GRAS

Processus discrets de Markov avec gains et politiques de contrôle

Publications des séminaires de mathématiques et informatique de Rennes, 1966-1967
« Séminaires de probabilités et statistiques », , exp. nº 6, p. 1-46
<http://www.numdam.org/item?id=PSMIR_1966-1967____A6_0>

UNIVERSITE DE RENNES
FACULTE DES SCIENCES

SEMINAIRE DE PROBABILITES

PROCESSUS DISCRETS DE MARKOV AVEC GAINS ET POLITIQUES DE CONTROLE

par

Monsieur GRAS

Année 1966-1967

FORMALISME DE LA THEORIE DE LA DECISION STATISTIQUE.

Soient 3 espaces mesurables

$(\Omega, \mathcal{A})$ espace des observations

$(\Theta, \mathcal{C})$ " des paramètres

$(\Delta, \mathcal{D})$ " des décisions.

Soit P une probabilité de transition relativement à $(\Theta, \mathcal{C})$ et $(\Omega, \mathcal{A})$; $P : (\theta, \times \mathcal{A}) \leadsto [0,1]$.

On appelle stratégie s une application mesurable de $(\Omega, \mathcal{A}) \leadsto (\Delta, \mathcal{D})$ qui consiste à décider $\delta \in \Delta$ après avoir observé $\omega \in \Omega$.

On appelle stratégie aléatoire S une probabilité de transition relativement à $(\Omega, \mathcal{A})$ et $(\Delta, \mathcal{D})$

$S(\omega, .)$ est ainsi une probabilité sur $(\Delta, \mathcal{D})$.

On appelle $W = \{w(\theta, \delta)\}$ la fonction mesurable réelle perte définie sur $(\Theta, \mathcal{C}) \times (\Delta, \mathcal{D})$: ainsi $W(\theta, \delta)$ est la perte consécutive à la décision δ quand θ est la valeur du paramètre.

Une stratégie aléatoire S entraîne une perte moyenne :

$$C(\theta, \omega) = \int_{\Delta} w(\theta, \delta)\, S(\omega, d\delta)$$

et un risque à l'instant θ :

$$R_S(\theta) = \int_{\Omega} P(d\omega, \theta) \int_{\Delta} w(\theta, \delta)\, S(\omega, d\delta).$$

Remarquons qu'une stratégie est une stratégie aléatoire pour laquelle $S(\omega,.)$ est concentrée en $s(\omega)$ ($s : \omega \rightsquigarrow s(\omega)$ et $S(\omega,A) = 1_A(s(\omega))$) Dans de telles conditions, la perte moyenne est $C(\theta,\omega) = w(\theta,s(\omega))$.

Un préordre ou préférence (resp. μ-préférence) peut être défini sur les stratégies aléatoires en posant :

$$S \prec S' \quad \text{si} \quad R_S(.) \leq R_{S'}(.) \text{ sur } \Theta$$

(resp. μ p.p. sur $(\Theta, \mathcal{C}, \mu)$).

Une s.a. est dite admissible (resp. μ admissible) si aucune s.a. ne peut lui être préférée strictement, i.e. si $S' \subset S \Longrightarrow S \subset S'$, i.e. encore s'il n'existe pas S' tel que $R_{S'}(.) \leq R_S(.)$ sur Θ et tel que $R_S \not\equiv R_{S'}$ (resp. tel que $R_{S'}(.) \leq R_S(.)$ μ -p.p. sur Θ et tel que $\mu(R_S \neq R_{S'}) = 0$).

Une s.a. de Bayes (pour μ) est une stratégie qui minimise $\int_\Theta R_S(\theta)\, \mu(d\theta)$ sur l'ensemble des s.a. S. Il va de soi qu'une telle stratégie est μ-admissible car si S_o minimise $\int_\Theta R_S(\theta)\, \mu(d\theta)$ alors :

$$\int R_{S_o}(\theta)\, \mu(d\theta) \leq \int R_S(\theta)\, \mu(d\theta) \qquad \forall\, S$$

$$\Longrightarrow \quad R_{S_o}(.) \leq R_S(.) \qquad \mu\text{-p.p.}$$

INTRODUCTION

L'objet de ce développement est l'étude des processus de Markov $\left(\Omega, \mathcal{F}, P, (x_t)_{t \in T}\right)$ avec gains, dans le cas particulier suivant :

l'ensemble des valeurs prises dans E par X_t est fini ; E est l'ensemble de N états : $A_1, A_2, \ldots A_N$ (que nous noterons encore 1,2,...N)

Le processus étant markovien, les probabilités conditionnelles vérifieront :

$$Pr\left[X_k \in A_j \mid X_1 \in A_q, X_2 \in A_q, \ldots X_{k-1} \in A_i\right] = Pr\left[X_k = j \mid X_{k-1} = i\right] = p_{ij}$$

Dans le chapître I, nous particulariserons cette chaîne, qualifié pour cette raison de discrète :

le temps sera défini sur un ensemble T discret, et nous supposerons, ce qui laisse cependant toute généralité au problème, que les éléments de T sont équidistants.

Dans le chapître II, T sera la demi-droite positive réelle.

Enfin, les matrices stochastiques de transition P (n, n-1) de l'état de la chaîne à l'instant n-1 à l'instant n, définiront une chaîne homogène. Nous prendrons donc pour tout n et tout k

$$P\ (n,\ n-1) = P\ (k,\ k-1) = P$$

Dans chacun de ces deux chapîtres, l'étude du processus est facilitée par l'utilisation d'une transformation :

- pour le cas discret par la z-transformation $\mathcal{Z}$:

  soit f définie sur $\mathbb{N}$, à valeurs scalaires ou vectorielles

$$f \overset{\mathcal{Z}}{\rightsquigarrow} \varphi(z) = \mathcal{Z}(f) = \sum_{0}^{\infty} f(n)\, z^n \quad \text{où } z \in \mathbb{C}$$

- pour le cas continu par la transformation de Laplace $\mathcal{L}$

  soit f définie sur $\mathbb{R}^+$:

$$f \overset{\mathcal{L}}{\rightsquigarrow} \mathcal{L}(f) = \varphi(z) = \int_{0}^{\infty} f(t)\, e^{-tz}\, dt$$

§ 1. QUELQUES NOTIONS ET APPLICATIONS RELATIVES A LA z-TRANSFORMATION (cf [2])

1.1. - Propriétés de $\mathcal{Z}$

$\mathcal{Z}$ particularise la fonction génératrice qui, au sens le plus général, est définie par :

$$\mathcal{G}(f) = \varphi(z) = \sum_{a}^{b} f(n)\, z^n \quad \text{où a et b sont des entiers relatifs}$$

finis ou non.

$\mathcal{Z}$ est linéaire et $\varphi(z)$ converge absolument pour

$$|z| < \frac{1}{\overline{\lim} \sqrt[n]{|f(n)|}}$$

De plus, dans un domaine de convergence que nous préciserons, $\mathcal{Z}$ est bijective :

Si $\varphi(z)$ est holomorphe dans le disque (o,R) alors

$$\varphi(z) = \sum_{0}^{\infty} a_n\, z^n \quad \text{où} \quad a_n = \frac{\varphi^{(n)}(o)}{n\,!}$$

$f(.)$ est donc la fonction unique définie par

$$f(o) = \varphi(o),\ f(1) = \varphi'(o), \dots\ f(n) = \frac{\varphi^{(n)}(o)}{n\ !},\ \dots$$

Nous pouvons dresser un tableau de correspondance entre antécédent et image par $\mathcal{Z}$.

antécédent $f(n)$	image $\mathcal{Z}(f) = \varphi(z)$
$\lambda_1\ f_1(n) + \lambda_2\ f_2(n)$	$\lambda_1\ \varphi_1(z) + \lambda_2\ \varphi_2(z)$
$f\ (n+1)$	$z^{-1}\left[\varphi\vert z) - f(o)\right]$
1	$\frac{1}{1-z}$
a^n	$\frac{1}{1-az}$
n	$\frac{z}{(1-z)^2}$
$n\ a^n$	$\frac{az}{(1-az)^2}$
$\frac{1}{n!}$	e^z
$-\frac{(-1)^n}{n}$	$\log\ (1+z)$
$\frac{n(n+1)\dots(n+k-1)}{k\ !}\ a^n$	$\frac{1}{(1-az)^{k+1}}$

1.2. Application.

a). - Considérons l'équation vectorielle récurrente suivante :

$\Pi(n+1) = \Pi(n)\, P \quad (1)$

P est une matrice stochastique i.e. $\sum_{j=1}^{N} p_{ij} = 1$ et $p_{ij} \geq o \ \forall i \ \forall j$

$$P = \begin{pmatrix} p_{11} \ldots\ldots & p_{1j} \ldots\ldots & p_{1N} \\ p_{i1} \ldots\ldots & p_{ij} \ldots\ldots & p_{iN} \\ p_{N1} \ldots\ldots & p_{Nj} \ldots\ldots & p_{NN} \end{pmatrix}$$

$\Pi(n) = (\Pi_1(n) \ \ldots\ldots \ \Pi_i(n) \ \ldots\ldots \ \Pi_N(n))$ est un vecteur stochastique.

Prenons les z-transformés des 2 membres de (1) :

or $\quad \Pi(n) \overset{\mathcal{Z}}{\rightsquigarrow} \mathcal{Z}(\Pi(n)) = \tau(z) = \sum_{o}^{\infty} \Pi(n)\, z^n$

$\Pi(n+1) \overset{\mathcal{Z}}{\rightsquigarrow} \mathcal{Z}(\Pi(n+1)) = z^{-1}\left[\tau(z) - \Pi(o)\right]$

$\Longrightarrow \left[\tau(z) - \Pi(o)\right] = z\,\tau(z)\, P \quad (1')$

et $\quad \tau(z)\left[I - z\,P\right] = \Pi(o)$

. Mais $|z| < 1 \Longrightarrow$ dét $\left[I - z\ P\right] \neq o$.

En effet :

- ce déterminant ne s'annule que si $z = \frac{1}{\lambda}$ (λ appartient au spectre de P)

- *toute valeur propre d'une matrice stochastique a son module borné par 1.* [1]

Donc $\left[I - z\,P\right]^{-1}$ existe et :

$$\boxed{\mathcal{Z}(\Pi(n)) = \tau(z) = \Pi(o)\left[I - z\,P\right]^{-1}}$$

$\Pi(o)$ étant une constante, nous pouvons considérer $\left[I - z P\right]^{-1}$ comme transformé d'une certaine fonction $H(n)$ telle que :

$$\left[I - z P\right]^{-1} = \sum_{o}^{\infty} H(n)\, z^n$$

soit en identifiant :

$$\Pi(n) = \Pi(o)\, H(n)$$

Interprétons :

Soit la chaîne de Markov définie par la matrice de transition P ; p_{ij} est la probabilité pour que le processus, étant dans l'état i à l'instant n, soit dans l'état j à l'instant n+1.

L'élément $\Pi_i(n)$ de $\Pi(n)$ représente la probabilité pour que le processus soit en l'état i à l'instant n. Alors en vertu de l'homogénéité de la chaîne :

$$\Pi(n) = \Pi(o)\, P^n$$

Ainsi $H(n) = P^n$

et $P^n = \mathcal{Z}^{-1}\left[(I - z P)^{-1}\right]$ et $\left[I - z P\right]^{-1} = \mathcal{Z}(P^n)$

Ce procédé indique un moyen efficace et rapide d'évaluer P^n, à condition, bien entendu, que l'image réciproque par $\mathcal{Z}^{-1}$ de $\left[I - z P\right]^{-1}$ soit aisée à calculer (1).

b) Exemple :

Un marchand de jouets dresse un bilan à la fin de chaque semaine : le dernier jouet produit a (état A_1) ou n'a pas (état A_2) la faveur du public. Dans le cas A_2, il sort un nouveau jouet. Les aléas de la faveur du public se traduisent par la matrice de transition de la chaîne supposée homogène :

---

(1) Nous savons d'ailleurs que si P, réelle ou complexe, admet les valeurs propres distinctes $\lambda_1, \lambda_2, \ldots \lambda_p$ avec les ordres de multiplicité $r_1, r_2, \ldots r_p$ $(\sum_1^p r_i = N)$ alors $P^n = A + \sum_{i+1}^{p} \lambda_i^n B_i(n)$ où

- A n'existe que si P admet la valeur propre o et A est nulle si $n \geq N$
- $B_i(n)$ est un polynôme matriciel.

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{2}{5} & \frac{3}{5} \end{pmatrix}$$

où par exemple $\frac{2}{5}$ est la probabilité pour que, placé dans l'état 2, le marchand passe dans l'état 1.

P est stochastique et la chaîne est ergodique (i.e. sans classe transitoire) et régulière (i.e. non cyclique : le P.G.C.D. des entiers k constituant l'ensemble N_{ii} fermé par rapport à l'addition, où k est un nombre de transitions permettant de passer de l'état i à lui-même, est égal à 1) (cf [1])

$$I - z\,P = \begin{pmatrix} 1-\frac{z}{2} & -\frac{z}{2} \\ -\frac{2z}{5} & 1-\frac{3z}{5} \end{pmatrix}$$

$\left[I - z\,P\right]^{-1}$ peut se décomposer ainsi :

$$\left[I - z\,P\right]^{-1} = \frac{1}{1-z} \begin{pmatrix} \frac{4}{9} & \frac{5}{9} \\ \frac{4}{9} & \frac{5}{9} \end{pmatrix} + \frac{1}{1-\frac{1}{10}\,z} \begin{pmatrix} \frac{5}{9} & -\frac{5}{9} \\ -\frac{4}{9} & \frac{4}{9} \end{pmatrix} \qquad (2)$$

Remarquons que $(I - z\,P)^{-1}$ est décomposée en fractions rationnelles en z admettant pour pôles :

$$z_o = \frac{1}{\lambda_o} = 1$$

$$z_1 = \frac{1}{\lambda_1} = 10$$

où λ_o et λ_1 sont les valeurs propres de P.

Ceci est général, que les racines soient imaginaires ou réelles, d'ordre 1 ou d'ordre p (il suffit de prendre le z-transformé de chaque membre de l'égalité matricielle de la note 1 page

Pour $|z| < 1$, ayant $|a| < 1$, $\frac{1}{1-az}$ est développable en série entière et nous pourrons prendre les z-transformées des 2 membres de (2).

$$P^n = \begin{pmatrix} \frac{4}{9} & \frac{5}{9} \\ \frac{4}{9} & \frac{5}{9} \end{pmatrix} + \left(\frac{1}{10}\right)^n \begin{pmatrix} \frac{5}{9} & -\frac{5}{9} \\ -\frac{4}{9} & \frac{4}{9} \end{pmatrix} = M + \left(\frac{1}{10}\right)^n D$$

Remarquons :

1) Quand $n \longrightarrow \infty$ $P^n \longrightarrow M = \begin{pmatrix} \frac{4}{9} & \frac{5}{9} \\ \frac{4}{9} & \frac{5}{9} \end{pmatrix}$

matrice stochastique qui est évidemment la limite de Cesaro de P^n, matrice à vecteurs lignes identiques, ce que nous savons être général (cf [1]).

D est appelé matrice différentielle (la somme des éléments ligne est nulle)

2) Dans le cas où P est matrice de transition d'une chaîne cyclique la matrice limite de Cesaro est exhibée par une z-transformation. Ainsi, si :

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad I - z P = \begin{pmatrix} 1 & -z \\ -z & 1 \end{pmatrix}$$

$$(I - z P)^{-1} = \frac{1}{1-z} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} + \frac{1}{1+z} \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

$$\text{et } P^n = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} + (-1)^n \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

$\{P^n\}$ ne converge donc pas simplement. Cette suite admet pour valeurs d'adhérence les matrices :

$$J = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{et} \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

§ 2. PROCESSUS DE MARKOV AVEC GAINS.

2.1. Etude théorique.

a). Formule récurrente.

Dans le cas d'une chaîne de Markov homogène, à une transition de l'état i (à l'instant n) vers l'état j (à l'instant n+1), nous avons associé une probabilité de passage p_{ij}.

Nous associerons maintenant, en plus, un gain g_{ij}. Ainsi un processus de Markov avec gains sera entièrement déterminé par la donnée :

- du vecteur initial $\Pi(o)$
- de la matrice stochastique de transition P
- de la matrice de gains $G = (g_{ij})$
- d'une valeur "potentielle" ou"capital" ou "valeur de vente" des états j que nous noterons $t_j(o)$ $j = 1 \ldots N$ $t_1(o)$ sera,par exemple la valeur de vente actuelle de l'usine du fabricant de jouets se trouvant à l'instant présent dans l'état 1.

Jusqu'alors, notre seule préoccupation a été l'examen de l'état du processus à l'instant n et nous avons pour cela déterminé une formule récurrente donnant l'état probable du processus à l'instant n+1.

Une préoccupation maintenant toute naturelle est de savoir quel est le gain total $t_i(n)$ que l'on peut s'attendre à percevoir lors des n prochaines transitions si le processus est actuellement dans l'état i.

Calculons $t_i(n)$. $t_i(1)$ est l'espérance mathématique, i.e. la valeur moyenne, de la variable aléatoire prenant la valeur $g_{ij} + t_j(o)$ avec la probabilité p_{ij}. En effet, si, placé dans l'état i, nous évaluons la valeur (comportant "valeur de vente" en j et gain de i à j) d'une telle transition aléatoire, nous obtiendrons : $g_{ij} + t_j(o)$

Ainsi $t_i(1) = \sum_{j=1}^{N} p_{ij} (g_{ij} + t_j(o))$

Plus généralement, évaluons, placés dans l'état i à un instant donné, le gain total probable en n transitions à venir : la valeur de la transition $i \longrightarrow j$ est égale au gain g_{ij}, auquel s'ajoute le gain total probable en (n-1) transitions suivantes, si nous partons l'instant suivant de j.

d'où $t_i(n) = \sum_{j=1}^{N} p_{ij} (g_{ij} + t_j(n-1))$

b) <u>Notation matricielle</u>

$$\text{Posons} \quad q = \begin{pmatrix} q_1 \\ \vdots \\ q_i \\ \vdots \\ q_N \end{pmatrix} \quad \text{avec } q_i = \sum_{1}^{N} p_{ij}\, g_{ij}$$

$$t(n) = \begin{pmatrix} t_1(n) \\ \vdots \\ t_i(n) \\ \vdots \\ t_N(n) \end{pmatrix} \quad \text{avec } t_i(n) = q_i + \sum_j p_{ij}\, t_j\,(n-1)$$

alors $\boxed{t(n) = q + P\ t(n-1)}$

$t(n)$ est le vecteur de gain total pour les n transitions à venir.

c). Application de la z-transformation à cette étude.

Posons $T(z) = \sum_{0}^{\infty} t(n)\ z^{n}$, $t(n)$ étant la fonction vectorielle définie précédemment sur $\mathbb{N}$.

Prenons le z-transformé des 2 membres de la relation récurrente : $t(n+1) = q + P\ t(n)$.

Nous obtenons :

$$z^{-1}\left[T(z) - t(o)\right] = \frac{1}{1-z}\ q + P\ t(z)$$

Soit $(I - z\ P)\ T(z) = \frac{z}{1-z}\ q + t(o)$

et pour $|z| < 1$ $\quad T(z) = \left[I-z\ P\right]^{-1} \frac{z}{1-z}\ q + \left[I-z\ P\right]^{-1}\ t(o)$

Or, nous avons vu que $\left[I - z\ P\right]^{-1}$ se développe en fractions rationnelles en z suivant les pôles de dét.$(I - z\ P)$. 1 étant toujours valeur propre d'une matrice stochastique P, $\left[I - z\ P\right]^{-1}$ est de la forme : $\frac{1}{1-z}\ M + \mathcal{F}(z)$ où $\mathcal{F}(z)$ est une somme de fractions rationnelles dont les coefficients $D_1 \ldots D_p$ sont des matrices différentielles.

Pour simplifier l'écriture, et la généralisation est facile, nous supposerons que $\mathcal{F}(z)$ se réduit à :

$$\mathcal{F}(z) = \frac{1}{1-az}\ D \quad \text{où } a \in \text{spectre de } P.$$

Remarquons alors que :

$$[I - z P]^{-1} \frac{z}{1-z} = \frac{z}{(1-z)^2} M + \frac{z}{(1-z)(1-az)} D$$

et que : $$\frac{z D}{(1-z)(1-az)} = \frac{1}{1-a} \left[\frac{1}{1-z} - \frac{1}{1-az}\right] D$$

$$= \left(\frac{1}{1-z} - \frac{1}{1-az}\right) \mathcal{F}'(1)$$

alors $$\boxed{T(z) = \frac{z}{(1-z)^2} Mq + \left(\frac{1}{1-z} - \frac{1}{1-az}\right) \mathcal{F}'(1)\ q + \frac{1}{1-z} Mt(o) + \mathcal{F}(z)\ t(o)} \quad (3)$$

Prenons, pour $|a| \leq 1$ et $|z| < 1$, les images réciproques par $\mathcal{Z}^{-1}$ des 2 membres de (3). Il vient :

$$t(n) = n\ Mq + (1-a^n)\ \mathcal{F}'(1)\ q + Dt(o) + a^n\ Dt(o)$$

Si P *est une matrice régulière, donc converge simplement, toutes ses valeurs propres sont, en dehors de 1 valeur simple, de module strictement inférieur à 1 (cf [3]).*

Nous le supposerons pour étudier le comportement asymptotique de $t(n)$, qui lorsque $n \longrightarrow \infty$ est équivalent à :

$$\boxed{t'(n) = n\ Mq + \mathcal{F}'(1)\ q + D\ t(o)}$$

Posons $Mq = b$ et $\mathcal{F}'(1)\ q + M\ t(o) = c$;

b est interprétable en remarquant que :

$$t_i'(n) - t_i'(n-1) = b_i$$

$\Longrightarrow$ b est le gain que l'on peut s'attendre à percevoir lors d'une transition d'un processus parti de l'état i et durant jusqu'à l'instant n (n suffisamment grand).

La formule réduite donnant t est, dans le cas où P est régulière :

$$\boxed{t(n) = nb + c}$$

Pour tout i, $t_i(n)$ est l'espérance mathématique asymptotique de gains au bout de n transitions quand le processus part de l'état i. Remarquons que $t_i(n)$ est fonction affine de n.

Si P n'est pas régulière, t(n) ne converge pas, le terme $a^n \left[\mathcal{F}'(1)\, q - Dt(o)\right]$ étant cyclique ; cependant dans de bonnes conditions ce terme est petit devant $n\, Mq + \mathcal{F}'(1)\, q + Dt(o)$.

2.2. Exemple.

Reprenons l'exemple du fabricant de jouets. Nous donnerons pour déterminer le processus :

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{2}{5} & \frac{3}{5} \end{pmatrix} \qquad G = \begin{pmatrix} 9 & 3 \\ 3 & -7 \end{pmatrix} \qquad t(o) = \begin{pmatrix} o \\ o \end{pmatrix}$$

9 représente le gain (par ex. 900 F) que peut percevoir le fabricant lors d'une étape pendant laquelle le jouet garde la faveur du public.

Ici :

$$M = \begin{pmatrix} \frac{4}{9} & \frac{5}{9} \\ \frac{4}{9} & \frac{5}{9} \end{pmatrix} \;;\; a = \frac{1}{10} \;;\; D = \begin{pmatrix} \frac{5}{9} & -\frac{5}{9} \\ -\frac{4}{9} & \frac{4}{9} \end{pmatrix} \;;\; q \begin{cases} q_1 = \frac{1}{2} \times 9 + \frac{1}{2} \times 3 = 6 \\ q_2 = \frac{2}{5} \times 3 - \frac{3}{5} \times 7 = -3 \end{cases}$$

$$T(z) = \frac{z}{(1-z)^2} \quad Mq + \frac{z}{(1-z)(1-\frac{1}{10}z)} \quad Dq$$

$$= \frac{z}{(1-z)^2} \quad Mq + \underbrace{\frac{10}{9}}_{\mathcal{F}(1)} \left(\frac{1}{1-z} - \frac{1}{1-\frac{1}{10}z}\right) \quad Dq$$

Prenons l'image inverse par $\mathcal{Z}^{-1}$ de cette égalité :

$$t(n) = n\ Mq + \frac{10}{9} \left[1 - (\frac{1}{10})^n\right] \quad Dq$$

$$= n\ \binom{1}{1} + \frac{10}{9} \left[1 - (\frac{1}{10})^n\right] \begin{pmatrix} 5 \\ -4 \end{pmatrix}$$

Lorsque n est très grand, t(n) se comporte comme :

$$t'(n) = n \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{10}{9} \begin{pmatrix} 5 \\ -4 \end{pmatrix} \Longrightarrow \begin{cases} t'_1(n) = n + \frac{50}{9} \\ t'_2(n) = n - \frac{40}{9} \end{cases}$$

Remarquons :

1) Pour tout n :

$t'_1(n) - t'_2(n) = 10$ donc constant. Ceci signifie que l'état de départ a une influence de 1 000 F sur le gain total asymptotique.

2) Pour tout n :

$t'_1(n+1) - t'_1(n) = t'_2(n+1) - t'_2(n) = 1$. Autrement dit, on peut s'attendre à gagner 100 F lors de chaque transition, quel que soit l'état initial.

§ 3. PROCESSUS DE MARKOV AVEC GAINS ET POLITIQUES DE CONTRÔLE. (cf [5])

Il faut bien avouer, cependant, que dans la réalité économique ou démographique, un processus de Markov n'a pas un déroulement immuable à partir de l'état initial. L'homogénéité est exceptionnelle et il est même heureux qu'une part d'initiative soit accordée à l'exécutant (industriel, commerçant, ouvrier,...). Placé dans un état à un instant donné, cet exécutant aura à choisir entre plusieurs politiques lui permettant de contrôler le processus, d'infléchir le hasard en régularisant le déroulement ultérieur de la chaîne, et, en particulier, en "optimisant", de viser un ou plusieurs états particuliers à l'étape suivante.

3.1. Etude théorique.

Attachons à chaque état A_i :

- le vecteur ligne de probabilité V_i représentant la distribution de probabilité sachant que A_i est réalisé.

- le vecteur de gain w_i

i étant fixé, à la suite finie $S_i = (1,2,\dots k_i)$ nombres que nous appellerons décisions (ou politiques) dans l'état i, associons les couples $\left[(V_i^{(1)}, W_i^{(1)}) ; (V_i^{(2)}, W_i^{(2)}), \dots (V_i^{(k_i)}, W_i^{(k_i)})\right]$. Considérant les N suites associées aux N états, nous aurons $k_1 \times k_2 \times \dots \times k_N$ couples possibles de matrices telles que (P,G) de transition et de gain. La chaîne n'est donc plus homogène : le choix d'une décision introduit une modification du processus.

Relativement à un problème économique, pour un esprit équilibré, il est tout naturel de songer à rendre maxima ses propres gains. Par exemple, nous trouvant dans l'état i, à un instant donné, et supposant l'arrêt du processus après n étapes, nous chercherons à rendre maximum la quantité :

$$t_i(n) = \sum_j p_{ij} \left(g_{ij} + t_j(n-1)\right)$$

Supposons que nous ayons rendu, par récurrence sur n, $t_j(n-1)$ maximum, nous "optimiserons" $t_i(n)$ en choisissant la politique $r_1 \in [1,\ldots k_i]$ qui rend $t_i(n)$ maximum, c'est-à-dire puisque $t_i(n)$ dépend des matrices P et G, politique telle que :

$$t_i^{r_1}(n) = \max_{\ell \in [1,\ldots k_i]} t_i^{\ell}(n)$$

Ainsi, de proche en proche, il aura été défini des couples de matrices $(P_{r_1}, G_{r_1}) \ldots (P_{r_n}, G_{r_n})$ telles que le gain visé au bout de n transitions soit maximum.

Soit d le vecteur de décisions dont les N composantes sont N nombres extraits des N suites S_i. A chaque nombre d'étapes n du processus correspond d(n) qui rend $t_i(n)$ maximum pour tout i :

$$d(n) = \begin{pmatrix} d_1(n) \\ \vdots \\ d_N(n) \end{pmatrix} \quad \text{où } d_i(n) \in S_i$$

3.2. Exemple.

Supposons que le marchand de jouets ait à choisir entre les deux politiques :

placé dans l'état 1 :

- ou bien il fait de la publicité (1)

- ou bien il ne fait pas de publicité (2)

placé dans l'état 2, il aura le même choix. Alors $S_1 = S_2 = \left[1,2\right]$.

Il est bien évident que les frais de publicité vont absorber une part des bénéfices, mais le risque d'avoir un jouet qui a la faveur du public augmente. Soient donc les vecteurs :

		Probabilités de transition	Gains
A_1	décision 1 d_1	$V_1^{(1)} = (\frac{1}{2} \quad \frac{1}{2})$	$W_1^{(1)} = (9 \quad 3)$
	d_2	$V_1^{(2)} = (\frac{4}{5} \quad \frac{1}{5})$	$W_1^{(2)} = (4 \quad 4)$
A_2	d_1	$V_2^{(1)} = (\frac{2}{5} \quad \frac{3}{5})$	$W_2^{(1)} = (3 \quad -7)$
	d_2	$V_2^{(2)} = (\frac{7}{10} \quad \frac{3}{10})$	$W_2^{(2)} = (1 \quad -19)$

Déterminons à chaque instant les gains possibles, relativement à chaque politique, par la relation :

pour d_1 : $t_1(n) = g_{11}^{(1)}\ p_{11}^{(1)} + g_{12}^{(1)}\ p_{12}^{(1)} + p_{11}^{(1)}\ t_1(n-1) + p_{12}^{(1)}\ t_2(n-1)$

pour d_2 : $t_1(n) = g_{11}^{(2)}\ p_{11}^{(2)} + g_{12}^{(2)}\ p_{12}^{(2)} + p_{11}^{(2)}\ t_1(n-1) + p_{12}^{(2)}\ t_2(n-1)$

La comparaison des 2 nombres nous indique le gain maximum et la décision à prendre au moment du passage de l'instant n à l'instant n-1 précédent l'arrêt.

		n = o	n = 1	n = 2	n = 3	n = 4
$t_1(n)$	d_1	o	$\boxed{6}$	7,5	9,25	11,24
	d_2	o	4	$\boxed{8,2}$	$\boxed{10,22}$	$\boxed{12,22}$
$t_2(n)$	d_1	o	$\boxed{-3}$	-2,4	-0,74	1,24
	d_2	o	-5	$\boxed{-1,7}$	$\boxed{0,25}$	$\boxed{2,22}$
choix		$\binom{1}{1}$	$\binom{2}{2}$	$\binom{2}{2}$	$\binom{2}{2}$	$\binom{2}{2}$

Remarquons la convergence du vecteur décision vers $\binom{2}{2}$.

Nous conclurons sur cette remarque : il est bien évident que rares sont les entreprises envisageant de s'arrêter délibérément (ou même de péricliter) à un instant déterminé. Aussi, il serait intéressant de voir si généralement pour n croissant indéfiniment, un choix asymptotique de décisions ne serait pas possible, ceci permettant de prévoir une ligne de conduite à très longue échéance.

(cf [5] et [4])

----

§ 4. METHODE ITERATIVE POUR LA RESOLUTION DU PROBLEME D'OPTIMISATION DES PROCESSUS A DECISIONS SEQUENTIELLES.

Soit une chaîne de Markov à N états définie pour une politique

$$d(k) = \begin{pmatrix} d_1(k) \\ \vdots \\ d_N(k) \end{pmatrix}$$

par la matrice $P^{d(k)}$ de probabilités de transition et par la matrice $G^{d(k)}$ de gains. Cette fois, nous nous intéresserons à un déroulement du processus se prolongeant indéfiniment dans le temps.

4.1. Position du problème.

Nous avons vu, dans § 3, que, si une politique, optimisant les gains à l'instant k-1 précédant l'arrêt, a été choisie, alors il était possible d'optimiser le gain $t_i(k)$ par le choix d'une politique $d_i(k)$; soit :

$$t_i(k) = \sum_{j=1}^{N} p_{ij}^{t_i(k)} \left(g_{ij}^{d_i(k)} + t_j(k-1)\right)$$

$$= q_i^{d_i(k)} + \sum_{j=1}^{N} p_{ij}^{d_i(k)} t_j(k-1).$$

Or (cf. § 2), t(k) prend la forme asymptotique, pour k suffisamment grand :

$$t(k) \sim kb + c \qquad \text{où} \qquad b = \begin{pmatrix} g \\ \vdots \\ g \end{pmatrix} \qquad \text{et} \qquad c = \begin{pmatrix} t_1 \\ \vdots \\ t_N \end{pmatrix}$$

les quantités négligées étant de la forme $\sum_{\lambda_i \in \text{spectre}}^{\Gamma} (\lambda_i)^k K$ où $|\lambda_i| < 1$ et K dépend de P et t(o), donc $t_j(k-1) \sim (k-1)g + t_j$ et

$$t_i(k) \sim kg + t_i \sim q_i^{d_i(k)} + \sum_{j=1}^{N} p_{ij}^{d_i(k)} \left[(k-1)g + t_j\right]$$

$$= q_i^{d_i(k)} + \sum_{j=1}^{N} p_{ij}^{d_i(k)} t_j + (k-1)g$$

Supprimant l'indice k dans ce qui suit indice qui devient indépendant de l'arrêt du processus :

$$\boxed{g + t_i = q_i^{d_i} + \sum_{j=1}^{N} p_{ij}^{d_i}\, t_j \qquad (1)}$$

Pour optimiser $t_i(k)$, il suffit donc d'optimiser $f(i) = q_i^{d_i} + \sum_{j=1}^{N} p_{ij}^{d_i}\, t_j$ et ceci pour tout i, en choisissant parmi les décisions figurant dans S_i, celle qui rend maximum f(i). En effet :

$$q_i^A + \sum_j p_{ij}^A\, t_j + (k-1)g > q_i^B + \sum_j p_{ij}^B\, t_j + (k-1)g$$

$$\Longleftrightarrow q_i^A + \sum_j p_{ij}^A\, t_j > q_i^B + \sum p_{ij}^B\, t_j$$

Il sera nécessaire que cette inégalité ne soit pas douteuse, c'est-à-dire que la différence entre les 2 expressions ne soit pas, en module, supérieure à la somme des erreurs de méthode provenant des termes exponentiels négligés et des erreurs de calcul dans la détermination des valeurs t_j. La méthode qui suit présuppose donc une évaluation préalable des erreurs de méthode.

4.2. Détermination des valeurs t_j.

Pour une politique d donnée, nous aurons N équations du type (1) :

$$g + t_i = q_i + \sum_{j=1}^{N} p_{ij}\, t_j$$

permettant de définir le vecteur limite t (N composantes) et g. Le problème devient un simple problème de résolution de système linéaire à N équations et N inconnues en posant arbitrairement $t_N = 0$ ou $v_j = t_j - t_N$ pour $j = 1..N$

$v_1, v_2, \ldots v_{N-1}, g$ seront déterminés par l'équation vectorielle

$$\begin{pmatrix} g \\ \vdots \\ g \end{pmatrix} + \begin{pmatrix} v_1 \\ \vdots \\ v_{N-1} \\ 0 \end{pmatrix} = \begin{pmatrix} q_1 \\ \vdots \\ q_N \end{pmatrix} + P \begin{pmatrix} v_1 \\ \vdots \\ v_{N-1} \\ 0 \end{pmatrix} \qquad (2)$$

Remarquons que l'optimisation des t_j équivaut à celle des v_j qui ne diffèrent des t_j que par une constante. D'ailleurs, les v_j ont un sens concret plus clair que les t_j. En effet, v_j permet la comparaison, au même instant, des totaux de gains en faxtion de l'état dans lequel se trouve le processus à cet instant car :

$$\left.\begin{aligned} t_j(k) &= kg + t_j \\ t_N(k) &= kg + t_N \end{aligned}\right\} \Longrightarrow t_j(k) - t_N(k) = v_j$$

v_j indique ainsi l'"avantage" que l'on peut tirer d'un passage dans l'état j.

4.3. Méthode itérative.

• Ainsi partant d'une politique arbitraire $d_1 = \begin{pmatrix} d_{11} \\ d_{1i} \\ d_{1N} \end{pmatrix}$, par exemple celle qui optimise $q_i = \sum_{j=1}^{N} p_{ij}\, g_{ij}$, nous déterminons les valeurs v_j sous cette politique au moyen du système (3).

$$(3) \qquad \left\{\begin{aligned} g + v_i &= q_i^{d_{1i}} + \sum_{j=1}^{N} p_{ij}^{d_{1i}}\, v_j \qquad i = 1,\dots N-1 \\ g &= q_N^{d_{1N}} + \sum_{j=1}^{N} p_{ij}^{d_{1N}}\, v_j \end{aligned}\right.$$

• Puis nous chercherons, à l'aide des v_j et g obtenus, le vecteur décision d_2 qui optimise $q_i + \sum_{j=1}^{N} p_{ij}\, v_j$ pour tout i.

Nous calculerons alors les nouvelles valeurs v_j sous la politique d_2 à l'aide de (3), etc...

Moyennant la convergence de d_2 vers d (décision optimale asymptotique), si lors de 2 cycles consécutifs on trouve le même vecteur d_1, alors $d_1 = d$.

En résumé le cycle itératif s'établit ainsi :

1° pour une politique donnée A (p_{ij}^A et q_i^A sont donc donnés) on résout le système (3) en v_i et g ($v_N = 0$)

$$g + v_i = q_i^A + \sum_{j=1}^{N} p_{ij}^A \, v_j$$

On trouve v_j^A et g_A.

2° pour chaque i, on cherche la décision B_i qui optimise

$$q_i + \sum_{j=1}^{N} p_{ij} \, v_j^A$$

ce qui détermine une politique $B((p_{ij}^B),(q_i^B))$, puis 1°...

Il est recommandé, afin de circonscrire la politique asymptotique d'observer ceci : si dans la phase 2° d'un cycle on constate que la décision du cycle précédent fournit la même valeur $q_i + \sum_{j=1}^{N} p_{ij} \, v_j$ que d'autres décisions on conservera la décision antérieure.

4.4. Proposition.

1° La méthode itérative conduit, après un nombre fini de cycles, au vecteur décision limite.

2° Ce vecteur limite représente la politique de meilleur gain.

Lemme 1. Soient deux politiques A et B obtenues consécutivement et dans cet ordre par la méthode itérative. Alors $g^B \geq g^A$.

Considérons les quantités tests calculées lors du cycle k (phase

2°) consécutif au cycle k-1 conclu par le choix de la politique A.

$$q_i^B + \sum_{j=1}^{N} p_{ij}^B \, v_j^A \qquad \text{(politique B)}$$

$$q_i^A + \sum_{j=1}^{N} p_{ij}^A \, v_j^A \qquad \text{(politique A).}$$

Le fait que B soit préférée à A dans le k^{ème} cycle, suppose que,

$$\delta_i = q_i^B - q_i^A + \sum_{j=1}^{N} p_{ij}^B \, v_j^A - \sum_{j=1}^{N} p_{ij}^A \, v_j^A \geq 0 \qquad i = 1,\dots N.$$

Pour chacune des politiques A et B, seront définies les quantités

$$g^A + v_i^A = q_i^A + \sum_{j=1}^{N} p_{ik}^A \, v_j^A$$

$$g^B + v_i^B = q_i^B + \sum_{j=1}^{N} p_{ij}^B \, v_j^B$$

Considérons la différence :

$$g^B - g^A + v_i^B - v_i^A = q_i^B - q_i^A + \sum_{j=1}^{N} p_{ij}^B \, v_j^B - \sum_{j=1}^{N} p_{ij}^A \, v_j^A$$

$$= \delta_i - \sum_{j=1}^{N} p_{ij}^B \, v_j^A + \Sigma \, p_{ij}^A \, v_j^A + \Sigma \, p_{ij}^B \, v_j^B - \Sigma \, p_{ij}^A \, v_j^A$$

$$= \delta_i + \Sigma \, p_{ij}^B \, (v_j^B - v_j^A)$$

Soit

$$\boxed{\Delta g + \Delta \, v_i = \delta_i + \sum_{j=1}^{N} p_{ij}^B \, \Delta \, v_j} \qquad (4) \qquad \underline{i = 1\dots N}$$

Or, nous avons vu que, une chaîne homogène (sous la même politique B) régulière étant donnée, de telles équations (4) admettaient en Δg la forme asymptotique :

$$\Delta g = \sum_{i=1}^{N} \Pi_i^B \, \delta_i \qquad (\delta_i \text{ joue le rôle des } q_i \text{ précédents figurant dans les équations (1)})$$

Π_i^B est la probabilité limite de l'état i.

En effet, (cf. p. 11), les équations (1) sont équivalentes à :

$$t'(k) = kMq + \mathcal{H}(1)q + Dt(0) = kb + c$$

où M est la matrice limite de P^n à N vecteurs lignes, de probabilité, identiques :

$$\begin{pmatrix} \Pi_1 \ldots \Pi_i \ldots \Pi_N \\ \ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots \\ \Pi_1 \ldots\ldots\ldots \Pi_N \end{pmatrix}$$

Mais $\Pi_i^B > 0$, $\forall$ i , i = 1...N, puisque B est régulière.

Or le choix de B au lieu de A signifie qu'au moins l'une des valeurs-tests δ_i est strictement positive. Donc :

$$\Delta_g = \sum_{i=1}^{N} \Pi_i^B \, \delta_i > 0$$

$$\Longrightarrow g^B > g^A$$

Lemme 2. Soit A la politique optimale asymptotique. Alors il n'existe pas de meilleure politique de gain que A. (la fonction de décision est de type croissant).

Supposons qu'il existe B telle que $g^B > g^A$. Puisque le vecteur décision converge vers le vecteur A alors

$$q_i^B + \sum_{j=1}^{N} p_{ij}^B \; v_j^A - q_i^A - \sum_j p_{ij}^A \; v_j^A = \delta_i \leq 0 \qquad i = 1 \ldots N$$

pour toutes les politiques telles que B, envisagées lors de chaque cycle.

Alors : $g^B - g^A = \Delta g = \sum_{i=1}^{N} \Pi_i^B \; \delta_i \leq 0$

d'où : $g^B \leq g^A$ ce qui est contraire à l'hypothèse.

Démonstration de la proposition.

Le processus ne comporte qu'un nombre fini de politiques (pour N états et $k_1, k_2, \ldots, k_N$ décisions pour les états 1,2,...N, il y a au plus $k_1 \times k_2 \times \ldots \times k_N$ politiques différentes). Donc

- Si lors des itérations nous obtenons 2 politiques consécutives identiques, le lemme 2 montre que cette politique finale est celle du meilleur gain et qu'elle est, en même temps, la politique asymptotique des valeurs puisque celle-ci conduit au meilleur gain. Nous arretons donc ici la méthode.

- Il n'est pas possible qu'une même politique apparaisse 2 fois non consécutivement d'après le lemme 1 qui prouve que la fonction gain croît avec le nombre d'itérations.

Ainsi un nombre fini d'itérations conduit à la décision asymptotique.

4.5. Exemples.

4.5.1. Reprenons l'exemple de l'étudiant pouvant inviter deux jeunes filles A et B et espérant maximiser le nombre de danses asymptotiques (!) en recherchant la meilleure politique :

invitation de B : politique (1)

invitation de A : politique (2)

Rappelons que

$$\begin{cases} q_1^{(1)} = \frac{1}{2} \times 17 + \frac{1}{2} \times 13 = 15 \\ q_1^{(2)} = \frac{4}{5} \times 24 + \frac{1}{5} \times 10 = 21{,}2 \text{ —— } \boxed{(2)} \end{cases}$$

$$\begin{cases} q_2^{(1)} = \frac{2}{5} \times 12 + \frac{3}{5} \times 22 = 18 \text{ —— } \boxed{(1)} \\ q_2^{(2)} = \frac{7}{10} \times 17 + \frac{3}{10} \times 18 = 17{,}3 \end{cases}$$

1er cycle : Pour la 1ère phase de la méthode itérative, nous choisirons les gains immédiats maxima : $q_1^{(2)}$ et $q_2^{(1)}$ et déterminerons g et v_1 (avec $v_2=0$)

$$\left.\begin{aligned} g + v_1 &= 21{,}2 + \frac{4}{5} v_1 \\ g &= 18 + \frac{2}{5} v_1 \end{aligned}\right\} \Longrightarrow \boxed{\begin{aligned} v_1 &= 16/3 \\ v_2 &= 0 \end{aligned} \quad g = 20{,}13}$$

Déterminons la politique optimisant $q_i + \sum_j p_{ij} v_j$:

$$\text{état } 1 \begin{cases} \text{politique (1) : } 15 + \frac{1}{2} \times \frac{16}{3} \sim 17{,}6 \\ \quad " \qquad (2) : 21{,}2 + \frac{4}{5} \times \frac{16}{3} \sim 25{,}2 \longrightarrow \boxed{(2)} \end{cases}$$

$$\text{état 2} \begin{cases} \text{politique (1) : } 18 + \frac{2}{5} \times \frac{16}{3} \sim 20{,}1 \\ \text{politique (2) : } 17{,}3 + \frac{7}{10} \times \frac{16}{3} \sim 21 \longrightarrow \boxed{(2)} \end{cases}$$

d'où l'adoption du vecteur décision $\binom{2}{2}$.

__2ème cycle__ : $g + v_1 = 21{,}2 + \frac{4}{5} v_1$ $\qquad v_1 = 4{,}33$

$g = 17{,}3 + \frac{7}{10} v_1 \implies v_2 = 0 \qquad g = 20{,}33$

$$\text{état 1} \begin{cases} \text{politique (1) : } 15 + \frac{1}{2} \times 4{,}33 \sim 17{,}17 \\ \text{politique (2) : } 21{,}2 + \frac{4}{5} \times 4{,}33 \sim 24{,}66 \longrightarrow \boxed{(2)} \end{cases}$$

$$\text{état 2} \begin{cases} \text{politique (1) : } 18 + \frac{2}{5} \times 4{,}33 \sim 19{,}73 \\ \text{politique (2) : } 17{,}3 + \frac{7}{10} \times 4{,}33 \sim 20{,}33 \longrightarrow \boxed{(2)} \end{cases}$$

Nous arrêtons donc ici la méthode puisque le vecteur décision $\binom{2}{2}$ est apparu 2 fois consécutives. Ainsi la décision asymptotique est celle présentée dans le § 3.

__4.5.2.__ Considérons le cas du remplacement d'une voiture usagée. La durée de vie maximum supposée, égale à 10 ans pour tous les véhicules, sera divisée en trimestres. Pour le possesseur d'une automobile, un choix se présentera à la fin de chaque trimestre :

- vendre au prix T_i sa voiture d'âge i ($0 \leq i \leq 40$) et acheter au prix C_j un véhicule d'âge j ($0 \leq j \leq 40$) = politique k = j+2
- ou bien conserver le sien (politique k=1) et s'attendre à dépenser durant le trimestre suivant la somme E_i pour entretenir la voiture.

Ainsi $1 \leq k \leq 41$.

Soit p_i la probabilité pour qu'une voiture d'âge i survive jusqu'à i+1 sans une dépense excessive en réparation. Nous supposerons que toute voiture subissant des dégâts importants, quel que soit son âge, sera immédiatement portée dans l'état 40 et $p_{40} = 0$. Il est évident qu'hélas p_i décroisse avec le temps.

On a donc :

pour k = 1 $\quad g + t_i = -E_i + p_i\, v_{i+1} + (1-p_i)\, t_{40}$

pour k > 1 $\quad g + t_i = T_i - C_{k-2} - E_{k-2} + p_{k-2}\, t_{k-1} + (1-p_{k-2})\, t_{40}$

et l'on retrouve les précédentes notations en posant :

$$q_i^{(k)} = -E_i \quad \text{pour } k = 1 \qquad q_i^{(k)} = T_i - C_{k-2} - E_{k-2} \text{ pour } k > 1$$

$$p_{ij}^{(1)} = \begin{cases} p_i & \text{pour } j = i+1 \\ 1-p_i & \text{pour } j = 40 \\ 0 & \text{pour tout autre } j \end{cases}$$

$$\underset{k>1}{p_{ij}^{(k)}} = \begin{cases} p_{k-2} & \text{pour } j = k-1 \\ 1-p_{k-2} & \text{pour } j = 40 \\ 0 & \text{pour tout autre } j. \end{cases}$$

On a alors à résoudre un problème d'optimisation asymptotique par itérations successives à l'aide de la relation :

$$g + v_i = q_i^{(k)} + \sum_{j=1}^{39} p_{ij}^{(k)}\, v_j \qquad \text{avec } v_{40} = 0$$

$$v_i = t_i - t_{40}$$

Il va de soi qu'un tel calcul est effectué par une machine. Dans Howard ([5]) moyennant la suite des valeurs "américaines" E_i, C_i, T_i, p_i, le problème après 7 itérations conduit à la politique suivante :

"Si nous possédons une voiture de plus de 6 mois et de moins de 6 1/2 ans, gardons-la. Si notre voiture a un âge différent de celui-ci, vendonsla et achetons une voiture de 3 ans que nous garderons jusqu'à l'âge de 6 1/2 ans"... (à ne pas divulguer).

§ 5. PROCESSUS A DECISIONS SEQUENTIELLES AVEC ESCOMPTE.

5.1. Exposé du problème.

Les problèmes précédents n'ont pas tenu compte de la rentabilité des revenus acquis (ou acquérables) à chaque étape du processus. Nous affecterons ces revenus d'un taux d'intérêt : ainsi le total des gains $t_j(n)$ pour n transitions précédant l'arrêt du processus sera différencié d'un même total à percevoir en p transitions ($p \neq n$).

Notons α = valeur au début d'une transition, devenant par le jeu des intérêts 1 à la fin de cette transition (par exemple, si la durée d'une transition est une année, le taux annuel d'intérêt sera : $r = \frac{100(1-\alpha)}{\alpha}$). Donc avec l'interprétation donnée : $\alpha \in [0,1[$.

Nous établirons une formule récurrente du type § 2 - 1 a).

$t_j(n)$ étant le gain total probable en n transitions à venir, si le processus part de l'état j, une étape auparavant (n+1$^{\text{ème}}$) ce gain sera considéré valant $\alpha\ t_j(n)$. Ainsi :

$$t_i(n+1) = \sum_{j=1}^{N} p_{ij}\ (g_{ij} + \alpha\ t_j(n)) = q_i + \alpha \sum_{j=1}^{N} p_{ij}\ t_j(n).$$

Le vecteur gain à l'instant n+1 (n+1 étapes avant la fin du processus) satisfait donc à l'équation vectorielle récurrente :

$$\boxed{t(n+1) = q + \alpha\ P\ t(n)}$$

En prenant les z transformés des 2 membres sous l'hypothèse présentée dans § 2 - 1 c), $\mathcal{D}(z) = \frac{1}{1-az} D$:

$$\boxed{T(z) = \left[I - \alpha zP\right]^{-1} \frac{z}{1-z} P + \left[I - \alpha zP\right]^{-1} t(0)}$$

5.2. <u>Forme asymptotique de $\tau(n)$.</u>

Remarquons que si λ est valeur propre de P alors

$$\det(I - \alpha zP) = 0 \quad \text{pour} \quad z = \frac{1}{\alpha\lambda}$$

1 étant valeur propre simple de P stochastique et régulière, α est valeur propre de αP. Il en est de même pour les autres valeurs propres. D'où :

$$\left[I-zP\right]^{-1} = \frac{M}{1-z} + \frac{1}{1-az} D \implies \left[I- zP\right]^{-1} = \frac{M}{1-\alpha z} + \frac{1}{1-a\alpha z} D$$

et

$$T(z) = \frac{z}{1-z} \; \frac{1}{1-\alpha z} Mq + \frac{z}{1-q} \; \frac{1}{1-a\alpha z} Dq + \frac{1}{1-\alpha z} Mt(0) + \frac{1}{1-a\alpha z} Dt(0)$$

$$= \frac{1}{1-\alpha} \left(\frac{1}{1-z} - \frac{1}{1-\alpha z}\right) Mq + \frac{1}{1-a\alpha} \left(\frac{1}{1-z} - \frac{1}{1-\alpha az}\right) Dq + \frac{1}{1-\alpha z} Mt(0) + $$

$$+ \frac{1}{1-a\alpha z} Dt(0)$$

$$\boxed{\begin{aligned} t(n) = \frac{1}{1-\alpha} Mq + \frac{1}{1-a\alpha} Dq - \frac{1}{1-\alpha} (\alpha)^n Mq - \frac{1}{1-a\alpha} (a\alpha)^n Dq + \\ + \alpha^n Mt(0) + (a\alpha)^n Dt(0) \end{aligned}}$$

et lorsque $n \longrightarrow \infty$, pour $|a| < 1$:

$$t(n) \sim t = \frac{1}{1-\alpha} Mq + \frac{1}{1-a\alpha} Dq.$$

<u>Remarquons</u> que contrairement aux résultats trouvés dans le § 2, t(n) n'est plus fonction affine de n. Bien entendu la convergence est lente car α est

généralement très voisin de 1.

Remarquons également que :

$$t(n) \sim \left[I - zP\right]^{-1} q.$$

5.3. Résolution par la méthode itérative.

Nous procèderons de la même manière que précédemment en utilisant comme politique de départ, celle qui optimise les gains q_i immédiats.

Montrons que les 2 lemmes démontrant l'essentiel de la proposition 4.4. restent vrais, la proposition s'en déduisant immédiatement.

Lemme 1. A et B sont 2 politiques optimisantes consécutives obtenues par la méthode itérative. Alors $N_i^B \geq v_i^A$ pour tout i.

B ayant été préférée à A :

pour tout i : $\delta_i = q_i^B + \alpha \sum_{j=1}^{N} p_{ij} v_j^A - q_i^A - \alpha \sum_{j=1}^{N} p_{ij}^A v_j^A \geq 0.$

Pour les politiques A et B :

$$v_i^A = q_i^A + \alpha \sum_{j=1}^{N} p_{ij}^A v_j^A \quad , \quad v_i^B = q_i^B + \alpha \sum_{j=1}^{N} p_{ij}^B v_j^B$$

alors

$$v_i^B - v_i^A = q_i^B - q_i^A + \alpha \Sigma p_{ij}^B v_j^B - \alpha \Sigma p_{ij}^A v_j^A$$

$$= \delta_i - \alpha \Sigma p_{ij}^B v_j^A + \alpha \Sigma p_{ij}^A v_j^A + \alpha \Sigma p_{ij}^B v_j^B - \alpha \Sigma p_{ij}^A v_j^A$$

$$\Longrightarrow \Delta v_i = \delta_i + \alpha \Sigma p_{ij}^B \Delta v_j$$

relation entre les accroissements de valeurs, identique à celle qui existe entre les valeurs et qui est donc susceptible de la représentation vectorielle :

$$\Delta v = \left[I - \alpha P^B\right]^{-1} \delta$$

Or $\exists$ i tel que $\delta_i > 0$, sinon B n'aurait pas été préférée à A ; de plus aucune ligne de $[I-\alpha P^B]^{-1}$ ne peut être identiquement nulle donc

$$\Delta v > 0 \Longrightarrow \exists i : v_i^B > v_i^A \text{ et pour tout autre indice : } v_i^B \geq v_i^A.$$

Lemme 2. Soit A la politique optimale asymptotique. Alors il n'existe pas de meilleure politique de gain que A.

Supposons qu'il existe B telle que :

$$\exists i \qquad v_i^B - v_i^A > 0$$

A étant la politique optimale asymptotique :

$$\forall i \qquad \delta_i = q_i^B + \alpha \sum p_{ij}^B v_j^A - q_i^A - \alpha \sum p_{ij}^A v_j^A \leq 0$$

Le vecteur δ a donc tous ses éléments ≤ 0 et

$$\Delta v = [I-\alpha P^B]^{-1} \delta \Longrightarrow v_i^B - v_i^A \leq 0 \qquad \forall i$$

d'où la contradiction.

5.4. Exemple : problème du remplacement des automobiles.

En reprenant les mêmes valeurs et $\alpha = 0,97$ (intérêt annuel voisin de 12%), on trouve que la meilleure politique est de conserver sa voiture jusqu'à 6 3/4 ans si elle a plus de 3 ans et racheter par la suite une voiture de 3 ans.

CHAPITRE II

ETUDE DE LA METHODE ITERATIVE DANS LE CAS D'UNE CHAINE MULTI ERGODIQUE

INTRODUCTION

Dans le chapitre I, nous avons limité notre étude des chaînes de Markov aux chaînes admettant une seule classe ergodique ou apériodique (donc non cyclique ou périodique). Nous allons étendre ici les résultats acquis, bénéficiant ainsi d'un cadre moins restrictif d'application aux problèmes concrets ; nous considérerons des chaînes admettant plusieurs classes ergodiques et transitoires :

ex : $$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{2}{5} & \frac{3}{5} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

§ 1. QUELQUES PROPRIETES RELATIVES A LA CONVERGENCE DE P^n.

En note du Chapitre I (p. 5), nous avons rappelé une relation générale donnant P^n, où P^n est une matrice carrée quelconque d'ordre N :

$$P^n = A^n + \sum_{j=1}^{p} \lambda_j^n M_j(n)$$

. λ_j est valeur propre avec l'ordre de multiplicité r_j et $\sum_1^p r_j = N$.

. $A_n = 0$ si $n \geq N$

. $M_j(n)$ est un polynôme matriciel en n de degré $p_j \leq r_j - d_j$ (d_j = dimension du sous-espace propre relatif à λ_j).

Pour toutes les valeurs propres simples : $p_j = 0$. Une valeur propre multiple λ_j a un spectre dit simple si $r_j = d_j$, i.e. si $p_j = 0$. Alors nous aurons les :

<u>Théorème 1. Toutes les valeurs propres de module 1 d'une matrice stochastique sont à spectre simple.</u> (cf. [3], p. 75-76).

C'est une conséquence de la propriété : P^n est à termes bornés.

<u>Application</u> : Si $|\lambda_j| = 1$, alors $M_j(n)$ est de degré 0 en n, donc indépendant de n.

Décomposons alors les p valeurs propres en :

- la valeur propre 1.
- l'ensemble J_1 des valeurs propres différentes de 1, mais de module 1.
- l'ensemble J_2 des valeurs propres de module < 1.

$$P^n = A_n + M_1 + \sum_{j \in J_1} \lambda_j^n M_j(n) + \sum_{j \in J_2} \lambda_j^n M_j(n).$$

<u>Théorème 2. Si $J_1 \neq \emptyset$, P^n ne converge pas simplement mais seulement au sens de Cesaro vers M_1.</u>

<u>Si $J_1 = \emptyset$, P^n converge simplement vers M_1.</u>

La démonstration en est triviale.

Nous savons d'ailleurs que si P est régulière (1 est valeur propre simple), alors M_1 a ses lignes identiques.

Ainsi, si P n'est plus constituée d'une seule classe ergodique, mais n'admet pas d'autre valeur propre de module 1 que 1, t(n) admet une forme asymptotique :

$$t(n) = n\ M\ q + \mathcal{O}(1)\ q + Mt(0)$$

M est stochastique, mais n'a pas ses lignes identiques.

<u>Théorème 3. Si P comporte r classes ergodiques (cycliques ou non), l'ordre de multiplicité r_1 de la valeur propre 1 est égal à r.</u>

P peut toujours être ordonné de la façon suivante :

$$P = \begin{pmatrix} E_1 & & & 0 & \\ & E_2 & & & 0 \\ & 0 & \ddots & E_r & \\ \hline & T_1 & & & T_2 \end{pmatrix}$$

où $E_1, E_2 \ldots E_r$ sont matrices des classes ergodiques du processus donc matrices stochastiques. $\begin{pmatrix} T_1 & T_2 \end{pmatrix}$ représente les vecteurs lignes relatifs aux classes transitoires.

Alors : $\det(\lambda I-P) = \det(\lambda I_1-E_1) \times \ldots \times \det(\lambda I_r-E_r) \times \det(\lambda I_{N-r}-T_2)$ où I_j (resp. I_{N-r}) est matrice de l'application identique de même ordre que E_j (resp. T_2).

Or E_j, j = 1...r, admet 1 pour valeur propre simple [3] donc l'ordre de multiplicité <u>r_1 de 1 est $\geq$ r.</u>

D'autre part, si V est vecteur propre de P relatif à 1 alors $PV = V \Longrightarrow P^n V = V \Longrightarrow M_1 V = V$ où $M_1 = \lim\limits_{n \to \infty} P^n$ (resp. $\lim\limits_{n} \frac{1}{n} \sum\limits_{1}^{n} P^k = M_1$ si P a des classes cycliques), donc M_1 admet les mêmes vecteurs propres relatifs à 1 que P et par conséquent le même ordre de multiplicité pour celle valeur propre.

Or au bout d'un temps suffisamment long la probabilité pour que le processus ait quitté toutes les classes transitoires est 1, donc :

$$M_1 = \left(\begin{array}{ccc|c} E'_1 & & O & \\ & E'_2 & & O \\ O & & \ddots \; E'_r & \\ \hline & T'_1 & & O \end{array}\right)$$

Ainsi $(\lambda I - M_1)$ admet 0 comme racine avec l'ordre de multiplicité $N-r$ et par conséquent $\underline{\lambda = 1}$ est valeur propre <u>d'ordre $\leq r$.</u>

Finalement $r_1 = r$.

<u>Théorème 4</u>. <u>Si P est une matrice monoergodique, ses matrices limites au sens de Cesaro et d'Euler existent, coïncident et sont régulières.</u>

L'existence et l'identité des 2 matrices limites sont des résultats classiques (cf. [1]). De plus si $Q = kI + (1-k)P$, $k \in]0,1[$, alors $q_{ii} = k + (1-k)p_{ii}$ (où q_{ii} et p_{ii} dont des éléments diagonaux de Q et P). $\Longrightarrow q_{ii} \neq 0 \quad \forall\, i = 1 \ldots N$,

donc Q est régulière et par conséquent Q^n converge vers $M_1 = \lim\limits_{n \to \infty} \mathrm{Ces}(P^n)$, matrice stochastique, à lignes identiques.

<u>Remarque</u>. Soit Π le vecteur ligne commun :

$$\Pi Q = \Pi = \Pi\,[kI + (1-k)P] \Longrightarrow (1-k)\Pi P = (1-k)\Pi \Longrightarrow \Pi P = \Pi$$

Mais ce résultat ne préjuge en rien, bien entendu, quant à l'unicité et le caractère "limite" de Π relativement à P (en particulier si P est cyclique).

Considérons maintenant une chaîne non homogène telle que nous l'avons définie dans § 3, chapitre I : il n'est pas nécessaire que les différentes politiques admettent des matrices de transition comportant les mêmes classes ergodiques.

§ 2. OPTIMISATION DES GAINS PAR LA METHODE ITERATIVE.

2.1. Détermination des valeurs asymptotiques.

Soient $Mq = \begin{pmatrix} g_1 \\ \vdots \\ g_i \\ \vdots \\ g_N \end{pmatrix}$ et $\mathscr{P}(1)q + Mt(0) = \begin{pmatrix} t_1 \\ \vdots \\ t_i \\ \vdots \\ t_N \end{pmatrix}$

alors $t_i(k) = kg_i + t_i \qquad i = 1 \ldots N.$

Or comme dans le chapitre I , § 2 :

$$t_i(k+1) = q_i + \sum_{j=1}^{N} p_{ij}\, t_j(k)$$

et en utilisant les valeurs asymptotiques des $t_i(k)$:

$$(k+1)\, g_i + t_i = q_i + \sum_{j=1}^{N} p_{ij}\, (kg_j + t_j)$$

ou

$$k(g_i - \sum_j p_{ij}\, g_j) + t_i + g_i - q_j - \sum_{j=1}^{N} p_{ij}\, t_j = 0$$

relation vérifiée pour tout k, donc :

$$\left\{ \begin{array}{|l|l} \hline g_i = \sum_{j=1}^{N} p_{ij}\, g_j & (1) \\ \hline t_i + g_i = q_i + \sum_{j=1}^{N} p_{ij}\, t_j & (2) \\ \hline \end{array} \right. \qquad \underline{i = 1 \ldots N}$$

soit matriciellement

$$\boxed{\begin{cases} g = Pg & (1) \\ t + g = q + Pt & (2) \end{cases}} \qquad \text{ou} \qquad \boxed{\begin{cases} (I-P)\, g = 0 \\ (I-P)\, t + g = q \end{cases}}$$

Mais (1) $\Longrightarrow$ g est vecteur propre relativement à la valeur propre

1 pour la politique asymptotique ; puisque l'espace propre relatif à 1 a pour dimension l'ordre de multiplicité r de 1, le système linéaire (1) et (2), à 2N équations scalaires, est de rang 2N-r. Il dépend de r arbitraires par exemple, r valeurs t_i prélevées chacune dans une des r classes ergodiques, et que nous égalerons à 0.

Nous retrouvons, en définitive, le cas déjà étudié dans le chapitre I, § 4, où dans une classe ergodique, il suffit de comparer les valeurs t_i à l'une d'entre elles : pour le problème asymptotique qui nous intéresse cette comparaison est suffisante pour nous permettre de choisir une politique à longue échéance puisque le séjour dans une classe transitoire est "probablement" fini et suivi d'une entrée dans une classe ergodique d'où le processus ne sortira pas.

2.2. Méthode itérative pour le choix d'une politique.

Notre but est d'améliorer $t_i(k+1)$ après avoir optimisé à chaque étape $t_j(k)$. Dans l'expression $t_i(k+1)$ pour la politique $d_i(k)$:

$$q_i^{d_i(k)} + \sum_{j=1}^{N} p_{ij}^{d_i(k)} \quad t_j(k)$$

remplaçons $t_j(k)$ par sa forme asymptotique : k g_j+t_j. Il vient

$$q_i^{d_i(k)} + k \sum_{j=1}^{N} p_{ij}^{d_i(k)} \quad g_j + \sum_{j=1}^{N} p_{ij}^{d_i(k)} \quad t_j$$

et nous rendrons optimum cette expression, pour k suffisamment grand, en maximisant le coefficient de k soit

$$f(i) = \sum_{j=1}^{N} p_{ij}^{d_i(k)} \quad g_j$$

Si 2 politiques conduisent à la même valeur f(i), il est nécessaire de choisir celle qui rend maximum la 2ème partie de l'expression donnant $t_j(k+1)$, à savoir :

$$q_i^{d_i(k)} + \sum_{j=1}^{N} p_{ij}^{d_i(k)} \; t_j$$

Nous adopterons alors de nouveau la méthode itérative :

Etant donnée une politique A admettant r_A classes ergodiques :

1° résoudre en g_i et t_i les 2N équations :

$$\begin{cases} g_i = \sum_{j=1}^{N} p_{ij}^{A} \, g_j \\ t_i + g_i = q_i^A + \sum_{j=1}^{N} p_{ij}^{A} \, t_j \end{cases}$$

r_A valeurs t_i prélevées chacune dans une classe ergodique sont nulles.

On trouvera g_i^A et t_i^A $\quad i = 1\ldots N$

2° Pour chaque état, rechercher la décision maximisant :

$\Sigma\, p_{ij}^{d_i} \, g_j^A$, d_i appartient à l'ensemble S_i des décisions possibles en i.

Soit B_i telle que : $\Sigma\, p_{ij}^{B_i} \, g_j^A \geq \Sigma\, p_{ij}^{d_i} \, g_j^A$, $\forall\, d_i \in S_i$

(en cas d'égalité utiliser le test : $q_i^{d_i} + \Sigma\, p_{ij}^{d_i} \, t_j$).

Arreter la méthode dès que la même décision apparaîtra 2 fois consécutives pour tous les états.

<u>Proposition</u>. <u>La méthode itérative conduit en un nombre fini de cycles au vecteur décision limite.</u>

2.3. <u>Lemme 1.</u> <u>Soient 2 politiques A et B obtenues consécutivement dans cet ordre par la méthode itérative.</u> Alors $g^B \geq g^A$.

Le choix de la politique A précède le calcul des valeurs afférentes à A : t_i^A et g_i^A. Dans la phase test suivante, supposons que nous ayons opté pour la politique B. C'est donc que :

$$\varepsilon_i = \sum_{j=1}^{N} p_{ij}^B g_j^A - \sum_{j=1}^{N} p_{ij}^A g_j^A \geq 0$$

ou que si $\varepsilon_i = 0$: $\delta_i = q_i^B + \sum_j p_{ij}^B t_j^A - (q_i^A + \sum p_{ij}^A t_j^A) \geq 0$

Comparons les valeurs g_i et t_i obtenues après choix des politiques A et B. Elles satisfont aux systèmes linéaires suivants :

$$\text{politique A} \quad \begin{cases} g_i^A = \sum_j p_{ij}^A g_j^A \\ g_i^A + t_i^A = q_i^A + \sum_j p_{ij}^A t_j^A \qquad t_i^A = 0 \text{ pour } r_A \text{ indices} \end{cases}$$

$$\text{politique B} \quad \begin{cases} g_i^B = \sum_j p_{ij}^B g_j^B \\ g_i^B + t_i^B = q_i^B + \sum_j p_{ij}^B t_j^B \qquad t_i^B = 0 \text{ pour } r_B \text{ indices} \end{cases}$$

alors $g_i^B - g_i^A = \sum_j p_{ij}^B g_j^B - \sum_j p_{ij}^A g_j^A = \varepsilon_i + \sum_j p_{ij}^B (g_j^B - g_j^A)$

ou

$$\boxed{\begin{array}{c} \Delta g_i = \varepsilon_i + \sum_j p_{ij}^B \Delta g_j \\ i = 1 \ldots N \end{array}} \iff \boxed{\Delta g = \varepsilon + P^B \Delta g} \qquad (3)$$

et de même :

$$g_i^B + t_i^B - (g_i^A + t_i^A) = \delta_i + \sum_j p_{ij}^B (t_j^B - t_j^A)$$

ou

$$\boxed{\Delta g_i + \Delta t_i = \delta_i + \sum_j p_{ij}^B \Delta t_j} \iff \boxed{\Delta g + \Delta t = \delta + P^B \Delta t} \qquad (4)$$

La relation (4) est identique à la relation (2) de § 2, 1 (page 5). Par contre, la relation (3) diffère de la relation (1) du § 2-1 par la présence inopportune de ε_i. Interprétons cette relation (3).

La politique B est par hypothèse représentée par une matrice de transition P^B dont le nombre de classes ergodiques est $r = r_B$ ($r_B \leq N$).

P^B pourra donc se mettre sous la forme suivante, tenant compte de ses classes ergodiques et transitoires :

$$P^B = \left(\begin{array}{c|c|c|c|c} E_1 & 0 & & 0 & 0 \\ \hline 0 & E_2 & & 0 & 0 \\ \hline & & & & \\ \hline 0 & 0 & & E_r & 0 \\ \hline T_1 & T_2 & & T_r & T_{r+1} \end{array}\right)$$

Certains éléments des matrices rectangulaires T_i ($i \leq r+1$) sont différents de 0

Opérons le même découpage sur les vecteurs Δg, Δt, ε , δ , à N composantes Δg_i, Δt_i, ε_i, δ_i, $i = 1 \ldots N$ et notons :

$$\Delta g = \begin{pmatrix} \Delta_1 g \\ \Delta_2 g \\ \vdots \\ \Delta_r g \\ \Delta_{r+1} g \end{pmatrix} \qquad \Delta t = \begin{pmatrix} \Delta_1 t \\ \vdots \\ \Delta_r t \\ \Delta_{r+1} t \end{pmatrix} \qquad \varepsilon = \begin{pmatrix} {}_1\varepsilon \\ \vdots \\ {}_r\varepsilon \\ {}_{r+1}\varepsilon \end{pmatrix} \qquad \delta = \begin{pmatrix} {}_1\delta \\ \vdots \\ {}_r\delta \\ {}_{r+1}\delta \end{pmatrix}$$

Par exemple, le vecteur $\Delta_i g$ est le vecteur d'écart de gain relatif à la $i^{ème}$ classe.

Soit $\Pi = ({}^1\Pi \; {}^2\Pi \; \ldots \; {}^r\Pi \; {}^{r+1}\Pi)$ le vecteur dont chaque élément est un vecteur ligne extrait des vecteurs lignes de la matrice $M_1^B = \lim\limits_n (P^B)^n$ ($= \lim\limits_n \mathrm{Ces}(P^B)^n$ cf. p. 37). ${}^i\Pi$ sera le vecteur stochastique ligne, extrait de M_i, limite de E_i (ou de $\mathcal{E}_i = kI + (1-k)E_i$) (il est le même d'ailleurs pour tous les éléments d'une même classe ergodique), si le processus est parti d'un état quelconque de la ième classe ergodique. On a donc :

$$^{i}\Pi = {}^{i}\Pi \, E_i \qquad \text{(cf. théo. 4, § 1).}$$

$^{r+1}\Pi$ sera nul puisque tout départ d'une classe transitive conduit à une transition dans une classe ergodique.

Les équations (3) et (4) prennant la nouvelle forme :

$$(3) \Longleftrightarrow \begin{cases} (3'_1) & \Delta_i g = {}_i\varepsilon + E_i \, \Delta_i g \qquad i = 1 \ldots r \\ (3'_2) & \Delta_{r+1} g = {}_{r+1}\varepsilon + \sum_{k=1}^{r+1} T_k \, \Delta_k g \end{cases}$$

$$(4) \Longleftrightarrow \begin{cases} (4'_1) & \Delta_i g + \Delta_i t = {}_i\delta + E_i \, \Delta_i t \qquad i = 1 \ldots r \\ (4'_2) & \Delta_{r+1} g + \Delta_{r+1} t = {}_{r+1}\delta + \sum_{k=1}^{r+1} T_k \, \Delta_k t \end{cases}$$

a) Amélioration des gains des classes ergodiques.

Multiplions $(3'_1)$ à gauche par $^{i}\Pi$:

$$^{i}\Pi \, \Delta_i g = {}^{i}\Pi \, {}_i\varepsilon + {}^{i}\Pi \, E_i \, \Delta_i g = {}^{i}\Pi \, {}_i\varepsilon + {}^{i}\Pi \, \Delta_i g$$

$$\Longrightarrow {}^{i}\Pi \, {}_i\varepsilon = 0 \qquad i = 1 \ldots r$$

Or $^{i}\Pi$ en tant que vecteur limite d'une chaîne régulière n'a pas d'élément nul $\Longrightarrow {}_i\varepsilon = 0$, $i = 1 \ldots r$

$\Longrightarrow \varepsilon_i = 0$ pour tous les états des r classes ergodiques.

Il s'ensuit que, pour tous ces états, notre décision permettant le choix de B_i (décision relative à l'état i) au lieu de A_i provient du 2ème test : celui des valeurs et non celui des gains.

Les équations $(3'_1)$ prennent la forme :

$$\Delta_i g = E_i \, \Delta_i g$$

- Or de telles équations vectorielles, où E_i est matrice d'une chaîne ergodique régulière, admettent la solution :

$\Delta_i g_j = 0$ ou bien $\Delta_i g_j = \alpha_i$ $\quad \forall j$ ensemble des indices de la $i^{ème}$ classe.

En effet, l'équation vectorielle

$$Pv = v \Longrightarrow P^n v = v \Longrightarrow (\lim P^n)\, v = Mv = v$$

où limite a le sens ordinaire de Cesaro et où M a ses lignes identiques $\Longrightarrow$ v a ses composantes égales.

Ainsi pour tous les états d'une même classe ergodique, on obtient le même accroissement de gain en préférant B à A.

Remarquons alors que dans $(4'_1)$:

$$^i\Pi\ \Delta_i g + {}^i\Pi\ \Delta_i t = {}^i\Pi\ {}_i\delta + ({}^i\Pi\ E_i)\ \Delta_i t$$
$$= {}^i\Pi\ {}_i\delta + {}^i\Pi\ \Delta_i t$$

mais $^i\Pi\ \Delta_i g = {}^i\Pi\ E_i\ \Delta_i g = \Delta_i g$

$$\Longrightarrow \Delta_i g = {}^i\Pi\ {}_i\delta.$$

$$\left.\begin{array}{l}\text{Or B préférée à A}\\ \text{et } {}_i\varepsilon = 0\end{array}\right\} \Longrightarrow {}_i\delta \geq 0 \text{ pour toutes les classes ergodiques.}$$

donc $\Delta_i g$ a toutes ses composantes positives ou nulles (à moins que A et B ne soient équivalentes i.e. que ${}_i\delta = 0$). Nous allons fournir 2 conditions suffisantes d'amélioration stricte du gain pour tous les états des classes ergodiques par la méthode itérative.

Cas particuliers. 1) Remarquons que si toutes les politiques admettent les mêmes r classes ergodiques :

$$g_i = \sum_{j=1}^{N} p_{ij}^A\, g_j = \sum_{j\in J_i} p_{ij}^A\, g_j$$

où J. est l'ensemble d'indices relatif aux états de E.. Et de même :

A £ A +•

A y A
= •. + . , p.. g.

La politique globale A est decomposable en décisions indépendan-

t e s A » «A

l"'" V pour les r classes ergodiques et en une décision A p+,| (dépen-

dant par contre des valeurs t^ et trouvés lors de l'optimisation des

valeurs tests des classes ergodiques].

Ainsi pour optimiser globalement le processus, il suffit d'opti-

miser séparément les r sous-processus ergodiques fournissant J^+„.+J^ com-

posantes dans le vecteur décision asymptotique.

Nous savons alors (cf. chap. I, 5 4, lemme 1) que :

- Si E....E sont régulières alors *i*
1 r

A.g > G i - 1. . „r
i

- S'il existe i tel que E^ soit régulière alors :
A > 0

2)Si toutes les politiques n'admettent que des classes ergodiques,

la fonction gain est strictement croissante d'un cycle au suivant.

En effet, si B est une politique préférée à A lors d'un cycle :

- = 0 pour tout i

- ^6 > 0 pour au moins un indice

B A
^ àg > 0 ^ g > g .

b) Amélioration des sains des états transitoires.

Les équations (3^) s'écrivent :

$$(I_{r+1} - T_{r+1})\ \Delta_{r+1}g = {}_{r+1}\varepsilon + \sum_{k=1}^{r} T_k\ \Delta_k g$$

Or la valeur propre 1 (d'après § $, théorème 3) n'étant relative qu'aux matrices ergodiques :

$I_{r+1} - T_{r+1}$ est inversible et :

$$\Delta_{r+1}g = (I_{r+1} - T_{r+1})^{-1}\ ({}_{r+1}\varepsilon + \sum_{k=1}^{r} T_k\ \Delta_k g).$$

On montre ([5]) qu'une matrice telle que $(I_{r+1} - T_{r+1})^{-1}$ a tous ses éléments positifs ou nuls. Cependant, ancune de ses lignes ne peut être identiquement nulle. ($I_{r+1} - T_{r+1}$ est une application régulière).

Nous savons de plus que ${}_{r+1}\varepsilon$ a tous ses éléments non négatifs et qu'il en est de même pour T_k et $\Delta_k g$ (pour $k = 1...r$). Donc $\Delta_{r+1}g$ a tous ses éléments non négatifs. Précisons :

- Si pour une seule classe ergodique ${}_i\delta > 0$ alors $\Delta_i g > 0$ et les états de la $i^{ème}$ classe ergodique pouvant être atteints à partir des états transitoires le vecteur $T_i\ \Delta_i g$ est positif ; par conséquent $\Delta_{r+1}g > 0$ (au moins une de ses composantes est positive) donc $g^B > g^A$.

- Si pour toutes les classes ergodiques ${}_i\delta = 0$ donc $\Delta_{ig} = 0$ alors

$$\Delta_{r+1}g = (I_{r+1} - T_{r+1})^{-1}\ {}_{r+1}\varepsilon.$$

Si le premier test a suffi pour choisir B, alors ${}_{r+1}\varepsilon$ a au moins une composante > 0 et $g^B > g^A$.

Si le 2ème test est nécessaire (${}_{r+1}\varepsilon \equiv 0$), on ne peut pas conclure. (à moins que cette éventualité soit absurde de qui entraînerait le choix de B par 1er test d'où $g^B > g^A$).

<u>Remarque</u>. Dans l'hypothèse 1) du a) précédent

$$\sum_{k=1}^{r} T_k \, \Delta_k g > 0 \Longrightarrow \Delta_{r+1} g > 0$$

et par conséquent $g^B > g^A$. Autrement dit, la méthode itérative améliore strictement les gains.

2.4. <u>Lemme 2.</u> <u>La méthode itérative conduit à la meilleure politique</u> (autrement dit, il n'existe pas de meilleure politique de gains que la politique asymptotique).

Soit A la politique asymptotique et supposons qu'une politique B conduise à un meilleur gain, dans un état quelconque k, que la politique A.

Alors A étant politique asymptotique :

$$\begin{cases} \varepsilon_i = \sum_{i=1}^{N} p_{ij}^B \, g_j^A - \sum_{i=1}^{N} p_{ij}^A \, g_j^A \leq 0 \\ \text{si } \varepsilon_i = 0 \text{ alors } \delta_i \leq 0 \end{cases}$$

Mais si k est un état ergodique :

$\Delta_k g = {}^k\Pi \; {}_k\delta$, relation dans laquelle $\Delta_k g$ a ses éléments > 0 par hypothèse, alors que ${}_k\delta$ a ses composantes ≤ 0, ce qui est incompatible.

Si k est un état transitoire :

$$\Delta_{r+1} g = (I_{r+1} - T_{r+1})^{-1} \left({}_{r+1}\varepsilon + \sum_{1}^{r} T_p \, \Delta_p g \right)$$

prouve que ${}_{r+1}\varepsilon$ et $\Delta_p g$ ayant leurs composantes ≤ 0, $\Delta_{r+1} g$ a tous ses éléments non positifs, ce qui est encore incompatible avec $\Delta g_k > 0$.

Ainsi B ne peut conduire à de meilleurs gains quels que soient les états.

<u>Démonstration de la proposition</u>. Dans le cas où la fonction gain est strictement croissante, la démonstration est identique à celle du chapitre 1, § 4.

BIBLIOGRAPHIE

R. BELLMANN - Dynamic Programming 1957.

R. BELLMANN - Adaptive control processes.

R. BELLMANN - Introduction to matrix analysis 1960.

R. BELLMANN et DREYFUS - Programmation dynamique et ses applications 1965.

[1] BOCLE - Cours de Probabilités II - 1962-63.

FEL'BAUM - Optimal control systems.

FORTET - Algèbres des tenseurs et des matrices. pb. de valeurs propres, C.D.U. 1961.

[2] FREEMAN - Discrete. time systems 1965.

[3] GORDON - Théorie des chaînes de Markov finies et ses applications. 1965.

[4] HENNEQUIN ET TORTRAT - Théorie des probabilités et quelques applications 1965.

J. HERNITER ET J. MAGEE - Customer behavior as a Markov process (operations Research, 1961, Volume 9).

[5] HOWARD - Dynamic programming and Markov processes 1960.

MORLAT - Statistique et théorie de la décision (Math. et S. humaines 1964).

NEVEU - Calcul des probabilités 1965.

SOURIAN - Calcul linéaire (P U F 1959).