REVUE DE STATISTIQUE APPLIQUÉE

VIDAL COHEN

Processus à états marquants

Revue de statistique appliquée, tome 18, n° 3 (1970), p. 87-99

<http://www.numdam.org/item?id=RSA_1970__18_3_87_0>

PROCESSUS A ÉTATS MARQUANTS

Vidal COHEN

I - INTRODUCTION

On sait qu'un processus de Markov est un processus stochastique dans lequel "le futur ne dépend du passé que par l'intermédiaire du présent". Il est facile de ramener à un tel processus tout processus dans lequel le futur ne dépendrait du passé que par l'intermédiaire du présent et d'un passé récent de durée fixée. Si cette durée n'est pas limitée, on aboutit à un processus à liaisons complètes.

Cependant certains processus pourraient se situer entre ces deux extrêmes. Le fait de passer par un état que nous qualifierons de marquant peut avoir une influence sur les transitions successives même longtemps plus tard. Nous signalons plus loin des applications de tels processus.

Définitions et notations

Soit un processus stochastique discret $\{X_n \mid n \geqslant 1\}$ défini sur un espace de probabilité $(\Omega, \mathfrak{A}, \mathfrak{P})$ et à valeurs dans un "espace des états" J. Nous supposerons qu'il existe une partition de J en

$$J = \mathcal{O} + \mathfrak{M}$$

$\mathcal{O}$: ensemble des états que nous qualifions d'"ordinaires"
$\mathfrak{M}$: ensemble des états que nous qualifions de "marquants".

Soient $t_1 < t_2 \ldots < t_n$ les instants d'observation du processus.

A ces instants correspondent des variables aléatoires

$$X_1, X_2, \ldots X_n$$

Rappelons que le processus est une chaîne simple de Markov (à temps discret), si la probabilité conditionnelle

$$P\left(X_n = i_n \;\middle|\; \begin{array}{l} X_k = i_k \text{ pour } k = 1,2, \ldots, n-1 \\ i_k \in J \quad \text{pour } k = 1,2, \ldots, n \end{array}\right)$$

n'est en réalité fonction que de i_{n-1} et de i_n. On construit alors la matrice de transition $[p_{ij}(n)]$. Si elle ne dépend pas de n, la chaîne est dite stationnaire.

Nous dirons par convention que le processus est une chaîne à états marquants si à l'instant t_n

$$P\left(X_{n+1} = i_{n+1} \;\middle|\; \begin{array}{l} X_k = i_k \text{ pour } k = 1,2, \ldots, n \\ i_k \in J \quad \text{pour } k = 1,2, \ldots, n+1 \end{array}\right) =$$

$$P\left(X_{n+1} = i_{n+1} \;\middle|\; \begin{array}{l} X_n = i_n \\ X_k = i_k \text{ pour } k = \sup \left\{ j \;\middle|\; \begin{array}{l} j \leqslant n \\ i_j \in \mathcal{M} \end{array} \right\} \end{array}\right)$$

Si cette dernière probabilité ne dépend pas de l'instant t_n, alors la chaîne sera dite stationnaire à états marquants. Dans une telle chaîne, nous supposerons, autrement dit, que, sachant qu'on est en i à un instant donné, la probabilité de passer à un certain état j à l'instant suivant ne dépend pas effectivement de tout ce qui a précédé mais seulement du dernier état marquant visité, ce dernier état marquant pouvant être éventuellement l'état i dans lequel on se trouve à l'instant donné. S'il y a stationnarité, il est inutile, quand on commence à étudier le processus, de savoir quand il a débuté. Il est par contre nécessaire de savoir quel est l'état marquant le plus récemment visité.

Si le processus est stationnaire, discret et si

- $\mathcal{O}$ est constitué d'un nombre fini k d'états ordinaires

1,2 ... k

- $\mathcal{M}$ est constitué de deux états marquants a, b, on construira les matrices de transition

$$P^{(a)} = \begin{bmatrix} \alpha_{11} & \cdots\cdots & \alpha_{1k} & \alpha_{1a} & \alpha_{1b} \\ & & & & \\ \alpha_{k1} & \cdots\cdots & \alpha_{kk} & \alpha_{ka} & \alpha_{kb} \\ \alpha_{a1} & \cdots\cdots & \alpha_{ak} & \alpha_{aa} & \alpha_{ab} \\ 0 & \cdots\cdots & 0 & 0 & 0 \end{bmatrix}$$

$$P^{(b)} = \begin{bmatrix} \beta_{11} & \cdots\cdots & \beta_{1k} & \beta_{1a} & \beta_{1b} \\ & & & & \\ \beta_{k1} & \cdots\cdots & \beta_{kk} & \beta_{ka} & \beta_{kb} \\ 0 & \cdots\cdots & 0 & 0 & 0 \\ \beta_{b} & \cdots\cdots & \beta_{bk} & \beta_{ba} & \beta_{bb} \end{bmatrix}$$

$P^{(a)}$ et $P^{(b)}$ étant substochastiques.

"sous l'influence de a" ou "dans (a)" signifiera : "a étant le dernier état marquant visité".

Conséquence de cette définition

A un processus à états marquants, on associe immédiatement un processus de Markov $\{Z_n \mid n \geqslant o\}$ dont l'espace des états est $\mathcal{M} \times J$. Mais ce processus n'est que partiellement observable : la composante $Y_n \in \mathcal{M}$ n'est observable que si la composante X_n dans J se trouve aussi être dans $\mathcal{M}$ au temps t_n.

Ici deux états seront entièrement observables :

$$(a, a) \quad \text{et } (b, b)$$

Par contre (a, b) et (b, a) n'existent pas puisque si b est l'état observable actuel il est aussi le plus récent état marquant visité. Nous n'utiliserons pas explicitement dans la suite la matrice de transition associée à $\{Z_n\}$ peu pratique (et qui se prête mal aux généralisations au cas d'ensembles $\mathcal{O}$ et $\mathcal{M}$ infinis).

Remarque : plusieurs jeux se rattachent à des processus à états marquants. Ainsi le jeu d'échecs : on sait qu'un joueur n'a plus le droit de roquer s'il a déplacé précédemment le roi ou sa tour, même si les pièces ont, depuis, regagné leurs places initiales. Donc si le roi est hors de sa place initiale, le jeu est dans un état marquant. Il en restera une trace éventuellement pendant toute la partie (tant que les tours ne seront pas perdues !) en ce sens qu'à chaque instant, les possibilités du joueur se trouvent modifiées (puisqu'il est privé ou non de la possibilité de roquer). Si, arrivé au cours d'une partie, un observateur voit un joueur se priver de roquer alors que cela semblerait avantageux, l'observateur pourra en inférer qu'il a déplacé son roi ou sa tour et qu'il est sous l'influence d'un certain état marquant (il pourra être difficile ou même impossible de préciser lequel).

II - UN PROBLEME DE DECISION DANS UN PROCESSUS A ETATS MARQUANTS

Problème : Soit un processus à états marquants (stationnaire, discret) dont on connait les états ordinaires, les états marquants et les matrices de transition associées. On suppose qu'au temps T on aura à prendre une décision qui sera d'autant mieux adaptée que sera meilleure la prévision de l'état du processus au temps T + 1.

Observer le processus au temps T seulement fait courir le risque de ne pas savoir quelle matrice de transition utiliser pour la prévision de l'état au temps T + 1.

Pour diminuer ce risque il apparait qu'on doit observer le processus avant le temps T.

Le problème est alors de déterminer l'entier n minimum tel que l'observation depuis le temps T - n abaisse au-dessous d'un seuil ε (> o, donné) le risque d'avoir à prendre une décision au temps T sans connaître l'état marquant qui le régit.

1) Cas de deux états marquants a et b

Les matrices de transition associées $P^{(a)} = (\alpha_{ij})$ et $P^{(b)} = (\beta_{ij})$ sont supposées connues (d'ordre $k + 2$, s'il y a k états ordinaires). L'état à un instant t est un couple

Z_t à deux composantes X_t et Y_t :

X_t composante observable qui peut prendre les valeurs : 1,2 ... i, ... k, a, b.

Y_t composante non observable (sauf si X_t vaut a ou b) qui peut prendre les valeurs a ou b.

Le problème précédent est donc de découvrir Y_T.

A - Méthode Bayésienne

Le théorème de Bayes donne

$$P\left(Y_t = a / X_t = i\right) = \frac{P\left(X_t = i / Y_t = a\right) P\left(Y_t = a\right)}{P\left(X_t = i / Y_t = a\right) P\left(Y_t = a\right) + P\left(X_t = i / Y_t = b\right) P\left(Y_t = b\right)} \quad (1)$$

et bien sûr

$$P\left(Y_t = b / X_t = i\right) = 1 - P\left(Y_t = a / X_t = i\right)$$

Il est évident en particulier que

$$P\left(Y_t = a / X_t = a\right) = P\left(Y_t = b / X_t = b\right) = 1$$

$$P\left(Y_t = a / X_t = b\right) = P\left(Y_t = b / X_t = a\right) = 0$$

et aussi que s'il est impossible de passer de a à i (càd)

$$P\left(X_t = i / Y_t = a\right) = 0$$

$$\text{alors } P\left(Y_t = a / X_t = i\right) = 0$$

Dans les autres cas, (1) n'est pas d'un grand secours. Supposons cependant, qu'on observe s étapes successives du temps t au temps $t + s - 1$ et que n'apparaissent que des états ordinaires :

$$\underbrace{i\ j\ k\ \ldots\ cf}_{s \text{ éléments}}$$

Alors le théorème de Bayes nous donne (2) :

$$P\left(Y_{t+s-1} = a \Big/ \begin{matrix} X_t = i \\ X_{t+1} = j \\ \vdots \\ X_{t+s-1} = f \end{matrix} \right) = \frac{\alpha_{ij}\alpha_{jk} \dots \alpha_{cf} P\left(X_t = i / Y_t = a\right) P\left(Y_t = a\right)}{\alpha_{ij}\alpha_{jk} \dots \alpha_{cf} P\left(X_t = i / Y_t = a\right) P(Y_t = a) + \beta_{ij} \dots \beta_{cf} P\left(X_t = i / Y_t = b\right) P(Y_t = b)} \quad (2)$$

Au second membre de (2) les probabilités à priori

$$P(Y_t = a) \quad \text{et} \quad P(Y_t = b)$$

sont généralement mal connues.

Le terme $P\left(X_t = i / Y_t = a\right)$ désigne la probabilité, partant de a, d'atteindre l'état ordinaire i sans passer par l'autre état marquant b.

Il pourra être évalué si l'on sait à quelle date τ on est passé par le dernier état marquant a. Si l'on ne connait pas τ mais que l'on sait que le processus a débuté depuis longtemps, on pourra évaluer directement le produit

$$P\left(X_t = i / Y_t = a\right) P(Y_t = a) = P\begin{pmatrix} X_t = i \\ Y_t = a \end{pmatrix} = P\left(Z_t = (i, a)\right)$$

en adoptant pour ce terme la probabilité limite (si elle existe) d'atteindre (i, a) dans le processus de Markov $\{Z\}$.

Ces méthodes ne sont pas satisfaisantes et l'on se heurte aux difficultés classiques de détermination des probabilités à priori. (elles ne seraient levées que dans des cas particuliers)

Cependant les termes

$$A_s = \alpha_{ij}\alpha_{jk} \dots \alpha_{cf} \text{ et } B_s = \beta_{ij}\beta_{jk} \dots \beta_{cf}$$

sont connus.

Si l'on pose alors

$$x_a = P\begin{pmatrix} X_t = i \\ Y_t = a \end{pmatrix} \text{ et } x_b = P\begin{pmatrix} X_t = i \\ Y_t = b \end{pmatrix}$$

on aura

$$y = P\left(Y_{t+s-1} = a \Big/ \begin{matrix} X_t = i \\ \vdots \\ X_{t+s-1} = f \end{matrix} \right) = \frac{A_s x_a}{A_s x_a + B_s x_b} \quad (2)$$

où $x_a \geqslant o$, $x_b \geqslant o$, $x_a + x_b \leqslant 1$

A_s et B_s ne peuvent être nuls simultanément. Sinon, on ne pourrait pas avoir observé la séquence i j k ... cf.

a) Cas où $A_s . B_s = o$ (càd l'un de ces deux termes est nul, et l'un seulement).

Si l'un de ces deux termes et nul, le problème est résolu puisque y vaut o ou 1, quelles que soient les valeurs attribuées à x_a et x_b (pourvu qu'elles ne contredisent pas les valeurs trouvées pour A_s et B_s)

b) Cas où $A_s . B_s \neq 0$. Nous supposerons également $x_a \neq 0$. Les probabilités à priori (mal connues) x_a et x_b jouent alors un rôle essentiel et déterminent y.

Etudions l'effet d'une indétermination sur x_a et x_b (étude différentielle)

1) Si x_a et x_b sont indépendants (au sens de l'analyse).

$$dy = \frac{\partial y}{\partial x_a} dx_a + \frac{\partial y}{\partial x_b} dx_b$$

or $$\frac{\partial y}{\partial x_a} = \frac{A_s B_s x_b}{(A_s x_a + B_s x_b)^2} = \frac{\lambda x_b}{(\lambda x_a + x_b)^2} \text{ en posant } \lambda = \frac{A_s}{B_s}$$

Comparons $\frac{\partial y}{\partial x_a}$ à 1 :

$$\frac{\partial y}{\partial x_a} > 1 \Longleftrightarrow \lambda^2 x_a^2 + \lambda (2\, x_a x_b - x_b) + x_b^2 < o$$

Calculons le discriminant Δ de ce trinôme en λ

$$\Delta = x_b^2 (1 - 4\, x_a)$$

Donc : si $x_a > \frac{1}{4}$, $\frac{\partial y}{\partial x_a} < 1 \quad \forall \lambda$

si $x_a \leqslant \frac{1}{4}$, ce trinôme s'annule pour

$$\lambda' = \frac{x_b}{2\, x_a^2} \left[- 2\, x_a + 1 - \sqrt{1 - 4\, x_a} \right] = \frac{4\, x_b}{(1 + \sqrt{1 - 4\, x_a})^2}$$

$$\lambda'' = \frac{4\, x_b}{(1 - \sqrt{1 - 4\, x_a})^2}$$

On voit facilement que si (Cas I)

$$x_a \frac{1 - \sqrt{1 - 4\, x_a}}{1 + \sqrt{1 - 4\, x_a}} \leqslant x_b \leqslant x_a \frac{1 + \sqrt{1 - 4\, x_a}}{1 - \sqrt{1 - 4\, x_a}}$$

on a la disposition

$$\begin{array}{cccc} 0 & \lambda' & 1 & \lambda'' \end{array} \longrightarrow$$

si (Cas II) :

$$x_b < x_a \frac{1 - \sqrt{1 - 4\,x_a}}{1 + \sqrt{1 - 4\,x_a}}, \ \frac{x_b}{x_a} < 1 \text{ donc } \lambda' \lambda'' < 1 \ ;$$

alors on a la disposition

$$\begin{array}{cccc} 0 & \lambda' & \lambda'' & 1 \end{array} \longrightarrow$$

si (Cas III)

$$x_b > x_a \frac{1 + \sqrt{1 - 4\,x_a}}{1 - \sqrt{1 - 4\,x_a}}, \ \frac{x_b}{x_a} > 1 \text{ donc } \lambda' \lambda'' > 1$$

$$\begin{array}{cccc} 0 & 1 & \lambda' & \lambda'' \end{array} \longrightarrow$$

Interprétation :

1) Si $x_a > 1/4$ l'utilisation de la formule de Bayes "affaiblit l'incertitude" sur y par rapport à celle sur la probabilité à priori x_a quelle que soit la séquence observée (càd $\forall \lambda$).

2) Si $x_a \leqslant 1/4$, la séquence observée ijk... cf joue un rôle important. Par exemple, dans le cas I :

$$\frac{\partial y}{\partial x_a} > 1 \qquad \text{si} \qquad \lambda' < \frac{A_s}{B_s} < \lambda'' \qquad (\text{avec } o < \lambda' < 1 < \lambda'')$$

cela signifie que si la séquence observée n'est pas nettement spécifique d'un état marquant, la formule de Bayes aggrave l'incertitude sur la nature du dernier état marquant visité :

Dans le cas III, la disposition

$$\begin{array}{cccc} 0 & 1 & \lambda' & \lambda'' \end{array} \longrightarrow$$

s'interprète comme suit : quand A_s/B_s dépasse λ' sans dépasser λ'', cela signifie que la séquence observée semble plutôt régie par l'état a alors qu'à priori on pensait que c'est b qui était le dernier état marquant visité. (Il faut d'ailleurs noter que λ' et λ'' sont proportionnels à x_b). Cette contradiction entre opinion à priori et observation fait que l'incertitude sur y se trouve accrue (puisque $\partial y/\partial x_a > 1$).

Cependant au-delà de λ'', la séquence observée apparait comme suffisamment spécifique de l'état a pour "contrebalancer" le choix à priori et diminuer l'incertitude sur y résultant d'une méconnaissance de x_a.

L'étude de l'incertitude sur x_b est analogue.

2) Supposons x_a et x_b liés (par exemple), par la relation $x_a + x_b = C$ (constante $C \leqslant 1$)

Dans ces conditions

$$\frac{dy}{dx_a} = \frac{A_s B_s C}{[(A_s - B_s)\, x_a + B_s C]^2}$$

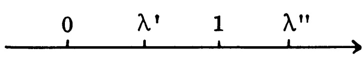
en posant encore $\frac{A_s}{B_s} = \lambda$ on voit que

$$\frac{dy}{dx_a} < 1 \Longleftrightarrow \lambda^2 x_a^2 + \lambda(2\, x_a x_b - x_a - x_b) + x_b^2 > o$$

$$\Delta = (x_a + x_b)\ (x_a + x_b - 4\, x_a x_b) = C(C - 4\, x_a x_b)$$

On voit que $\Delta > o \quad \forall\ x_a,\ x_b$ distincts vérifiant $x_a + x_b = C$.
et que Δ ne s'annule que pour $x_a = x_b = \frac{1}{2}$.

On a la disposition



(λ' et λ'' se confondant en 1 seulement pour $x_a = x_b = 1/2$)

Ces résultats s'interprètent comme plus haut en termes d'influence de l'incertitude concernant x_a sur l'incertitude concernant y, suivant la séquence observée. En particulier, si à priori $x_a = x_b \neq 1/2$ il pourra exister des séquences dont l'observation conduit à une aggravation de l'incertitude, puisque $\lambda' \neq \lambda''$. Au contraire si $x_a = x_b = 1/2$ alors $\lambda' = \lambda'' = 1$ ce qui veut dire que seule une séquence (s) vérifiant $\frac{A_s}{B_s} = 1$ ne diminuera pas l'incertitude sur y résultant d'une incertitude sur x_a. (S'il apparait une séquence (s) vérifiant $\frac{A_s}{B_s} = 1$ on peut dire qu'aucun élément nouveau n'est intervenu "en faveur" de l'un ou l'autre des états marquants a, b : $\frac{dy}{dx_a} = 1$).

Remarques :

1) La relation de liaison $x_a + x_b = C$ a été choisie arbitrairement. Elle serait vérifiée par exemple si l'on connaissait la date de début du processus et que l'on savait qu'au temps t

$$P\left(X_t = i / Y_t = a\right) = P\left(X_t = i / Y_t = b\right) = C$$

2) Supposer que $C = 1$ pour tout $t > t_o$ signifierait que le processus va s'enfermer dans i. Nous avons seulement supposé $C = 1$ au temps t (ce qui, pratiquement, nécessite un processus assez particulier et que l'on connaîtrait bien).

3) Quand $x_a = x_b = \frac{C}{2} < \frac{1}{2}$, la zône ($\lambda'$, λ'') ne se réduit pas au point $\lambda = 1$. L'explication est la suivante : Si la séquence observée (s) est peu spécifique de l'un ou l'autre état, l'incertitude sur y est plus grande que l'incertitude sur x_a ; c'est qu'il s'agit d'incertitude absolue et qu'au temps T on "répartit" entre

$$y = P(Y_T = a) \text{ et } \bar{y} = P(Y_T = b) = 1 - y$$

une probabilité égale à 1 et donc plus grande, quand $C < 1$, que celle que l'on répartit entre x_a et x_b.

Cela met l'accent sur la différence de nature de x_a et x_b d'une part, y et $\bar{y}$ d'autre part. Les probabilités à priori x_a et x_b portent sur la valeur du couple (X_t, Y_t) où $t = T-s+1$. y et $\bar{y}$ portent sur Y_T puisque X_T est observé. Cependant l'étude de leurs variations relatives serait moins adaptée au problème traité ici.

Dans le cas présent si l'on s'intéresse au rapport des incertitudes relatives on trouve

$$\frac{dy}{dx_a} \cdot \frac{x_a}{y} = \frac{2}{\lambda + 1} \qquad \text{(quand } x_a + x_b = C)$$

Comme prévu ce rapport vaut 1 pour $\frac{A_s}{B_s} = 1$.

Si $\lambda < 1$, à un accroissement relatif de x_a correspond un accroissement relatif plus grand de y qui peut s'interpréter comme une augmentation de la "non-concordance" entre probabilités à priori et séquence observée. A une diminution relative de x_a correspond une diminution relative de y ; en valeurs absolues celle sur y est supérieure à celle sur x : cette fois la diminution de x_a va dans le sens de l'observation de (s) qui a conduit à $\lambda < 1$. Si $\lambda > 1$, il peut y avoir encore renforcement ou contradiction entre l'"à priori" et l'"observé" mais le rapport des accroissements relatifs de x_a et y est inférieur à 1.!

B - Cas où l'on ne sait rien des probabilités à priori

Supposons maintenant qu'on ignore complètement les valeurs des probabilités à priori, donc qu'on ne sait rien de x_a et de x_b.

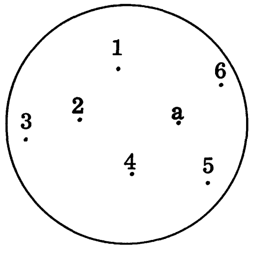
Dans ces conditions, seul l'examen de la séquence (s) observée pourrait nous permettre au temps $t + s - 1 = T$ de "choisir" entre les deux états marquants a, b. Chacun seulement des trois événements suivants nous permet d'y parvenir.

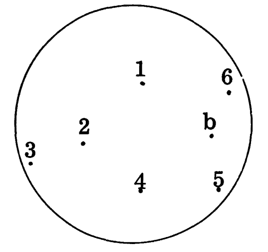
E 1. Un état marquant au moins apparait dans (s) : en effet, dès qu'un état marquant apparait, on connait parfaitement l'état du processus et il suffit alors de le suivre pas à pas jusqu'au temps T.

E 2. Il apparait dans (s) une transition $i \to j$ spécifique d'un certain état marquant càd dont la probabilité n'est non nulle que lorsque l'on est sous l'influence de cet état marquant.

E 3. Il apparait dans (s) une famille d'états ordinaires spécifique d'un certain état marquant, càd dont l'apparition simultanée dans (s) n'a une probabilité > o que sous l'influence d'un seul état marquant.

L'algorithme de recherche de l'état marquant au temps T en découle. Pour plus de clarté nous le préciserons sur un exemple. Supposons qu'il y ait deux états marquants a et b et 6 états ordinaires 1, 2, 3, 4, 5, 6.





Rappelons que (a, i) et (b, i) pour i = 1,2 ... 6 sont indiscernables et que pour passer de (a, i) à (b, j) (i et j étant ordinaires) il est nécessaire de passer par (b, b).

L'algorithme consiste simplement à chercher un majorant de la probabilité de l'événement $(\overline{E_1 \cup E_2 \cup E_3})$ (càd événement contraire de $(E_1 \cup E_2 \cup E_3)$.

1ère étape : On supprime les 2 colonnes et les 2 lignes relatives aux 2 états marquants. Il reste alors deux matrices (6×6): $\tilde{P}^{(a)}$ et $\tilde{P}^{(b)}$.

2ème étape : on interdit dans (a) toute transition impossible dans (b) (et vice-versa). Cela revient à annuler dans $\tilde{P}^{(a)}$ tout élément dont l'homologue est nul dans $\tilde{P}^{(b)}$ (et inversement). Par exemple si $4 \rightarrow 3$ est impossible dans (b), $\beta_{43} = 0$.
Alors on remplacera α_{43} par o car cette transition "trahirait" par son apparition l'état marquant a. On obtient alors les matrices $\hat{P}^{(a)}$ et $\hat{P}^{(b)}$.

Conséquences de ces deux transformations

Ces transformations étant faites, on cherche les classes finales de (a) et de (b). Il peut en exister plus qu'initialement, pour deux raisons simples que l'on comprendra sur un exemple.

1) La sortie d'un sous-ensemble de (a) pouvait être possible par l'intermédiaire de l'état b ce qui est maintenant interdit. On pouvait peut-être passer tout à l'heure de (a, 2) à (b, b) à (a, a) à (a, 4), chemin qui a été brisé.

2) La communication entre deux sous-ensembles de (a) peut avoir été brisée parce qu'elle s'opérait par une transition maintenant interdite parce qu'elle était impossible dans (b) (exemple de la transition $4 \rightarrow 3$).

Proposition : A la fin de la 2ème étape de l'algorithme tout état ordinaire inessentiel dans (a) est inessentiel dans (b) (et réciproquement).

Démonstration : Soit en effet un état x inessentiel dans (a) ; les états appartenant aux classes finales de (a) ne conduisent pas à x. Donc, après la transformation indiquée dans la 2ème étape, si x n'était pas inessentiel dans (b), il ne pourrait de toute façon qu'appartenir à une classe finale de (b) formée d'états différents de ceux qui sont essentiels dans (a). En d'autres termes cette classe finale dans (b) serait composée d'états tous inessentiels dans (a). Or c'est impossible car, comme on ne pourrait pas sortir de la classe finale de x dans (b), on aurait annulé les probabilités de transition dans (a) au cours de la 2ème étape de l'algorithme et les états de cette classe finale dans (b) formeraient aussi classe finale dans (a)

Corollaire : A la fin de la 2ème étape de l'algorithme, tout état ordinaire essentiel dans (a) est essentiel dans (b) (et réciproquement).
(découle immédiatement de la proposition précédente).

Proposition : A l'issue de la 2ème étape de l'algorithme, les classes finales dans (a) et dans (b) sont "appariées" et contiennent les mêmes états ordinaires (mais avec des probabilités de transition différentes en général).

Démonstration : si x, état ordinaire, appartient à une classe finale $C_x^{(a)}$ dans (a) (à l'issue de la 2ème étape de l'algorithme), alors $\forall_y \in C_x^{(a)}$, $\forall_z \notin C_x^{(a)}$ la transition $y \to z$ est impossible. Mais alors les $y \in C_x^{(a)}$ forment dans (b) une ou plusieurs classes finales (d'après le corollaire). En particulier, il existe donc une classe finale $C_x^{(b)}$ contenant x et dont tous les états faisaient partie de $C_x^{(a)}$ dans (a). Ce raisonnement pouvant être repris en partant de (b), la proposition est établie.

Conséquences :

1) Les matrices $\hat{P}^{(a)}$ et $\hat{P}^{(b)}$ ont la même structure par blocs. Les blocs sur la diagonale principale correspondant aux classes finales ; ceux qui sont "rectangulaires" correspondant aux états inessentiels. Les zéros situés à l'intérieur de ces blocs se retrouvent dans les 2 matrices.

2) Si aucune transition spécifique d'un certain état marquant n'est possible, il devient également impossible de voir apparaître dans (s) une famille d'états ordinaires spécifique d'un état marquant autrement dit $\overline{E}_2 \Rightarrow \overline{E}_3$. Ceci semble d'ailleurs intuitif.

3ème étape de l'algorithme

On élève $\hat{P}^{(a)}$ et $\hat{P}^{(b)}$ aux puissances entières successives. Dès que la plus petite des deux plus grandes sommes par ligne des matrices $[\hat{P}^{(a)}]^n$ et $[\hat{P}^{(b)}]^n$ tombe au-dessous de ε, on peut affirmer qu'une séquence observée de longueur n nous assure (au risque ε) de "dépister" l'état marquant (même s'il n'apparait pas).

En définitive si l'on veut prendre une décision au temps T, en limitant à ε le risque de ne pas connaître à cette date l'état marquant qui régit les transitions, on observera à partir du temps T - s + 1 avec

$$s = \inf \left\{ n \in N \;\middle|\; \inf_{u \in \{a,b\}} \; \sup_{i \in \mathcal{O}} \sum_j \left(\hat{p}_{ij}^{(u)} \right)^{(n)} < \varepsilon \right\}$$

où $(\hat{p}_{ij}^{(u)})^{(n)}$ est élément de $[\hat{P}^{(u)}]^n$

Remarques importantes :

1) Le procédé n'aboutira pas dans le cas où il existe, dès le début, au moins une classe finale dans (a) et une classe finale dans (b) qui contiennent les mêmes états ordinaires avec les mêmes transitions possibles (càd les zéros éventuels aux mêmes endroits dans les matrices associées). Dans ce cas on aura en effet

$$\inf_{u\in\{a,b\}} \ \sup_{i\in\mathscr{C}} \sum_j \left(\hat{p}_{ij}^{(u)}\right)^{(n)} = 1$$

En d'autres termes, l'entrée dans l'une ou l'autre de ces deux classes amènera une séquence qui, quelle que soit sa longueur, ne comportera ni états marquants ni transitions spécifiques.

On pourrait alors examiner le rapport $\lambda = A_s/B_s$ quand s grandit, pour l'ensemble des séquences de longueur s possibles.

Si λ ne devenait pas significativement différent de 1, cela indiquerait que le processus une fois entré dans ces classes est markovien et pratiquement de même nature, qu'il "provienne" de a ou de b.

2) La "stratégie d'observation" mentionnée ici est celle du risque minimax. Il serait possible d'en atténuer la "prudence" (et donc de diminuer s) en supposant qu'au temps $T - s + 1$ les probabilités d'être dans tel ou tel état sont connues et égales par exemple aux probabilités limites du processus si celles-ci existent et que le processus est ancien.

3) Une fois obtenues les matrices $\hat{P}^{(a)}$ et $\hat{P}^{(b)}$, nous avons vu que nous devions les élever aux puissances entières successives.

On peut trouver une valeur approchée de s par le procédé suivant. On démontre que la probabilité de ne pas avoir quitté les états inessentiels au bout d'un temps n "décroît géométriquement vers zéro" quand n grandit. On peut alors choisir n_1 pour que la probabilité de s'y trouver encore soit suffisamment faible. La suite de l'étude se fera alors simplement sur les blocs correspondant aux différentes classes, d'où la détermination d'un deuxième entier n_2. Un majorant commode de s est alors $\inf (n_1 + n_2)$ déterminé dans (a) et (b).

2) Cas de plus de deux états marquants

A - L'étude bayésienne est tout à fait analogue. Si les états marquants sont a, b, ... r les formules seront du type

$$P(Y_T = a) = \frac{A_s x_a}{A_s x_a + B_s x_b + \ldots + R_s x_r}$$

B - Si l'on ne sait rien des probabilités à priori

On peut mettre en oeuvre l'algorithme précédent :

1) On obtiendra d'abord les matrices $\tilde{P}^{(a)}$, $\tilde{P}^{(b)}$ ... $\tilde{P}^{(r)}$.

2) On annulera dans $\tilde{P}^{(a)}$ les probabilités de transition qui sont nulles dans toutes les autres matrices. Même opération sur les autres matrices. On obtient $\hat{P}^{(a)}$ ... $\hat{P}^{(r)}$.

3) On élève les matrices $\hat{P}$ aux puissances entières successives et on achève comme dans le cas de deux états marquants. On obtient un entier $s(\varepsilon)$ si l'algorithme aboutit.

Remarque importante : les matrices $\hat{P}^{(a)}$ ... $\hat{P}^{(r)}$ n'ont pas entre elles, en général, la même structure d'états inessentiels et de classes finales.