - \frac{t^2}{2\lambda} + o(1/\lambda) - 1 - i\frac{t}{\sqrt{\lambda}}\right) \\
&= -\frac{t^2}{2} + o(1)
\end{aligned}$$

Donc $\lim_{\lambda\to+\infty} \varphi_{Y_\lambda}(t) = e^{-t^2/2}$ *pour tout* $t \in \mathbb{R}$.

Conclusion : $\frac{X_\lambda - \lambda}{\sqrt{\lambda}}$ *converge en loi vers* $\mathcal{N}(0,1)$ *quand* $\lambda \to +\infty$

2. Application du TCL aux intervalles de confiance

Soit $X_1, \ldots, X_n$ des v.a. indépendantes de loi de Bernoulli $b(p)$, représentant le résultat d'une expérience.

2 cas de figures :

– on connaît p, on voudrait savoir *a priori* où se situeront les $X_1, \ldots, X_n$.

– on ne connaît pas p, on cherche à le déterminer à l'aide des observations.

1er cas : p connu

Un fournisseur fabrique des pièces, défectueuses avec une proba p. S'il vend un lot contenant trop de pièces défectueuses, le client peut demander à être remboursé.

$M_n = \frac{X_1+\cdots+X_n}{n}$: proportion de pièces défectueuses parmi n.

On fixe $\varepsilon > 0$. On veut trouver un intervalle I_n (intervalle de confiance) tel que $M_n \in I_n$ avec une probabilité d'au moins $1 - \varepsilon$.

Ici, on veut surtout majorer M_n : on cherche un intervalle de forme $]-\infty, m]$.

$E(X_i) = p$ et $\mathrm{Var}(X_i) = p(1-p)$. Le théorème central limite donne :

$$\lim_{n\to+\infty} P\left(a \leq \frac{M_n - p}{\sqrt{\frac{p(1-p)}{n}}} \leq b\right) \longrightarrow \frac{1}{\sqrt{2\pi}}\int_a^b e^{-t^2/2}\,dt.$$

Soit b tel que $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^b e^{-t^2/2}\,dt > 1-\varepsilon$.

Si n est assez grand,

$$P\left(\frac{M_n - p}{\sqrt{\frac{p(1-p)}{n}}} \leq b\right) > 1-\varepsilon.$$

$$\frac{M_n - p}{\sqrt{\frac{p(1-p)}{n}}} \leq b \Longleftrightarrow M_n \leq p + b\sqrt{\frac{p(1-p)}{n}}$$

On pose $I_n = \left]-\infty, p + b\sqrt{\frac{p(1-p)}{n}}\right]$.

Si n est assez grand, $M_n \in I_n$ avec probabilité au moins $1-\varepsilon$.

Application numérique :
$n = 1000$, $p = 0,1$, $\varepsilon = 0,05$.
On trouve $b \simeq 1.6$ et $I_n =]-\infty, 0,115]$.
Autrement dit, avec 95% de chance, un lot de 1000 pièces contient au plus 11,5% de pièces défectueuses.

2ème cas. Exemple : sondage

[Barbe-Ledoux p146]

p : probabilité de voter A dans la population.

On cherche à déterminer p (inconnu). Pour cela, on interroge n personnes au hasard.

Modélisation : $X_1, \ldots, X_n$ des v.a. indépendantes de loi de Bernoulli $b(p)$, où X_i est la réponse de la i-ème personne.

Soit $M_n = \frac{1}{n}(X_1 + \cdots + X_n)$. Par la loi forte des grands nombres, M_n tend p.s. vers p.

On fixe $\varepsilon > 0$. On veut trouver un intervalle I_n (intervalle de confiance) tel que $p \in I_n$ avec une probabilité d'au moins $1 - \varepsilon$.

$E(X_i) = p$ et $\mathrm{Var}(X_i) = p(1-p)$. Le théorème central limite donne :

$$P\left(a \leq \frac{M_n - p}{\sqrt{\frac{p(1-p)}{n}}} \leq b\right) \longrightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\, dt.$$

M_n converge p.s. vers p, donc $M_n(1-M_n)$ converge p.s. vers $p(1-p) > 0$, et

$$P\left(a \leq \frac{M_n - p}{\sqrt{\frac{M_n(1-M_n)}{n}}} \leq b\right) \longrightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\, dt.$$

Soit $\varepsilon > 0$. Soit c tel que $\frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-t^2/2}\, dt > 1 - \varepsilon$.

Si n est assez grand,

$$P\left(-c \leq \frac{M_n - p}{\sqrt{\frac{M_n(1-M_n)}{n}}} \leq c\right) > 1 - \varepsilon.$$

$-c \leq \frac{M_n - p}{\sqrt{\frac{M_n(1-M_n)}{n}}} \leq c \Longleftrightarrow M_n - c\sqrt{\frac{M_n(1-M_n)}{n}} \leq p \leq M_n + c\sqrt{\frac{M_n(1-M_n)}{n}}$

On pose $I_n = \left[M_n - c\sqrt{\frac{M_n(1-M_n)}{n}}, M_n + c\sqrt{\frac{M_n(1-M_n)}{n}}\right]$.
Si n est assez grand, $p \in I_n$ avec probabilité au moins $1 - \varepsilon$.
Remarque : p est fixé, c'est la valeur de M_n qui varie.

Pour ε fixé, la marge d'erreur décroît en $1/\sqrt{n}$.

Application numérique :

On interroge 1000 personnes, le sondage donne 52% en faveur de A. On cherche un intervalle de confiance à 5%.

$n = 1000$, $\varepsilon = 0,05$, $M_n(\omega) = 0,52$
(M_n = v.a., $M_n(\omega)$ = résultat d'une expérience précise)

On trouve $c \simeq 2$ et $c\sqrt{\frac{M_n(\omega)(1-M_n(\omega))}{n}} \simeq 0,03$.

p appartient à $I_n = [0,49; 0,55]$ avec 95% de chance.

La marge d'erreur est de 3% ($p = M_n \pm 0,03$).

Si $n = 10\,000$, la marge d'erreur est de 1%.

Remarque.
L'inégalité de Bienaymé-Tchebychev donne aussi une marge d'erreur, mais plus mauvaise (avec les données ci-dessus, on obtient une marge d'erreur de 7% pour $n = 1000$).

3. Méthode de Monte-Carlo (utilise LGN et éventuellement TCL)

Soit B un pavé de $\mathbb{R}^d$ et $f\colon B \to \mathbb{R}$ une fonction mesurable intégrable. On souhaite approximer $\int_B f(x) d\lambda(x)$, où λ = mesure de Lebesgue sur $\mathbb{R}^d$.

Soit $(X_n)_{n\geq 1}$ des v.a. indépendantes de loi uniforme sur B : $P_{X_n} = \frac{1}{\lambda(B)}\lambda$.

Soit $Y_n = f(X_n)$. Ce sont des v.a. indépendantes de même loi.

$$E(|Y_n|) = \int_B |f(x)| dP_{X_n}(x) = \frac{1}{\lambda(B)} \int_B |f(x)| d\lambda(x)$$

donc Y_n est L^1 et $E(Y_n) = \frac{1}{\lambda(B)} \int_B f(x) d\lambda(x)$.

Loi des grands nombres appliquée à $(Y_n)_{n\geq 1}$:

$$\lim_{n\to+\infty} \frac{Y_1 + \cdots + Y_n}{n} = E(Y_1) \text{ p.s.,}$$

autrement dit :

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{k=1}^n f(Xk) = \frac{1}{\lambda(B)} \int_B f(x) d\lambda(x) \text{ p.s..}$$

Vitesse de convergence sans TCL.

On note $Z_n = \dfrac{Y_1 + \cdots + Y_n}{n} - E(Y_1)$. On veut évaluer $|Z_n|$. *(écart entre l'intégrale et son approximation)*

On suppose que f est L^2 (donc les Y_n aussi), on pose $\sigma^2 = \mathrm{Var}(Y_1)$.

L'inégalité de Tchebitchev donne :

$$\forall \varepsilon > 0,\ P(|Z_n| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

1) On se fixe une marge d'erreur ε sur le résultat. On sait que $|Z_n| \leq \varepsilon$ avec une probabilité tendant vers 1, avec une vitesse en $O\left(\frac{1}{n}\right)$.

2) On se fixe une marge d'erreur δ sur la probabilité. Soit ε_n tel que $\delta = \frac{\sigma^2}{n(\varepsilon_n)^2}$. Ainsi $P\left(|Z_n| \geq \frac{\sigma}{\sqrt{n\delta}}\right) \leq \delta$. On sait que, avec une probabilité $1-\delta$, $|Z_n|$ converge vers 0 avec une vitesse en $O\left(\frac{1}{\sqrt{n}}\right)$.

Application numérique avec $\delta = 5\%$: avec proba 95%, $|Z_n| \leq \frac{1}{\sqrt{\delta}} \frac{\sigma}{\sqrt{n}} \simeq 4,5 \frac{\sigma}{\sqrt{n}}$ pour n assez grand.

Vitesse de convergence avec TCL. *On a aussi une vitesse de convergence en* $O\left(\frac{1}{\sqrt{n}}\right)$*, comme dans 2), mais avec une meilleure constante.*

Par le TCL, $\frac{Z_n\sqrt{n}}{\sigma} \to \mathcal{N}(0,1)$ *en loi, donc*

$$P\left(-c \leq \frac{Z_n\sqrt{n}}{\sigma} \leq c\right) \to \frac{1}{\sqrt{2\pi}}\int_{-c}^{c} e^{-t^2/2}dt.$$

On fixe $\delta > 0$ *(marge d'erreur sur la probabilité). Soit* c *tel que* $\frac{1}{\sqrt{2\pi}}\int_{-c}^{c} e^{-t^2/2}dt > 1-\delta$.

Si n *est assez grand,*

$$P\left(-c \leq \frac{Z_n\sqrt{n}}{\sigma} \leq c\right) > 1-\delta.$$

$-c \leq \frac{Z_n\sqrt{n}}{\sigma} \leq c \Longleftrightarrow |Z_n| \leq \frac{c\sigma}{\sqrt{n}}$

Donc, avec une probabilité $1-\delta$, $|Z_n| \leq \frac{c\sigma}{\sqrt{n}}$ *(pour* n *assez grand).*

On a de nouveau une convergence en $O\left(\frac{1}{\sqrt{n}}\right)$*, mais avec une meilleure constante :*

Application numérique avec $\delta = 5\%$ *: on trouve* $c \simeq 2$*, donc, avec proba 95%,* $|Z_n| \leq c\frac{\sigma}{\sqrt{n}} \simeq 2\frac{\sigma}{\sqrt{n}}$ *pour* n *assez grand.*

X. Modèle de v.a. indépendantes

On lance un dé et une pièce.

Modèles types :
dé : $\Omega_1 = \{1,\ldots,6\}$*, proba* P_1 *(uniforme ou non).*
pièce : $\Omega_2 = \{0,1\}$*, proba* P_2.

On veut définir les 2 expériences sur un même ensemble Ω.
On prend $\Omega = \Omega_1 \times \Omega_2$.

résultat du dé : $X_1(\omega_1,\omega_2) = \omega_1$
résultat de la pièce : $X_2(\omega_1,\omega_2) = \omega_2$

($X_1 : \Omega \to \Omega_1$ *et* $X_2 : \Omega \to \Omega_2$ *sont les projections sur les 1ère et 2ème coordonnées)*

Quelle probabilité P *mettre sur* Ω *?*

Soit $A_1 \subset \Omega_1, A_2 \subset \Omega_2$*. On veut que*
$P(X_1 \in A_1) = P_1(A_1)$ *et* $P(X_2 \in A_2) = P_2(A_2)$.

X_1 *et* X_2 *sont indépendantes, donc il faut que*
$\underbrace{P(X_1 \in A_1, X_2 \in A_2)}_{P(A_1\times A_2)} = \underbrace{P(X_1 \in A_1)}_{P_1(A_1)}\underbrace{P(X_2 \in A_2)}_{P_2(A_2)}$.

Donc la bonne probabilité sur Ω *est* $P = P_1 \otimes P_2$.

On lance n *fois une pièce ayant la probabilité* p *de tomber sur face=1. Soit* X_k *le résultat du* k*-ième lancer.* $X_1,\ldots,X_n$ *doivent être des v.a. indépendantes, chacune prenant ses valeurs dans* $\{0,1\}$ *en suivant une loi de Bernoulli* $b(p)$.

Les X_k *sont des fonctions, elles doivent être définies sur le même ensemble sinon on ne peut pas parler d'indépendance.*

Quel modèle peut-on prendre ?

De nouveau, il suffit de prendre le produit :

$$\Omega_n = \{0,1\} \times \{0,1\} \times \cdots \times \{0,1\} = \{0,1\}^n.$$

Si μ *est la probabilité de loi* $b(p)$ *sur* $\{0,1\}$ *:*

$$P_n = \mu \otimes \mu \otimes \cdots \otimes \mu = \mu^{\otimes n}.$$

Soit $X_k : \Omega \to \{0,1\}$ *la projection sur la* k*-ième coordonnée.*

Alors $X_1,\ldots,X_n$ *sont des v.a. indépendantes de loi* $b(p)$.

Alice lance une pièce jusqu'à ce qu'elle tombe sur face=1. Combien de lancers va-t-elle faire en moyenne ?
(voir l'exercice 11)

Quel modèle peut-on prendre ?

Le nombre de lancers n'est pas borné. On va modéliser un nombre infini de lancers, X_k *étant le résultat du* k*-ième lancer pour tout* $k \geq 0$.

Remarque. *Alice ne tire pas une infinité de fois.*
Si $X_0 = X_1 = 0$ *et* $X_2 = 1$*, alors Alice s'arrête et on n'utilise pas les* X_k *pour* $k \geq 3$.

Soit $\Omega = \{0,1\}^{\mathbb{N}}$ *et* $X_k : \Omega \to \{0,1\}$ *la projection sur la* k*-ième coordonnée :* $X_k((\omega_n)_{n\in\mathbb{N}}) = \omega_k$.

On veut mettre une probabilité P *sur* Ω *de sorte que les* $(X_k)_{k\geq 0}$ *soient indépendants et de loi* $b(p)$.
On a envie de définir P *comme étant un produit infini de mesures* μ *de loi* $b(p)$.

Par le théorème de Kolmogorov, on peut définir proprement un tel produit infini de mesures. On note parfois $P = \mu^{\otimes\mathbb{N}}$.

Théorème (conséquence du théorème de Kolmogorov).
Soit $(\Omega_k, \mathcal{F}_k, P_k)$ *des espaces probabilisés,*

$$\Omega = \prod_{n=0}^{+\infty} \Omega_n = \Omega_0 \times \Omega_1 \times \cdots$$

et $X_k : \Omega \to \Omega_k$ *la projection sur la* k*-ième coordonnée.*

On peut définir une tribu $\mathcal{F}$ *sur* Ω *(produit infini des tribus* $\mathcal{F}_k$*) et une probabilité* P *sur* Ω *(produit infini des probabilités* P_k*) tels que* X_k *est de loi* P_k *et la famille de v.a.* $(X_k)_{k\geq}0$ *est indépendante.*

De façon plus précise :
$\mathcal{F}$ *est la tribu de* Ω *engendrée par les ensembles ("cylindres")*

$$A_0 \times A_1 \times \cdots \times A_k \times \prod_{n=k+1}^{+\infty} \Omega_n,\ A_i \in \mathcal{F}_i \text{ pour } 0 \leq i \leq k.$$

On définit $P(A_0 \times \cdots \times A_k \times \prod_{n=k+1}^{+\infty} \Omega_n) = P_0(A_0)\cdots P_k(A_k)$.
Alors P *s'étend en une mesure de probabilité sur* Ω *(théorème de Kolmogorov)*

$(\Omega, \mathcal{F}, P)$ permet également de modéliser un nombre fini de lancers, il suffit de ne considérer que $X_1, \ldots, X_n$ (qui sont bien des v.a. indépendantes de loi $b(p)$). Cela peut être pratique quand on veut parler de convergence quand $n \to +\infty$.

Exemple. *On lance n fois une pièce et on regarde le nombre moyen de face obtenu.*

1ère modélisation.
Pour tout n, on définit les v.a. $X_1^n, \ldots, X_n^n \colon \Omega_n \to \{0,1\}$ (avec $\Omega_n = \{0,1\}^n$, $P_n = \mu^{\otimes n}$ et X_k^n la projection sur la k-ième coordonnée, comme précédemment).

Le nombre moyen de lancers est $M_n = \dfrac{X_1^n + \cdots + X_n^n}{n}$, c'est une v.a. définie sur Ω_n (càd une fonction $\Omega_n \to \mathbb{R}$).

2ème modélisation.
On définit les v.a. $X_k \colon \Omega \to \{0,1\}$ pour tout $k \geq 1$, indépendamment de n.

Le nombre moyen de lancers est $M_n = \dfrac{X_1 + \cdots + X_n}{n}$, c'est une v.a. définie sur Ω.

Expérimentalement, le nombre moyen de face tend vers p quand n tend vers l'infini. On a envie de dire que M_n converge vers p. Il faut faire attention aux objets mathématiques qu'on manipule : M_n est une fonction. Dans le 1er cas, les M_n ne sont pas définies sur le même ensemble, ça pose problème pour parler de convergence de fonctions. Dans le 2ème cas, parler de la convergence de M_n vers la fonction constante égale à p (sur Ω) a un sens.

Beaucoup de livres commencent par définir les probabilités dans les espaces dénombrables. Dans ces espaces, il n'y a pas besoin de théorie de la mesure pour parler de probabilité (inutile de parler de tribu, d'existence de mesure, etc). C'est donc parfait pour s'initier aux probabilités sans avoir besoin d'un bagage mathématique important.

Mais cette approche interdit d'introduire un espace comme $\Omega = \{0,1\}^{\mathbb{N}}$ qui n'est pas dénombrable. Dans l'exemple précédent, il faut donc utiliser la 1ère modélisation, et cela conduit parfois à des "contorsions" pour parler de choses qui paraissent naturelles si on se place dans un espace commun.

Vous avez le bagage mathématique nécessaire pour utiliser la 2ème modélisation. Elle est assez naturelle, assez facile à manipuler et elle facilite beaucoup les choses dans certains cas. Je vous conseille plutôt d'utiliser celle-là.

En tout cas, regardez les livres avec en tête la différence entre ces 2 approches et choisissez en connaissance de cause.

XI. Tests en statistiques
(tests non paramétriques)

1. Tester H_0 contre H_1

Exemple de situation : On veut comparer un nouveau médicament B avec un médicament connu A. On administre A à n malades, B à m malades, et on mesure le temps de guérison.

On modélise le test par des v.a. indépendantes $X_1, \ldots, X_n$ de même loi P_A (malades prenant A) et $Y_1, \ldots, Y_m$ de même loi P_B (malades prenant B).

Définition.
Un **n-échantillon** de loi Q est un n-uplet $(X_1, \ldots, X_n)$ tels que $X_1, \ldots, X_n$ sont des v.a. indépendantes de même loi Q.

Si X est une v.a., un n-échantillon de X est un n-échantillon de loi P_X (répétition d'une même expérience représentée par X).

On veut savoir si B est mieux que A. On fait 2 hypothèses :

- $\boldsymbol{H_0}$: B agit pareil que A, c'est-à-dire $P_B = P_A$.
- $\boldsymbol{H_1}$: B est mieux que A (dans un certain sens, par exemple les fonctions de répartition vérifient $F_B < F_A$).

On dit qu'on teste H_0 contre H_1.

Idée d'un test

Soit $Z = (X_1, \ldots, X_n, Y_1, \ldots, Y_m)$.

Si $P_B = P_A$ (hypothèse H_0), alors la v.a. Z est un $(n+m)$-échantillon de loi P_A. L'idée est d'écarter H_0 si Z est suspect.

On définit une zone de rejet $D \subset R^{n+m}$. On veut ne rejeter Z à tort qu'avec une faible probabilité, c'est-à-dire que si H_0 est vraie alors $P(Z \in D) < \varepsilon$.

Le choix de la zone de rejet comporte une certaine dose d'arbitraire dès que la situation n'est pas extrêmement simple.

Exemple. Avant se présenter à une élection, le candidat A fait faire un sondage : Sur 1000 personnes interrogées, 549 disent vouloir voter A. Que peut-on en déduire sur les chances de A ?

Soit p la probabilité qu'un électeur préfère A.

$H_0 : p = 1/2.$ $\qquad H_1 : p > 1/2.$

A veut se présenter à l'élection uniquement si plus de 50 % des électeurs comptent voter pour lui, c'est-à-dire si H_1 est vraie. Il accepte une marge d'erreur de 5 % (c'est-à-dire 5 % de chance de se présenter même si on n'est pas dans le cas H_1).

Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi $b(p)$, et
$$\overline{X} = \frac{X_1 + \cdots + X_n}{n}$$
la moyenne empirique.

Ici $n = 1000$ et $\overline{X} = 549/1000 = 0,549$.

Par le théorème central limite,
$$\frac{n\overline{X} - np}{\sqrt{np(1-p)}} \xrightarrow{loi} \mathcal{N}(0,1).$$
Donc
$$P\left(\frac{n\overline{X} - np}{\sqrt{np(1-p)}} \geq a\right) = P\left(\overline{X} \geq p + \frac{a\sqrt{p(1-p)}}{\sqrt{n}}\right) \simeq \int_a^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

$\int_a^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \simeq 0,05$ pour $a \simeq 1,65$.

Pour $p = 1/2$ et $n = 1000$, $p + a\sqrt{\frac{p(1-p)}{n}} \simeq 0,526$.

Donc, si H_0 est vraie, $P(\overline{X} \geq 0,526) \leq 0,05$. Or $\overline{X} = 0,549$, donc on rejette H_0 (très peu probable) et on conclut que H_1 est vraie. Le candidat A décide donc de se présenter.

Si seulement 515 personnes sur 1000 avaient répondu oui, on aurait conclu que H_0 était vraie. Le choix entre H_0 et H_1 dépend de la marge d'erreur choisie.

2. Tests du χ^2 [Dacunha-Castelle p135]

Soit X une v.a. à valeurs dans $\{1, \ldots, N\}$. On cherche la loi de X.

Soit $X_1, \ldots, X_n$ un n-échantillon de X.

Estimateur de $P(X = k)$:
$$\bar{p}_n(k) = \frac{\text{Card}\{i \in \{1, \ldots, n\} \mid X_i = k\}}{n}$$

Loi empirique de X :
$$\bar{\mu}_n = \sum_{k=1}^{N} \bar{p}_n(k) \delta_k$$

Remarque.
$\sum_{k=1}^{N} P(X = k) = 1$ donc la loi de X est entièrement déterminée par $(P(X = k))_{1 \leq k \leq N-1}$.
Il n'y a que $N - 1$ inconnues.

De même, les $\bar{p}_n(k)$ sont liés par la relation $\sum_{k=1}^{N} \bar{p}_n(k) = 1$.

"Distance" entre 2 mesures :
si $\mu = \sum_{k=1}^{N} p(k)\delta_k$ et $\nu = \sum_{k=1}^{N} q(k)\delta_k$, on pose
$$\chi^2(\mu, \nu) = \sum_{k=1}^{N} \frac{(p(k) - q(k))^2}{p(k)}.$$

Remarque. Ce n'est pas symétrique, donc ce n'est pas une vraie distance.

Si μ est la loi de X, $\bar{p}_n(k) \to p(k)$ quand $n \to +\infty$ en raison de la loi des grands nombres.
(pour k fixé, considérer les v.a. $Y_i = \mathbb{1}_{\{X_i = k\}}$)

Donc $\lim_{n \to +\infty} \chi^2(\mu, \bar{\mu}_n) \to 0$.

χ^2 d'ajustement :
$$\chi_n^2(\mu, \bar{\mu}_n) \stackrel{\text{def}}{=} n\chi^2(\mu, \bar{\mu}_n).$$

Cette quantité sert à préciser à quelle vitesse $\chi^2(\mu, \bar{\mu}_n)$ tend vers 0. Le théorème suivant montre que c'est la bonne normalisation, autrement dit $\chi^2(\mu, \bar{\mu}_n)$ est un $o(1/n)$.

La **loi du χ^2 à d degrés de liberté**, notée $\chi^2(d)$, est la loi de $N_1^2 + \cdots + N_d^2$, où $N_1, \ldots, N_d$ sont des v.a. indépendantes de loi $\mathcal{N}(0,1)$.

Théorème. Soit $(X_n)_{n\geq 1}$ des v.a. indépendantes de même loi, à valeurs dans $\{1,\ldots,N\}$, et $\bar{\mu}_n$ la loi empirique de $(X_1,\ldots,X_n)$. Si la loi des X_n est μ alors

$$\chi^2_n(\mu,\bar{\mu}_n) \xrightarrow{loi} \chi^2(N-1) \quad \text{quand } n \to +\infty.$$

Si la loi des X_n est différente de μ, alors

$$\chi^2_n(\mu,\bar{\mu}_n) \xrightarrow{p.s.} +\infty$$

Test d'ajustement.
Permet de tester "la loi de X est-elle égale à μ ?".
– oui si $\chi^2_n(\mu,\bar{\mu}_n)$ tend vers $\chi^2(N-1)$,
– non si $\chi^2_n(\mu,\bar{\mu}_n)$ tend vers $+\infty$.

Tester l'indépendance de 2 v.a.

Soit X une v.a. à valeurs dans $\{1,\ldots,N\}$ et Y une v.a. à valeurs dans $\{1,\ldots,M\}$. On cherche à savoir si $X \perp Y$. Comme ce sont des lois discrètes :

$$X \perp Y \Leftrightarrow \forall i,j, P(X=i,Y=j) = P(X=i)P(Y=j).$$

Soit $(X_1,Y_1),\ldots,(X_n,Y_n)$ un échantillon de (X,Y).

Estimateur de $P(X=i,Y=j)$:

$$\bar{p}_{ij}(n) = \frac{\text{Card}\{k \in \{1,\ldots,n\} \mid X_k = i, Y_k = j\}}{n}$$

Estimateur de $P(X=i)$: $\quad \bar{p}_i^X(n) = \sum_j \bar{p}_{ij}(n)$

(car $P(X=i) = P(X=i, Y \in \{1,\ldots,M\})$)

Estimateur de $P(Y=j)$: $\quad \bar{p}_j^Y(n) = \sum_i \bar{p}_{ij}(n)$

Soit $\bar{\mu}_n$ la loi empirique de (X,Y), c'est-à-dire c'est la loi donnée par

$$\bar{\mu}_n(i,j) = \bar{p}_{ij}(n).$$

Soit $\bar{\nu}_n$ le produit des lois empiriques de X et de Y, c'est-à-dire la loi donnée par

$$\bar{\nu}_n(i,j) = \bar{p}_i^X(n)\bar{p}_j^Y(n).$$

($\bar{\mu}_n$ et $\bar{\nu}_n$ sont à valeurs dans $\{1,\ldots,N\}\times\{1,\ldots,M\}$)

Loi des grands nombres : $\bar{\mu}_n(i,j) \to P(X=i,Y=j)$ et $\bar{\nu}_n(i,j) \to P(X=i)P(Y=j)$.

Pour savoir si $X \perp Y$, il faut comparer $\bar{\mu}_n$ et $\bar{\nu}_n$.

Théorème.
Si X et Y sont indépendantes, alors

$$\chi^2_n(\bar{\nu}_n,\bar{\mu}_n) \xrightarrow{loi} \chi^2((N-1)(M-1)).$$

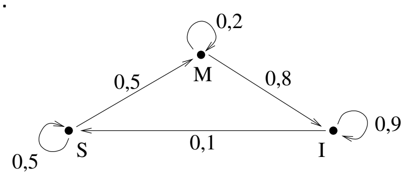
Sinon, $\chi^2_n(\bar{\nu}_n,\bar{\mu}_n) \xrightarrow{p.s.} +\infty$.

$\chi^2_n(\bar{\nu}_n,\bar{\mu}_n)$ est appelé le **χ^2 d'indépendance**.

XII. Chaîne de Markov

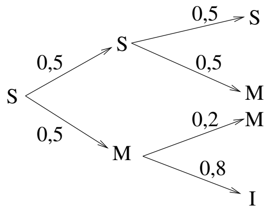
1. Exemple et définition

On s'intéresse à une maladie véhiculée par des insectes. Un individu peut être soit malade (M), soit immunisé (I), soit sain (S). D'un mois à l'autre, il peut passer d'un état à l'autre avec les probabilités suivantes :



Au départ, il est sain. On veut modéliser l'évolution future de son état de santé.

Soit X_n la v.a. à valeurs dans $\{M,S,I\}$ représentant l'état de santé de l'individu au mois n.
(Si on considère une population entière au lieu d'un individu, la loi de X_n donne les proportions de sains, malades et immunisés ; la loi de X_0 est la répartition initiale)
$P(X_0 = S) = 1$.
$P(X_1 = S) = 0,5$; $P(X_1 = M) = 0,5$; $P(X_1 = I) = 0$.



$P(X_2 = S) = 0,5^2 = 0,25$
$P(X_2 = M) = 0,5^2 + 0,5 \times 0,2 = 0,35$
$P(X_2 = I) = 0,5 \times 0,8 = 0,4$

Vu la modélisation adoptée pour l'évolution de la maladie, les $(X_n)_{n\geq 0}$ ne sont pas indépendantes.

$P(X_{n+1} = M \mid X_n = S) = 0,5$
$P(X_{n+1} = M \mid X_n = M) = 0,2$
$P(X_{n+1} = M \mid X_n = I) = 0$, etc.

Il suffit de savoir l'état de l'individu au temps n pour prévoir son état au temps $n+1$, il n'est pas nécessaire de connaître l'état de l'individu aux temps $0, \dots, n-1$.

Définition. Soit $(X_n)_{n\geq 0}$ des v.a. définies sur Ω, à valeurs dans E (ensemble fini ou dénombrable). La suite de v.a. $(X_n)_{n\geq 0}$ est une **chaîne de Markov** si elle vérifie :

$$(*) \quad P(X_{n+1} = i_{n+1} \mid X_0 = i_0, \dots X_n = i_n) = P(X_{n+1} = i_{n+1} \mid X_n = i_n)$$

pour tous $i_0, \dots, i_{n+1} \in E$ tels que
$P(X_0 = i_0, \dots, X_n = i_n) > 0$
(sinon la proba conditionnelle n'est pas définie)

La propriété $(*)$ est appelée **propriété de Markov**.

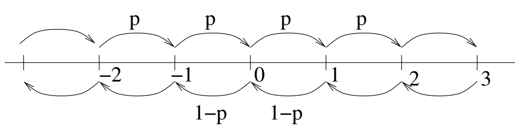
E est l'**espace des états**. La loi de X_0 est la **loi initiale** (ou probabilité initiale).

La chaîne de Markov est **homogène** si, $\forall i, j$, $P(X_{n+1} = j \mid X_n = i)$ est indépendant de n.

Exemple.
Dans l'exemple de modélisation de maladie, $(X_n)_{n\geq 0}$ est une chaîne de Markov homogène.
Si le nombre d'insectes évolue avec le temps, la probabilité de tomber malade change ($P(X_{n+1} = M \mid X_n = S) = p_n$). $(X_n)_{n\geq 0}$ est encore une chaîne de Markov, mais pas homogène.

Exemple. *Marche aléatoire sur $\mathbb{Z}$.*

On part de 0, on se déplace sur $\mathbb{Z}$ en faisant à chaque étape un pas vers la droite avec probabilité p ou vers la gauche avec probabilité $1-p$.



X_n est la position au bout de n pas ($X_0 = 0$). C'est une chaîne de Markov homogène.

Probabilités de transition :
$P(X_{n+1} = i+1 \mid X_n = i) = p$
$P(X_{n+1} = i-1 \mid X_n = i) = 1-p$
$P(X_{n+1} = j \mid X_n = i) = 0$ *si* $j \neq i+1, i-1$.

Une marche aléatoire sur n'importe quel graphe, fini ou dénombrable, donne une chaîne de Markov homogène.

2. Matrice de transition

Définition.
Si $(X_n)_{n\geq 0}$ est une chaîne de Markov homogène, la **matrice de transition** associée est $M = (m_{ij})_{i,j\in E}$ définie par $m_{ij} = P(X_{n+1} = j \mid X_n = i)$.

Définition.
$A = (a_{ij})_{i,j\in E}$ est une **matrice stochastique** si $A \geq 0$ et la somme de chaque ligne vaut 1.
(càd $\forall i, j \in E, a_{ij} \geq 0$ et $\forall i \in E, \sum_{j\in E} a_{ij} = 1$)

La matrice de transition M d'une chaîne de Markov est toujours une matrice stochastique car $\sum_{j\in E} m_{ij}$ est la probabilité d'être dans un état j quelconque en partant de l'état i, donc cette somme vaut 1 ; de plus, $m_{ij} \geq 0$ car c'est une probabilité.

Dans la représentation sous forme de graphe probabiliste, la somme des probabilités sur les flèches partant d'un état doit être égale à 1.

Définition.
Un **vecteur de probabilité** est un vecteur ligne $v = (v_i)_{i \in E}$ avec $v \geq 0$ et $\sum\limits_{i \in E} v_i = 1$.

Si X est une v.a. à valeurs dans E et $\pi_i = P(X = i)$, le vecteur $\pi = (\pi_i)_{i \in E}$ est un vecteur de probabilité.

On identifie une loi et le vecteur de probabilité qui lui correspond.

Exemple . Dans l'exemple de modélisation d'une maladie, si on prend dans l'ordre M, I, S, la matrice de transition est

$$\begin{pmatrix} 0,2 & 0,8 & 0 \\ 0 & 0,9 & 0,1 \\ 0,5 & 0 & 0,5 \end{pmatrix}$$

et le vecteur de proba de la loi initiale est $(0, 0, 1)$.

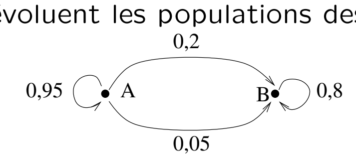
Proposition. Si $\pi = (\pi_i)_{i \in E}$ est la loi de X_0 alors

$$P(X_0 = i_0, \ldots, X_n = i_n) = \pi_{i_0} m_{i_0 i_1} m_{i_1 i_2} \ldots m_{i_{n-1} i_n}.$$

3. Comportement asymptotique (quand l'espace d'états est fini)

Exemple. La ville A est moins chère mais la ville B est plus agréable. Chaque année, 5% des habitants de A partent habiter B et 20% des habitants de B partent habiter A.

Comment évoluent les populations des 2 villes ?



Sur chaque flèche est indiquée la probabilité d'un habitant de passer d'une ville à une autre ou de ne pas changer de ville. Au niveau global, ce sont les taux de transfert de population.

Remarque. Les populations sont stables ssi les flux s'équilibrent, c'est-à-dire $0,05 N_A = 0,2 N_B$, avec N_A, N_B les nombres d'habitants dans A, B. Le seul état stable est donc quand $N_A = 4 N_B$, càd quand 80% de la population est dans A et 20% dans B.

On a une chaîne de Markov homogène, de matrice de transition $M = \begin{pmatrix} 0,95 & 0,05 \\ 0,2 & 0,8 \end{pmatrix}$.

Soit a_n et b_n les pourcentages de la population totale dans les villes A et B l'année n.

$a_{n+1} = a_n \times 0,95 + b_n \times 0,2$
$b_{n+1} = a_n \times 0,05 + b_n \times 0,8$

Si $V_n = (a_n \ b_n)$, on a la relation $V_{n+1} = V_n \cdot M$.

Remarque. On utilise des vecteurs lignes et on multiplie les vecteurs à gauche. *C'est la convention en probabilité, qui découle de la convention définissant les matrices de transition.*
Cela reviendrait au même de considérer des vecteurs colonnes pour la matrice transposée ${}^tM$.

On voit immédiatement que $V_n = V_0 \cdot M^n$.

M^n donne les probabilités de transition au bout de n années.

Si $(X_n)_{n \geq 0}$ et la chaîne de Markov associée à ce problème, la loi de X_n est :
$P(X_n = A) = a_n$,
$P(X_n = B) = b_n$.
(probabilité qu'un individu soit dans A ou dans B la n-ième année)

La loi de X_n est donnée par le vecteur V_n.

De façon générale, on a le résultat suivant :

Proposition.
Soit $(X_n)_{n \geq 0}$ une chaîne de Markov homogène de matrice de transition M, et V_n le vecteur ligne donnant la loi de X_n. Alors

$$V_n = V_{n-1} \cdot M = V_0 \cdot M^n.$$

On est ramené à étudier le comportement asymptotique de $V_0 \cdot M^n$.

Valeurs propres de M dans l'exemple ?

$$M \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

car M est une matrice stochastique.
Donc 1 est valeur propre de M.

$\det M = 0,75 \neq 1^2$ donc M a une autre valeur propre $\lambda = \frac{\det M}{1} = 0,75$.

On veut utiliser des vecteurs lignes à gauche.
V vecteur ligne propre à gauche pour $M \iff$
${}^tV$ vecteur colonne propre à droite pour ${}^tM$.

M et ${}^tM$ ont les mêmes valeurs propres. Il existe des vecteurs propres à gauche U_1 et U_2 associées aux valeurs propres 1 et λ, qui forment une base.

Tout vecteur ligne V peut s'écrire $V = \alpha U_1 + \beta U_2$ donc $V \cdot M^n = \alpha U_1 + \beta \lambda^n U_2 \to \alpha U_1$ car $\lambda^n \to 0$.

Dans l'exemple : $U_1 = (0,8; 0,2)$ (on normalise U_1 pour obtenir un vecteur de probabilité),
$U_2 = (1; -1)$ et $V_0 = \alpha U_1 + \beta U_2$ avec $\alpha = 1$.

$V_0 = \alpha U_1 + \beta U_2$ est un vecteur de proba. Or $(U_1)_1 + (U_1)_2 = 1$ et $(U_2)_1 + (U_2)_2 = 0$ donc $(V_0)_1 + (V_0)_2 = 1 = 1\alpha + 0\beta$.

Donc $V_n = V_0 \cdot M^n$ converge vers U_1.
Si $V_0 = U_1$, alors $V_n = U_1$ pour tout n.
(c'est le cas d'équilibre vu au début)

Les populations dans les villes A et B convergent vers des populations stables correspondant à 80% et 20% de la population totale, indépendamment de la répartition initiale entre les 2 villes.

Définition.
Un vecteur de probabilité V est **invariant** (ou stationnaire) si $V = VM$.
Une chaîne de Markov $(X_n)_{n\geq 0}$ est **stationnaire** si la loi de X_n est la même pour tout n.
$(X_n)_{n\geq 0}$ est stationnaire ssi la loi de X_0 est invariante.

Définition. Un graphe orienté est **fortement connexe** si pour tous sommets i, j, il existe une suite de flèches allant de i à j.

Une matrice stochastique A est **irréductible** si le graphe associé à A est fortement connexe (on ne dessine que les flèches avec proba > 0).
De façon équivalente, A est irréductible si

$$\forall i, j, \text{ il existe } n \geq 1 \text{ tel que } (A^n)_{ij} > 0.$$

Un graphe orienté est **apériodique** si le pgcd des longueurs des boucles vaut 1.
*Il est **périodique** de période $p > 1$ si toutes les boucles ont une longueur multiple de p (avec p minimal).*

Une matrice stochastique A irréductible et apériodique est dite **primitive**. De façon équivalente, A est primitive si

$$\text{il existe } n \geq 1 \text{ tel que } A^n > 0.$$

Théorème de Perron-Frobénius (pour matrices stochastiques).
Soit A une matrice stochastique primitive.
– 1 est valeur propre de A, c'est une racine simple du polynôme caractéristique de A.
– Si λ est une autre valeur propre de A, $|\lambda| < 1$.
– 1 possède un vecteur propre à gauche dont les coordonnées sont strictement positives.
– Pour tout vecteur ligne v à coordonnées positives, la suite $v \cdot A^n$ converge exponentiellement vite vers un vecteur propre associé à 1 ;
la vitesse de convergence est donnée par le module de la 2ème plus grande valeur propre.

Conséquences.

Si $(X_n)_{n\geq 0}$ est une chaîne de Markov homogène avec un espace d'états fini et une matrice de transition primitive, alors il existe une unique probabilité invariante μ, et la loi de X_n converge en loi vers μ, indépendamment de la loi initiale.

Comme $\mu(i) > 0$ pour tout état i, la suite $(X_n)_{n\geq 0}$ passe p.s. par tous les états, une infinité de fois.

4. Comportement asymptotique quand le nombre d'états est dénombrable

Soit $(X_n)_{n\geq 0}$ une chaîne de Markov homogène à valeurs dans E, avec E dénombrable.

Définition.
Soit $i \in E$. On dit que l'état i est **récurrent** si, en partant de i, la chaîne de Markov repasse en i p.s. *(c'est-à-dire $P(\exists n \geq 1, X_n = i \mid X_0 = i) = 1$).*
Sinon, on dit que i est **transient**.

Théorème.
Si i est récurrent, alors la chaîne de Markov partant de i repasse une infinité de fois par i p.s.

Exemple.
Marche aléatoire sur $\mathbb{Z}$, avec probabilité p de faire un pas sur la droite.
– 0 est transient si $p \neq 1/2$.
– 0 est récurrent si $p = 1/2$.

Définition.
La chaîne de Markov homogène $(X_n)_{n\geq 0}$ est **irréductible** si
$\forall i, j \in E, \exists n > 0, P(X_n = j \mid X_0 = i) > 0$.
(si on représente la chaîne de Markov par un graphe avec des flèches ayant des probabilités non nulles, il est équivalent de demander que pour tous i, j on peut aller de i à j en suivant les flèches).

Théorème.
Si $(X_n)_{n\geq 0}$ est irréductible, alors soit tous les états sont récurrents, soit tous les états sont transients.

Théorème.
Si $(X_n)_{n\geq 0}$ est irréductible et si E est fini, alors tout $i \in E$ est récurrent.

Définition.
Soit $i \in E$ un état récurrent. Si l'espérance du temps de retour en i de la chaîne de Markov partant de i est fini, on dit que i est **récurrent positif**.
Sinon, on dit qu'il est **récurrent nul**.

Théorème.
Si $(X_n)_{n\geq 0}$ est irréductible et récurrente, alors soit tous les états sont récurrents positifs, soit tous les états sont récurrents nuls.

Théorème.
Si $(X_n)_{n\geq 0}$ est irréductible et si E est fini, alors tout $i \in E$ est récurrent positif.

Exemple.
La marche aléatoire sur $\mathbb{Z}$, avec probabilité $p = 1/2$ de faire un pas sur la droite est récurrente nulle.