Probas : principales définitions et théorèmes

Laure Danthony

Probas discrètes et lois des grands nombres

DÉFINITION 1
On se donne un ensemble Ω d'**événements élémentaires**. Pour chaque événement de Ω on définit $p(\omega) \geq 0$. La probabilité de $A \subseteq \Omega$ est définie par $p(A) = \sum_{\omega \in \Omega} p(\omega)$.

Loi faible des grands nombres

THÉORÈME 1 (LOI FAIBLE DES GRANDS NOMBRES)

$$\forall \varepsilon > 0, \lim_{n \to +\infty} p\Big(\omega / \left| \frac{1}{n} \sum_{i=1}^{n} d_i(\omega) - \frac{1}{2} \right| \Big) \geq \varepsilon) = 0$$

Il se démontre à l'aide de la :

PROPOSITION 1 *Soit $f : \Omega \to \mathbb{R}$ une fonction en escalier : $f(x) = c_i$ pour $x \in]a_i, a_{i+1}]$ et $a_0 = 0 < a_1 < \ldots < a_n = 1$. Alors l'ensemble $\{\omega / f(\omega) \geq \alpha\}$ est une union finie d'intervalles.*

Loi forte des grands nombres

DÉFINITION 2
$\omega \in \Omega$ est dit **normal** si sa fréquence de 1 dans son développement binaire est égale à sa fréquence de 0, ou plus formellement : si $\frac{1}{n} \sum_{i=1}^{n} d_i(\omega) \underset{n \to +\infty}{\longrightarrow} \frac{1}{2}$.

NOTATION : On note $\mathbf{N}$ l'ensemble des nombres normaux.

REMARQUE 1 Il y a un petit problème car $\mathbf{N}$ n'est pas une union finie d'intervalles. On montre donc que son complémentaire $\mathbf{N}^C$ est *négligeable*

DÉFINITION 3
Une partie $A \subseteq \Omega$ est négligeable si quel que soit $\varepsilon > 0$, il existe une suite finie ou infinie d'intervalles $(I_k)_{k \geq 1}$ telle que A soit inclus dans leur réunion et $\sum_{k \geq 1} p(I_k) \leq \varepsilon$.

THÉORÈME 2 (LOI FORTE DES GRANDS NOMBRES) *Le complémentaire $\mathbf{N}^C$ dans Ω de l'ensemble des nombres normaux est négligeable.*

Sigma-algèbres

DÉFINITION 4
On se donne Ω. Une famille $\mathcal{F}$ de parties de Ω est une **σ-algèbre** ou **tribu** si elle vérifie :

1. $\Omega \in \mathcal{F}$;
2. si $A \in \mathcal{F}$, alors $\Omega \setminus A \in \mathcal{F}$;
3. (clôture par union dénombrable) Si $(A_n)_{n\geq 1} \in \mathcal{F}$, alors $\bigcup_{n\in\mathbb{N}^*} A_n \in \mathcal{F}$.

REMARQUES 2
- Si $\mathcal{F}$ est une σ-algèbre, elle est également close par union dénombrable.
- La même définition où on remplace *clôt par union dénombrable* par *clôt par union finie* est celle d'une **algèbre**.

EXEMPLES 1
- *L'ensemble des partires de Ω est une σ-algèbre.*
- *L'ensemble des unions finies d'intervalles de $\Omega =]0, 1]$ est une algèbre, mais pas une σ-algèbre.*

DÉFINITION 5
Soit $\mathcal{A}$ une famille de parties de Ω. On note $\sigma(\mathcal{A})$ la **sigma-algèbre engendrée** par $\mathcal{A}$, c'est à dire l'intersection de toutes les sigma-algèbres contenant $\mathcal{A}$.

IMPORTANT ! Si $\Omega =]0, 1]$ et $\mathcal{A}$ est l'ensemble des intervalles de la forme $]a, b]$. Alors $\sigma(\mathcal{A})$ est appelée **la sigma-algèbre des boréliens de** $]0, 1]$.

EXEMPLES 2
- *tout ouvert de Ω est un borélien. (et tout fermé)*
- *l'ensemble des nombres normaux est un borélien.*

Mesures de Probabilités

DÉFINITION 6
Soit $\mathcal{F}$ une algèbre de Ω. Une mesure de probabilité P associe à chaque $A \in \mathcal{F}$ un réel $P(A)$ de sorte que :

1. $0 \leq P(A) \leq 1$ pour tout A de $\mathcal{F}$.
2. $P(\emptyset) = 0, P(\Omega) = 1$.
3. additivité dénombrable : Soit $(A_n)_{n\geq 1}$ une suite d'éléments de $\mathcal{F}$ disjoints, alors :
$$\textit{si} \bigcup_{n\geq 1} A_n \in \mathcal{F}, \text{ alors } P\Big(\bigcup_{n\geq 1} A_n\Big) = \sum_{n=1}^{\infty} P(A_n)$$

Remarque 3 La dernière condition (stabilité par union dénombrable) est inutile si $\mathcal{F}$ est une σ-algèbre.

On obtient en conséquences les propriétés suivantes :

- additivité finie
- $P(\Omega \setminus A) = 1 - P(A)$ si $A \in \mathcal{F}$.
- $A \subseteq B$ implique $P(A) \leq P(B)$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- sous additivité : $P\Big(\bigcup_{k=1}^{n} A_k\Big) \leq \sum_{k=1}^{n} P(A_k)$; et si $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$, alors $P\Big(\bigcup_{k=1}^{\infty} A_k\Big) \leq \sum_{k=1}^{\infty} P(A_k)$

Théorème 3 *Soit $(A_n)_{n\geq 1}$ une suite d'éléments de $\mathcal{F}$:*

1. *si $\bigcup_{n\geq 1} A_n \in \mathcal{F}$ et $A_n \subseteq A_{n+1}$, alors $P(\bigcup_{k\geq 1} A_k) = \underset{n\to\infty}{\longrightarrow} P(A_k)$.*
2. *si $\bigcap_{n\geq 1} A_n \in \mathcal{F}$ et $A_{n+1} \subseteq A_n$, alors $P(\bigcap_{k\geq 1} A_k) = \underset{n\to\infty}{\longrightarrow} P(A_k)$.*

Mesure de Lebesgue

Il existe une unique mesure de probabilité sur l'ensemble des boréliens de $]0,1]$, telle que pour $]a,b] \subseteq]0,1]$, $P(]a,b]) = b - a$ (mesure de Lebesgue).

Théorème 4 *Une mesure de probabilité P sur une algèbre $\mathcal{F}$ a une extension unique à la sigma-algèbre $\sigma(\mathcal{F})$ engendrée par $\mathcal{F}$.*

On applique ce théorème plus général à la mesure de Lebesgue : on prend pour $\mathcal{F}$ l'algèbre des unions finies d'intervalles $]a,b] \in]0,1]$. Si $A = \bigcup_{i=1}^{n}]a_i, b_i]$ (union disjointe), on pose $P(A) = \lambda(A) = \sum_{i=1}^{n}(b_i - a_i)$.

Variables aléatoires

Définition 7
On se donne $(\Omega, \mathcal{F}, P)$, où $\mathcal{F}$ est une sigma-algèbre sur Ω, P une mesure de probabilité sur $\mathcal{F}$. Une **variable aléatoire** est une fonction $X : \Omega \to \mathbb{R}$ telle que $\forall x \in \mathbb{R}, \{\omega / X(\omega) \leq x\} \in \mathcal{F}$.

Définition 8
La **fonction de distribution** est alors : $F_X : \mathbb{R} \to [0,1]$, $x \mapsto P(\{\omega/X(\omega) \leq x\})$.

PROPOSITION 2 *Quelques propriétés des fonctions de distribution*

- F *est croissante*
- $F \underset{x \to +\infty}{\longrightarrow} = 1, F \underset{x \to -\infty}{\longrightarrow} = 0.$
- F *est continue à droite.*

THÉORÈME 5 *Une fonction* $F : \mathbb{R} \to [0,1]$ *est une fonction de distribution ssi elle vérifie les trois propriétés précédentes.*

PROPOSITION 3 *Si* $X : \Omega \to \mathbb{R}$ *est une variable aléatoire,* $X^{-1}(A) \in \mathcal{F}$ *pour tout borélien* A *de* $\mathbb{R}$.

DÉFINITION 9
Considérons $(\Omega, \mathcal{F}, P)$ et soit X une variable aléatoire sur cet espace. Soit $A \subseteq \mathbb{R}$ un borélien et soit $Q(A) = P(\omega, X(\omega) \in A)$. Alors Q est une mesure de proabailité sur l'ensemble des boréliens de $\mathbb{R}$ appelée **la mesure image de P par X**.

Indépendance

DÉFINITION 10
Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. On dit que deux éléments A et B de $\mathcal{F}$ sont **indépendants** si $P(A \cap b) = P(A)P(B)$. On peut généraliser à une suite d'événements de $\mathcal{F}$.

DÉFINITION 11
Deux variables aléatoires $X, Y : \Omega \to \mathbb{R}$ sont dites **indépendantes** si les événements $X \in A, Y \in B$ sont indépendants pour tous boréliens A et B de $\mathbb{R}$. On peut généraliser à une suite de variables aléatoires.

Espérance

DÉFINITION 12
On dit que $X : \Omega \to R$ est une **variable aléatoire discrète** si $X(\Omega)$ est dénombrable.

DÉFINITION 13

- Soit $A = X(\Omega)$. Si la série $\sum_{a \in A} |a| \, P(X = a)$ converge, on dit que l'**espérance** existe et sa valeur est alors :
$$E(X) = \sum_{a \in A} aP(X = a).$$
- Si en plus $A \subseteq \mathbb{R}^+$ et que la série $\sum_{a \in A} aP(X = a)$. diverge, alors on dit que $E(x) = +\infty$.

PROPOSITION 4 *Supposons que* $E(X)$ *et* $E(Y)$ *existent* $(X, Y : \Omega \to R)$, *alors :*

$$E(X+Y) = E(X) + E(Y).$$

REMARQUE 4 Attention, une telle propriété n'existe pas dans le cas général pour un produit.

Par contre on a :

PROPOSITION 5 *Si* X *et* Y *sont des variables indépendantes, alors*

$$E(X * Y) = E(X) * E(Y).$$

PROPOSITION 6 *Supposons* $X : \Omega \to A$ *avec* $A \subseteq R+$ *fini. Alors on a :*

$$E(X) = \int_0^{+\infty} P(X \geq x)dx$$

Cas particulier :

COROLLAIRE 1 *Soit* $X : \Omega \to \mathbb{N}$ *une variable aléatoire. Si* $E(X)$ *existe, alors*

$$E(X) = \sum_{n \geq 1} P(X \geq n).$$

Quelques résultats (inégalités) sur les variables aléatoires discrètes :

PROPOSITION 7 *Si* X *et* y *sont des variables aléatoires discrètes :*

- *si* $0 \leq x \leq y$, *alors* $0 \leq E(X) \leq E(Y)$, *cette inégalité étant même vraie si il y a divergence.*
- *si* $X \geq 0$, *on a* $P(X \geq \alpha) \leq \frac{1}{\alpha}E(X)$.
- *(Cauchy Schwarz) :* $E(XY)^2 \leq E(X^2)E(Y^2)$.

PROPOSITION 8 (INÉGALITÉ DE MARKOV)

$$\forall \alpha > 0, \forall k, P(|X| \geq \alpha) \leq \frac{1}{\alpha}E(|X|^k).$$

On utilise cette dernière proposition pour montrer le :

COROLLAIRE 2 (TCHÉBITCHEV) $P(X - m \geq \alpha) \leq \dfrac{\sigma^2}{\alpha^2}$

Variance

DÉFINITION 14
Soit X une variable aléatoire telle que $E(X)$ est définie. Alors **la variance** de X : $Var(X) = E\big((X-m)^2\big)$ avec $m = E(X)$. Cette quantité mesure donc un "écart à la moyenne".

DÉFINITION 15 (ECART-TYPE)
$Var(X) = \sigma^2$, σ étant **l'écart type**.

PROPOSITION 9 *$Var(X) = E(X^2) - E(X)^2$.*

PROPOSITION 10 *Quelques propriétés sur la variance :*

- *$Var(\lambda X) = \lambda^2 Var(X)$ (bilinéarité).*
- *si X et Y sont des variables aléatoires indépendantes, alors $Var(X+Y) = Var(X) + Var(Y)$.*
- *si les variables aléatoires X_i sont indépendantes deux à deux, alors :*

$$Var(X_1 + \ldots + X_n) = \sum_{i=1}^{n} Var(X_i).$$

Lois générales des grands nombres

Loi faible des grands nombres

DÉFINITION 16
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires discrètes : $\Omega \to \mathbb{R}$ de même espérance μ finie. Cette suite **vérifie la loi faible** si :

$$\forall \epsilon > 0, \lim_{n\to+\infty} p\Big(\Big|\frac{X_1 + \ldots + X_n}{n} - \mu\Big| \geq \epsilon\Big) = 0.$$

THÉORÈME 6 *Soient $X_1, \ldots X_n$ des variables aléatoires indépendantes deux à deux avec la même espérance m finie et la même espérance σ^2 finie. Alors :*

$$P\Big(\Big|\frac{X_1 + \ldots + X_n}{n} - m\Big| \geq \alpha\Big) \leq \frac{\sigma^2}{n\alpha^2}$$

Loi forte des grands nombres

DÉFINITION 17
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires discrètes : $\Omega \to \mathbb{R}$ de même espérance μ finie. Cette suite **vérifie la loi forte** si :

$$\forall \epsilon > 0, \lim_{n\to+\infty} p\Big(\Big|\frac{X_1 + \ldots + X_n}{n} = \mu\Big|\Big) = 1.$$

REMARQUE 5 Si une suite de variables aléatoires discrètes vérifie la loi forte, elle vérifie la loi faible. La démo n'est pas si évidente que cela, elle utilise le lemme suivant.

LEMME 1 *Soit A_n des événements dépendants de n. $A = limsup(A_n)$ est l'événement "A_n se produit pour une infinité de valeurs de n". Alors :*

$$\text{si } P(A) = 0, \text{ alors } \lim_{n\to+\infty} P(A_n) = 0$$

REMARQUE 6 $B = liminf(A_n)$ est l'événement "A_n se produit toujours à partir d'un certain rang".

THÉORÈME 7 *Soit X_n une suite de variables aléatoires discrètes indépendantes, de même espérance finie μ avec $E(X_n^4) \leq A < +\infty$ pour un certain A indépendant de n. Alors :*

$$P\Big(\lim_{n\to+\infty} \frac{X_1 + \ldots + X_n}{n} = \mu\Big) = 1$$

REMARQUE 7 Ce théorème est aussi valable si il n'y a qu'indépendance 4 à 4. Si les variances sont identiquement distribuées, la seule hypothèse à vérifier est $\mu = E(X_i)$ fini.

Fonctions génératrices

DÉFINITION 18
Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. La fonction génératrice de probabilité de X est la fonction g_X définie par : $g_X(z) = E(z^X)$. Plus simplement :

$$g_X(z) = \sum_{n=0}^{+\infty} P(X = n) z^n.$$

FAIT 1 *De manière générale, on peut seulement dire que le rayon de convergence de la série entière est ≥ 1, ce qui pose des problèmes de dérivation.*

PROPOSITION 11 *La fonction génératrice est très utile car :*

- $g'_X(1) = E(X)$,
- $g''_X(1) = E(X^2) - E(X)$,
- *donc* $E(X) = g'_X(1)$ *et* $Var(X) = g''_X(1) + g'_X(1) - (g''_X(1))^2$.

REMARQUE 8 En toute rigueur il faudrait parler de dérivée "à gauche".

FAIT 2 *X et Y étant deux variables indépendantes : $G_{X+Y} = G_X G_Y$, son rayon de convergence vérifiant $\rho_{X+Y} \geq \text{Inf}\{\rho_X, \rho_Y\}$*

REMARQUE 9 Ceci permet de montrer facilement que la somme de deux lois de Poisson de paramètres λ et μ est une loi de Poisson de paramètre $\lambda + \mu$, que la somme de deux lois géométriques est géométrique, ...

Bornes de Chernoff

Soient $X_1, \dots X_n$ une suite de variables aléatoires indépendantes, telles que $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$, et soit $S_n = \sum_{i=1}^{n} X_i$.

Théorème 8 $\forall a > 0, P(S_n > a) = P(S_n < -a) < e^{-\frac{a^2}{2n}}$

Variantes : Soit Y une somme de variables aléatoires indicatrices (c'est-à-dire à valeurs dans $0, 1$) indépendantes. Soit $\mu = E(Y)$.

Lemme 2 $P\big(Y < (1-\varepsilon)\mu\big) < e^{-\frac{\mu\varepsilon^2}{2}}$

Lemme 3 $P\big(Y > (1-\varepsilon)\mu\big) < \Big(\frac{e^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\Big)^{\mu}$

Avec des deux lemmes, on montre le suivant :

Lemme 4 $P\big(\,|Y-\mu| > \varepsilon\mu\big) < 2^{-c_\varepsilon \mu}$, *$c_{eps}$ ne dépend que de ε.*

Chaînes de Markov

Définition 19 (Chaîne)
Une **chaîne de Markov** est la donnée d'un ensemble d'états S (fini ou dénombrable), et de probabilités de transition $p_{i,j}$ $(i,j \in S)$, telles que $\forall i,j, p_{i,j} \geq 0$ et $\forall i, \sum_{j \in S} p_{i,j} = 1$.

On note X_n l'état au rang n. On dispose ainsi d'une suite de variables aléatoires.

Remarquons que :

- L'état au rang n ne dépend que de l'état au rang $n-1$:

  $P(X_n = i_n / X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1} = P(X_n = i_n / X_{n-1} = i_{n-1}.$

- L'évolution est totalement déterminée par les $p_{i,j}$ et la distribution initiale $\pi^0(j)$: $P(X_{N+1} = j) = \sum_{i \in S} P(X_n = i) p_{i,j}$, ce qui s'écrit sous **forme matricielle** $\pi^{n=1} = P.\pi^n$ si P désigne la matrice de terme $a_{i,j} = p_{j,i}$ *(la probabilité $p_{i,j}$ est à la ligne j et à la colonne i)*.

Remarques 10

- π^n est appelé **vecteur d'état**.
- Cela marche aussi avec S dénombrable, on travaille alors avec des matrices "infinies" ...
- On a aussi $\pi^{n+k} = P^k \pi^n$, d'où l'intérêt de savoir calculer efficacement des exposants de matrices ...

Définition 20 (état persistant/transitoire)
On dit qu'un état i est **persistant** si, partant de l'état i, on retourne en i avec une probabilité 1. Dans le cas contraire, on dit que i est **transitoire**.

Notations :

- $P_i(A) = P(A / X_0 = i)$.
- la probabilité de se retrouver un jour dans l'état j sachant que $X_0 = i$ est : $P_i\Big(\bigcup_{n=0}^{+\infty} X_n = j\Big) = f_{i,j}$.

Alors on a : $f_{i,j} = \sum_{n=0}^{+\infty} f_{i,j}^{(n)}$ où $f_{i,j}^{(n)}$ désigne la probabilité du premier passage en l'état j à l'instant n, sachant que $X_0 = i$: $f_{i,j}^{(n)} = P_i\big(X_1 \neq j, X_2 \neq j, \ldots, X_{n-1} \neq j, X_n = j\big)$.

Par définition, i est persistant ssi $f_{i,i} = 1$, i transitoire si $f_{i,i} < 1$.

Proposition 12

$$P_i\big(X_n = j \text{ pour une infinité de valeurs de } n\,\big) = \begin{cases} 0 & \text{si } j \text{ transitoire} \\ f_{i,j} & \text{si } j \text{ persistant} \end{cases}$$

On a donc en particulier :

PROPOSITION 13

$$P_i\big(X_n = i \textit{ pour une infinité de valeurs de } n\ \big) = \begin{cases} 0 & \textit{si } i \textit{ transitoire} \\ 1 & \textit{si } i \textit{ persistant} \end{cases}$$

THÉORÈME 9

$$i \textit{ transitoire } \Leftrightarrow \sum_{n=0}^{+\infty} p_{i,i}^{(n)} \textit{ converge .}$$

EXEMPLE 3 *Une marche au hasard sur $\mathbb{Z}^k$ où on a une probabilité $\dfrac{1}{2k}$ d'aller sur chaque voisin. Alors si $k \leq 2$, tous les états sont persistants, et si $k \geq 3$, tous les états sont transitoires. Pour démontrer ce résultat, on utilise le critère précédent.*

DÉFINITION 21 (CHAÎNE IRRÉDUCTIBLE)
Une chaîne est dite **irréductible** si $\forall i, j, f_{i,j} > 0$.
Autrement dit, $\forall n \geq 1, p_{i,j}^{(n)} > 0$.

Dans ce cadre, on obtient l'important théorème suivant :

THÉORÈME 10 *Pour une chaîne de Markov irréductible, l'une des 2 propriétés suivantes est vraie :*

1. *Tous les états sont transitoires, $\forall i, j, P_i\big(X_n = j \textit{ infiniment souvent}\ \big) = 0$ et $\sum_{n\geq 1} p_{i,j}^{(n)}$ converge. (cela n'est possible que si S est infini).*

2. *Tous les états sont persistants, $\forall i, j, P_i\big(X_n = j \textit{ infiniment souvent}\ \big) = 1$ et $\sum_{n\geq 1} p_{i,j}^{(n)}$ diverge.*

Dans le premier cas, on dit que la chaîne est **transitoire**, dans le second qu'elle est **persistante**.

DÉFINITION 22 (DISTRIBUTIONS STATIONNAIRES)
Une **distribution stationnaire** est une distribution de probabilités π sur l'ensemble des états telle que $P\pi = \pi$.

REMARQUE 11 De l'intérêt de savoir calculer des vecteurs propres !

Méthode matricielle d'analyse

Provisoirement, on suppose S fini. Le but est de trouver π tel que $P\pi = \pi$, autrement dit, π est un vecteur propre associé à la valeur propre 1. A quelle condition 1 est-elle valeur propre ? En fait, 1 est toujours valeur propre pour les matrices considérées (1 est valeur propre de ${}^tP$). On peut aussi montrer que 1 est la valeur propre de plus grand module .

DÉFINITION 23
Une matrice A est dite **irréductible** si par permutation de lignes et de colonnes elle peut se mettre sous la forme :

$$\begin{array}{cc} \begin{array}{ccc} I & & J \end{array} & \\ \begin{pmatrix} * & \vdots & * \\ \cdots & \vdots & \cdots \\ 0 & \vdots & * \end{pmatrix} & \begin{array}{c} I \\ \\ J \end{array} \end{array}$$

avec $I \cup J = [\![1, n]\!] = S$ et les étoiles représentent n'importe quoi. !! par permutation de lignes et de colonnes on entend que si on permute deux lignes, on permute les deux colonnes correspondantes.

THÉORÈME 11 (PERRON-FRÖBENIUS) *Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice positive (au sens où toutes ses entrées sont positives ou nulles). Alors $\rho(A)$ est une valeur propre de A associée à un vecteur propre ≥ 0. De plus, si A est irréductible, $\rho(A)$ est une valeur propre simple et le vecteur propre associé est strictement positif.*

Conséquence pour les chaînes de Markov : on a toujours une distribution stationnaire. Si la chaîne de Markov est irréductible, cette distribution stationnaire est unique et $\forall j \in S, \pi_j > 0$ (car la matrice de transition est irréductible).

Convergence

On prend ici S fini ou dénombrable.

DÉFINITION 24
La période de S est $\text{pgcd}\,\{n \geq 1, p_{i,i} > 0\}$. Elle peut être infinie.

REMARQUE 12 $\left(\forall n \geq 1, p_{i,i}^{(n)} > 0 \Rightarrow t|n\right) \Leftrightarrow t$ divise la période de i.

LEMME 5 *Pour une chaîne de Markov irréductible, la période est la même pour tous les états. On l'appelle* **période de la chaîne**.

DÉFINITION 25
Une chaîne est dite **apériodique** si sa période est 1.

LEMME 6 *Soit une chaîne de Markov irréductible et apériodique. Alors pour tous i, j de S, $p_{i,j}^{(n)} > 0$ pour n suffisamment grand.*

THÉORÈME 12 (CONVERGENCE) *Supposons qu'il existe pour une chaîne de Markov irréductible et apériodique une distribution stationnaire π. Alors :*

- *π est unique ;*
- *la chaîne est persistante ;*

- $\forall j, \pi_j > 0$;
- $\lim\limits_{n\to+\infty} p_{i,j}^{(n)} = \pi_j$.

REMARQUE 13 Il se peut qu'une telle distribution stationnaire n'existe pas sous les hypothèses données, voir l'exemple de la file d'attente.

THÉORÈME 13 *On considère une chaîne de Markov finie, irréductible. Soit* $h_{i,j} = \sum_{t=1}^{+\infty} t.f_{i,j}^{(t)}$ *le temps moyen pour passer de* i *à* j. *Alors* $h_{i,j}$ *est toujous fini (et est même quadratique en la distance de* i *à* j.

Rappel : si il existe une distribution stationnaire π, alors $\lim\limits_{n\to+\infty} p_{i,j}^{(n)} = \pi_j$.

THÉORÈME 14 *Pour une chaîne de Markov apériodique et irréductible, sans distribution stationnaire,* $\forall i, j, \lim\limits_{n\to+\infty} p_{i,j}^{(n)} = 0$.

Rappel : le temps moyen de retour en j est $\mu_j = \sum_{n=1}^{+\infty} n.f_{j,j}^{(n)}$.

LEMME 7 *Soit* j *un état persistant. Supposons que* $\lim n{+}\infty p_{j,j}^{(n)} = u$ *(*u *est fini dans le cas irréductible+apériodique). Alors* $u > 0$ *ssi* μ_j *finie et dans ce cas,* $u = \dfrac{1}{\mu_j}$.

D'où le théorème important :

THÉORÈME 15 *Pour une chaîne de Markov irréductible et apériodique, il y a trois possibilités :*

- *La chaîne est* **transitoire**, $\forall i, j, \sum_n p_{i,j}^{(n)}$ *est finie.*
- *La chaîne est persistante et il n'y a pas de distribution stationnaire, alors* $\lim\limits_{n\to+\infty} p_{i,j} = 0$ *mais pas trop vite car* $\sum_n p_{i,j}^{(n)}$ *doit diverger. On a* $\mu_j = \infty$ *pour tout* j. *La chaîne est dite* **persistante nulle**.
- *La chaîne est persistante et il y a une distribution stationnaire* π *(démontrée persistante), alors* $\lim\limits_{n\to+\infty} p_{i,j} = \pi_j > 0$ *et* $\mu_j = \dfrac{1}{\pi_j}$. *La chaîne est dite* **persistante positive**.

REMARQUE 14 Les trois cas existent.