Modèles Semi-Markoviens

Jean-Brice ARNAUD

10 juin 2008

Table des matières

1 Introduction

Chebyshev (1821-1894) était un des rares mathématiciens à s'intéresser aux probabilités dans la deuxième moitié du XIXème siècle. Il s'intéressa notamment à la loi des grands nombres, dont il fit une preuve plus fine que celle de ses prédecesseurs, à l'aide de la célèbre inégalité qui porte son nom.

En 1878, Chebyshev propose au meilleur étudiant de sa promotion un poste d'enseignant. Ce dernier va reprendre les enseignements et les travaux de son professeur, et notamment améliorer l'inégalité de Chebyshev par une inégalité qui porte aujourd'hui son nom : l'inégalité de Andrei Andreevich Markov (1856-1922).

Une querelle l'opposa à Pavel Alekseevitch Nekrasov (1854-1924), un autre mathématicien moscovite qui s'intéressait aussi aux comportements asymptotiques des variables aléatoires. D'après Nekrasov, puisque la Loi Forte des Grands Nombres, qui entraîne des phénomènes de moyennisation, requiert l'indépendance des variables, et puisqu'on peut constater des comportements moyens dans les phénomènes sociaux, c'est donc que l'homme est parfaitement indépendant de ses congénères, et qu'il a donc son libre-arbitre.

Et c'est notamment pour des raisons théologiques qu'en 1907, Markov publie *Extension de la loi des grands nombres à des grandeurs dépendant les unes des autres* dans la revue *Le Bulletin de la Société Physico-Mathématique de Kazan*. C'est dans cet article que Markov va introduire des processus, dits aujourd'hui markoviens, et montrer que l'on peut appliquer la Loi Forte des Grands Nombres sans l'indépendance des variables. Il s'agit donc d'une condition suffisante, mais nullement nécessaire.

Suite au travail de Markov, l'étude des chaînes de Markov a pris une importance croissante dans l'étude des probabilités. Les chaînes de Markov ont aujourd'hui des applications dans de nombreux domaines, dont l'économie, la finance, l'informatique ou la physique.

On a notamment développé des outils qui permettent d'optimiser, par le biais de contrôles, de modèles aléatoires dont l'évolution est markovienne : les modèles markoviens contrôlés. Une première formulation vient de Bellman, dont les travaux seront repris notamment par Howard et Oliver.

À la suite des modèles markoviens, Lévy, Smith et Takács ont développé indépendamment en 1954 ce que l'on nomme aujourd'hui les processus semi-markoviens. Les processus semi-markoviens sont une généralisation des processus markoviens, où le temps de séjour dans un état peut suivre une loi aléatoire quelconque. En effet, on verra que, dans les processus markoviens, le temps de séjour dans un état suit nécessairement une loi géométrique pour le cas discret (ou une loi exponentielle pour le cas continu).

À partir de ces processus, on a alors pu étudier des modèles semi-markoviens contrôlés, qui ont notamment permis de modéliser des expériences toujours plus variées. Ce sont ces modèles semi-markoviens contrôlés qui sont l'objet du présent document.

Nous allons tout d'abord commencer par de brefs rappels sur les chaînes de Markov, en nous appuyant sur : Baldi P., Priouret P. et Mazliak L., *Martingales et Chaînes de Markov*, pp. 85-107, 1998.

Ensuite, nous nous intéresserons aux modèles markoviens contrôlés, à travers l'étude de Hernadez-Lerma O. et Lasserre J., dans *Discrete-Time Markov Control Processes*, Stochastic Modelling and Applied Probability vol. 30, pp. 1-42, 1996.

Nous nous attaquerons après aux processus semi-markoviens, grâce aux travaux de Barbu V.S. et Limnios N., dans *Semi-Markov Chains and Hidden Semi-Markov Models towards Applications*, pp.1-10 et 45-61, 2008.

Nous finirons enfin par les modèles semi-markoviens contrôlés, via Jewell W., *Markov Renewal Programming*, Operations Research, pp. 938-953, 1963.

Cette introduction, quant à elle, s'inspire largement de Mazliak L., *Markov et ses Chaînes*, Pour la science vol. 339, Janvier 2006.

2 Rappels sur les Chaînes de Markov

2.1 Introduction

Commençons par deux exemples introductifs.

Considérons, d'une part, Indiana Jones et le Temple Maudit. À la fin, Indiana fuit à travers les dédales d'une mine, poursuivi par des aborigènes. À chaque embranchement, il choisit un passage au hasard, mais ne peut retourner sur ses pas, où l'attendent les aborigènes.

Considérons d'autre part une grenouille, dans une mare, qui saute de nénuphar en nénuphar de façon aléatoire. À chaque étape, elle saute sur un des nénuphars avoisinants de façon équiprobable.

La différence fondamentale de ces deux processus est que, dans le premier, Indiana Jones ne peut repasser là où il est déjà passé : à chaque intersection, sa trajectoire dépend donc de l'endroit où il est, mais aussi des endroits où il est déjà passé. À l'inverse, à chaque saut, la trajectoire de la grenouille ne dépend que du nénuphar où elle se trouve.

La trajectoire de la grenouille est dite sans mémoire, puisqu'à chaque étape, son évolution future ne dépend que du nénuphar où elle se trouve, et non de ceux par lesquels elle est passée auparavant.

L'objet des chaînes de Markov est l'étude de ces processus dits sans mémoire, dont l'évolution future ne dépend du passé qu'en ce qu'il a conduit à l'état présent.

Commençons par introduire les objets qui nous permettront d'étudier les dits processus.

2.2 Le Modèle

On se donne un ensemble $E = \{1, \ldots, s\}$, désigné par la suite comme l'ensemble des états du processus. On pourrait prendre E infini dénombrable, mais le cas fini suffira pour notre propos.

On munit E de la tribu discrète $\mathscr{F}$, qui contient l'ensemble des parties de E, ainsi que d'une filtration $(\mathscr{F}_n)_{n\geq 0}$ et d'une probabilité $\mathbb{P}$.

Intuitivement, la filtration $(\mathscr{F}_n)_{n\geq 0}$ représente le temps. $\forall n \in \mathbb{N}$, $\mathscr{F}_n$ contient l'ensemble des informations disponibles au temps n, i.e. l'évolution du processus jusqu'au temps n inclus.

2.3 Les Matrices de Transition

Définition 1 ***Matrices de transition***

Une matrice de transition P sur E est une famille de réels $(P(i,j))_{i,j\in E}$ telle que, $\forall i, j \in E$:

i) $P(i, j) \geq 0$

ii) $\sum_{j \in E} P(i,j) = 1$

On remarque avant tout que, $\forall i \in E$, $P(i,\cdot)$ est une probabilité sur E.

Les $P(i,j)$ vont nous servir à exprimer la probabilité que le processus passe de l'état i au temps n à l'état j au temps $n+1$.

Définition 2 ***Produit de matrices de transition***

Si P est une matrice de transition sur E, on définit par récurrence la suite $P^n, \forall n \in \mathbb{N}$, par :

– $P^0(i,j) = \delta_{ij}$ *(i.e.* $P^0 = I_E$*)*
– $P^1(i,j) = P(i,j)$ *(i.e.* $P^1 = P$*)*
– $P^{n+1}(i,j) = \sum_{k \in E} P^n(i,k)P(k,j)$, $\forall n \in \mathbb{N}$

Tout d'abord, on voit que, si P est une matrice de transition sur E, les P^n le sont aussi, $\forall n \in \mathbb{N}$.

De plus, si $P(i,j)$ est la probabilité de passer de i en j en une unité de temps, on voit que $P^2(i,j) = \sum_{k \in E} P(i,k)P(k,j)$ est la probabilité de passer de i en j en deux unités de temps, en passant par l'état intermédiaire k.

Plus généralement, $P^n(i,j)$ représentera la probabilité, partant de l'état i, d'être dans l'état j n unités de temps plus tard.

2.4 Les Chaînes de Markov

Définition 3 ***Chaînes de Markov***

Une chaîne de Markov homogène $(X_n)_{n \geq 0}$, de loi initiale μ (où μ est une loi à valeurs dans E) et de matrice de transition P, est un processus adapté, à valeurs dans E, et tel que :

i) $X_0 \sim \mu$

ii) $\mathbb{P}(X_{n+1} \in A \ / \ \mathscr{F}_n) = P(X_n, A)$, $\mathbb{P} - p.s.$, $\forall A \in \mathscr{F}$, $\forall n \geq 0$

On voit donc que l'état initial du processus suit la loi μ et que, dans l'état i au temps n, $n \geq 0$, le processus passe à l'état j au temps $n+1$ avec la probabilité $P(i,j)$.

La propriété (ii) caractérise l'absence de mémoire des processus que l'on veut étudier : l'état du processus au temps $n+1$ ne dépend que de son état au temps n, et non des états antérieurs qui l'ont amené là. C'est la propriété dite de Markov.

Quand la loi initiale du processus est μ, on note $\mathbb{P}_\mu(X_n \in A)$ la probabilité que X_n appartienne à A sachant que $X_0 \sim \mu$.

Si $\mu = i$, $\mathbb{P} - p.s.$, on notera $\mathbb{P}_i(X_n \in A)$.

On fera de même pour les espérances respectives.

Définition 4 ***Temps de retour du processus***

Soit $\sigma_i = inf\{n \geq 1,\, X_n = i\}$.

σ_i *est clairement un temps d'arrêt, puisqu'à valeurs dans* $\mathbb{N}$ *et parce que,* $\forall n \in \mathbb{N}$, X_n *est* $\mathscr{F}_n$*-mesurable (le processus étant adapté), donc* $\{\sigma_i \leq n\} = \bigcup_{k=0}^{n} \{X_k = i\}$ *l'est aussi.*

On dit que σ_i *est le temps de retour du processus en i.*

Définition 5 ***Récurrence et transience***

i) Si $\mathbb{P}_i(\sigma_i < \infty) = 1$*, on dit que l'état i est récurrent.*

ii) Si $\mathbb{P}_i(\sigma_i < \infty) < 1$*, on dit que l'état i est transient.*

iii) Si tous les états sont récurrents (resp. transients), on dit que la chaîne est récurrente (resp. transiente).

On voit que, si $\mathbb{P}_i(\sigma_i < \infty) = 1$, cela signifie que le processus, une fois en i, y retourne $\mathbb{P} - p.s.$ en un temps fini. Il suit que le processus passera $\mathbb{P} - p.s.$ une infinité de fois en i.

À l'inverse, si $\mathbb{P}_i(\sigma_i < \infty) < 1$, le processus ne passera $\mathbb{P} - p.s.$ qu'un nombre fini de fois en i.

Définition 6 ***Irréductibilité***

i) On dit qu'une partie F de E est close si, $\forall i \in F$, $P(i,F) = 1$.

ii) Si une partie close F est réduite à un point i, on dit que i est un état absorbant.

iii) Si E est la seule partie close, on dit que la chaîne est irréductible.

Clairement, si $F \subsetneq E$ est une partie close, quand le processus y entre, il ne peut plus en sortir.

À l'inverse, si E est la seule partie close, $\forall i, j \in E$, $\mathbb{P}_i(\sigma_j < \infty) > 0$ et $\mathbb{P}_j(\sigma_i < \infty) > 0$. On dit alors que les états i et j communiquent.

Si la chaîne est irréductible, c'est donc que tous les états communiquent. Donc si une chaîne est irréductible, elle peut être soit récurrente, soit transiente, puisque si un état est récurrent, comme tous les états communiquent, ils le sont tous.

2.5 La propriété de Markov forte

Définition 7 ***Espace et chaîne canoniques***

On appelle espace canonique un terme $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geq 0}, (X_n)_{n\geq 0})$ *tel que :*

- $\Omega = E^{\mathbb{N}}$
- $\omega = (\omega_n)_{n\geq 0}$, $\forall \omega \in \Omega$
- $X_n(\omega) = \omega_n$, $\forall \omega \in \Omega$
- $\mathscr{F}_n = \sigma(X_k, k \leq n)$, $\forall n \in \mathbb{N}$
- $\mathscr{F} = \sigma(X_k, k \geq 0)$

On appellera alors chaîne de Markov canonique un terme $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geq 0}, (X_n)_{n\geq 0}, (\mathbb{P}_x)_{x\in E})$ *où* $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geq 0}, (X_n)_{n\geq 0})$ *est un espace canonique, au sens défini ci-dessus, et où,* $\forall x \in E$, $\mathbb{P}_x$ *est une probabilité sur* $(\Omega, \mathscr{F})$ *vérifiant,* $\forall (a_0, \ldots, a_n) \in E^{n+1}$ *:*

$\mathbb{P}_x(X_0 = a_0, \ldots, X_n = a_n) = \mathbb{1}_{\{x\}}(a_0) P(a_0, a_1) \ldots P(a_{n-1}, a_n)$

Définition 8 ***Opérateurs de translation***

On définit sur l'espace canonique l'application :

$$\theta : \begin{array}{ccc} \Omega & \to & \Omega \\ \omega = (\omega_n)_{n\geq 0} & \to & \omega = (\omega_{n+1})_{n\geq 0} \end{array}$$

puis par récurrence :

$$\theta_0 = Id,\ \theta_1 = \theta$$

$$\theta_{n+1} = \theta \circ \theta_n = \theta_n \circ \theta,\ \forall n \geq 1$$

Clairement, on voit que, $\forall n, p \geq 0$, $X_n \circ \theta_p = X_{n+p}$

Théorème 1 ***Propriété de Markov***

Soit X une chaîne de Markov canonique. Pour tout $n \geq 0$ et pour toute fonction φ mesurable bornée ou mesurable positive, on a :

$$\mathbb{E}_\mu(\varphi \circ \theta_n \ / \ \mathscr{F}_n) = \mathbb{E}_{X_n}(\varphi)$$

Et ceci se généralise aux temps d'arrêt avec :

Théorème 2 ***Propriété de Markov forte***

Soit X une chaîne de Markov canonique et τ un temps d'arrêt. Pour tout $n \geq 0$ et pour toute fonction φ mesurable bornée ou mesurable positive, on a :

$$\mathbb{E}_\mu(\mathbb{1}_{\{\tau<+\infty\}}\varphi \circ \theta_\tau \ / \ \mathscr{F}_\tau) = \mathbb{1}_{\{\tau<+\infty\}}\mathbb{E}_{X_\tau}(\varphi)$$

3 Les chaînes semi-markoviennes et les chaînes de renouvellement markovien

3.1 Introduction

Pour une chaîne de Markov, si N_i est le temps de séjour du processus dans l'état i, on voit que, $\forall k \in \mathbb{N}$, $\mathbb{P}_i(N_i = k) = (P(i,i))^k(1-P(i,i))$.
N_i suit donc une loi géométrique de paramètre $(1-P(i,i))$.

L'étude des chaînes semi-markoviennes va nous permettre d'étudier des processus dont le temps de séjour dans un état peut suivre une loi discrète quelconque, mais qui gardent une évolution de type markovienne (processus sans mémoire).
Commençons par étudier un exemple concret.

3.2 Un Exemple : Usine de textiles

Soit $\mathscr{F}$ une usine de textiles, dont les déchets sont traités par l'unité de traitement $\mathscr{U}$ avant d'être jetés dans la rivière $\mathscr{R}$.

En cas de panne de $\mathscr{U}$, pour ne pas devoir arrêter la production, on a créé un dépôt de stockage des déchets $\mathscr{D}$ qui sert de zone tampon entre $\mathscr{F}$ et $\mathscr{U}$.
Si une panne intervient en $\mathscr{U}$ et qu'elle est réparée avant que $\mathscr{D}$ ne soit plein, alors $\mathscr{F}$ continue à tourner normalement, et on suppose que $\mathscr{D}$ se vide instantanément quand $\mathscr{U}$ est réparée. Si la panne n'est pas réparée à temps, alors $\mathscr{F}$ doit s'arrêter et la production cesse jusqu'à ce que $\mathscr{U}$ soit réparée.

Modélisons à présent le système décrit ci-dessus.
Soit $E = \{1,2,3\}$ l'ensemble des états possibles du système :
- 1 : tout va bien ($\mathscr{F}$ et $\mathscr{U}$ marchent, $\mathscr{D}$ est vide)
- 2 : panne de $\mathscr{U}$ mais $\mathscr{D}$ n'est pas encore plein (donc $\mathscr{F}$ marche normalement)
- 3 : panne de $\mathscr{U}$, $\mathscr{D}$ est plein, $\mathscr{F}$ ne marche pas

On voit que :
- $1 \rightsquigarrow 2$ (si panne de $\mathscr{U}$)
- $2 \rightsquigarrow 3$ (si $\mathscr{D}$ plein)
- $2 \rightsquigarrow 1$ (si panne de $\mathscr{U}$ résolue avant que $\mathscr{D}$ ne soit plein)
- $3 \rightsquigarrow 1$ (quand $\mathscr{U}$ est réparée)

On a donc un processus J_n, qui représente les états successifs du système, et qui est défini par :
- une loi initiale $\alpha = (\alpha_1, \alpha_2, \alpha_3)$, où $0 < \alpha_1, \alpha_2, \alpha_3$ et $\alpha_1 + \alpha_2 + \alpha_3 = 1$
- une matrice de transition (comme 1 ne communique qu'avec 2, 2 avec 1 et 3, et 3 avec 1) :

$$p = \begin{pmatrix} 0 & 1 & 0 \\ a & 0 & b \\ 1 & 0 & 0 \end{pmatrix} \text{ où } 0 < a, b,\ a+b = 1$$

Toutefois, pour bien modéliser le système qui nous intéresse, les changements d'état ne suffisent pas : il nous faut en effet considérer le temps X_n que le système passe dans chaque état (d'où découlera par exemple une estimation de la production). Supposons donc que l'on peut discrétiser le temps (par exemple en prenant pour unité de temps une heure), et que les changements d'état ne peuvent intervenir qu'à ces moments là.

Introduisons alors, $\forall k \in \mathbb{N}$:

$$f(k) = \begin{pmatrix} 0 & f_{12}(k) & 0 \\ f_{21}(k) & 0 & f_{23}(k) \\ f_{31}(k) & 0 & 0 \end{pmatrix}$$

où :

- $f_{12}(\cdot)$ est la loi du temps que met $\mathscr{U}$ à tomber en panne
- $f_{21}(\cdot)$ est la loi du temps que met $\mathscr{U}$ à être réparée
- $f_{12}(\cdot)$ est la loi du temps de remplissage de $\mathscr{D}$
- $f_{12}(\cdot)$ est la loi du temps nécessaire pour réparer $\mathscr{U}$, et donc que $\mathscr{F}$ fonctionne à nouveau, une fois $\mathscr{D}$ plein

Le temps de séjour dans un état dépend donc de l'état du système. On peut alors ≪ choisir ≫ les lois discrètes (Poisson, géométrique,...) qui s'adaptent le mieux à l'expérience réelle (via des estimations statistiques, par exemple).

On peut alors poser $S_n = \sum_{k=0}^{n} X_k$, qui est le temps du n-ième changement d'état.

Le système est donc défini par deux données :

- Un processus markovien J_n, de loi initiale α et de matrice de transition p, qui représente l'état du système après le n-ième changement d'état
- Un processus S_n qui représente le temps du n-ième changement d'état

Sous certaines hypothèses, que nous verrons ci-après, **le couple** (J_n, S_n) forme ce que l'on appelle une **chaîne de renouvellement markovien**. **Le processus** $Z_n = J_{N(n)}$, où $N(n) = max\{k \in \mathbb{N}, S_k \leq n\}$, qui représente l'état du système au temps n, sera alors ce que l'on appelle une **chaîne semi-markovienne**.

Introduisons à présent les objets qui vont permettre l'étude de tels processus.

3.3 Le Modèle

On reprend notre ensemble des états $E = \{1, \dots, s\}$.

Soit Z_n l'état du processus au temps n.

On note S_n le temps du n-ième changement d'état, $X_n = S_{n+1} - S_n$ le temps de séjour du processus dans le n-ième état, et J_n l'état du processus après le n-ième changement d'état.

On voit tout d'abord que $J_n = Z_{S_n}$, i.e. $Z_n = J_{N(n)}$, où $N(n) = max\{k \in \mathbb{N}, S_k \leq n\}$.

Définition 9 ***Chaînes semi-markoviennes***

On dit que (J_n, S_n) *est une chaîne de renouvellement markovien (ou Markov Renewal Chain, notée MRC) et que* Z_n *est une chaîne semi-markovienne (ou Semi Markov Chain, notée SMC) si,* $\forall n \in \mathbb{N}$ *:*

$$\mathbb{P}(J_{n+1} = j, X_n = k \;/\; J_0, S_0, \ldots, J_n, S_n) = \mathbb{P}(J_{n+1} = j, X_n = k \;/\; J_n) \tag{1}$$

Cette condition signifie que l'évolution future d'un processus semi-markovien (temps de séjour dans l'état présent et état suivant) ne dépend que de l'état présent de celui-ci, et non des états antérieurs ou temps de séjour en ces derniers.

Si le terme de droite de (1) ne dépend pas de n, on parle de chaînes homogènes, et on peut alors introduire la matrice de transition semi-markovienne suivante :

$$q_{ij}(k) = \mathbb{P}(J_{n+1} = j, X_n = k \;/\; J_n = i) \tag{2}$$

3.4 Matrices de transition semi-markoviennes

Soit $\mathscr{M}_E$ l'ensemble des matrices $E \times E$.
Soit $\mathscr{M}_E(\mathbb{N})$ l'ensemble des applications de $\mathbb{N}$ dans $\mathscr{M}_E$.

Définition 10 ***Matrices de transition semi-markoviennes***

On dit que $q \in \mathscr{M}_E(\mathbb{N})$ *est une matrice de transition semi-markovienne si :*

i) $q_{ij}(k) \geq 0$, $\forall i, j \in E$, $\forall k \in \mathbb{N}$

ii) $q_{ij}(0) = 0 \;\forall i, j \in E$

iii) $\sum_{k \in \mathbb{N}} \sum_{j \in E} \; q_{ij}(k) = 1 \;\forall i \in E$

On voit clairement, dans (2), que $q_{ij}(k)$ s'interprète comme la probabilité de passer de i en j après un temps de séjour de k unités de temps en i.

Définition 11 ***Produit de convolution sur*** $\mathscr{M}_E(\mathbb{N})$

$\forall A, B \in \mathscr{M}_E(\mathbb{N})$, *on pose :*

$$(A * B)_{ij}(k) = \sum_{r \in E} \sum_{l=0}^{k} A_{ir}(k-l) B_{rj}(l)$$

On note δI *l'élément neutre de ce produit de convolution, tel que*

$$\delta I(k) = \begin{cases} I_E & si\ k = 0 \\ 0_E & sinon \end{cases}$$

Définition 12 ***Puissance de matrices semi-markoviennes***

Si $q \in \mathscr{M}_E(\mathbb{N})$ *est une matrice de transition semi-markovienne, on définit par récurrence les puissances matricielles suivantes :* $\forall i, j \in E$, $\forall k \in \mathbb{N}$

$$\begin{aligned} q_{ij}^{(0)}(k) &= \begin{cases} 1 & si\ i = j\ et\ k = 0 \\ 0 & sinon \end{cases} \quad (i.e.\ q^{(0)} = \delta I) \\ q_{ij}^{(1)}(k) &= q_{ij}(k) \;\; (i.e.\ q^{(1)} = q) \\ \forall n \in \mathbb{N},\ q_{ij}^{(n+1)}(k) &= (q^{(n)} * q)_{ij}(k) \;\; (= \sum_{r \in E} \sum_{l=0}^{k} q_{ir}^{(n)}(k-l) q_{rj}(l)) \end{aligned}$$

On vérifie aisément que, si q est une matrice de transition semi-markovienne, il en est de même pour $q^{(n)}$, $\forall n \in \mathbb{N}$.

De plus, on voit que $q_{ij}^{(2)}(k) = \sum_{r \in E} \sum_{l=0}^{k} q_{ir}(k-l) q_{rj}(l)$ est la probabilité de passer de i en j en deux changements d'état, et en un temps de séjour cumulé de k unités de temps : $(k-l)$ passées en i et l passées en r.

De même, $q_{ij}^{(n)}(k)$ est la probabilité de passer de i en j en n changements d'état et en un temps de séjour cumulé de k unités de temps, i.e. $q_{ij}^{(n)}(k) = \mathbb{P}(J_n = j, S_n = k \ / \ J_0 = i)$.

Il en découle que, $\forall n \geq k+1$, $q_{ij}^{(n)}(k) = 0$.

3.5 Temps de séjour

On remarque que, si (J_n, S_n) est une MRC homogène, J_n est une chaîne de Markov de matrice de transition $p_{ij} = \sum_{k \in \mathbb{N}} q_{ij}(k)$.

Introduisons à présent les objets suivants :

Définition 13 ***Lois du temps de séjour***

$\forall i, j \in E$, on appelle :

i) loi conditionnelle de X_n la suite $f_{ij}(\cdot)$ telle que, $\forall k \in \mathbb{N}$,

$$f_{ij}(k) = \mathbb{P}(X_n = k \ / \ J_n = i, J_{n+1} = j) = \begin{cases} \frac{q_{ij}(k)}{p_{ij}} & si\ p_{ij} \neq 0 \\ \mathbb{1}_{k=\infty} & sinon \end{cases}$$

ii) fonction de répartition conditionnelle de X_n la suite $F_{ij}(\cdot)$ telle que, $\forall k \in \mathbb{N}$, $F_{ij}(k) = \sum_{l=0}^{k} f_{ij}(k)$

iii) loi du temps de séjour en i la suite $h_i(\cdot)$ telle que, $\forall k \in \mathbb{N}$, $h_i(k) = \mathbb{P}(X_n = k \ / \ J_n = i) = \sum_{j \in E} q_{ij}(k)$

iv) fonction de répartition du temps de séjour en i la suite $H_i(\cdot)$ telle que, $\forall k \in \mathbb{N}$, $H_i(k) = \sum_{l=0}^{k} h_i(k)$

3.6 Équations de Renouvellement Semi-Markovien

On a défini au (2.4) les temps de retour d'un processus markovien, qui permettent notamment d'étudier les renouvellements d'un processus, i.e. les temps où le processus revient à son point de départ. Ces renouvellements permettent notamment d'appliquer aux chaînes de Markov des théorèmes limites, comme la Loi Forte des Grands Nombres.

On peut aussi étudier les renouvellements des processus semi-markoviens, et nous allons ici en esquisser une première approche.

Commençons par poser $P_{ij}(n) = \mathbb{P}(Z_n = j \ / \ Z_0 = i)$.

On a, $\forall k \in \mathbb{N}$,

$$
\begin{aligned}
P_{ij}(k) &= \underbrace{\delta_{ij} \times [1 - H_i(k)]}_{\text{si } i=j\text{, rester plus de } (k) \text{ en } i} + \overbrace{\sum_{r \in E} \sum_{l=0}^{k} q_{ir}(l) P_{rj}(k-l)}^{\text{rester } (l) \text{ en } i\text{, puis } r\text{, puis passer de } r \text{ en } j \text{ en } (k-l)} \\
&= G_{ij}(k) + (q * P)_{ij}(k) \quad (\text{où } G_{ij}(k) = \delta_{ij} \times [1 - H_i(k)]) \qquad (3)
\end{aligned}
$$

Définition 14 ***Équation de Renouvellement semi-markovien***
Si $L, G \in \mathscr{M}_E(\mathbb{N})$, on appelle équation de renouvellement semi-markovien (ERSM) l'équation :

$$
\begin{aligned}
L(k) &= G(k) + (q * L)(k) \\
\text{i.e. } [(\delta I - q) * L](k) &= G(k)
\end{aligned}
$$

Définition 15 ***Inverse de gauche pour le produit de convolution***
Soit $A \in \mathscr{M}_E(\mathbb{N})$.
S'il existe $B \in \mathscr{M}_E(\mathbb{N})$ tel que $B * A = \delta I$, on dit que B est l'inverse de gauche de A (pour le produit de convolution), noté $A^{(-1)}$.

Propriété 1 *Soit $A \in \mathscr{M}_E(\mathbb{N})$.*
$A^{(-1)}$ existe ssi $det(A(0)) \neq 0$
De plus, $\forall k \in \mathbb{N}$

$$
A^{(-1)}(k) = \begin{cases} (A(0))^{-1} & \text{si } k = 0 \\ -(A(0))^{-1} \times \sum_{l=0}^{k-1} A^{(-1)}(l) A(k-l) & \text{sinon} \end{cases}
$$

Propriété 2 *On peut résoudre une ERSM si $(\delta I - q)$ est inversible, i.e. si $det((\delta I - q)(0)) \neq 0$.*

Or

$$
(\delta I - q)(0) = \underbrace{\delta I(0)}_{=I_E} - \underbrace{q(0)}_{=0_E} = I_E
$$

Donc on peut résoudre les ERSM, et en particulier (3).
On peut alors vérifier que $(\delta I - q)^{(-1)}(n) = \sum_{k=0}^{n} q^{(k)}(n)$,
i.e. $\forall i, j \in E,\ (\delta I - q)_{ij}^{(-1)}(n) = \mathbb{P}(\bigcup_{k=0}^{n} \{J_k = j,\ S_k = n\} \ / \ J_0 = i)$, qui est la probabilité, partant de i, de faire un saut en j au temps n.

Donc $\forall j \in E,\ (\delta I - q)_{jj}^{(-1)}(n)$ est la probabilité, partant de j, que la chaîne se renouvelle au temps n.

3.7 Récurrence, Irréductibilité

On note $(S_n^j)_{n \geq 0}$ les temps de retour successifs du processus en j, où (S_0^j) est le temps du premier passage en j.

$\forall i \neq j$, on note :

– $g_{ij}(n) = \mathbb{P}_i(S_0^j = n)$ la probabilité, partant de i, que le processus passe en j au temps n.
– $g_{jj}(n) = \mathbb{P}_j(S_1^j = n)$ la probabilité, partant de j, que le processus repasse en j au temps n.
– $G_{jj}(n) = \sum_{k=0}^{n} g_{jj}(k)$ la probabilité, partant de j, que le processus repasse en j au temps n au plus tard.

On remarque alors que $G_{ij}(\infty)$ est la probabilité que la chaîne passe de i en j.

Définition 16 ***Chaînes semi-markoviennes irréductibles***

Si $G_{ij}(\infty) \times G_{ji}(\infty) > 0$, on dit que i et j communiquent. Cette relation, notée $i \leftrightarrow j$, est une relation d'équivalence sur E. On appelle classes les éléments de l'espace quotienté $E \ / \ \leftrightarrow$.

S'il n'existe qu'une seule classe, i.e. si tous les états communiquent, on dira que la chaîne semi-markovienne est irréductible.

Définition 17 ***Chaînes semi-markoviennes récurrentes***

Si $G_{jj}(\infty) = 1$, on dit que l'état j est récurrent, sinon qu'il est transient.

On parle de chaîne semi-markovienne récurrente (resp. transiente) si tous les états sont récurrents (resp. transients).

4 Modèles Markoviens Contrôlés

4.1 Exemple introductif

Considérons l'entretien du moteur d'une voiture. Supposons pour simplifier que l'état du moteur peut s'évaluer suivant une grille allant de 0, pour un moteur à l'état neuf, jusqu'à s pour un moteur cassé. On note X_n l'état du moteur après la n-ième utilisation du véhicule, et on suppose que son évolution est markovienne, d'état initial 0 et de matrice de transition P sur $\{1 \dots s\}$. Un coût $\alpha(i)$ apparaît à chaque utilisation du moteur, qui dépend de l'usure du moteur (par exemple les frais d'essence).

À chaque utilisation, on peut faire réviser le moteur, ramenant son état à 0 pour un coût R. Cette possibilité donne à l'utilisateur un moyen de contrôle, grâce auquel il peut influencer l'expérience à chaque étape.

On cherche alors à déterminer la stratégie d'entretien optimale en termes de coûts, à un horizon fini N (i.e. quand effectuer des révisions pour minimiser les coûts). On cherche donc à minimiser l'espérance des coûts cumulés pour N unités de temps.

Nous allons voir, après une première formalisation, que l'on a des méthodes générales de résolution de ce type de problèmes d'optimisation de coûts de processus à évolution markovienne contrôlés : les modèles markoviens contrôlés.

4.2 Le Modèle

Définition 18 ***Modèle Markovien Contrôlé***

Un modèle markovien contrôlé est formé du quintuplé $\{X, A, \{A(x), x \in X\}, Q, c\}$, où :

- *X est l'ensemble des états possibles*
- *A est l'ensemble des contrôles possibles*
- *$\{A(x)/x \in X\}$ est une famille où $A(x)$ est l'ensemble des contrôles possibles dans un état x donné. On note $\mathbb{K} = \{(x,a)/x \in X, a \in A(x)\}$*
- *Q est une matrice de transition sur X sachant $\mathbb{K}$*
- *c est une fonction de coût/gain pour une étape*

4.3 Première formalisation

À l'étape n, si le système est dans l'état X_n et qu'on lui applique le contrôle A_n :

- un coût $c(X_n, A_n)$ apparaît
- le système passe à l'état X_{n+1} au temps $n+1$, suivant la loi $Q(\cdot/x,a)$, i.e. $\mathbb{P}(X_{n+1} \in B \;/\; X_n = x,\ A_n = a) = Q(B \;/\; x,a), \forall B \in \mathscr{B}(\{1 \dots s\})$

On a souvent $X_{n+1} = f(X_n, A_n, \varepsilon_n)$, où les (ε_n) sont des v.a.r. i.i.d. à valeurs dans un espace S, de loi μ indépendante de X_0. On a alors :

$$\begin{aligned} Q(B \;/\; x,a) &= \mu(\{s \in S \;/\; f(x,a,s) \in B\}) \\ &= \int_S \mathbb{1}_B[f(x,a,s)]\mu(ds) \\ &= \mathbb{E}_\mu[\mathbb{1}_B(f(x,a,\varepsilon_1))] \end{aligned}$$

On cherche alors à minimiser les coûts à un horizon N, via une politique de contrôles notée π, i.e. à minimiser $\mathbb{E}_x^\pi[\sum_{k=0}^{N-1} c(X_k, A_k) + c(X_N)]$, pour un état initial x.

On cherche donc une politique de contrôles optimale et l'espérance de coûts optimale qui lui est associée.

4.4 Le cadre canonique

Soit $(\Omega, \mathscr{F})$ un espace probabilisable tel que $\Omega = (X \times A)^\infty$ et $\mathscr{F}$ soit la tribu engendrée par Ω.

$\forall \omega \in \Omega$, ω est de la forme $\omega = (x_0, a_0, x_1, a_1, \dots)$.

Ω contient donc l'ensemble des histoires possibles, i.e. $\mathbb{K}^\infty$.

On note à présent $H_n = (X \times A)^n \times X$ l'ensemble des histoires possibles au temps n.

Définition 19 ***Politiques de contrôle***

Une politique de contrôle $\pi = (\pi_n)_{n \in \mathbb{N}}$ est une famille de matrices de transition sur A sachant H_n, telle que $\pi_n(A(X_n)/h_n) = 1 \ \forall h_n \in H_n$

Intuitivement, cela signifie que la politique de contrôle détermine au temps n, et de façon aléatoire, un contrôle $A(x)$ en fonction du temps et de l'histoire du système jusqu'au temps n inclus : h_n.

Définition 20 ***Politiques markoviennes***

Une politique de contrôle $\pi = (\pi_n)_{n \in \mathbb{N}}$ est :

i) une politique (ou stratégie) markovienne (aléatoire) s'il existe une famille de matrices de transition $\varphi = (\varphi_n)_{n \in \mathbb{N}}$ sur A sachant X telle que
   - $\varphi_n(A(x)/x) = 1$, $\forall x \in X$
   - $\pi_n(\cdot/h_n) = \varphi_n(\cdot/x_{n)}$, $\forall h_n \in H_n$, $\forall n \in \mathbb{N}$.

ii) une politique (ou stratégie) markovienne stationnaire s'il existe une matrice de transition φ sur A sachant X telle que :
   - $\varphi(A(x)/x) = 1$, $\forall x \in X$
   - $\pi_n(\cdot/h_n) = \varphi(\cdot/x_{n)}$, $\forall n \in \mathbb{N}$.

Une politique markovienne est donc une politique de contrôle qui détermine un contrôle en fonction du seul état présent, et non des états et contrôles antérieurs, contenus dans h_n.

Une politique markovienne stationnaire ne varie pas en fonction du temps, d'où son appellation.

Propriété 3 *Soit ν une loi sur X.*

Si $X_0 \sim \nu$ et si $\pi = (\pi_n)_{n \in \mathbb{N}}$ est une politique markovienne (resp. markovienne stationnaire), alors X_n est une chaîne de Markov inhomogène (resp. homogène) de matrices de transition $(Q(\cdot\ /\ \cdot, \pi_n))_{n \in \mathbb{N}}$ (resp. $Q(\cdot\ /\ \cdot, \pi)$).

Cela signifie notamment que, $\forall B \in B(X)$, $\forall n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{P}_\nu^\pi(X_{n+1} \in B\ /\ X_0, \dots, X_n) &= \mathbb{P}_\nu^\pi(X_{n+1} \in B\ /\ X_n) \\ &= Q(B\ /\ X_n, \varphi_n) \end{aligned}$$

4.5 La programmation dynamique : l'algorithme de Bellman

Ici, nous ne considérerons que les stratégies markoviennes.

Rappellons que l'objectif des modèles markoviens est de minimiser, quand c'est possible, $\mathbb{E}^\pi[\sum_{k=0}^{N-1} c_k(X_k,A_k)+c_N(X_N)]$.

Théorème 3 ***Algorithme de Bellman***

Soient $J_0,\ldots,J_N$ des fonctions définies sur X par la récurrence backward suivante :

$$J_N(x) = c_N(x)$$

$$J_n(x) = \min_{a\in A(x)}[c_n(x,a)+\int_X J_{n+1}(y)\; Q(dy/x,a)]$$

Si le minimum existe et est atteint en $f_n(x)$ $\forall n\in\{0,\ldots,N-1\}$, alors le coût optimal est $J^*(X_0)=J_0(X_0)$ et f_n est une politique de contrôle optimale.

Les fonctions J_n seront ci-après désignées comme les fonctions de valeur.

Preuve :

Soit π une stratégie arbitraire.

$J_N(x)=c_N(x)$

Supposons que $T_{n+1}(\pi,x)=\mathbb{E}^\pi[\sum_{k=n+1}^{N-1} c_k(X_k,A_k)+c_N(X_N)\;/\;X_{n+1}=x]\geq J_{n+1}(x)$.

Alors :

$$\begin{aligned}
T_n(\pi,x) &= \mathbb{E}^\pi[c_n(X_n,A_n)+\sum_{k=n+1}^{N-1} c_k(X_k,A_k)+c_N(X_N)\;/\;X_n=x]\\
&= \int_A [c_n(x,a)+\int_X T_{n+1}(\pi,y)Q(dy\;/\;x,a)]\pi_n(da\;/\;x)\\
&\geq \int_A [c_n(x,a)+\int_X J_{n+1}(y)Q(dy\;/\;x,a)]\pi_n(da\;/\;x) \text{ par hypothèse de récurrence}\\
&\geq \min_{a\in A(x)}[c_n(x,a)+\int_X J_{n+1}(y)Q(dy\;/\;x,a)] = J_n(x)
\end{aligned}$$

Donc, par une récurrence backward, on voit bien que :

$$\mathbb{E}^\pi[\sum_{k=0}^{N-1} c_k(X_k,A_k)+c_N(X_N)]\geq J_0(x), \text{ d'où la conclusion.}$$

4.6 Martingales et programmation dynamique

4.6.1 Martingales et optimalité

On a, $\forall n\leq N-1$:

$$J_n(x) = \inf_{a\in A(x)}[c(x,a)+\int_X J_{n+1}(y)\; Q(dy\;/\;x,a)]$$

On remarque aussi que

$$c_n(X_n,a)+\int_X J_{n+1}(y)Q(dy\ /\ x,a)=\mathbb{E}[c_n(X_n,a)+J_{n+1}(X_{n+1})\ /\ \mathscr{F}_n]$$

Soit $\pi=(\pi_0,\dots,\pi_N)$ une stratégie markovienne aléatoire, au sens où, dans l'état x au temps n, on prend le contrôle $\pi_n(x)$. Soit X_n^π le processus contrôlé associé.

Posons $\Gamma_n^\pi=\sum_{k=0}^{n-1}c_k(X_k^\pi,\pi_k(X_k^\pi))+J_n(X_n^\pi)$.

$$J_n(X_n^\pi)\leq\mathbb{E}[c_n(X_n^\pi,\pi_n(X_n^\pi))+J_{n+1}(X_{n+1}^\pi)\ /\ \mathscr{F}_n] \tag{4}$$

Donc :

$$J_n(X_n^\pi)+\sum_{k=0}^{n-1}c_k(X_k^\pi,\pi_k(X_k^\pi))\leq\mathbb{E}[\sum_{k=0}^{n}c_k(X_k^\pi,\pi_k(X_k^\pi))+J_{n+1}(X_{n+1}^\pi)\ /\ \mathscr{F}_n]$$

car $\sum_{k=0}^{n-1}c_k(X_k^\pi,\pi_k(X_k^\pi))$ est $\mathscr{F}_n$-mesurable.

D'où il vient que $\Gamma_n^\pi\leq\mathbb{E}[\Gamma_{n+1}^\pi\ /\ \mathscr{F}_n]$
On en déduit donc que Γ_n^π **est une sous-martingale.**

Proposition 1 *π est une stratégie optimale ssi Γ_n^π est une martingale.*

<u>*Preuve :*</u>
- Si π est une stratégie optimale, (4) devient une égalité, d'où Γ_n^π est une martingale.
- Si Γ_n^π est une martingale, alors :

$$\mathbb{E}_x[\Gamma_N^\pi]=\mathbb{E}_x[\Gamma_0^\pi]=J_0(x)$$

Soit encore :

$$\mathbb{E}_x[\sum_{k=0}^{N-1}c_k(X_k^\pi,\pi_k(X_k^\pi))+J_N(X_N^\pi)]=J_0(x)$$

Ce parallèle entre les martingales et la programmation dynamique est en fait un moyen de validation de résultats : l'algorithme de Bellman donne une condition suffisante d'optimalité, mais ne prétend pas donner toutes les stratégies optimales. On peut donc tester des stratégies ne découlant pas de l'algorithme de Bellman via les Γ_n.

4.6.2 Un exemple

Nous allons ici esquisser rapidement un exemple d'utilisation de l'approche martingale des processus markoviens contrôlés.

Plaçons nous dans un cadre linéaire gaussien de la forme :

$$X_{n+1}=aX_n+bU_n+\varepsilon_n$$

Où les ε_n sont des gaussiennes standard indépendantes et indépendantes de l'état initial X_0.

On se donne également des fonctions de coût quadratique, de la forme :

$$C_n(x,u) = \alpha x^2 + \beta u^2$$

$$C_N(x) = \gamma x^2$$

On peut alors prouver que les fonctions de valeur sont de la forme $J_n(x) = \mu_n x_n^2 + \rho_n$.

Donc, pour que $\Gamma_n = \sum_{k=0}^{n-1}(\alpha X_k^2 + \beta U_k^2) + J_n(X_n)$ soit une martingale, il faut et il suffit que :

$$\mathbb{E}[\alpha X_n^2 + \beta U_n^2 + \mu_{n+1} X_{n+1}^2 + \rho_{n+1} \ / \ \mathscr{F}_n] = J_n(X_n) = \mu_n X_n^2 + \rho_n$$

$$\Leftrightarrow \alpha X_n^2 + \beta U_n^2 + \mu_{n+1}\mathbb{E}[X_{n+1}^2 \ / \ \mathscr{F}_n] + \rho_{n+1} = J_n(X_n) = \mu_n X_n^2 + \rho_n$$

$$\Leftrightarrow \alpha X_n^2 + \beta U_n^2 + \mu_{n+1}[(aX_n + bU_n)^2 + 1] + \rho_{n+1} = J_n(X_n) = \mu_n X_n^2 + \rho_n$$

Car $X_{n+1} = aX_n + bU_n + \varepsilon_n$.

On peut alors montrer que les contrôles optimaux sont de la forme $U_n = \delta_n X_n$.

5 Modèles Semi-Markoviens Contrôlés

5.1 Introduction

On l'a vu, les modèles markoviens contrôlés sont un système dans lequel un processus évolue dans un ensemble d'états possibles, avec un gain à chaque changement d'état. Remarquons que pour les modèles markoviens, on peut parler de transition à un même état (p_{ii}), ce qui n'est pas le cas pour les chaînes semi-markoviennes.

À chaque changement d'état, donc, une décision est prise, qui affecte la probabilité de transition et le gain. Il s'agissait alors de maximiser la récompense totale espérée.

Dans le cadre semi-markovien, un autre élément est à prendre en compte : le temps de séjour dans un état. Reprenons rapidement l'exemple de l'usine de textiles : admettons que l'on se donne un contrôle sur le nombre de lots en cours de production, et que l'on cherche à maximiser un gain, celui des quantités produites. Produire beaucoup rapporte donc plus, mais engendre aussi plus de panne de l'unité de traitement $\mathscr{U}$, tout en augmentant la probabilité que le dépôt $\mathscr{D}$ ne sature avant que $\mathscr{U}$ ne soit réparée (d'où un arrêt de l'usine).

Pour bien rendre compte de l'expérience, il nous faut donc prendre en compte un gain au changement d'état, qui va être aussi fonction du temps de séjour dans cet état.

De plus, on peut envisager plusieurs sortes d'horizon fini, dans un modèle semi-markovien. S'agira-t'il d'un nombre fini de changements d'états ou d'un nombre fini d'unités de temps (ce qui colle mieux à notre exemple) ?

Commençons par introduire les outils qui vont nous servir à étudier ces modèles semi-markoviens contrôlés.

5.2 Le Modèle

5.2.1 Les états

On se donne un ensemble fini d'états $E = \{1, \dots, N\}$, un processus markovien $(i_n)_{n\geq 0}$ représentant les états successifs du système.

Le processus markovien $(i_n)_{n\geq 0}$ est déterminé par son état initial i_0 et une matrice de transition $p_{ij} = \mathbb{P}(i_{n+1} = j \ / \ i_n = i)$.

5.2.2 Les temps de séjour

On se donne des lois des temps de séjour $(\tau(i_n, i_{n+1}))_{n\geq 0}$, fonctions de l'état présent et de l'état futur.

Les temps de séjour sont déterminés par leur fonction de répartition $F_{ij}(t) = \mathbb{P}(\tau(i, j) \leq t)$.

On note $\nu_{ij}^{(n)} = \mathbb{E}[(\tau(i, j))^n]$ les moments de ces temps de séjours. On suppose que, $\forall i, j \in E$, $F_{ij}(0) = 0$, ce qui entraîne que $\nu_{ij}^{(n)} > 0 \ \forall n \in \mathbb{N}$.

On pose alors $Q_{ij}(t) = p_{ij} \times F_{ij}(t)$, qui est la probabilité que le système passe de i en j après un temps de séjour inférieur ou égal à t.

5.2.3 Le gain

Considérons à présent le gain. Il doit dépendre de l'état présent, de l'état suivant, mais aussi du temps passé dans l'état présent. Quand le système est dans l'état i, en route vers j, il accumule un gain $R_{ij}(t\ /\ \tau)$, qui varie avec i, j, $\tau(i,j)$ et $t \in [0;\tau(i,j)]$, le temps déjà passé en i.

On suppose que $R_{ij}(0\ /\ \tau) = 0, \forall i, j \in E$.

Le gain total, à la fin du séjour en i, est noté $R_{ij}(\tau)$.

On peut considérer, par exemple, un gain fixe R_{ij} perçu initialement, ainsi qu'un gain variable de r_{ij} i.e. :

$$R_{ij}(t\ /\ \tau) = \begin{cases} 0 & \text{si } t = 0 \\ R_{ij} + t \times r_{ij} & \text{sinon} \end{cases}$$

Le gain moyen actualisé du séjour en i, en route vers j, est, pour un taux d'actualisation continu α :

$$\rho_{ij}(\alpha) = \int_0^\infty dF_{ij}(\tau) \int_0^\tau e^{-\alpha x} R_{ij}(x\ /\ \tau)dx$$

Le gain moyen, partant de i, pour une transition, est donc :

$$\rho_i(\alpha) = \sum_{j=1}^{N} p_{ij}\rho_{ij}(\alpha)$$

5.2.4 Les contrôles

Soit $\mathscr{Z} = \{1,\dots,Z\}$ l'ensemble des contrôles possibles.

Une procédure de contrôle interviendra de la façon suivante :

- Le système entre dans l'état i.
- On prend la décision $z(i)$, fonction de i et, éventuellement, du temps restant.
- En fonction de $z(i)$, le futur état du système suit la loi $p_{ij}^{z(i)}$, et le temps de séjour en i suit la loi $F_{ij}^{z(i)}(t)$.
- Le gain $R_{ij}^{z(i)}[t\ /\ \tau(i,j,z(i))]$ est accumulé.
- Le système entre dans l'état j.

5.3 Horizon fini : nombre fini de transitions

On se donne un horizon de n transitions.

Il s'agit ici de maximiser $\mathbb{E}_{i_0}^{z(i_k)}[\sum_{k=0}^{n-1} \rho_{i_k}^z + \gamma(i_n)]$, où γ est le gain final, fonction de l'état final, et où z est une politique de contrôle.

On note $J_i(n,\alpha)$ les fonctions de valeur, qui représentent le gain total espéré actualisé (au taux continu α) pour les n transitions, partant de i, et avec une stratégie optimale $z^*(i,n)$.

On note $J_i(0,\alpha) = \gamma$ la fonction de gain final.

La relation de récurrence s'écrit alors, $\forall i \in E,\ \forall n \geq 1$:

$$J_i(n,\alpha) = \max_{z\in\mathscr{Z}}[\rho_i^z(\alpha) + \sum_{j=1}^{n} p_{ij}^z \tilde{f}_{ij}^z(\alpha) J_j(n-1,\alpha)]$$

Où on a noté $\tilde{f}_{ij}^z(\alpha)$ la transformée de Laplace-Stieltjes de $F_{ij}(t)$, ou la transformée de Laplace de sa dérivée $f_{ij}(t)$, si elle existe, i.e. :

$$\tilde{f}_{ij}^z(\alpha) = \int_0^\infty e^{-\alpha t} dF_{ij}(t) = \int_0^\infty e^{-\alpha t} f_{ij}(t)dt$$

$\tilde{f}_{ij}^z(\alpha)$ désigne donc l'actualisation moyenne qu'il faut appliquer aux gains futurs $J_j(n-1,\alpha)$, qui dépend du temps passé en i.

On remarque qu'ici, la récurrence n'est pas une récurrence backward, puisque n représente le nombre de transitions qu'il reste, et non le nombre de transitions déjà effectuées.

5.4 Horizon fini : temps fini

Nous allons considérer ici un horizon de temps fini, qui modélise mieux certaines expériences, comme celle de l'usine de textiles que nous avons déjà vue.

Il n'est pas facile, dans le cas général, d'exprimer le gain total cumulé, c'est-à-dire la quantité que l'on cherche à maximiser. Paradoxalement, on sait résoudre ce type général de problèmes par le biais de l'algorithme décrit ci-dessous.

On se donne tout d'abord un horizon de temps $t \in \mathbb{R}^+$.

La stratégie optimale $z^*(i,t)$ dépend cette fois-ci du temps qu'il reste.

Les contrôles ne peuvent être appliqués qu'au moment des transitions, mais il est à présent possible que l'expérience s'arrête entre deux transitions (d'où des problèmes de coupure que nous allons voir).

On définit de nouvelles fonctions de valeur, les $v_i(t,\alpha)$, qui représentent le gain actualisé d'un processus de durée t, partant de i, et sous une stratégie optimale $z^*(i,t)$.

On note aussi $S_{ij}^z(t,\tau)$ le gain final, quand le processus est interrompu dans l'état i, en route vers j, après un temps de séjour de t en i (pour un temps de séjour total de τ).

Il y a deux possibilités : soit le système est resté en i tout du long (i.e. $\tau \geq t$), soit il y a eu au moins une transition, et il est passé en j au temps τ. On peut donc poser :

$$v_i(t,\alpha) = \sum_{j=1}^{N} p_{ij} \left[\begin{array}{ll} & \int_t^\infty F_{ij}(d\tau)[\int_0^t e^{-\alpha x} R_{ij}(x\ /\ \tau)dx + e^{-\alpha t} S_{ij}(t,\tau)] \\ + & \int_0^t F_{ij}(d\tau)[\int_0^\tau e^{-\alpha x} R_{ij}(x\ /\ \tau)dx + e^{-\alpha\tau} V_j(t-\tau,\alpha)] \end{array} \right] \quad (5)$$

Clairement, la ligne haute de (5) considère le cas où on est resté en i tout du long ($\tau \geq t$) : on calcule alors le gain actualisé du séjour en i, du temps 0 au temps t, auquel s'additionne la récompense finale $S_{ij}(t,\tau)$ actualisée.

La ligne basse de (5) s'intéresse au cas où il y a eu au moins une transition ($\tau < t$) : on cumule alors le gain actualisé du séjour en i (d'une durée τ), auquel s'additionne le gain futur espéré actualisé, partant de j, et pour une durée restante $t-\tau$: $V_j(t-\tau,\alpha)$.

On en déduit donc la relation de récurrence suivante, pour les fonctions de valeur :

$$v_i(t,\alpha) = \max_{z\in\mathscr{Z}}\{\sigma_i^z(\alpha,t) + \sum_{j=1}^{N} p_{ij}^z[\int_0^t e^{-\alpha\tau} v_j(t-\tau,\alpha) F_{ij}^z(d\tau)]\} \qquad (6)$$

Où, $\forall t > 0$, $\forall i$, $\sigma_i^z(\alpha,t)$ est le gain moyen du séjour en i, soit ρ_i auquel on rajoute, quand on passe tout le séjour en i, le gain final $S_{ij}^z(t,\tau)$ moins le manque à gagner dû à l'interruption du séjour, soit :

$$\sigma_i^z(\alpha,t) = \rho_i^z(\alpha) + \sum_{j=1}^{N} p_{ij}^z \int_t^{\infty} F_{ij}^z(d\tau)[e^{-\alpha t} S_{ij}^z(t,\tau) - \int_t^{\tau} e^{-\alpha x} R_{ij}(x \mathbin{/} \tau) dx]$$

Il n'y a pas de façon de résoudre (6), si ce n'est par le biais d'un ordinateur.

On peut aussi « dicrétiser » le temps, en posant par exemple $t = k\Delta,\ k \in \mathbb{N}$, puis faire tendre Δ vers 0.