9 Chaînes de Markov dénombrables.

Nous nous intéressons maintenant au cas où la chaîne de Markov (X_n) est à valeurs dans un ensemble E non plus fini mais dénombrable.

Parmi les propriétés des chaînes de Markov finies, beaucoup seront préservées, mais deux choses fondamentales ont changé : l'existence d'une mesure de probabilité invariante n'est plus assurée pour les chaînes récurrentes irréductibles, et le statut des points récurents et transitoires n'est plus le même.

9.1 Récurrence et transience.

Dans tout ce qui suit, (X_n) sera une chaîne de Markov homogène, à valeurs dans l'ensemble dénombrable E (muni de la tribu $\mathcal{P}(E)$), et de noyau de transition $P = (P(x,y), (x,y) \in E \times E)$ qui représente la loi $\mathcal{L}(X_{n+1}/X_n)$, c'est à dire que $P(x,y) = \mathbf{P}(X_{n+1} = y/X_n = x)$. Comme d'habitude, $\mathcal{F}_n$ désigne la tribu $\sigma(X_0, \cdots, X_n)$, et nous supposerons le processus construit sur l'espace canonique $\hat{\Omega}$ des trajectoires à valeurs dans $E \cup \delta$, où δ est une poubelle ajoutée à l'espace E pour y jeter les trajectoires usagées.

P est une "matrice markovienne de taille infinie", c'est à dire que

1. $\forall (x,y) \in E \times E,\ P(x,y) \geq 0\,;$
2. $\forall x \in E,\ \sum_y P(x,y) = 1.$

Nous lui associons les opérateurs $f \mapsto P(f)$ et $\mu \mapsto \mu P$, définis le premier sur les fonctions f positives ou bornées, le second sur les mesures μ positives, par

$$P(f)(x) = \sum_y P(x,y)f(y);\ \mu P(y) = \sum_x \mu(x)P(x,y).$$

Par définition, nous supposerons que les mesures mettent toujours des masses finies sur tous les points (en d'autres termes, ce sont des mesures σ-finies).

Ainsi, si μ est la loi de X_n, μP est la loi de X_{n+1}, et μP^k est la loi de X_{n+k}. De même, si f est une fonction positive ou bornée de $E \mapsto \mathbb{R}$, alors $Pf(X_n) = \mathbf{E}(f(X_{n+1})/\mathcal{F}_n)$, et $P^k(f)(X_n) = \mathbf{E}(f(X_{n+k})/\mathcal{F}_n)$.

Commençons par étudier le problème des mesures invariantes.

Définition 9.1. *Comme plus haut, nous dirons qu'une mesure positive sur E*

est invariante si et seulement si $\mu P = \mu$. Remarquons que nous ne demanderons pas à la mesure μ d'être une probabilité.

Exemple. — Considérons sur $\mathbb{Z}$ de processus défini par $X_{n+1} = X_n + 1$. C'est un processus de Markov, bien qu'il soit déterministe. Les seules mesures invariantes pour ce processus sont les mesures proportionnelles à la mesure uniforme sur $\mathbb{Z}$.

Considérons le même processus, mais à valeurs dans $\mathbb{N}$, cette fois-ci : il n'admet plus de mesure invariante non nulle, car une telle mesure devrait mettre la masse infinie au point 0.

Nous pouvons comme dans le cas fini classifier les états en états récurrents et transitoires, à partir de la relation $x \ll y$ définie en 2. Comme nous allons donner ci-dessous une autre définition de la récurrence et de la transience, nous convenons d'appeler **point permanent** un point x tel que $x \ll y \Longrightarrow y \ll x$ et **points éphémères** les autres (ce que nous appelions points récurrents et points transitoires). Les **classes de communication** sont les classes d'équivalence pour la relation $x \ll y$ sur l'ensemble des points permanents (ce que nous appelions classes de récurrence).

Tout d'abord, il se peut qu'il n'existe aucun point permanent (prendre l'exemple précédent de la chaîne sur $\mathbb{Z}$), ou bien que la probabilité de revenir à un point permanent soit strictement inférieure à 1. Les points **récurrents** seront ceux tels que la probabilité d'y revenir est égale à 1 :

Définition 9.2. *Un point x est dit récurrent si $\mathbf{P}_x(T_x < \infty) = 1$, où $T_x = \inf\{n > 0 \mid X_n = x\}$.*

Si ce n'est pas le cas, on dira que le point x est transitoire. Un point récurrent est forcément permament. La chaîne sera dite **irréductible** si tous les points sont permaments et qu'il n'y a qu'une seule classe de communication (anciennement récurrente irréductible).

Pour y voir plus clair, entre points permanents et points récurrents, nous introduisons le noyau potentiel.

Définition 9.3. *Soit $P^n(x,y)$ le noyau P composé n fois avec lui-même. Le noyau potentiel est égal à*

$$U(x,y) = \sum_{n=0}^{\infty} P^n(x,y).$$

C'est une quantité finie ou infinie. C'est l'espérance du temps passé en y partant de x.

La probabilité que le temps d'atteinte de y soit fini est liée à la valeur de ce noyau potentiel :

Proposition 9.4. *Soit $T_x = \inf\{n > 0 |\ X_n = x\}$ le temps de retour en x.*

1. *Un point x est récurrent (c'est à dire $\mathbf{P}_x(T_x < \infty) = 1$) si et seulement si $U(x,x) = \infty$.*
2. *Soit $N^y = \sum_{n\geq 0} \mathbf{1}_{X_n=y}$ le temps total passé en y. Alors, si $x \neq y$,*
$$U(x,y) = \mathbf{E}_x(N^y) = \mathbf{P}_x(T_y < \infty)U(y,y).$$
3. *En particulier, $U(x,y) \leq U(y,y)$.*

Démonstration. — Remarquons que $U(x,x) = 1 + G(x)$, où la fonction G a été définie dans le corollaire 8.7. Rappelons qu'on a $\mathbf{P}_x(T_x < \infty) = \frac{G(x)}{1+G(x)}$, et donc que $\mathbf{P}_x(T_x < \infty) < 1$ si et seulement si $U(x,x) < \infty$.

On a d'autre part, si $X_0 \neq y$, $N^y = \mathbf{1}_{\{T_y<\infty\}}[\theta_{T_y} N^y]$. En appliquant la propriété de Markov forte, nous avons
$$U(x,y) = \mathbf{P}_x(T_y < \infty)U(y,y).$$

∎

Corollaire 9.5. *Si $U(x,x) < \infty$, alors $\forall y \in E$, $U(y,x) < \infty$. Par conséquent, si un point x est transitoire, la chaîne n'y passe qu'un nombre fini de fois, quel que soit son point de départ y.*

C'est immédiat d'après ce qui précède ; remarquons cependant qu'on peut avoir le temps d'atteinte de x presque sûrement fini partant de y, mais un temps de retour en x infini avec probabilité positive.

La propriété pour la fonction $U(x,x)$ d'être finie ou non est indépendante du point choisi dans la classe de communication de x.

Proposition 9.6.
1. *Si x est éphémère, $U(x,x) < \infty$ (x est transitoire).*
2. *Si x est permanent et que $U(x,x) = \infty$ (c'est à dire si x est récurrent), alors $U(y,z) = \infty$, pour tous les points y et z de la classe de communication de x.*

3. *En particulier, dans une classe de communication, tous les points sont soit tous récurrents, soit tous transitoires.* **La propriété d'être récurrent est une propriété de la classe de communication.**

   On dira donc qu'une classe de communication est récurrente ou transitoire.

4. *De plus, si x est récurrent, la chaîne partant de x passe une infinité de fois par tous les points de la classe de x.*

Dans le cas d'une chaîne irréductible, nous dirons que la chaîne est **récurrente** *si tous ses points sont récurrents, et* **transiente** *sinon. De même, une classe de communication sera dite* **récurrente** *ou* **transiente***.*

Démonstration. — Nous avons déjà vu que, si x est éphémère, $\mathbf{P}_x(T_x < \infty) < 1$, et donc que $U(x,x) < \infty$.

Soit x est permanent avec $U(x,x) = \infty$, et soient y et z deux points choisis dans la classe de x. Alors

$$U(y,z) = \sum_n P^n(y,z).$$

Mais, si y et z sont dans la classe de x, il existe des entiers k et k_1 tels que $P^k(y,x) > 0$ et $P^{k_1}(x,z) > 0$; or $P^{k_1+n+k}(y,z) \geq P^{k_1}(y,x)P^n(x,x)P^k(x,y)$, et la série qui définit $U(y,z)$ diverge dès que celle qui définit $U(x,x)$ diverge.

Si un point x est récurrent, presque sûrement, le temps de retour T_x au point x est fini. En appliquant la propriété de Markov au temps T_x, on voit que le second temps de retour est fini presque sûrement, et de même pour le k-ième temps de retour en x.

Si y est un autre point de la classe de x, appelons c la probabilité de ne pas visiter y avant le premier temps de retour T_x : $c = \mathbf{P}_x(T_x < T_y)$. Nous savons que $c < 1$ car la chaîne est irréductible.

Soit $T_x^{(k)} = \inf\{n > T_x^{(k-1)} \mid X_n = x\}$ le k^{ieme} retour en x. On a

$$\{T_y > T_x^{(k+1)}\} = \{T_y > T_x^{(k)}\} \cap \theta_{T_x^{(k)}}(\{T_y > T_x\}),$$

et en prenant l'espérance

$$\mathbf{P}(T_y > T_x^{k+1}) = c\mathbf{P}(T_y > T_x^k).$$

Donc, $\mathbf{P}(T_y > T_x^k) = c^k$ et $\mathbf{P}_x(T_y = \infty) = 0$.

Maintenant, sachant qu'on atteint y avec probabilité 1, et que, partant de y, on y revient presque sûrement une infinité de fois, on voit que, partant de x, on visite y une infinité de fois avec probabilité 1.

∎

Contrairement à ce qui se passe dans le cas fini, on peut avoir dans le cas irréductible des fonctions invariantes non constantes. Mais ce ne sera pas le cas si la chaîne est récurrente.

Corollaire 9.7. *Soit f une fonction positive telle que $P(f) = f$. Alors f est constante sur les classes de communication récurrentes. Ce résultat reste vrai si f est minorée, ou bien négative majorée, mais faux sans cette hypothèse. De plus, il est faux sur une classe de communication transitoire.*

Démonstration. — La suite $f(X_n)$ est une martingale positive. Pour la probabilité $\mathbf{P}_x$, elle est d'espérance $f(x)$, donc bornée dans L^1. Elle converge donc presque sûrement. Mais la suite X_n repasse une infinité de fois par tous les points de la classe de x, et donc $f(X_n)$ ne peut converger presque sûrement que si elle est constante sur cette classe.

On passe au cas de f minorée en ajoutant une constante à f, ce qui ne change pas son caractère invariant.

Nous verrons plus bas un contre exemple avec la marche aléatoire simple sur $\mathbb{Z}$, qui est récurrente et pour laquelle la fonction $f(x) = x$ est invariante.

Pour voir un contre exemple de fonction invariante bornée pour une chaîne irréductible transiente, considérons de cas $E = \mathbb{N}$, avec $P(n, n+1) = 1 - \alpha_n$, $P(n, 0) = \alpha_n$. Si on fait l'hypothèse que la série $\sum_n \alpha_n$ est convergente, et qu'on pose

$$\beta_0 = 1,\ \beta_n = \prod_{i=0}^{n-1} (1 - \alpha_i)^{-1}\ (n \geq 1), s_n = 1 - \sum_{i=0}^{n-1} \frac{\alpha_k}{\beta_k},$$

toutes ces quantités sont bornées d'après l'hypothèse de convergence, et la fonction $f(n) = \beta_n s_n$ est une fonction invariante non constante.

On voit que sur cet exemple, la chaîne est irréductible, et on en conclut qu'elle est transiente (mais on peut le voir de façon beaucoup plus simple en considérant l'événement de probabilité positive $\{\forall n, X_n = n\}$). ∎

Dans la suite de ce chapitre, nous supposerons que la chaîne est irréductible.

Nous pouvons avoir une chaîne irréductible avec $U(x,x) < \infty$ partout (c'est à dire transiente), comme le montre l'exemple suivant :

Exemple. — Soit d un entier supérieur ou égal à 1, et appelons e_i le point de $\mathbb{Z}^d$ dont toutes les composantes sont nulles sauf la i-ième qui vaut 1 (e_i est le i-ième vecteur de la base canonique de $\mathbb{R}^d$). Soit (Y_n) une suite de variables aléatoires indépendantes et de même moi, telle que, pour $i = 1, \cdots, d$, $\mathbf{P}(Y_n = e_i) = \mathbf{P}(Y_n = -e_i) = 1/(2d)$. La suite $X_n = Y_1 + \cdots + Y_n$ ($X_0 = 0$) s'appelle la **marche aléatoire simple** sur $\mathbb{Z}^d$. On vérifie immédiatement qu'elle est irréductible.

Proposition 9.8. *La marche aléatoire simple sur $\mathbb{Z}^d$ est récurrente pour $d = 1, 2$ et transiente si $d \geq 3$.*

De même, la marche aléatoire sur $\mathbb{Z}^d$ formée de d copies indépendantes de la marche aléatoire sur $\mathbb{Z}$ est récurrente pour $d = 1, 2$, transiente pour $d \geq 3$.

Démonstration. —

Puisque qu'il n'y a qu'une seule classe de communication, il suffit d'étudier $P^n(0,0)$. C'est la probabilité d'être de retour en 0 au temps n. On voit immédiatement que $P^n(0,0) = 0$ si n est impair. Si $n = 2p$, c'est la probabilité que, pour chaque $i = 1, \cdots, d$, il y ait autant de k pour lesquels $Y_k = e_i$ que de k pour lesquels $Y_k = -e_i$. Cette probabilité n'est pas si facile à estimer : appelons la $a_p^{(d)}$.

Commençons par le cas $d = 1$. Alors $a_p^{(1)} = C_{2p}^p \simeq 1/\sqrt{\pi p}$ par la formule de Stirling : la série est donc divergente : le chaîne est récurrente.

Nous pouvons ensuite facilement passer au cas de d copies indépendantes. De façon générale, si l'on a une chaîne X_n de noyau P sur E et qu'on considère d copies indépendantes de la chaîne X_n dans E^d, c'est une chaîne de Markov de noyau

$$P((x_1, \cdots, x_d), (y_1, \cdots, y_d)) = P(x_1, y_1) \dots P(x_d, y_d).$$

On a pour tout n

$$P^n((x_1, \cdots, x_d), (y_1, \cdots, y_d)) = P^n(x_1, y_1) \dots P^n(x_d, y_d).$$

(C'est immédiat par la propriété d'indépendance, mais se vérifie aussi facilement par le calcul.

Un point de la diagonale $\{(x, x, \dots, x),\ x \in E\}$ est donc récurrent si et seulement si $\sum_n P^n(x,x)^d = \infty$.

Dans l'exemple de $\mathbb{Z}^d$, si on considère d copies indépendantes de la marche aléatoire simple dans $\mathbb{Z}$, on voit que le point $(0,\ldots,0)$ est récurrent pour $d=1,2$ et transitoire si $d \geq 3$, puisque $\mathbf{P}_0(X_n=0) \sim c/\sqrt{n}$.

Mais bien que ce processus ressemble à la marche aléatoire simple dans $\mathbb{Z}^d$, il ne faut pas le confondre avec celle-ci.

Revenons au cas de la marche aléatoire sur $\mathbb{Z}^d$

Pour $d=2$, si l'on prend deux copies indépendantes de la marche aléatoire sur $\mathbb{Z}$, en tournant la figure de $\pi/4$, on voit qu'on obtient une marche aléatoire sur $\mathbb{Z}^2$. Plus exactement, $(Y^1_n+Y^2_n)/2, (Y^1_n-Y^2_n)/2$ est une marche aléatoire sur $\mathbb{Z}^2$ (ici, X^1 et X^2 désignent les deux coordonnées du vecteur X). Elle est donc encore récurrente.

Le calcul explicite dans le cas $d \geq 3$ est beaucoup plus difficile, car la marche aléatoire sur $\mathbb{Z}^d$ et le produit de d marches aléatoires sur $\mathbb{Z}$ ont des comportements différents (la première se déplace sur $2d$ points voisins, la seconde sur 2^d points voisins). Nous pouvons nous en sortir par l'astuce suivante : ni $n \in \mathbb{Z}^d$, et en désignant par $x.y$ le produit scalaire euclidien de $\mathbb{R}^d$, nous avons

$$\frac{1}{(2\pi)^d}\int_{(-\pi,\pi)^d} \exp(in.\theta)d\theta = \mathbf{1}_{\{n=0\}}.$$

En intégrant par rapport à la mesure $\mathbf{P}$, on obtient, pour toute variable aléatoire X à valeurs dans $\mathbb{Z}^d$, et en notant $\phi_X(\theta)=\mathbf{E}[\exp(iX.\theta)]$ (la transformée de Fourier de la loi de X),

$$\mathbf{P}(X=0) = \frac{1}{(2\pi)^d}\int_{(-\pi,\pi)^d} \mathbf{E}(\exp(iX.\theta))d\theta = \frac{1}{(2\pi)^d}\int_{(-\pi,\pi)^d} \phi_X(\theta)d\theta.$$

Or, en posant $f(\theta)=\mathbf{E}[\exp(iY_1.\theta)] = \frac{1}{d}(\sum_{i=1}^d \cos(\theta_i))$, on a $\phi_{X_n}(\theta)=f^n(\theta)$, et donc

$$\mathbf{P}(X_{2p}=0) = \frac{1}{(2\pi)^d}\int_{(-\pi,\pi)^d} f^{2p}(\theta)d\theta.$$

Finalement, en sommant, ce qui est licite puisque tout est positif,

$$\sum_p \mathbf{P}(X_{2p}=0) = \frac{1}{(2\pi)^d}\int_{(-\pi,\pi)^d} \frac{1}{1-f^2(\theta)}d\theta.$$

Il reste à travailler un peu pour montrer que cette intégrale est convergente si $d \geq 3$, ce qu'on peut faire en observant que les seuls endroits où cette intégrale pose problème est en $(0,\cdots,0)$, $(\pi,\cdots,\pi)$, et $(-\pi,\cdots,-\pi)$. Il faut alors faire un développement en coordonnées polaires au voisinage de chacun de ces points : ceci est laissé à titre d'exercice.

∎

Remarque. — Dans l'exemple ci-dessus, la mesure uniforme sur $\mathbb{Z}^d$ est invariante, dans tous les cas. Mais il ne peut y avoir de probabilité invariante si la chaîne est transiente, comme on le verra un peu plus bas (proposition 9.13).

Dans le cas de la dimension 1, la fonction $f(p) = p$ est invariante ($P(f) = f$), mais non constante. C'est un exemple de fonction invariante non constante pour une chaîne récurrente. (Elle n'est bien sûr pas positive, en vertu du corollaire 9.7.)

Comme dans le cas fini, lorsque la chaîne est irréductible et que μ est une mesure invariante non nulle, elle charge tous les points. Si $\mu(x)$ est nulle en un point, en utilisant l'équation $\mu(x) = \sum_y \mu(y)P(y,x)$, alors $P(x,y) > 0 \Longrightarrow \mu(y) = 0$. Puis on conclut en utilisant l'irréductibilité. (Nous avions déjà fait ce raisonnement dans le cas des chaînes finies.)

Nous pouvons alors introduire le noyau adjoint $Q(x,y) = \mu(y)P(y,x)\frac{1}{\mu(x)}$. On a

Proposition 9.9. *Si μ est une mesure invariante et que P est irréductible, alors le noyau adjoint Q est un noyau markovien irréductible. Il est récurrent si et seulement si P l'est.*

De plus, comme dans le cas fini,

$$\frac{d\nu P}{d\mu} = Q(\frac{d\nu}{d\mu}).$$

Démonstration. — Le fait que le noyau adjoint soit markovien et irréductible a déjà été vu dans le cas des chaînes finies (il n'y a rien à changer à la démonstration). Pour voir qu'il est récurrent en même temps que P, il suffit d'observer que

$$Q^n(x,y) = \mu(y)P^n(y,x)\frac{1}{\mu(x)}.$$

L'équation des densités est la même que dans le cas fini.

∎

Proposition 9.10. *Si la chaîne est irréductible et récurrente, une mesure invariante est unique à une constante multiplicative près et charge tous les* *points. En particulier, s'il existe une probabilité invariante, celle ci est unique et toute autre mesure invariante est de masse totale finie et est proportionnelle à cette probabilité.*

Remarque. — Nous verrons plus bas qu'une chaîne récurrente irréductible admet toujours une mesure invariante.

Démonstration. —

On peut donc supposer que μ ne s'annule pas. Il nous reste à voir que si μ et μ_1 sont deux mesures invariantes, μ_1 est proportionnelle à μ. Introduisons le noyau adjoint Q :

$$Q(y,x) = \mu(x)P(x,y)\frac{1}{\mu(y)}.$$

Soit alors f la densité de μ_1 par rapport à μ. Nous savons que $Q(f) = f$. La fonction f est donc constante sur les classes de communication de Q, qui sont les mêmes que celles de P.

∎

Remarque. — Il n'est pas vrai que la mesure invariante soit unique sans l'hypothèse d'être récurrente, pour une chaîne irréductible. Par exemple, sur $\mathbb{Z}$, la chaîne de probabilités de transition $P(x,x+1) = \alpha$, $P(x,x-1) = 1-\alpha$ admet comme mesure invariante la mesure uniforme ainsi que la mesure

$$\mu(n) = (\frac{\alpha}{1-\alpha})^n.$$

Si la chaîne est récurrente, elle admet une mesure invariante, qui sera unique d'après ce qu'on vient de voir. Cette mesure est caractérisée à une constante près par les espérances du nombre de passages en y avant un retour en x, comme dans le cas fini.

Théorème 9.11. *Soit (X_n) une chaîne récurrente irréductible ; alors, elle admet une mesure invariante. Plus précisément : si x et y sont deux points de E et que N_x^y désigne le nombre de passages en y avant le premier retour en x, nous avons*

1. *Si μ est une mesure invariante, alors $\mathbf{E}_x(N_x^y) = \frac{\mu(y)}{\mu(x)}$.*
2. *Réciproquement, la variable N_x^y est intégrable pour la probabilité $\mathbf{P}_x$ et si on définit $\mu_x(y) = \mathbf{E}_x(N_x^y)$, alors cette expression définit une mesure invariante.*

Démonstration. — Commençons par le premier point, qui est le plus facile. Nous allons invoquer un argument de retournement du temps. Nous savons déjà que la mesure μ est strictement positive partout. Considérons le noyau adjoint $Q(y,x) = \mu(x)P(x,y)/\mu(y)$, dont on sait déjà que c'est un noyau markovien, irréductible et récurrent.

Nous savons que, si X_0 a la loi μ, la suite $(X_n, \cdots, X_0)$ a la même loi que celle d'une chaîne de Markov de mesure initiale μ et de noyau Q. Plus précisément, puisque qu'ici la mesure μ n'est pas a priori une probabilité, nous savons que

$$\mu(x_0)\mathbf{P}[(X_0, \cdots, X_n) = (x_0, \cdots, x_n)] = \mu(x_n)\mathbf{P}[(Y_0, \cdots, Y_n) = (x_n, \cdots, x_0)],$$

où Y est une chaîne de Markov de noyau Q. (C'est une évidence si on écrit ces probabilités avec les noyaux $P(x,y)$ d'un côté, et avec les noyaux $Q(x,y)$ de l'autre).

Cette formule se généralise aux fonctions bornées $F(x_0, \ldots, x_n)$:

$$\begin{aligned}\mu(x_0)\mathbf{E}_{x_0}[\mathbf{1}_{\{X_0=x_0\}}F(X_0, \ldots, X_n)\mathbf{1}_{\{X_n=x_n\}}] &= \\ \mu(x_n)\hat{\mathbf{E}}_{x_n}[\mathbf{1}_{\{Y_0=x_n\}}F(Y_n, \ldots, Y_0)\mathbf{1}_{\{Y_n=x_0\}}]\end{aligned}$$

Notons $T_x = \inf\{n > 0 | \ X_n = x\}$ le temps de retour en x et

$$\hat{T}_x = \inf\{n > 0 | \ Y_n = x\}.$$

Notons ausi $\hat{\mathbf{E}}_y(.)$ l'espérance pour la loi de la chaîne de Markov de matrice Q partant de y. On a

$$\mathbf{E}_x[\mathbf{1}_{\{X_n=y\}}\mathbf{1}_{\{n\leq T_x\}}] = \frac{\mu(y)}{\mu(x)}\hat{\mathbf{E}}_y[\mathbf{1}_{\{n\leq \hat{T}_x\}}\mathbf{1}_{\{Y_n=x\}}] = \frac{\mu(y)}{\mu(x)}\hat{E}_y(\mathbf{1}_{\hat{T}_x=n}).$$

En effet,

$$\mathbf{1}_{X_0=x}\mathbf{1}_{\{n\leq T_x}\mathbf{1}_{X_n=y} = I_{X_0=x}\mathbf{1}_{\{X_1\neq x\}}\ldots\mathbf{1}_{\{X_{n-1}\neq x\}}\mathbf{1}_{X_n=y},$$

et il suffit d'appliquer la formule précédente.

En sommant de 1 à l'infini, on obtient

$$\mathbf{E}_x(N_x^y) = \frac{\mu(y)}{\mu(x)}\hat{\mathbf{P}}_y(\hat{T}_x < \infty) = \frac{\mu(y)}{\mu(x)},$$

la dernière identité venant du fait que le noyau Q étant récurrent, $\hat{T}_x$ est fini presque sûrement.

Passons à la réciproque. Pour cela, fixons x et appelons plus haut N_x^y le nombre de passages en y avant le premier retour en x.

Notre premier travail va être de montrer que cette quantité est intégrable. Fixons les points x et y, appelons A_p la quantité $A_p = \mathbf{P}_x(N_x^y \geq p)$ et c la quantité $\mathbf{P}_y(N_y^x \geq 1)$, c'est à dire la probabilité d'aller de y à x sans repasser par y. L'hypothèse d'irréductibilité nous dit que $c > 0$. Si $c = 1$, alors N_x^y est bornée par 1 (car, au premier passage en y, le processus revient avec probabilité 1 en x sans repasser par y), et donc il n'y a rien à démontrer. Nous pouvons donc supposer $0 < c < 1$.

Pour $p \geq 1$, appelons $T_y^{(p)}$ le p-ième passage en y : dire que $N_x^y = p$ revient à dire que $T_y^{(p)} < T_x^{(1)}$ et qu'après $T_y^{(p)}$, il n'y a pas de retour en y avant le premier passage en x : nous pouvons donc décomposer l'événement $\{N_x^y = p\}$:

$$\{N_x^y = p\} = \{N_x^y \geq p\} \cap \theta_{T_y^{(p)}}\{N_y^x \geq 1\}.$$

En appliquant la propriété de Markov forte au temps $T_y^{(p)}$, on a

$$\mathbf{P}_x(N_x^y = p) = \mathbf{P}_x(N_x^y \geq p)\mathbf{P}_y(N_y^x \geq 1),$$

ou encore que

$$A_p - A_{p+1} = cA_p.$$

Ceci nous donne

$$A_p = (1-c)^{p-1}A_1,$$

et donc N_x^y suit une loi géométrique (au moins à partir de $p = 1$), de raison $1 - c$. C'est donc une variable intégrable.

Remarquons que cette démonstration nous donne une information plus précise. Puisque $\mathbf{E}(N) = \sum_n \mathbf{P}(N \geq n)$, nous obtenons

$$\mathbf{E}_x(N_x^y) = \frac{\mathbf{P}_x(T_y < T_x)}{\mathbf{P}_y(T_x < T_y)}.$$

Appelons $\mu_x(y)$ la quantité $\mathbf{E}_x(N_x^y)$, et montrons que c'est une mesure invariante. Nous allons proposer deux méthodes.

Commençons par la première : rappelons que $T_y^{(p)}$ désigne le p-ième passage en y, c'est à dire $T_y^{(0)} = 0$ et $T_y^{(p)} = \inf\{n > T_y^{(p-1)} | \; X_n = y\}$, alors

$$N_x^y = \sum_{p \geq 1} \mathbf{1}_{\{T_y^{(p)} < T_x^{(1)}\}}, \quad \text{pour } y \neq x,$$

et $N_x^x = 1$.

Mettons nous sous la loi $\mathbf{P}_x$ (c'est à dire que nous supposerons que $X_0 = x$). Si $y \neq x$, à chaque passage en y de X_n avant $T_x^{(1)}$, X_{n-1} était en un point z, ce qui revient à dire que $(n-1)$ correspond à l'un des temps $T_z^{(p)}$, pour un $z \neq x$ (sauf si $n = 1$ auquel cas $X_0 = x$).

Pour $y = x$, nous pouvons écrire la même décomposition en regardant la valeur prise par X au temps $T_x^{(1)} - 1$.

Au bout du compte, nous avons, pour tous les y de E (y compris le point x)

$$\begin{aligned} N_x^y &= \mathbf{1}_{\{X_1=y\}} + \sum_{z\neq x}\sum_{p\geq 1} \mathbf{1}_{\{T_z^{(p)}<T_x^{(1)}\}}\mathbf{1}_{\{X_{T_z^{(p)}+1}=y\}} \\ &= \mathbf{1}_{\{X_1=y\}} + \sum_{z\neq x}\sum_{p\geq 1} \mathbf{1}_{\{T_z^{(p)}<T_x^{(1)}\}}\theta_{T_z^{(p)}}(\mathbf{1}_{\{X_1=y\}}). \end{aligned}$$

Appliquons la propriété de Markov au temps $T_z^{(p)}$ dans ce qui précède, en remarquant que $X_{T_z^{(p)}} = z$ sur l'ensemble $\{T_z^{(p)} < T_x^{(1)}\}$ ($T_x^{(1)} < \infty$ car le point est récurrent). Il vient

$$\begin{aligned} \mu_x(y) &= P(x,y) + \sum_{z\neq x}\sum_{p\geq 1} \mathbf{E}_x[\mathbf{1}_{\{T_z^{(p)}<T_x^{(1)}\}}\mathbf{E}_z(\mathbf{1}_{\{X_1=y\}})] \\ &= P(x,y) + \sum_{z\neq x}\mu_x(z)P(z,y). \end{aligned}$$

Finalement, on a

$$\mu_x(y) = \sum_z \mu_x(z)P(z,x),$$

et la mesure $\mu_x(y)$ est invariante.

Une autre façon, plus élégante mais moins intuitive, de démontrer que la quantité définie par $\mu_x(y) = \mathbf{E}_x(N_x^y)$ est invariante est la suivante : en appelant comme toujours $T_x^{(1)}$ le premier temps de retour en x, pour une fonction f à support borné, on voit que

$$\int f(y)\mu_x(dy) = \mathbf{E}_x[\sum_{i=0}^{T_x-1} f(X_i)] = \mathbf{E}_x[\sum_{i=1}^{T_x} f(X_i)],$$

la dernière inégalité venant que, dans les deux expressions, on ne compte $f(x)$ qu'une seule fois, au début ou à la fin de la somme.

Montrons que $\int P(f)\mu_x(dy) = \int f\mu_x(dy)$. On a

$$\begin{aligned}
\int P(f)\mu_x(dy) &= \mathbf{E}_x[\sum_{i\geq 0} \mathbf{1}_{\{i<T_x\}} P(f)(X_i)] = \mathbf{E}_x[\sum_{i\geq 0} \mathbf{1}_{\{i<T_x\}} \mathbf{E}_{X_i}(f(X_1))] \\
&= \mathbf{E}_x[\sum_{i\geq 0} \mathbf{1}_{\{i<T_x\}} f(X_{i+1})] = \mathbf{E}_x[\sum_{p=1}^{T_x} f(X_p)] \\
&= \int f\mu_x(dy)
\end{aligned}$$

∎

Nous avons vu que la récurrence implique l'existence d'une mesure invariante, unique à une constante près ; on a aussi vu des exemples de chaînes transientes ayant des mesures invariantes (la marche aléatoire sur Z^d, $d \geq 3$), ou ayant plusieurs mesures invariantes distinctes. Mais il existe aussi des chaînes irréductibles sans mesure invariante (elles sont évidemment transientes).

Exemple. — Considérons $E = \mathbb{N}$, avec le noyau $P(n,0) = \alpha_n$, $P(n, n+1) = 1 - \alpha_n$. Alors, il existe une mesure invariante si et seulement si la série $\sum_n \alpha_n$ diverge.

En effet, si on appelle $R_n = \Pi_0^{n-1}(1-\alpha_k)$, $(R_0 = 1)$, alors une mesure invariante satisfait $\mu(n) = R_n\mu(0)$ et

$$\mu(0) = \mu(0)\sum_n R_n\alpha_n = \mu(0)\sum_n (R_n - R_{n+1}),$$

ce qui est impossible si la suite R_n converge vers une valeur non nulle.

Remarquons que la condition est exactement équivalente à ce que le temps de retour en 0 soit presque sûrement fini, car

$$\mathbf{P}_0(T_0 = n) = \mathbf{P}_0(X_1 = 1, \ldots, X_{n-1} = n-1, X_n = 0) = \alpha_{n-1}R_{n-1} = R_{n-1} - R_n,$$

$\mathbf{P}_0(T_0 < \infty) = 1 - \lim_n R_n$.

Remarque. — Nous avons vu qu'une chaine récurrente irréductible repasse une infinité de fois par tous les points. De même, une chaîne irréductible transiente ne repasse (presque sûrement) qu'au plus un nombre fini de fois par chaque point. En effet, si $T_y^{(k)}$ désigne le $k$ième passage en y en partant de x, nous avons, par la propriété de Markov au tmps $T_y^{(1)}$,

$$\mathbf{P}_x(T_y^{(k)} < \infty) = \mathbf{P}_x(T_y^{(1)} < \infty)\mathbf{P}_y(T_y^{(1)} < \infty)^{k-1},$$

ce qui est le terme général d'une série convergente. Il suffit ensuite d'appliquer le lemme de Borel-Cantelli.

Un corollaire est le suivant :

Corollaire 9.12. *Si X_n est une chaîne irréductible transiente, elle converge presque sûrement vers l'infini, quelle que soit la loi du point de départ.*

Plus précisément, quelle que soit la loi du point de départ, il existe un ensemble Ω_0 de probabilité 1 tel que, pour tout $\omega \in \Omega_0$, tout A de cardinal fini, il existe un entier n_A tel que, si $n \geq n_A$, $X_n \notin A$.

En d'autres termes, une chaîne est transiente si et seulement si elle converge vers l'infini avec probabilité 1. Pour que ce soit le cas, il suffit qu'elle converge vers l'infini avec une probabilité non nulle.

Démonstration. — Comme toutes les propriétés qu'on veut démontrer presque sûrement, on peut se ramenerà la loi $\mathbf{P}_x$. Ensuite, il suffit de la faire pour $A = \{y\}$, puisqu'ensuite on peut choisir $n_A = \max_{y\in A} n_{\{y\}}$.

Enfin, il suffit alors d'appliquer la remarque qui précède, puisqu'après un certain temps, on ne retourne pas en $\{y\}$.

∎

Remarque. — Puisqu'un ensemble dénombrable non fini est en bijection avec $\mathbb{N}$, on peut toujours se ramener au cas $E = \mathbb{N}$, auquel cas la convergence précédente à l'infini est bien la notion usuelle de convergence vers l'infini dans $\mathbb{N}$.

Nous avons vu que la récurrence implique l'existence d'une mesure invariante. En retour, l'existence d'une **probabilité** invariante assure la récurence :

Théorème 9.13. *Soit X_n une chaîne irréductible. Les propositions suivantes sont équivalentes :*

1. *La chaîne admet une probabilité invariante.*
2. *Pour tout x de E, $\mathbf{E}_x(T_x^{(1)}) < \infty$.*
3. *Pour un point x de E, $\mathbf{E}_x(T_x^{(1)}) < \infty$.*

Dans ces conditions, l'unique probabilité invariante vaut

$$\mu(x) = \frac{1}{\mathbf{E}_x(T_x^{(1)})}.$$

En particulier, une chaîne irréductible admettant une probabilité invariante est récurrente.

Définition 9.14. *On dit que la chaîne est* **récurrente positive** *si elle admet une probabilité invariante. Si elle est récurrente mais que les mesures invariantes non nulles sont infinies, on dit qu'elle est récurrente nulle. (Rappelons que les mesures invariantes sont toutes proportionnelles dans le cas récurrent)*

Démonstration. — Soit μ une probabilité invariante. Montrons d'abord que la chaîne est récurrente : pour tout n, on a donc $\mu(y) = \sum_x \mu(x)P^n(x,y)$. Supposons que la chaîne soit transiente. On a donc $P^n(x,y) \to 0$ $(n \to \infty)$, puisque la série $\sum_n P^n(x,y)$ est convergente. Par convergence dominée, nous en déduisons que $\mu(y) = 0$, ce qui est absurde.

Appelons T_x le premier temps de retour en x. La mesure

$$\mu_x(y) = \mathbf{E}_x(N_x^y) = \mathbf{E}_x(\sum_1^{T_x} \mathbf{1}_{\{X_n=y\}})$$

est invariante, comme nous l'avons vu plus haut. Si la chaîne admet une probabilité invariante, puisqu'elle est récurrente, toutes les mesures invariantes sont bornées.

Mais $\sum_{y\in E} N_x^y = T_x$, et donc

$$\mathbf{E}_x(T_x) = \sum_{y\in E} \mu_x(y) < \infty.$$

D'autre part, si, pour un x de E, nous avons $\mathbf{E}_x(T_x) < \infty$, alors la mesure μ_x est de masse totale finie. Donc la chaîne admet une probabilité invariante.

Nous avons donc démontré l'équivalence des trois propositions.

Pour obtenir la formule donnant $\mu(x)$, il suffit de remarquer que $\mu_x(x) = 1$ et que $\mu_x(E) = \mathbf{E}_x(T_x)$. L'unique probabilité invariante étant donnée par

$$\mu(y) = \frac{\mu_x(y)}{\mu_x(E)},$$

on obtient le résultat en prenant $y = x$.

∎

Nous avons vu que le temps de retour en x est intégrable pour une chaîne récurrente positive. Il en va de même des temps d'atteinte de tous les points.

Théorème 9.15. *Soit X_n une chaîne récurrente positive, alors, pour tous les couples de points (x, y), $\mathbf{E}_x(T_y) < \infty$.*

Démonstration. —

Choisissons $y \neq x$ et appelons $T_x^{(k)}$ les temps de retour successifs en x, avec $T_x^{(0)} = 0$. Nous écrivons

$$\mathbf{E}_x(T_y) = \sum_{k=1}^{\infty} \mathbf{E}_x(T_y \mathbf{1}_{\{T_x^{(k)} \leq T_y < T_x^{(k+1)}\}}).$$

En prenant l'espérance conditionnelle au temps $T_x^{(k)}$, nous voyons que

$$\mathbf{E}_x(T_y \mathbf{1}_{\{T_x^{(k)} \leq T_y < T_x^{(k+1)}\}}) = \mathbf{E}_x(T_y \mathbf{1}_{T_y < T_x^{(1)}})\mathbf{P}_x(T_y \geq T_x^{(k)}) \leq \mathbf{E}_x(T_x)\mathbf{P}_x(T_y \geq T_x^{(k)}).$$

En utilisant une fois de plus la propriété de Markov, nous obtenons

$$\mathbf{P}_x(T_y \geq T_x^{(k)}) = \mathbf{P}_x(T_y \geq T_x^{(1)})^k = c^k,$$

où $c < 1$. On en déduit que

$$\mathbf{E}_x(T_Y) \leq \frac{\mathbf{E}_x(T_x)}{1 - c}.$$

∎

Remarque. — On voit qu'en fait

$$\mathbf{E}_x(T_y) = \frac{\mathbf{E}_x(T_y \mathbf{1}_{T_y \leq T_x}}{\mathbf{P}_x(T_y) \leq T_x},$$

où T_x est le premier temps de retour en x.

9.2 Convergence vers la mesure invariante.

Avant d'aborder le problème de la convergence de $P^n(x, y)$ vers la mesure invariante, nous énonçons un lemme qui donne une condition suffisante pour que l'image d'une chaîne de Markov est une chaîne de Markov.

Lemme 9.16. *(Critère de Dynkin.) Soit X_n une chaîne de Markov de noyau P sur E et f une application de E dans F. Alors, si, pour tout $z \in F$ $P(x, f^{-1}(z))$ ne dépend que de $f(x) = t$, et vaut $Q(t, z)$, alors $f(X_n)$ est une chaîne de Markov de noyau Q.*

Démonstration. — L'hypothèse dit que la probabilité d'aller de x dans l'ensemble $f^{-1}(z)$ est constante sur l'ensemble des points x qui ont même image par f. Le lecteur peut se représenter une fonction f de E dans F comme une partition de E, qui représente les ensembles $f^{-1}(t)$. Si on appelle B_t les éléments de cette partition, le critère s'énonce sous la forme $\forall t,\ P(x, B_t)$ est constant sur les classes $B_{t'}$. Une autre façon de reformuler l'hypothèse est de demander que, pour toute fonction bornée $h : F \mapsto \mathbb{R}$, $P(h(f))$ est une fonction de f, qui n'est alors rien d'autre que $Q(h)(f)$. (Une telle fonction est une combinaison linéaire d'indicatrice de $f^{-1}(z)$.) L'opérateur Q est alors automatiquement markovien : ce sera le noyau de la chaîne image.

L'hypothèse étant traduite sous la forme $P(h(f)) = Q(h)(f)$, la démonstration est presque immédiate. Si on appelle $\mathcal{G}_n$ la tribu $\sigma(f(X_0), \cdots, f(X_n))$ et $\mathcal{F}_n = \sigma(X_0, \cdots, X_n)$, alors

$$
\begin{aligned}
\mathbf{E}(h(f(X_{n+1}))/\mathcal{G}_n) &= \mathbf{E}(h(f(X_{n+1}))/F_n/\mathcal{G}_n)) \\
&= \mathbf{E}(Q(h)(f)(X_n)/G_n \\
&= Q(h)(f(X_n)).
\end{aligned}
$$

Ceci à la fois montre le caractère markovien de $f(X_n)$ et donne son noyau.

∎

Pour les chaînes irréductibles, on peut définir la notion de période, de la même manière que pour les chaînes finies.

Définition 9.17. *Soit X_n une chaîne irréductible. La période d'un point est le PGCD de l'ensemble des n tels que $P^n(x, x) > 0$. Une chaîne est apériodique si la période de tous ses points est égale à 1.*

Comme dans le cas fini, tous les points ont même période. De même, si un point est apériodique, il existe un entier $n_0(x)$ tel que si $n \geq n_0$, alors $P^n(x, x) > 0$. Mais la différence avec le cas fini est qu'on ne peut pas s'assurer qu'on peut choisir le même entier n_0 pour tous les points x de E.

Nous avons encore un théorème de convergence vers la probabilité invariante :

Théorème 9.18. *Soit X_n une chaîne irréductible, récurrente positive et apériodique. Si μ désigne sa probabilité invariante, alors,*

$$\forall x \in E, \lim_{n\to\infty} P^n(x,y) = \mu(y).$$

On peut même écrire un peu mieux : nous allons voir qu'il y a convergence des mesures $\mu_0 P^n$ vers la mesure invariante μ en variation totale.

Si μ et ν sont deux probabilités, appelons

$$\|\mu - \nu\| = \sup_{|f|\leq 1} |\int f d\mu - \int f d\nu|.$$

Cette distance s'appelle la **distance en variation totale** de μ à ν. (Ce n'est pas la définition classique de la distance en variation totale, qui est $\sup_A |\mu(A) - \nu(A)|$, mais elle en diffère pare un facteur 2.)

Sur un ensemble dénombrable, en prenat $f = \mathbf{1}_{\mu(x)\geq\nu(x)} - \mathbf{1}_{\mu(x)<\nu(x)}$, nous voyons que

$$\|\mu - \nu\| = \sum_x |\mu(x) - \nu(x)|.$$

Si μ_0 est une probabilité sur E et μ est la probabilité invariante, nous allons montrer que

$$\lim_{n\to\infty} \|\mu_0 P^n - \mu\| = 0.$$

Plus généralement, nous avons

Théorème 9.19. *Soit X_n une chaîne irréductible, récurrente apériodique, de noyau P.*

Supposons ou bien que la chaîne soit récurrente positive, ou bien que, pour au moins un point x de E, $\sum_n P^n(x,x)^2 = \infty$.

Alors, si μ_1 et μ_2 sont deux probabilités sur E, nous avons

$$\lim_{n\to\infty} \|\mu_1 P^n - \mu_2 P^n\| = 0.$$

Remarque. — Le premier théorème s'en déduit en appliquant le second avec la probabilité invariante μ à la place de μ_1 et δ_x à la place de μ_2. Comme conséquence de ce théorème, nous voyons que si la chaîne est récurrente positive apériodique, alors $\lim_n P^n(x,x) = \mu(x) > 0$, et donc dans ce cas

$\sum_n P^n(x,x)^2 = \infty$, puisque le terme général de la série ne converge pas vers 0. Le premier cas n'est donc qu'un cas particulier du second.

Démonstration. — Nous allons démontrer ce dernier théorème. Nous nous servirons du second cas plus bas pour traiter le cas des mesures invariantes infinies.

Nous ne pouvons pas utiliser la méthode utilisée pour les chaînes finies, car l'espace des mesures de probabilité sur E n'est pas compact (si l'on munit E de la topologie discrète, les parties compactes sont les parties finies).

Nous allons utiliser une méthode astucieuse, **le couplage**.

Pour commencer, rappelons que si (X_n) et (X'_n) sont deux chaînes de Markov indépendantes dur E de même noyau P, alors la chaîne $Z_n = (X_n, X'_n)$ est aussi une chaîne de Markov sur $E \times E$, de noyau Q donné par

$$Q((x,x'),(y,y')) = P(x,y)P(x',y'),$$

qu'on note $Q = P \otimes P$, car pour tout couple (x,x'), la mesure $(y,y') \mapsto Q((x,x'),(y,y'))$ est le produit $P(x,\bullet) \otimes P(x',\bullet)$ des mesures $y \mapsto P(x,y)$ et $y' \mapsto P(x',y')$.

Appelons alors T le premier temps où les deux composantes sont égales :

$$T = \inf\{n|\ X_n = X'_n\}.$$

Ce temps peut être éventuellement infini, mais c'est un temps d'arrêt. T est appelé le temps de couplage de la chaîne ; c'est aussi le temps d'atteinte de la diagonale dans $E \times E$.

Nous pouvons alors définir un nouveau processus Y_n sur $E \times E$ de la façon suivante

$$Y_n = Z_n \mathbf{1}_{\{n \leq T-\}} + (X_n, X_n)\mathbf{1}_{\{n>T\}}.$$

C'est le processus couplé : après le temps de couplage, on impose aux deux coordonnées de rester égales.

Alors, ce processus est aussi une chaîne de Markov, et son noyau de transition est

C'est le noyau de transition

$$Q_1((x,x'),(y,y')) = \begin{cases} P(x,y)P(x',y') & \text{si } x \neq x' \\ 0 & \text{si } x = x' \text{ et } y \neq y' \\ P(x,y) & \text{si } x = x' \text{ et } y = y' \end{cases}$$

Pour le voir, on remarque que la tribu $\bar{\mathcal{F}}_n = \sigma(Y_0, \ldots, Y_n)$ est incluse dans la tribu $\hat{\mathcal{F}}_n = \sigma(Z_0, \ldots, Z_n)$, et que le temps d'atteinte de la diagonale pour Y_n ou pour Z_n est le même, de telle façon que T est aussi un temps d'arrêt de la filtration $\bar{\mathcal{F}}_n$.

Ensuite, on remarque que

$$\mathbf{E}(G(Y_{n+1})/\bar{\mathcal{F}}_n) = \mathbf{E}(G(Y_{n+1})/\hat{\mathcal{F}}_n/\bar{\mathcal{F}}_n),$$

puis on calcule $\mathbf{E}(G(Y_{n+1})/\hat{\mathcal{F}}_n)$:

$$\mathbf{E}(G(Z_{n+1})/\hat{\mathcal{F}}_n)\mathbf{1}_{\{T\leq n\}} + \mathbf{E}(G(X_{n+1}, X_{n+1})/\hat{\mathcal{F}}_n)\mathbf{1}_{T\geq n+1}.$$

On applique la propriété de Markov de (Z_n) sur le premier terme et celle de (X_n) sur le second, car, par indépendance de (X_n) et (X'_n),

$$\mathbf{E}(h(X_{n+1})/\hat{\mathcal{F}}_n) = \mathbf{E}(h(X_{n+1})/\sigma(X_0, \ldots, X_n)).$$

Le calcul du noyau Q_1 s'obtient immédiatement d'après cette formule.

Si μ est une mesure invariante pour P, on voit aisément que Q_1 admet comme mesure invariante $\nu(x, x') = \mathbf{1}_{\{x=x'\}}\mu(x)$. En effet

$$\sum_{x,x'} \nu(x, x')Q((x, x'), (y, y')) = \sum_x \mu(x)P(x, y)\mathbf{1}_{\{y=y'\}} = \mu(y)\mathbf{1}_{\{y=y'\}}.$$

De même, la mesure $\mu \otimes \mu$ est une mesure invariante pour $Q = P \otimes P$.

Nous voulons tout d'abord montrer que la chaîne Z_n est irréductible sur $E \times E$: c'est ici qu'intervient l'hypothèse d'apériodicité.

En effet, pout tout quadruplet $((x, x'), (y, y'))$, nous pouvons trouver un indice n_0 tel que, pour tout $n \geq n_0$, on ait $P^n(x, y) >$ et $P^n(x', y') > 0$. On voit d'autre part immédiatement que $Q^n((x, x'), (y, y')) = P^n(x, y)P^n(x', y')$.

Nous allons montrer que, sous les conditions données, le temps T est fini presque sûrement : dans le cas où la chaîne X est récurrente positive, avec comme probabilité invariante μ, nous voyons que Z_n admet comme probabilité invariante $\mu \otimes \mu$, et donc qu'elle est récurrente positive, puisqu'elle est irréductible.

Sinon, choisissons un point de z de $E \times E$ de la forme (x, x) (un point de la diagonale) : par hypothèse,

$$\sum_n Q^n(z, z) = \sum_n P^n(x, x)^2 = \infty,$$

et la chaîne est bien récurrente.

On en déduit que le temps d'atteinte d'un point (x,x) est fini presque sûrement, quel que soit le point de départ, et donc il en va de même du temps d'atteinte T de la diagonale, partant de n'importe quel point (x,x'). Il en va de même pour toutes les lois initiales ν sur $E\otimes E$, puisque $\mathbf{P}_\nu(A)=\int \mathbf{P}_z(A)\nu(dz)$. En particulier, pour toute loi initiale ν, on a

$$\mathbf{P}_\nu(T>n)\to 0 \ (n\to\infty).$$

Remarquons que pour la chaîne (Y_n), les points hors diagonale sont éphémères, puisqu'on ne revient jamais d'un point de la diagonale vers un point hors diagonale.

Les deux projections X_n et $\hat{X}_n$ de Y_n sur E sont aussi des chaînes de Markov de noyau P. C'est évident pour la première (sans utiliser le critère, puisque c'est X_n), et par symétrie de la loi de Y_n pour la seconde.

Appelons T le temps d'atteinte de la diagonale. Choisissons comme mesure initiale de la chaîne couplée $\mu_1\otimes\mu_2$, c'est à dire que la première composante a pour loi μ_1 et que la seconde μ_2, les deux étant indépendantes. Alors, X_n de loi μ_1P^n et $\hat{X}_n$ est de loi μ_2P^n, puisque ce sont toutes les deux des chaînes de Markov de matrice P.

Nous avons

$$\mathbf{P}(X_n\neq\hat{X}_n)=\mathbf{P}(T>n)\to 0(n\to\infty).$$

Soit f une fonction bornée par 1. Posons $\mu_n^1=\mu_1P^n$ et $\mu_n^2=\mu_2P^n$. Nous pouvons écrire

$$|\int fd\mu_n^1-\int fd\mu_n^2|=|\mathbf{E}(f(X_n))-\mathbf{E}(f(\hat{X}_n))|\leq 2\mathbf{P}(X_n\neq\hat{X}_n),$$

la dernière inégalité venant de ce que f est bornée par 1.

Nous avons donc montré la propriété.

∎

Le résultat équivalent pour les chaînes récurrentes nulles est un peu plus compliqué (il y a alors convergence vers 0). Nous le traiterons dans la prochaîne section, comme un résultat de renouvellement.

9.3 Renouvellement

L'étude du renouvellement sur $\mathbb{N}$ est un cas particulier de l'étude des sommes de variables aléatoires indépendantes : on se donne une suite (T_i) de variables aléatoires à valeurs dans $\mathbb{N}^* = \{n \in \mathbb{N},\ n > 0\}$, indépendantes et de même lois, et on considère leur somme $S_n = T_1 + \cdots + T_n$: c'est une suite strictement croissante d'entiers, et, fixant un point n, on considère l'événement $A_n = \cup_k \{S_k = n\}$. Cette réunion est disjointe. Imaginer que les T_i sont les instants successifs qui séparent les passages de l'autobus, et A_n est la présence ou non de d'un autobus à l'instant n.

On appelle ce système un système de renouvellement, et la loi de T est la loi du renouvellement. On dit qu'il est **apériodique** si la loi de T_1 n'est pas contenue dans un sous-groupe de $\mathbb{N}$ pour l'addition. Ce qui revient à dire que, si on appelle M le support de T, c'est à dire l'ensemble des points n tels que $\mathbf{P}(T = n) > 0$, alors le PGCD de M est 1.

Puisque la suite S_n ne prend ses valeurs que dans le sous-groupe additif engendré par les valeurs prises par T, le PGCD de M est aussi le PGCD des entiers tels que $P(A_n) \neq 0$.

Lorsqu'on a une chaîne de Markov récurrente irréductible, si on fixe un point x, et qu'on considère les intervalles entre les temps de passage en x, on a un système de renouvellement. Il est apériodique si la période de ce point est 1, c'est à dire si la chaîne elle même est apériodique.

En fait, tout système de renouvellement se représente par des passages en un point d'une chaîne de Markov :

Proposition 9.20. *Considérons un système de renouvellement de loi μ et soit $r_n = \mathbf{P}(T > n)$. Posons $\rho_n = r_{n+1}/r_n$. Considérons la chaîne de Markov sur $\mathbb{N}$ de transition $P(n, n+1) = \rho_n$, $P(n, 0) = 1 - \rho_n$.*

Cette chaîne est récurrente, la loi du temps de retour en 0 est μ. Elle admet comme unique mesure invariante (à une constante près) $\mu(n) = r_n$. Elle est récurrente positive si et seulement si la série $\sum r_n$ est convergente. Dans ce cas, $\sum_n r_n = \mathbf{E}(T)$.

Nous avions déjà considéré ce modèle un peu plus haut en remarquant qu'il n'avait pas de mesure invariante si r_n ne convergeait pas vers 0.

Ce n'est pas la seule chaîne qu'on peut construire qui représente ce système de renouvellement. (Exercice : en trouver une autre !)

Démonstration. — Il n'y a pas grand chose à montrer. Partant de 0, la chaîne ne peut que croître par pas de 1 jusqu'à retomber en 0. On peut aller de tout les points à 0 et aller de 0 à tous les points : la chaîne est irréductible. De plus,

$$\mathbf{P}_0(T_0 > n) = \mathbf{P}_0(X_n = n) = \rho_0\rho_1 \cdots \rho_{n-1} = r_n = \mathbf{P}(T > n).$$

Donc la loi de T_0 sous $\mathbf{P}_0$ est bien μ. Puisque T est fini presque sûrement, il en a de même de T_0, et donc la chaîne est récurrente.

Vérifions que r_n est une mesure invariante : cela s'écrit

$$\forall n,\ r_n P(n, n+1) = r_{n+1},\ r_0 = \sum_{n\geq 0} r_n(1-\rho_n),$$

ce qui est immédiat si on remarque que $r_n(1-\rho_n) = r_n - r_{n+1}$.

Enfin, nous savons déjà que le temps de retour est intégrable si et seulement si les mesures invariantes non nulles sont finies, ce qu'on retrouve ici sous la forme d'un résultat bien connu :

$$\sum_{n\geq 0} \mathbf{P}(T > n) = \mathbf{E}(\sum_{n=0}^{\infty} \mathbf{1}_{\{0\leq n\leq T-1\}}) = \mathbf{E}(T).$$

∎

Nous pouvons alors énoncer le théorème principal du renouvellement :

Théorème 9.21. *Considérons un système de renouvellement de loi μ.*

1. *Si μ est intégrable (c'est à dire si le temps T est intégrable) et si le système est apériodique, alors*

$$\lim_{n\to\infty} \mathbf{P}(A_n) = 1/\mathbf{E}(T).$$

2. *Si μ n'est pas intégrable,*

$$\lim_{n\to\infty} \mathbf{P}(A_n) = 0.$$

En d'autre termes, lorsque $n \to \infty$, la probabilité que n soit l'un des temps de passage S_k converge vers l'inverse de la longueur moyenne consécutive entre deux intervalles, ce qui est intuitivement raisonnable puisqu'il y a une valeur occupée par intervalle.

Démonstration. — Commençons par le cas récurrent positif apériodique : c'est alors une application simple de ce que nous venons de démontrer sur les chaînes de Markov récurrente positives apériodiques :

$$\mathbf{P}(A_n) = \mathbf{P}_0(X_n = 0) = P^n(0,0) \to \nu(0),$$

où ν est la probabilité invariante. Maintenant, si μ est la mesure invariante décrite dans la proposition précédente ($\mu(0) = 1$, $\mu(n) = r_n$), alors $\nu(n) = r_n/R$, où $R = \sum_n r_n = \mathbf{E}(T)$, et $\nu(0) = 1/R$. C'est ce que nous voulions démontrer.

Le cas où $\mathbf{E}(T) = \infty$ est un peu plus délicat. Cela correspond à une chaîne de Markov récurrente nulle. Commençons par le cas apériodique.

Si la série $\sum_n \mathbf{P}(A_n)^2$ converge, alors $\mathbf{P}(A_n) \to 0$, et il n'y a rien à démontrer. Sinon, nous sommes dans le second cas du théorème 9.19. Considérons alors la mesure μ_k qui est la mesure μ restreinte à $[0, \cdots, k]$, c'est à dire qu'on remplace μ par 0 sur $\{k+1, k+2, \ldots\}$.

Appelons a_k sa masse totale, c'est à dire $a_k = \sum_0^k r_i$, qui converge vers l'infini par hypothèse. Nous savons que, pour tout k,

$$|\frac{\mu_k}{a_k} P^n(0) - P^n(0,0)| \to 0, \ (n \to \infty).$$

Par ailleurs, il est facile de calculer $\mu_k P$: nous avons

$$\mu_k P = \mu_{k+1} - r_{k+1}\delta_0,$$

c'est à dire que les masses sont inchangées sur tous les points, sauf le $(k+1)$-ième dont la masse passe de 0 à r_{k+1} et le premier dont la masse passe de 1 à $1 - r_{k+1}$. (La masse totale reste inchangée, bien sûr.)

On a donc $\mu_k P \leq \mu_{k+1}$, et par suite $\mu_k P^n \leq \mu_{k+n}$, d'où l'on déduit que

$$\mu_k P^n(0) \leq \mu_{k+n}(0) = r_0.$$

D'où $\frac{\mu_k}{a_k} P^n(0) \leq 1/a_k$.

Finalement

$$\limsup_n P^n(0,0) \leq \limsup_n \frac{\mu_k}{a_k} P^n(0) \leq \frac{r_0}{a_k},$$

et il nous reste à faire converger k vers l'infini pour conclure.

Passer du cas apériodique au cas général est facile. Il suffit de remarquer que, si π est la période, en dehors des points de la forme $k\pi$, $\mathbf{P}(A_n) = 0$, et sur l'ensemble des multiples de π, on se ramène au cas précédent en divisant tout par π.

∎

Comme conséquence, nous obtenons le résultat équivalent pour les chaînes de Markov :

Théorème 9.22. *Soit X_n une chaîne de Markov irréductible, récurrente nulle. Alors, pour tout x,*

$$\lim_{n\to\infty} P^n(x,x) = 0.$$

Démonstration. — Il n'y a rien à démontrer : il suffit d'appliquer le résultat précédent au système de renouvellement formé par le temps de retour en x. Remarquer que la démonstration que nous avons faite en choisissant une chaîne de Markov particulière qui représente ce système de renouvellement ne s'applique pas immédiatement. Nous avons en fait construit une autre chaîne de Markov ayant le même système de renouvellement au point x, mais pour lequel on puisse approximer la mesure invariante par des mesures bornées d'une façon contrôlée. ∎

Remarques

1. On voit comme conséquence qu'on a aussi $P^n(x,y) \to 0$ dans les mêmes conditions. En utilisant le premier temps de passage T_y de la chaîne en y, et en décomposant

$$\{X_n = y\} = (\{X_n = y\} \cap \{n \le T_y\}) \cup (\{X_n = y\} \cap \{T_y < n\}),$$

on peut utiliser la propriété de Markov au temps T_y pour le second terme et obtenir une majoration

$$\mathbf{P}_x(X_n = y) \le \mathbf{P}_x(n \le T_y) + \sum_{k\ge 1} \mathbf{P}_x(T_y = k) P^{n-k}(y,y).$$

On conclut en appliquant le théorème de convergence dominée.

2. Contrairement à ce qui se passe pour des chaînes récurrente positives, il n'est pas vrai que les chaînes récurrentes nulles vérifient $\mathbf{E}_x(T_y) = \infty$, si $x \neq y$. Il suffit pour s'en convaincre de considérer la chaîne sur $\mathbb{Z}$ de noyau $P(n,n+1) = P(n,n-1) = 1/2$, partout sauf en $n = 0$ ou en $n = 1$ où l'on pose $P(0,1) = 1$ et $P(1,2) = P(1,-1) = 1/2$. On

voit immédiatement que cette chaîne est récurrente (exercice !), qu'elle admet comme mesure invariante (unique !) la mesure $\mu(n) = 1$, si $n \neq 0$, et $\mu(0) = 1/2$; elle est donc récurrente nulle. Mais, partant de 0, le temps d'atteinte du point 1 est égal à 1.

3. Nous avons donné un critère pour que le produit de deux chaînes indépendantes irréductibles apériodiques soit récurrent. On voit aisément que c'est une CNS. On peut construire (à l'aide de marches aléatoires) deux chaînes indépendantes sur $\mathbb{Z}^2$, récurrentes irréductibles, dont le produit est transient. (Exercice !)

9.4 Critères de récurrence.

Comme nous venons de le voir, pour une chaîne irréductible apériodique, il n'existe que deux types de comportement de $P^n(x, y)$ lorsque $n \to \infty$. Soit X_n est récurrente nulle ou bien transiente, et dans ce cas, même si elle n'est pas apériodique, $P^n(x, y)$ converge vers 0 lorsque $n \to \infty$, soit elle est récurrente positive, et dans ce cas $P^n(x, y)$ converge vers $\mu(y)$, où μ est l'unique probabilité invariante. Ceci nous donne un premier critère simple de récurrence.

Proposition 9.23. *(***Critère de Doeblin***) Soit X_n une chaîne récurrente. Supposons qu'il existe un entier n, un ensemble fini A et un $\varepsilon > 0$ tels que, pour tout $x \in E$*

$$P^n(x, A) = \sum_{y \in A} P^n(x, y) \geq \varepsilon.$$

Alors, X_n est récurrente positive.

Démonstration. — En effet, sous les hypothèses précédentes, pour $m \geq n$, on peut écrire

$$P^m(x, A) = \sum_{z \in E} P^{m-n}(x, z) P^n(z, A) \geq \varepsilon.$$

En conséquence, $\liminf_m P^m(x, A) \geq \varepsilon$, ce qui est impossible si, pour tout $y \in A$, $\lim_m P^m(x, y) = 0$. ∎

D'autre part, la récurrence est caratérisée par la finitude des temps de retour aux points, et la récurrence positive par leur intégrabilité.

Il en va de même des temps de retour dans les ensembles finis.

Proposition 9.24. *Soit (X_n) une chaîne irréductible et A une partie finie de E. Soit $T_A = \inf\{n \geq 1 \mid X_n \in A\}$. Alors,*

1. *Si,* $\forall x \in A$*,* $\mathbf{P}_x(T_A < \infty) = 1$*, la chaîne est récurrente.*
2. *Si,* $\forall x \notin A$*,* $\mathbf{P}_x(T_A < \infty) = 1$*, la chaîne est récurrente.*
3. *Si,* $\forall x \in A$*,* $\mathbf{E}_x(T_A) < \infty$*, alors la chaîne est récurrente positive.*

Démonstration. — Remarquons que le temps T_A est l'infimum de tous les temps de retour T_x aux points de A. Il est a priori plus petit que chacun d'entre eux.

Prenons la première propriété. Appelons $T_A^{(n)}$ les temps de retour successifs en A. Avec nos hypothèses, en appliquant la propriété de Markov aux temps $T_A^{(n)}$ successifs, nous voyons que tous ces temps sont presque sûrement finis. Si la chaîne était transiente, elle ne passerait qu'un nombre fini de fois en A (proposition 9.12), ce qui donne une contradiction.

Pour la seconde assertion, nous appelons appliquer le critère précédent. Soit donc x un point de A, supposnons que $X_0 = x$, et soit $S = \inf\{n \geq 0, X_n \notin A\}$.

Si la chaîne était transiente, on aurait $S < \infty$ presque sûrement, car la chaîne ne passe qu'un nombre fini de fois dans chaque point de A. Si la chaîne est récurrente, alors S est fini presque sûrement car il est inférieur au temps d'atteinte de n'importe quel point de A, qui est fini. Donc, S est toujours presque sûrement fini.

Ecrivons alors

$$\mathbf{P}_x(T_A < \infty) = \mathbf{P}_x(S < T_A < \infty) + \mathbf{P}_x(S > T_A).$$

Sur $\{S < T_A\}$, on a $T_A = \theta_S(T_A)$, et, par la propriété de Markov,

$$\mathbf{P}_x(S < T_A < \infty) = \mathbf{E}_x[\mathbf{1}_{S<T_A}\mathbf{E}_{X_S}(\mathbf{1}_{T_A<\infty})].$$

Comme $X_S \notin A$, cette quantité est en fait égale d'après notre hypothèse à $\mathbf{P}_x(S < T_A)$, et on obtient ainsi

$$\mathbf{P}_x(T_A < \infty) = 1.$$

(Remarquons que l'événement $\{S < T_A\}$ ne peut se produire que si au temps 1, $X_1 \notin A$, si bien qu'en fait notre raisonnement revient à considérer les deux cas : soit $X_1 \in A$, soit $X_1 \notin A$).

Pour la dernière assertion, supposons que T_A soit intégrable, et appelons $Q(x, y) = \mathbf{P}_x(X_{T_A} = y)$, pour $(x, y) \in A \times A$. Alors, en appliquant la propriété de Markov aux temps $T_A^{(n)}$, on voit que $Y_n = X_{T_A^{(n)}}$ est une chaîne de Markov sur l'ensemble fini A de matrice Q. Si X est irréductible, alors il en va de même de Y. Soit S_x le temps de retour en x de cette chaîne, avec point de départ $x \in A$. On voit que le temps de retour T_x de la chaîne X n'est rien d'autre que $T_A^{(S_x)}$. On voit aisément par la propriété de Markov que, si $m = \sup_{x \in A} E_x(T_A)$, alors, pour tout $x \in A$, $\mathbf{E}_x(T_A^{(n)}) \leq mn$, et même que, si S est un temps d'arrêt de la filtration associée à Y, alors $\mathbf{E}_x(T_A^{(S)}) \leq m\mathbf{E}_x(S)$.

Comme S_x est intégrable (c'est le temps de retour en x de la chaîne Y, qui est récurrente irréductible sur un ensemble fini, donc récurrente positive), on a donc

$$\mathbf{E}_x(T_x) \leq m\mathbf{E}_x(S_x) < \infty.$$

∎

Une autre famille de critères est donnée par l'existence de fonctions de Lyapounov : ce sont des fonctions ayant des propriétés particulières dont l'existence assure la transience ou la récurrence.

Proposition 9.25. *Soit (X_n) une chaîne irréductible. Alors, (X_n) est récurrente si et seulement si il existe une fonction positive f qui converge vers l'infini à l'infini et un ensemble fini A tels que $P(f) \leq f$ sur A^c.*

Démonstration. — Commençons par la condition suffisante. Partons d'un point $x \notin A$, et soit T le temps d'entrée dans A : $T = \inf\{n \mid X_n \in A\}$. Soit $M_n = f(X_{n \wedge T})$. Alors, M_n est une sur-martingale positive. En effet,

$$\begin{aligned}
\mathbf{E}_x(M_{n+1}/\mathcal{F}_n) &= \mathbf{E}_x(f(X_{n+1})/\mathcal{F}_n)\mathbf{1}_{n<T} + f(X_T)\mathbf{1}_{T<n+1} \\
&= Pf(X_n)\mathbf{1}_{n<T} + f(X_T)\mathbf{1}_{T<n+1} \\
&\leq f(X_n)\mathbf{1}_{n<T} + f(X_T)\mathbf{1}_{T<n+1} \\
&= M_n,
\end{aligned}$$

l'inégalité venant de ce que, sur $\{n < T\}$, $X_n \in A^c$, et donc $P(f)(X_n) \leq f(X_n)$.

Or, une sur-martingale positive est toujours bornée dans L^1, et converge donc presque sûrement. Sur $\{T = \infty\}$, si X_n était transiente, on aurait $M_n \to \infty$, puisque $X_n \to \infty$. D'où une contradiction.

Pour la réciproque, nous pouvons sans perdre de généralité nous restreindre au cas où $E = \mathbb{N}$. Posons $T_n = \inf\{p \geq 0 \mid X_p \geq n\}$, et

$$S_0 = \inf\{n \geq 0 \mid X_n = 0\}.$$

Soit $g_n(i) = \mathbf{P}_i(T_n < S_0)$. Puisque la chaîne est supposée récurrente, on sait que, pour tout i, $g_n(i) \to 0$ $(n \to \infty)$. (La suite des T_n est croissante). D'autre part, $g_n(i) = 1$ si $i \geq n$, $g_n(0) = 0$, et $1 \leq g_n \leq 1$ partout. Si $0 < X_0 < n$, on

$$\{T_n < S_0\} = [\{0 < X_1 < n\} \cap \theta(\{T_n < S_0\}] \cup \{X_1 \geq n\},$$

d'où

$$g_n(i) = \mathbf{E}_x(\mathbf{1}_{\{0<X_1<n\}} g_n(X_1) + \mathbf{1}_{\{X_1 \geq n\}}) = P g_n(i).$$

Si $i \geq n$, on a $g_n(i) = 1 \geq P g_n(i)$.

On a donc bien $P g_n(i) \leq g_n(i)$ sur $i > 0$.

Puisque $g_n(i)$ converge vers 0 lorsque $n \to \infty$, on peut choisir une suite n_k telle que $g_{n_k}(i) \leq 2^{-k}$ pour tout $i \leq k$. Posons alors $f(i) = \sum_{k=1}^{\infty} g_{n_k}(i)$. La série converge pour tout i, par construction des entiers n_k.

De plus, nous savons que $g_n(p) \geq 1$ si $p \geq n$. Donc, $g(i) \geq k$ si $i \geq \sup_{p=1}^{k} n_p$. D'où l'on voit que $g(i)$ converge vers l'infini avec i.

D'autre part, on a bien, pour $i \neq 0$,

$$P(g)(i) = \sum_k P(g_{n_k})(i) \leq \sum_k g_{n_k}(i) = g(i),$$

ce qui achève la démonstration. ∎

Un corollaire particulièrement utile de ce résulat est le suivant.

Corollaire 9.26. *Soient (X_n) et (Y_n) deux chaînes irréductibles, de noyau P et Q respectivement. S'il existe un ensemble fini A tel que, pour tout $i \notin A$, et pour tout j, $P(i,j) = Q(i,j)$, alors elles sont ensemble récurrentes ou transientes.*

Démonstration. — Supposons (X_n) récurrente, et soit f une fonction (de Lyapounov) qui lui est associée par la proposition 9.25.

Alors, $Q(f) = P(f) \leq f$ sur A^c, et donc les conditions de la proposition 9.25 s'appliquent à (Y_n), et celle-ci est aussi récurrente. ∎

Une autre caractérisation est la suivante.

Proposition 9.27. *Une chaîne de Markov irréductible est transiente si et seulement si il existe une fonction positive f et un ensemble fini A tel que $P(f) \leq f$ sur A^c et*

$$\inf_{A^c} f < \inf_A f.$$

En particulier, il suffit que f soit strictement positive sur A et converge vers 0 à l'infini pour vérifier la dernière condition.

Démonstration. — Commençons par montrer que la condition est suffisante. Choisissons $x \in A^c$ tel que $f(x) < \inf_A f$. Nous choisissons $X_0 = x$ et prenons $T = \inf\{n \mid X_n \in A\}$. Si X est récurrente, alors T est fini presque sûrement. Comme plus haut, la suite $M_n = f(X_{n\wedge T})$ est une surmartingale positive, et converge donc. Sa limite est $f(X_T)$. Mais, pour une surmartingale positive de limite M_∞, on a $\mathbf{E}(M_0) \geq \mathbf{E}(M_\infty)$, en utilisant le lemme de Fatou et la propriété de surmartingale. On obtient donc

$$f(x) \geq \mathbf{E}_x(f(X_T)).$$

Mais $X_T \in A$, et $f(X_T) \geq \inf_A f$. Finalement, nous obtenons $f(x) \geq \inf_A f$, ce qui contredit le choix initial de x.

Pour la réciproque, ramenons nous comme dans la proposition 9.25 au cas où $E = \mathbb{N}$, et posons $S_0 = \inf\{n \geq 0 \mid X_n = 0\}$, puis $f(x) = \mathbf{P}_x(S_0 < \infty)$. On a $f(0) = 1$ et, pour $x \neq 0$,

$$\{S_0 < \infty\} = \{X_1 = 0\} \cup [\{X_1 \neq 0\} \cap \theta(S_0 < \infty\}).$$

On en tire aisément que $P(f)(x) = f(x)$ si $x \neq 0$.

De plus, puisque X est transiente, il existe un $x \neq 0$ tel que $f(x) < 1$. Ceci suffit à conclure, avec $A = \{0\}$. ∎

Enfin, on peut avoir aussi un critère équivalent our la récurrence positive :

Proposition 9.28. *Une chaîne irréductible est récurrente positive si et seulement si il existe une fonction positive f, un ensemble fini A et un $\varepsilon > 0$ tels que $P(f) \leq f - \varepsilon$ sur A^c, et $P(f) < \infty$ sur A.*

Démonstration. — Comme plus haut nous commençons par la condition suffisante. On introduit le temps $T = \inf\{n \geq 1 \mid X_n \in A\}$, et le processus $S_n = f(X_{n\wedge T})$.

On a par hypothèse, pour $x \notin A$,

$$\mathbf{E}_x(S_{n+1}/\mathcal{F}_n) \leq S_n - \varepsilon \mathbf{1}_{\{n<T\}}.$$

(Ceci ne s'applique qu'aux $x \notin A$ car on n'a pas la garantie que $X_n \notin A$ sur $\{n < T\}$ sinon.)

Ceci nous montre immédiatement que le processus $S_n + \varepsilon n \wedge T$ est une surmartingale positive, et

$$f(x) \geq \mathbf{E}_x(S_{n\wedge T}) + \varepsilon \mathbf{E}_x(n \wedge T) \geq \varepsilon \mathbf{E}_x(n \wedge T).$$

En passant à la limite, on obtient

$$E_x(T) \leq f(x)/\varepsilon.$$

Pour $x \in A$, en appliquant la propriété de Markov au temps 1, on obtient

$$\begin{aligned} \mathbf{E}_x(T_A) &= \sum_{y \in A} P(x,y) + \sum_{y \notin A} P(x,y)[\mathbf{E}_y(T_A)+1] \\ &\leq 1 + \sum_{y \notin A} P(x,y) f(y)/\varepsilon \\ &\leq 1 + P(f)(x)/\varepsilon. \end{aligned}$$

Donc, le temps de retour dans A est intégrable pour tous les points de départ. D'après la proposition 9.24, la chaîne est récurrente positive.

Passons à la réciproque. On se ramène par commodité au cas où $E = \mathbb{N}$. On pose $A = \{0\}$ et l'on pose $f(0) = 0$, et $f(x) = \mathbf{E}_x(T_0)$ pour $x \neq 0$. On a, pour $x \neq 0$,

$$T_0 = \mathbf{1}_{\{X_1=0\}} + \mathbf{1}_{\{X_1>0\}}[1+\theta(T_0)] = 1 + \mathbf{1}_{\{X_1>0\}}\theta(T_0).$$

On en tire, pour $x \neq 0$, en intégrant

$$f(x) = 1 + \mathbf{E}_x(f(X_1)\mathbf{1}_{\{X_1>0\}}) = 1 + P(f)(x).$$

Pour $x = 0$, on écrit, toujours en utilisant la même décomposition au temps 1,

$$\begin{aligned} P(f)(0) &= \sum_{y>0} P(0,y)\mathbf{E}_y(T_0) \\ &= \mathbf{E}_0[\mathbf{1}_{\{X_1>0\}}\mathbf{E}_{X_1}(T_0)] \\ &= \mathbf{E}_0[\mathbf{1}_{\{X_1>0\}}\theta(T_0)] \\ &\leq \mathbf{E}_0[\mathbf{1}_{\{X_1>0\}}(1+\theta(T_0))] \\ &= \mathbf{E}_0[\mathbf{1}_{\{X_1>0\}}T_0] \leq \mathbf{E}_0(T_0) \end{aligned}$$

Cette dernière quantité est finie par l'hypothèse de récurrence positive. ∎

On pourra trouver d'autres conditions nécessaires et suffisantes de ce type dans le livre [2].

9.5 Théorèmes ergodiques.

Dans le chapitre consacré aux chaînes de Markov finies, nous avions utilisé le théorème ergodique pour établir la propriété $\mathbf{E}_x(N_x^y) = \mu(y)/\mu(x)$.

Nous allons faire l'inverse ici.

Théorème 9.29. *Soit X_n une chaîne irréductible récurrente et soit μ mesure invariante μ. Alors, si f et g sont deux fonctions intégrables pour μ, et si $\int g\, d\mu \neq 0$, alors, quelle que soit la mesure initiale μ_0,*

$$\lim_{n\to\infty} \frac{\sum_1^n f(X_p)}{\sum_1^n g(X_p)} = \frac{\int f\, d\mu}{\int g\, d\mu}, \text{ presque sûrement.}$$

Remarque. — Nous savons déjà que la mesure μ existe et est unique à une constante près sous ces conditions.

De plus, si la chaîne est récurrente positive, que μ est la probabilité invariante et que nous choisissons $g = 1$, nous retrouvons le théorème classique

$$\frac{1}{n}\sum_1^n f(X_p) \to \int f\, d\mu.$$

Démonstration. — Il suffit comme d'habitude de démontrer la propriété lorsque la mesure initiale est δ_x, c'est à dire pour $\mathbf{P}_x$. La démonstration se fait en décomposant les trajectoires selon les excursions autour du point x. Nous introduisons comme plus haut la suite de temps d'arrêt $(T^{(p)})$ des temps de passage successifs en x. Ils ont tous finis presque sûrement puisque la chaîne est récurrente irréductible.

Appelons $Z(f)_k$ la somme

$$Z(f)_k = \sum_{i=T^{(k-1)}+1}^{T^{(k)}} f(X_i).$$

$Z(f)_k$ est une fonction de la k-ième excursion, et cette fonction est toujours la même : on somme toutes les valeurs de f le long de l'excursion. On voit donc que c'est une suite de variables aléatoires indépendantes et de même loi. Montrons qu'elles sont intégrables pour $\mathbf{P}_x$: l'expression que nous avons donnée plus haut de $\int f(y) d\mu_x(y)$ montre qu'on a exactement

$$\mathbf{E}_x(Z(f)_1) = \int f d\mu_x(y).$$

On a donc, en utilisant la loi forte des grands nombres

$$\frac{1}{k}\sum_{i=1}^{T^{(k)}} f(X_i) = \frac{1}{k}\sum_{i=1}^{k} Z(f)_i \to \int f\, d\mu_x,$$

presque sûrement pour la mesure $\mathbf{P}_x$.

Soit n un entier, et appelons k_n l'indice du plus grand $T^{(k)}$ tel que $T^{(k)} \leq n$. k_n converge vers l'infini avec n (k_n est aléatoire, bien sûr), et on a

$$\sum_{i=1}^{n} f(X_i) = \sum_{i=1}^{T^{(k_n)}} f(X_i) + R_n(f),$$

avec $R_n(f) = \sum_{T^{(k_n)}+1}^{n} f(X_i)$, et donc

$$|R_n| \leq Z(|f|)_{k_n+1}.$$

On a donc

$$\frac{\sum_{i=1}^{n} f(X_i)}{\sum_{i=1}^{n} g(X_i)} = \frac{\frac{1}{k_n}\sum_{i=1}^{T^{(k_n)}} f(X_i) + \frac{R_n(f)}{k_n}}{\frac{1}{k_n}\sum_{i=1}^{T^{(k_n)}} g(X_i) + \frac{R_n(g)}{k_n}},$$

expression qui s'écrit

$$\frac{a_n + b_n}{c_n + d_n},$$

avec $a_n \to \int f d\mu$, $c_n \to \int g d\mu \neq 0$, et $b_n = R_n(f)/k_n$, $d_n = R_n(g)/k_n$.

Il nous suffit donc pour conclure de voir que b_n et d_n convergent vers 0. Faisons le pour b_n : en majorant $R_n(f)$ par $Z(|f|)_{k_n+1}$, c'est une conséquence du lemme suivant, appliqué à $U_n = Z(|f|)_n$. ∎

Lemme 9.30. *Si U_n est une suite de variables aléatoires intégrables, de même loi, alors U_n/n converge vers 0 presque sûrement*

Démonstration. — Si U est une variable aléatoire intégrable, alors $\sum_n \mathbf{P}(|U| \geq n) < \infty$. (C'est même une CNS d'intégrabilité). On a donc, pour tout $a > 0$

$$\sum_{n\geq 1} \mathbf{P}(\frac{|U_n|}{n} \geq a) = \sum_{n\geq 1} \mathbf{P}(\frac{|U|}{a} \geq n) < \infty,$$

et donc, par le lemme de Borel Cantelli, $\limsup(\frac{|U_n|}{n}) \leq a$, presque sûrement. Ceci étant vrai pour tout $a > 0$, $\limsup(\frac{|U_n|}{n}) = 0$, presque sûrement. ∎

9.6 Exemple 1 : marches aléatoires sur $\mathbb{Z}$.

Une marche aléatoire (X_n) est (en général), la somme de n variables aléatoires indépendantes et de même lois. Si on pose $X_0 = x$, on obtient ainsi une chaîne de Markov : $X_{n+1} = X_n + Y_{n+1}$, où Y_{n+1} est une suite de variables aléatoires indépendantes. Si la suite Y_n prend ses valeurs dans $\mathbb{Z}$, on obtient ainsi une marche aléatoire à valeurs dans $\mathbb{Z}$.

Son noyau est donné par $Q(x, y) = Q(0, y - x) = \mu(y - x)$, où $\mu(y)$ est la loi commune des variables Y_n. Cette loi μ détermine donc complètement le comportement de la chaîne.

Réciproquement, on peut facilement voir qu'une chaîne de Markov sur $\mathbb{Z}$ ou $\mathbb{Z}^d$ dont le noyau Q est invariant par translation ($Q(x, y) = Q(0, y - x)$ est en fait une marche aléatoire (exercice!). On vérifie facilement que la mesure uniforme (sur $\mathbb{Z}$) est invariante.

Par translation, il est suffisant de considérer la chaîne de Markov issue de 0. La première remarque à faire et que la chaîne est récurrente si et seulement si le semigroupe (pour l'addition) par le support de μ est égal à $\mathbb{Z}$. (Attention à ce que ça ne se confond pas avec la condition d'apériodicité). C'est l'hypothèse que nous ferons désormais. La chaîne est donc irréductible. La mesure uniforme est toujours invariante (nous avons vu plus haut un exemple ou il y a une autre mesure invariante).

Considérons une variable Y de loi μ. Si Y est intégrable, alors soit $m = \mathbf{E}(Y) \neq 0$, auquel cas la suite X_n/n converge vers m, et X_n converge presque sûrement vers l'infini, et dans ce cas la chaîne est transiente.

Nous allons montrer que, si $m = 0$, alors la chaîne est récurrente. (Ce n'est plus le cas dans $\mathbb{Z}^d, d \geq 3$, comme nous l'avons vu plus haut.) Pour cela, introduisons le noyau potentiel

$$U(x, y) = \sum_n Q^n(x, y),$$

dont nous avons déjà vu que $U(x, y) \leq U(y, y)$. Ici, l'invariance par translation nous donne $U(x, x) = U(0, 0)$. Nous voulons montrer que $U(0, 0) = \infty$.

Pour une partie A de $\mathbb{Z}$, appelons $U(0, A) = \sum_{x \in A} U(0, x)$. Alors, si $|A|$ désigne le cardinal de l'ensemble A, nous avons $U(0, A) \leq |A| U(0, 0)$. Pour $\varepsilon > 0$ fixé, choisissons $A = [-\varepsilon n, \varepsilon n]$, ou plutôt son intersection avec $\mathbb{Z}$, dont le cardinal est majoré par $2\varepsilon n+1$. On a

$$\frac{1}{n}\sum_1^n \mathbf{P}_0(|X_p|\leq \varepsilon p)\leq \frac{1}{n}\sum_1^n \mathbf{P}_0(|X_p|\leq \varepsilon n)\leq \frac{1}{n}U(0,A)\leq \frac{1}{n}(2\varepsilon n+1)U(0,0).$$

Mais nous savons que X_n/n converge vers 0 presque sûrement, donc en probabilité, et par conséquent $\mathbf{P}_0(|X_p|\leq \varepsilon p)\to 1$ lorsque $p\to\infty$. On en conclut que

$$\frac{1}{n}\sum_1^n \mathbf{P}_0(|X_p|\leq \varepsilon p)\to 1,\ n\to\infty,$$

puis que $\varepsilon U(0,0)\geq 1$. Ceci étant vrai pour tout $\varepsilon>0$, nous en déduisons que la chaîne est récurrente.

Toujours dans le cas où $m=0$, alors la chaîne est récurrente nulle, puisque la mesure uniforme est invariante, et qu'elle est de masse infinie. On pouvait le voir directement en supposant pour simplifier que $\mathbf{E}(Y^2)=\sigma^2<\infty$. Alors, X_n est une martingale de carré intégrable, , et $\langle X\rangle_n=n\sigma^2$. On voit donc que, pour tout temps d'arrêt T, X_n^T est une martingale bornée dans L^2 dès que $\mathbf{E}(T)<\infty$. Dans ce cas, $X_n^T=\mathbf{E}(X_T/\mathcal{F}_n)$. Si T est le temps de retour en 0, alors, $X_T=0$, et comme X_n n'est pas identiquement nulle avant T, on voit que T ne peut pas être intégrable. On verra plus bas que le temps de retour en 0 est aussi celui du temps de retour en 0 pour une autre chaîne (la chaîne G/G/1), qui ne peut pas être récurrente positive dans ce cas, même si $\mathbf{E}(Y^2)=\infty$.

Enfin, dans le cas où $m>0$, nous pouvons calculer dans certains cas la probabilité que la chaîne partant de 0 n'y revienne jamais. Supposons que la variable Y prenne ses valeurs dans l'ensemble $\{-1,0,1,2,\cdots\}$ c'est à dire soit bornée inférieurement par -1, et soit d'espérance m strictement positive. Si $X_0>0$, et si T désigne le temps d'arrivée dans les nombres négatifs, ($T=\inf\{n>0|\ X_n\leq 0\}$), l'hypothèse nous assure qu'on a $X_T=0$ sur $\{T<\infty\}$.

D'autre part, la fonction $g(\lambda)=\log\mathbf{E}(\exp(\lambda Y))$, qui est définie au moins pour $\lambda\leq 0$, est telle que $g(0)=0$, $g'(0)=m>0$, et $g(\lambda)/\lambda\to -1$ lorsque $\lambda\to-\infty$. (En effet $\exp(-g(\lambda)/\lambda)$ converge vers $\|\exp(-Y)\|_\infty=e$.) Elle est de plus strictement convexe, car sa dérivée seconde est la variance de Y pour la mesure de probabilité de densité $\exp(\lambda Y-g(\lambda))$ par rapport à $\mathbf{P}$.

Il existe donc un unique $\lambda<0$ tel que $g(\lambda)=0$. Pour cette valeur, la fonction $h(x)=\exp(\lambda x)$ est invariante, et $h(X_n)=\exp(\lambda X_n)$ est une martingale, puisque

$$Q(h)(x)=\mathbf{E}_x[\exp(\lambda(x+Y_1))]=\exp(\lambda x)=h(x).$$

(Il n'y a aucun problème d'intégrabilité ici puisque $Y_n \geq Y_0 - n$ et que λ est négatif.)

Si on appelle T le temps de retour en 0 défini plus haut, alors $h(X_n)^T$ est une martingale bornée, (puisque $X_n \geq 0$ sur $[0, T]$) qui converge donc presque sûrement vers une limite qui vaut 1 sur $\{T < \infty\}$ et 0 sur $\{T = \infty\}$ et (car X_n converge presque sûrement vers $+\infty$), et on a donc $\mathbf{E}_x(h(X_T)) = h(x) = \mathbf{P}_x(T < \infty)$, ou encore $P_x(T = \infty) = 1 - h(x)$.

On a

$$\begin{aligned}
\mathbf{P}_0(T = \infty) &= \mathbf{E}_0[\mathbf{1}_{X_1 > 0}\theta(\mathbf{1}_{T=\infty}) \\
&= \mathbf{E}_0[\mathbf{1}_{X_1 > 0}\mathbf{E}_{X_1}(\mathbf{1}_{T=\infty})) \\
&= \mathbf{E}_0((1 - h(X_1))\mathbf{1}_{X_1 > 0}),
\end{aligned}$$

expression qui se calcule directement à partir de la loi μ.

Nous avons donc obtenu :

si λ est l'unique racine non nulle de l'équation $\mathbf{E}(\exp(\lambda Y)) = 1$, alors

$$\mathbf{P}_0(T_0 = \infty) = \mathbf{E}((1 - e^{\lambda Y})\mathbf{1}_{Y > 0}).$$

9.7 Exemple 2 : une file d'attente.

Un file d'attente très simple devant un guichet peut être décrite par la suite $(A_n, n \geq 1)$ des intervalles de temps entre les arrivées du n-ième et du $(n+1)$-ième client, et la suite $(B_n, n \geq 1)$ des durées de service du n-ième client. On suppose que ces deux suites sont chacune des suites de variables aléatoires indépendantes, à valeurs dans $\mathbb{N}$, finies, positives de même loi, notées respectivement α et β, et que ces deux suites sont indépendantes entre elles. (queue "G/G/1")

On appelle W_n la durée d'attente du n-ième client, $n \geq 1$; on a la relation

$$W_{n+1} = (W_n + B_n - A_n)_+.$$

En effet, le $(n+1)$-ième client n'attend pas si $A_n \geq W_n + B_n$, sinon il attend $W_n + B_n - A_n$. On suppose que W_1 est une variable aléatoire positive finie de loi quelconque, et indépendante des deux suites $(A_n, n \in \mathbb{N}^*)$ et $(B_n, n \in \mathbb{N}^*)$. (En fait, dans la plupart des cas, $W_1 = 0$ (sauf si l'opérateur au guichet est de mauvaise humeur et décide de faire attendre son premier client coûte que coûte). On posera

$$X_n = W_{n+1}, \mathcal{F}_n = \sigma(X_0, \cdots, X_n), \ (n \geq 0).$$

On a donc $X_n = (X_{n-1} + B_n - A_n)_+$. Alors les variables A_n et B_n sont indépendantes de $\mathcal{F}_{n-1}$. On note $C_n = A_{n-1} - B_{n-1}$, dont la loi est $\gamma = \beta * \hat{\alpha}$ où $\hat{\alpha}$ est la loi symétrique de α. On a donc $X_{n+1} = (X_n + C_{n+1})_+$, et on est bien dans le cadre des chaînes de Markov homogènes sur $\mathbb{N}$. Son noyau Q s'écrit $Q(x, 0) = \gamma(]-\infty, -x])$ et $Q(x, y) = \gamma(x - y)$ si $y > 0$.

Supposons que la loi γ admette une espérance m, différence entre les moyennes des durées de service et des inter-arrivées. Posons $T_0 = \inf\{n, n \geq 1, X_n = 0\}$. Ainsi, $T_0 + 1$ est l'indice du premier client qui n'attend pas (sans compter le premier de tous). En supposant $X_0 = 0$, on pose $S_n = C_1 + \cdots + C_n$, avec $S_0 = 0$, qui est la marche aléatoire associée partant de 0.

Lemme 9.31. *1. On a les relations*

$$X_n = S_n - \inf\{0, S_1, \cdots, S_n\} = S_n + [\inf_{1 \leq k \leq n} S_k]_-.$$

De plus, $X_n \geq S_n$.

2. *Si τ désigne le temps d'entrée de (S_n) dans $\mathbb{Z}_-$, on a $T_0 = \tau$.*
3. *Pour tout n, la loi de X_n est la même que celle de* $\sup_0^n(S_1, S_2, \cdots, S_n)$.

Démonstration. — On montre la première formule par une récurrence immédiate. En notant $\underline{S_n} = \inf(S_0, \ldots, S_n)$, on a

$$(S_{n+1} - \underline{S_n})_+ = S_{n+1} - \underline{S_{n+1}}.$$

La seconde vient de ce que $[\inf_{1 \leq k \leq n} S_k]_- = -\inf\{0, S_1, \cdots, S_n\}$.

La seconde formule montre que le temps d'entrée de S dans $\mathbb{Z}_-$ correspond au temps d'atteinte de 0 pour X.

Le troisième point utilise également la formule du premier point :
si $(x_1, x_2, \cdots, x_k)$ sont des réels, alors

$$\begin{aligned} x_1 + x_2 + \cdots + x_n - \inf_k(0, x_1, x_1 + x_2, \cdots, x_1 + x_2 + \cdots x_n) &= \\ \sup_k(0, x_n, x_n + x_{n-1}, \cdots, x_n + \cdots x_{n-k}, \ldots, x_0 + \ldots x_n). \end{aligned}$$

Mais la loi de $(C_1, C_2, \cdots, C_n)$ est la même que celle de $(C_n, C_{n-1}, \cdots, C_1)$, et donc X_n a même loi que $\sup_n(S_0, \cdots, S_n)$, avec $S_0 = 0$. ∎

En conséquence, il est facile de voir que avant τ, les processus X et S coïncident : $X_n = S_n^+$ sur $\{T_0 \geq n\}$. Au temps T_0, la suite X_n repart comme au temps initial grâce à la propriété de Markov.

D'après le paragraphe précédent, on connait le comportement de la marche aléatoire selon la valeur de m. On distingue donc les trois cas suivants :

- pour $m > 0$, on a vu que S_n est converge vers l'infini. Il en va de même de X_n. $X_n \geq S_n \to +\infty$ presque sûrement et $\mathbf{P}_0(T_0 < \infty) < 1$, $E(T_0) = +\infty$. La suite X_n repasse un nombre fini de fois par 0 et finit par converger vers $+\infty$.

- pour $m = 0$, $S_n/n \to 0$ presque sûrement et on a vu que 0, donc $\mathbb{Z}_-$, est récurrent pour la chaîne S donc 0 est aussi récurrent pour la chaîne X et donc $\mathbf{P}_0(T_0 < \infty) = 1$. Mais on a aussi vu que la loi de X_n est la même que celle de $\max(0, S_1, \cdots, S_n)$, et donc que $\mathbf{P}_0(X_n = 0)$ est la probabilité que le temps de retour en 0 pour la suite $-S_p$ soit supérieur à n. Ceci converge vers 0 lorsque $n \to \infty$. On voit aussi que X_n converge en loi vers ∞, tout en repassant une infinité de fois en 0. Dans ce cas, la chaîne est récurrente nulle.

- pour $m < 0$, la chaîne S est transiente et tend vers $-\infty$. Ce qui fait que $\tau = T_0$ est presque sûrement fini aussi dans ce cas et presque sûrement une infinité de clients n'attend pas. Ici, de plus, l'espérance de T_0 est finie. La loi de X_n converge vers la loi du $\sup_n S_n$, qui est une variable presque sûrement finie. Cette loi est la mesure invariante de la chaîne : la chaîne est récurrente positive. (Si la loi de X_n converge vers une probabilité, ce ne peut être que la mesure invariante de la chaîne et nécessairement nous sommes dans le cas récurrent positif. Cette loi est difficile à calculer en général. Lorsque A_n est majorée par 1, alors la loi de $B_n - A_n$ est à valeurs dans $\{1, 0, -1, -2, \cdots\}$ et nous pouvons appliquer ce que nous avons vu pour les marches aléatoires pour calculer $\lim_n \mathbf{P}(X_n = 0)$ qui est égale à $\lim_n \mathbf{P}_0(\sup_{k\leq n} S_k = 0)$, et cette limite est la probabilité que la marche aléatoire $-S_n$ ne retourne jamais en 0. Nous l'avons calculée dans l'exemple précédent.

On peut aussi introduire un troisième processus qui est le nombre de clients dans la file d'attente non encore servis : c'est également une chaîne de Markov.

9.8 Exemple 3 : le processus de Galton-Watson

Le processus de Galton-Watson, introduit en 1874, est sans doute le plus simple des processus de population : on suppose que les X_n individus de la n-ième génération ont indépendament des descendants dont le nombre suit une loi μ sur $\mathbb{N}$. On peut le modéliser de la façon suivante : on se donne une double suite de variables aléatoires (Y_i^k) indépendantes et de loi μ, et une valeur initiale X_0, déterministe ou indépendante de la double suite Y. On peut alors définir

$$X_{n+1} = \sum_{i=1}^{X_n} Y_i^{n+1},$$

si $X_n \neq 0$, et $X_{n+1} = 0$ si $X_n = 0$.

D'autre part, si on désigne par X_n^1 le processus issu de $X_0 = 1$ on voit sur la définition que la loi de la suite X_n sachant $X_0 = k$ est celle de la somme de k suites indépendantes de même loi que X_n^1, et on peut donc se ramener à l'étude de la loi de X_n^1, que nous noterons X_n dans ce qui suit.

On voit immédiatement sur cette définition que 0 est un point absorbant, et un des premiers problèmes est de calculer la probabilité d'atteindre 0, c'est à dire la probabilité d'extinction. On note Y une variable de loi μ, et on introduit la fonction génératrice de la loi μ, c'est à dire $f(s) = \mathbf{E}(s^Y)$ qui est définie pour $|s| \leq 1$. Remarquons que $|f(s)| \leq 1$ si $|s| \leq 1$, et que $f'(1) = \mathbf{E}(Y) = m$ si Y est dans L^1, valeur éventuellement infinie.

Nous supposerons désormais que $f(0) = \mathbf{P}(Y = 0) > 0$ (sinon X_n ne peut que croître), et nous noterons f_n la fonction composée n fois de f avec elle même, c'est à dire $f_{n+1}(s) = f(f_n(s))$. On a

Proposition 9.32. *Soit E l'événement "extinction", c'est à dire*

$$E = \{\exists n > 0, \ X_n = 0\}.$$

1. $\mathbf{E}(s^{X_n}) = f_n(s)$;
2. $\mathbf{P}(X_n = 0) = f_n(0)$;
3. *on a* $\mathbf{P}_1(E) = \lim_n f_n(0) = q$;
4. *si $m < 1$, il existe une unique solution à l'équation $f(s) = s$ dans $[0, 1[$, et cette solution est égale à q ;*
5. *si $m \leq 1$, alors $q = 1$.*

On voit donc que si $m \leq 1$, il y a presque sûrement extinction.

Démonstration. —

Commençons par le premier point : calculons d'abord $\mathbf{E}(s^{X_{n+1}}/X_n = k)$. D'après la définition, pour $k > 0$, c'est égal à $\mathbf{E}(s^{Y_1 + \cdots + Y_k})$, où les variables Y_i

sont indépendantes et de loi μ. On obtient ainsi $\mathbf{E}(s^{X_{n+1}}/X_n = k) = f(s)^k$. Remarquons que cette formule reste vraie pour $k = 0$.

On obtient donc $\mathbf{E}(s^{X_{n+1}}) = \mathbf{E}(f(s)^{X_n})$, et le résultat par une récurrence immédiate.

Le second point vient de ce que, pour toute variable aléatoire X à valeurs dans $\mathbb{N}$, si on pose $g(s) = \mathbf{E}(s^X)$, alors $g(0) = \mathbf{P}(X = 0)$.

Ensuite, on remarque que les ensembles $\{X_n = 0\}$ sont croissants (0 est absorbant), et donc que

$$\mathbf{P}(\exists n,\ X_n = 0) = \mathbf{P}(\cup_n\{X_n = 0\}) = \lim_n \mathbf{P}(X_n = 0).$$

La suite vient de l'observation suivante : la fonction $f(s)$, envoie $[0,1]$ dans $[0,1]$, est telle que $f(0) = \mathbf{P}(Y = 0)$, $f(1) = 1$, et est croissante. Sa dérivée seconde vaut $\mathbf{E}(Y(Y-1)s^{Y-2}) \geq 0$, et donc elle est aussi convexe. Si $m > 1$, puisque $m = f'(1)$, alors il existe une unique solution à $f(s) = s$ dans $]0,1[$, et la suite $f_n(0)$ converge vers cette solution, en croissant. Si $m \leq 1$, la suite $f_n(0)$ converge vers 1.

∎

On peut aussi montrer que, si $m > 1$, alors la suite X_n converge presque sûrement vers l'infini là où elle ne converge pas vers 0. En effet, nous avons :

Proposition 9.33. *Supposons $m = \mathbf{E}(Y) > 1$ et soit E l'événement "extinction". Alors, X_n converge vers l'infini presque sûrement sur E^c. De plus, si $\mathbf{E}(Y) = m < \infty$, alors X_n/m^n converge presque sûrement vers une variable W. Si Y est de carré intégrable, la convergence a lieu dans L^2.*

Démonstration. — On commence par le premier point : la suite $M_n = q^{X_n}$ est une martingale, puique nous avons vu que, pour tout $s \in [0,1]$,

$$\mathbf{E}(s^{X_{n+1}}/\mathcal{F}_n) = \mathbf{E}(s^{X_{n+1}}/X_n) = f(s)^{X_n}.$$

Cette martingale est positive, majorée par 1, et donc converge presque sûrement vers une variable $M \in [0,1]$. Or, puisque 0 est absorbant, que $M_n = 1$ sur $\{X_n = 0\}$, nous voyons que $M = 1$ sur l'ensemble d'extinction E. Nous avons

$$\mathbf{E}(M) = \mathbf{E}(M_0) = q \geq \mathbf{P}(M = 1) \geq \mathbf{P}(E) = q,$$

et donc $M = 0$ presque sûrement sur $M < 1$. Ceci montre que X_n converge vers l'infini sur E^c.

Pour l'autre point, nous exhibons une seconde martingale $N_n = X_n/m^n$. C'est une martingale car

$$\mathbf{E}(X_{n+1}/X_n = k) = k\mathbf{E}(Y) = km,$$

et donc $\mathbf{E}(X_{n+1}/\mathcal{F}_n) = mX_n$. C'est une martingale positive, donc bornée dans L^1, et elle converge presque sûrement vers une variable W. Remarquons que si Y est de carré intégrable, alors la variance de X_{n+1} sachant que $X_n = k$ est égale à $k\sigma^2$, où σ^2 est la variance de Y, et donc

$$\sigma^2(X_{n+1}) = \mathbf{E}((X_{n+1} - \mathbf{E}(X_{n+1}/F_n))^2) + m^2\sigma^2(X_n) = \sigma^2\mathbf{E}(X_n) + m^2\sigma^2(X_n).$$

On en déduit une formule de récurrence pour la variance σ_n^2 de X_n

$$\sigma_{n+1}^2 = m^n\sigma^2 + m^2\sigma_n^2,$$

d'où la valeur de σ_n^2/m^{2n}, qui est la variance de N_n

$$\frac{\sigma_n^2}{m^{2n}} = \frac{\sigma^2(1-m^{-n})}{m(m-1)}.$$

Ceci montre que la martingale $N_n = X_n/m^n$ est bornée dans L^2, et donc que dans ce cas la convergence a lieu dans L^2.

On a alors $\mathbf{E}(W) = 1$ et $\sigma^2(W) = \sigma^2/[m(m-1)]$.

On peut montrer en plus montrer que $W > 0$ sur E^c, lorsque la variable X_1 est de carré intégrable. En effet, l'ensemble $\{W = 0\}$ est inclus dans E, et on remarque que la loi de la suite (X_{n+1}) sachant que $X_1 = k$ est la loi de $(X_n^1 + \cdots + X_n^k)$, où les (X_n^i) sont k suites indépendantes de même loi que celle de (X_n) lorsque $X_0 = 1$.

On en déduit que

$$\mathbf{P}(\lim_n X_n/m^n = 0/X_1 = k) = \mathbf{P}(\lim_n X_n/m^n = 0)^k.$$

Si q_1 est cette dernière probabilité, on a donc $q_1 = f(q_1)$, et puisque $q_1 \neq 1$, car sinon W serait identiquement nulle et ne serait pas d'espérance 1, alors $q_1 = q$, et les événements W et E, qui sont inclus l'un dans l'autre et qui ont même probabilité, coïncident. Nous ne nous sommes servis ici que du fait que N_n converge dans L^1 vers W, et on peut montrer que ceci a encore lieu si $\mathbf{E}(Y\log_+(Y)) < \infty$.

∎