

ANNALES DE L'I. H. P., SECTION B

M. C. BERT

Convergence d'un processus à liaisons complètes, et application à un processus d'apprentissage linéaire

Annales de l'I. H. P., section B, tome 4, n° 1 (1968), p. 1-24

http://www.numdam.org/item?id=AIHPB_1968__4_1_1_0

© Gauthier-Villars, 1968, tous droits réservés.

L'accès aux archives de la revue « Annales de l'I. H. P., section B » (<http://www.elsevier.com/locate/anihpb>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

Convergence d'un processus à liaisons complètes, et application à un processus d'apprentissage linéaire

par

M^{me} M. C. BERT

Chaire de Calcul des Probabilités (*)
Institut Henri Poincaré.

On présente, dans une première partie, la démonstration d'un théorème de convergence pour un processus à temps discret, à valeurs dans un espace mesurable quelconque, auquel on impose deux conditions du type suivant : la première est une condition de positivité — existence d'un ensemble privilégié dans lequel le processus aura une probabilité positive de se trouver, à chaque instant, quel que soit son passé ; la seconde est une condition de faible dépendance du passé lointain, dont la définition fait intervenir l'ensemble privilégié. On reconnaît là les conditions imposées à certaines chaînes à liaisons complètes étudiées par W. Doeblin et R. Fortet. Le théorème de convergence établi par ces auteurs dans le cas où les processus prennent leurs valeurs dans un espace fini [1], repris plus tard par J. Lamperti et P. Suppes [2], est ici généralisé au cas où les processus prennent leurs valeurs dans un espace mesurable quelconque.

Dans une seconde partie, on définit des processus d'apprentissage linéaires à valeurs dans des espaces mesurables quelconques, et l'on établit un théorème de convergence pour ces processus, en faisant certaines hypothèses sur les données paramétriques, et la règle d'expérimentation ; on suppose, en particulier, que la règle d'expérimentation a une dépendance du passé de longueur m , quelconque. Ce théorème généralise un théorème de convergence établi pour des modèles d'apprentissage linéaires à valeurs dans des espaces finis, par J. Lamperti et P. Suppes ([2], théorème 4.1). La démonstration utilise le théorème de convergence établi dans la première partie.

(*) Équipe de recherche n° 1, « Processus stochastiques et applications » associée au CNRS.

L'énoncé des deux théorèmes, dont on présente ici la démonstration, a fait l'objet d'une note [3].

NOTATIONS ET DÉFINITIONS PRÉLIMINAIRES

On désigne par \mathbb{N} l'ensemble des entiers naturels positifs ou nuls, et par \mathbb{Z} l'ensemble de tous les entiers naturels.

DÉFINITION 1. — On appellera *générateur de probabilité* sur une suite d'espaces mesurables $(A_n = (\underline{A}_n, \mathcal{A}_n), n \in \mathbb{N})$, un ensemble

$$\mathcal{P} = \{ P_0, (P_n(\cdot; \cdot), n \geq 1) \},$$

où P_0 est une probabilité sur A_0 , et $\forall n \geq 1, P_n(\cdot; \cdot)$ est une probabilité de transition

$$\prod_{i=0}^{n-1} (A_i) \prec A_n$$

DÉFINITION 2. — Soient $(A_n = (\underline{A}_n, \mathcal{A}_n), n \in \mathbb{N})$ une suite d'espaces mesurables, et \mathcal{P} un générateur de probabilité sur $(A_n, n \in \mathbb{N})$, on appellera *processus canonique engendré par \mathcal{P}* , le processus canonique

$$Y = (\underline{\Omega}, \mathcal{F}, P, (Y_n)_{n \in \mathbb{N}})$$

défini sur \mathbb{N} , pour lequel

$$(\underline{\Omega}, \mathcal{F}) = \prod_{n \in \mathbb{N}} (\underline{A}_n, \mathcal{A}_n),$$

$\forall n \in \mathbb{N}, Y_n$ est l'application coordonnée d'indice n de $\underline{\Omega}$ dans A_n , et P est l'unique probabilité sur $(\underline{\Omega}, \mathcal{F}) = \prod (\underline{A}_n, \mathcal{A}_n)$, associée par le théorème de Tulcea au générateur \mathcal{P} (voir [4], V, 1).

Soit $A = (\underline{A}, \mathcal{A})$ un espace mesurable, et $Y = (\underline{\Omega}, \mathcal{F}, P, (Y_n)_{n \in \mathbb{N}})$ un processus quelconque défini sur \mathbb{N} , à valeurs dans A .

$\forall n \in \mathbb{N}$, on désigne par \mathcal{F}_n la sous-tribu de \mathcal{F} engendrée par les $Y_k (0 \leq k \leq n)$, et par $W^n = (\underline{W}^n, w^n)$ l'espace mesurable produit

$$\prod_{k=0}^n A_k, \quad \text{où } \forall k \in \mathbb{N}, \quad A_k = A.$$

$\forall n \in \mathbb{N}$, on appellera *version régulière de la probabilité conditionnelle* $H \rightarrow P \{ Y_{n+1} \in H \mid \mathcal{F}_n \}$, une probabilité de transition

$$P_{n+1}(\cdot; \cdot) : W^n \prec A_{n+1},$$

telle que $\forall H \in \mathcal{A}$,

$$P_{n+1}(Y^n; H) = P \{ X_{n+1} \in H | \mathcal{F}_n \} \quad \text{P. p. s.}$$

en désignant par Y^n la v. a. $\omega \rightarrow (Y_k(\omega), 0 \leq k \leq n)$ de Ω dans W^n , et par $P_{n+1}(Y^n, H)$ l'application $\omega \rightarrow P_{n+1}(Y^n(\omega); H)$.

On appellera *version régulière du processus Y*, un générateur de probabilité $\mathcal{P} = \{ P_0, (P_n(\cdot; \cdot), n \geq 1) \}$ sur $(A_n, n \in \mathbb{N})$, tel que

$$P_0(H) = P \{ Y_0 \in H \}, \quad \forall H \in \mathcal{A}, \quad \text{et} \quad \forall n \in \mathbb{N}, \quad P_{n+1}(\cdot; \cdot)$$

est une version régulière de la probabilité conditionnelle

$$H \rightarrow P \{ X_{n+1} \in H | \mathcal{F}_n \}.$$

A partir d'une version régulière du processus Y,

$$\mathcal{P} = \{ P_0, (P_n(\cdot; \cdot), n \geq 1) \},$$

on définit, $\forall 0 \leq q < n$, les probabilités de transition $P_n^q(\cdot; \cdot) : W^q \rightarrow A_n$, en posant, $\forall w^q \in \underline{W}^q (q \leq n-1)$, et $\forall H \in \mathcal{A}$:

$$(1) \left\{ \begin{array}{l} P_n^{n-1}(w^{n-1}; H) = P_n(w^{n-1}; H) \\ \text{et par récurrence descendante sur } q, \\ P_n^q(w^q; H) = \int_A P_{q+1}(w^q; da) P_n^{q+1}(w^q, a; H) \end{array} \right.$$

On a alors, $\forall H \in \mathcal{A}$,

$$P \{ Y_n \in H | \mathcal{F}_q \} = P_n^q(Y^q; H) \quad \text{P. p. s. sur } \Omega.$$

PROPOSITION I. — Tout générateur de probabilité \mathcal{P} sur $(A_n; n \in \mathbb{N})$ est une version régulière du processus canonique qu'il engendre. Soit en effet $Y = (\Omega, \mathcal{F}, P, (Y_n)_{n \in \mathbb{N}})$ le processus canonique engendré par un générateur $\mathcal{P} = \{ P_0, (P_n(\cdot; \cdot), n \geq 1) \}$ on vérifie, en appliquant le théorème de Tulcea, que $\forall H \in \mathcal{A}$, $P \{ H_0 \in H \} = P_0(H)$, et $\forall n \geq 1$, $P_n(Y^{n-1}; H)$ est une version de la probabilité conditionnelle $H \rightarrow P \{ X_n \in H | \mathcal{F}_{n-1} \}$.

Soit enfin $A = (A, \mathcal{A})$ un espace mesurable, et $X = (\Omega, \mathcal{F}, P, (X_n)_{n \in \mathbb{Z}})$ un processus défini sur Z (et non plus sur \mathbb{N}), à valeurs dans A . On adoptera encore les notations \mathcal{F}_n, W^n, X^n ($\forall n \in \mathbb{Z}$), pour désigner, respectivement, la sous-tribu de \mathcal{F} engendrée par les $X_k (k \leq n)$, l'espace mesurable produit

$$\prod_{k \leq n} (A_k),$$

et la v. a. $\omega \rightarrow (X_k(\omega))_{k \leq n}$ de Ω dans W^n .

$\forall n \in Z$, on appellera alors *version régulière de la probabilité conditionnelle* $H \rightarrow P \{ X_{n+1} \in H | \mathcal{F}_n \}$, une probabilité de transition

$$P_{n+1}(\cdot; \cdot) : W^n \prec A_{n+1},$$

telle que $\forall H \in \mathcal{A}$,

$$P_{n+1}(X^n; H) = P \{ X_{n+1} \in H | \mathcal{F}_n \} \quad \text{P. p. s. sur } \Omega$$

Un ensemble $\mathcal{P} = \{ P_n(\cdot; \cdot), n \in Z \}$, où $\forall n \in Z$, $P_n(\cdot; \cdot)$ est une probabilité de transition $W^{n-1} \prec A_n$, sera appelé *version régulière du processus X*, si $\forall n \in Z$, $P_{n+1}(\cdot; \cdot)$ est une version régulière de la probabilité conditionnelle

$$H \rightarrow P \{ X_{n+1} \in H | \mathcal{F}_n \}.$$

A partir d'une version régulière du processus X,

$$\mathcal{P} = \{ P_n(\cdot; \cdot), n \in Z \},$$

on définit là encore, $\forall q < n$, les probabilités de transition

$$P_n^q(\cdot; \cdot) : W^q \prec A_n,$$

par des relations identiques, formellement, aux relations (1). Et l'on a, $\forall H \in \mathcal{A}$

$$P \{ X_n \in H | \mathcal{F}_q \} = P_n^q(X^q; H) \quad \text{P. p. s. sur } \Omega$$

Un processus, défini sur N ou sur Z, est dit *régulier*, s'il possède une version régulière.

Le théorème de convergence que l'on va établir dans la première partie, sera démontré pour un processus défini sur Z. Mais la correspondance des notations et des définitions que l'on a adoptées pour un processus défini sur N, d'une part, et pour un processus défini sur Z, d'autre part, permet de reprendre le même énoncé et la même démonstration, en changeant seulement Z en N. Dans la seconde partie, c'est à un processus défini sur N que l'on appliquera le théorème.

I

Soit $X = (\Omega, \mathcal{F}, P, (X_n)_{n \in Z})$ un processus défini sur Z, et à valeurs dans un espace mesurable quelconque $A = (\underline{A}, \mathcal{A})$.

DÉFINITION. — Soit $H_0 \in \mathcal{A}$, m entier ≥ 1 , $(k, l) \in Z \times Z$. On dira que $(\omega, \omega') \in \underline{\Omega} \times \underline{\Omega}$ satisfait aux conditions $C_{k,l}^m(H_0)$ si :

∃ un ensemble de m indices, soit $M = \{i_1, \dots, i_m\} \subset \mathbb{N}$, tel que :

$$\begin{cases} i_1 > i_2 > \dots > i_m \geq 0 \\ (X_{k-i}(\omega), X_{k+l-i}(\omega')) \in H_0 \times H_0 & \forall i \in M \\ X_{k-i}(\omega) = X_{k+l-i}(\omega') & \forall i \in \mathbb{N} - M \quad \text{et} \quad 0 \leq i < i_1 \end{cases}$$

On pose alors :

$$E_{k,l}^m(H_0) = \{(\omega, \omega') \in \Omega \times \Omega \text{ satisfaisant aux conditions } C_{k,l}^m(H_0)\}$$

Étant donné un ensemble $H_0 \in \mathcal{A}$, et un entier $m \geq 1$, on associe à une version régulière $\mathcal{P} = \{P_n(\cdot, \cdot), n \in \mathbb{Z}\}$ du processus X , une suite réelle

$$\{\varepsilon_m^{(p)}(H_0), p \geq 1\},$$

en posant, $\forall p \geq 1$:

$$(2) \quad \varepsilon_m^{(p)}(H_0) = \sup_{\substack{(k,l) \in \mathbb{Z} \times \mathbb{Z} \\ H \in \mathcal{A}}} \sup_{(\omega, \omega') \in E_{k,l}^m(H_0)} |P_{k+p}^k(X^k(\omega), H) - P_{k+l+p}^{k+l}(X^{k+l}(\omega'), H)|$$

THÉORÈME I.

Hypothèses : Il existe un ensemble $H_0 \in \mathcal{A}$, et une version régulière $\mathcal{P} = \{P_n(\cdot, \cdot), n \in \mathbb{Z}\}$ du processus X , tels que :

1. $\exists 0 < \delta < 1 : \forall k \in \mathbb{Z},$

$$P_{k+1}(X^k; H_0) \geq \delta$$

2. $\sum_{m=1}^{\infty} \varepsilon_m^{(1)}(H_0) < \infty,$

où $\varepsilon_m^{(1)}(H_0)$ est défini, $\forall m \geq 1$, à partir de l'ensemble H_0 et de la version régulière \mathcal{P} par la relation (2) (avec $p = 1$).

Conclusions : Il existe une probabilité Π sur A , telle que : $\forall H \in \mathcal{A}$ et $\forall k \in \mathbb{Z},$

$$\lim_{n \rightarrow \infty} P \{X_{k+n} \in H | \mathcal{F}_k\} = \Pi(H) \quad \text{P. p. s. sur } \Omega$$

et par suite

$$\lim_{n \rightarrow \infty} P \{X_n \in H\} = \Pi(H)$$

Se référant toujours au même ensemble $H_0 \in \mathcal{A}$ satisfaisant aux hypothèses 1. et 2., pour tout entier $m \geq 1$, on désignera pour $\varepsilon_m^{(p)}$, le nombre réel $\varepsilon_m^{(p)}(H_0)$, $\forall p$ entier ≥ 1 , et par $E_{k,l}^m$, l'ensemble $E_{k,l}^m(H_0)$, $\forall (k, l) \in \mathbb{Z} \times \mathbb{Z}$.

LEMME 1. — Sous les hypothèses du théorème,

$$\varepsilon_m^{(p)} \rightarrow 0 \quad (m \rightarrow \infty) \text{ uniformément en } p.$$

Démonstration : On pose

$$\Delta_{k,l,H}^p(\omega, \omega') = |P_{k+p}^k(X^k(\omega), H) - P_{k+l+p}^{k+l}(X^{k+l}(\omega'), H)|.$$

On a alors,

$$\varepsilon_m^{(p)} = \sup_{\substack{(k,l) \in Z \times Z \\ H \in \mathcal{A}}} \sup_{(\omega, \omega') \in E_{k,l}^m} \Delta_{k,l,H}^p(\omega, \omega').$$

On cherche d'abord, $\mathbf{V}(k, l) \in Z \times Z$ et $H \in \mathcal{A}$, à majorer $\Delta_{k,l,H}^p(\cdot; \cdot)$ sur $E_{k,l}^m$.
Par définition (voir les relations (1)), on a $\mathbf{V}(\omega, \omega') \in \underline{\Omega} \times \underline{\Omega}$,

$$\left\{ \begin{array}{l} P_{k+p}^k(X^k(\omega); H) = \int_{\Delta} P_{k+1}(X^k(\omega); da) P_{k+p}^{k+1}(X^k(\omega), a; H) \\ P_{k+l+p}^{k+l}(X^{k+l}(\omega'); H) = \int_{\Delta} P_{k+l+1}(X^{k+l}(\omega'); da) P_{k+l+p}^{k+l+1}(X^{k+l}(\omega'), a; H) \end{array} \right.$$

D'où

$$\Delta_{k,l,H}^p(\omega, \omega') \leq V_{k,l,H}^p(\omega, \omega') + W_{k,l,H}^p(\omega, \omega')$$

en posant :

$$\begin{aligned} V_{k,l,H}^p(\omega, \omega') &= \int_{\Delta} P_{k+1}(X^k(\omega), da) | P_{k+p}^{k+1}(X^k(\omega), a; H) - P_{k+l+p}^{k+l+1}(X^{k+l}(\omega'), a; H) | \\ W_{k,l,H}^p(\omega, \omega') &= \left| \int_{\Delta} P_{k+1}(X^k(\omega); da) P_{k+l+p}^{k+l+1}(X^{k+l}(\omega'), a; H) \right. \\ &\quad \left. - \int_{\Delta} P_{k+l+1}(X^{k+l}(\omega'); da) P_{k+l+p}^{k+l+1}(X^{k+l}(\omega'), a; H) \right| \end{aligned}$$

Pour majorer $V_{k,l,H}^p(\omega, \omega')$, quand $(\omega, \omega') \in E_{k,l}^m$, on remarque, d'après la définition de l'ensemble $E_{k,l}^m$ que :

$$\left. \begin{array}{l} (\omega, \omega') \in E_{k,l}^m \\ X_{k+1}(\omega) = X_{k+l+1}(\omega') = a \end{array} \right\} \Rightarrow (\omega, \omega') \in \begin{cases} E_{k+l+1,l+1}^{m+1} & \text{si } a \in H_0 \\ E_{k+l+1,l+1}^m & \text{si } a \notin H_0 \end{cases}$$

et par suite, si $(\omega, \omega') \in E_{k,l}^m$, l'application

$$a \rightarrow |P_{k+p}^{k+1}(X^k(\omega), a; H) - P_{k+l+p}^{k+l+1}(X^{k+l}(\omega'), a; H)|$$

est majorée par

$$\begin{cases} \varepsilon_{m+1}^{(p-1)} & \text{si } a \in H_0 \\ \varepsilon_m^{(p-1)} & \text{si } a \notin H_0 \end{cases}$$

d'après la définition des $\varepsilon_m^{(p)}$ (voir la relation (2)). On a donc, si $(\omega, \omega') \in E_{k,l}^m$,

$$V_{k,l,H}^p(\omega, \omega') \leq \varepsilon_{m+1}^{(p-1)} P_{k+1}(X^k(\omega); H_0) + \varepsilon_m^{(p-1)} (1 - P_{k+1}(X^k(\omega); H_0)),$$

et comme, d'après l'hypothèse 1, $P_{k+1}(X^k(\omega), H_0) \geq \delta \mathbf{V} \omega \in \Omega$, et que, d'après la définition des $\varepsilon_m^{(p)}$ et des ensembles $E_{k,l}^m$, $\varepsilon_{m+1}^{(p-1)} \leq \varepsilon_m^{(p-1)}$ ($\mathbf{V}m$ et $\mathbf{V}p$), on a

$$V_{k,l,H}^p(\omega, \omega') \leq \delta \varepsilon_{m+1}^{(p-1)} + (1 - \delta) \varepsilon_m^{(p-1)}.$$

Il reste à majorer $W_{k,l,H}^p(\omega, \omega')$, pour $(\omega, \omega') \in E_{k,l}^m$. Or, $\mathbf{V}(\omega, \omega') \in \underline{\Omega} \times \underline{\Omega}$ fixé, l'application

$$K \rightarrow P_{k+1}(X^k(\omega); K) - P_{k+l+1}(X^{k+l}(\omega'); K)$$

est une mesure de signe varié sur A, soit $Q_{k,l}(\omega, \omega'; \cdot)$; et, si $(\omega, \omega') \in E_{k,l}^m$, on a, $\mathbf{V}H \in \mathcal{A}$, $|Q_{k,l}(\omega, \omega'; H)| \leq \varepsilon_m^{(1)}$. $W_{k,l,H}^p(\omega, \omega')$ est la valeur absolue de l'intégrale, par rapport à cette mesure $Q_{k,l}(\omega, \omega'; \cdot)$ de l'application

$$a \rightarrow P_{k+l+1}^{k+l+1}(X^{k+l}(\omega'), a; H),$$

définie sur A, mesurable, à valeur dans $[0,1]$.

Il s'en suit que

$$W_{k,l,H}^p(\omega, \omega') \leq 2\varepsilon_m^{(1)}.$$

On a donc finalement, $\mathbf{V}(\omega, \omega') \in E_{k,l}^m$,

$$\Delta_{k,l,H}^p(\omega, \omega') \leq 2\varepsilon_m^{(1)} + \delta \varepsilon_{m+1}^{(p-1)} + (1 - \delta) \varepsilon_m^{(p-1)}$$

On voit que cette majoration est indépendante de $(k, l) \in Z \times Z$ et de $H \in \mathcal{A}$. Ce qui permet d'écrire, $\mathbf{V}m \geq 1$ et $\mathbf{V}p \geq 2$:

$$\varepsilon_m^{(p)} \leq 2\varepsilon_m^{(1)} + \delta \varepsilon_{m+1}^{(p-1)} + (1 - \delta) \varepsilon_m^{(p-1)}.$$

Le développement de cette équation de récurrence conduit à l'inégalité

$$\varepsilon_m^{(p)} \leq \frac{1}{\delta} \sum_{l=0}^{p-1} \varepsilon_{m+l}^{(1)}, \quad \mathbf{V}p \geq 2;$$

l'hypothèse

$$\sum_{m=1}^{\infty} \varepsilon_m^{(1)} < \infty,$$

entraîne alors que $\varepsilon_m^{(p)} \rightarrow 0$ ($m \rightarrow \infty$) uniformément en p , ce qui démontre le lemme 1.

On considère maintenant, $\mathbf{V}(k, l) \in Z \times Z$, $H \in \mathcal{A}$, $(\omega, \omega') \in \underline{\Omega} \times \underline{\Omega}$, la suite $\{\Delta_{k,l,H}^n(\omega, \omega'), n \geq 1\}$, où $\mathbf{V}n \geq 1$,

$$\Delta_{k,l,H}^n(\omega, \omega') = |P_{k+n}^k(X^k(\omega); H) - P_{k+l+n}^{k+l}(X^{k+l}(\omega'); H)|,$$

et, sans imposer cette fois de conditions à (ω, ω') , on étudie la limite de cette suite quand $n \rightarrow \infty$.

LEMME 2. — Sous les hypothèses du théorème,

$$\Delta_{k,l,H}^n(\omega, \omega') \rightarrow 0 \quad (n \rightarrow \infty)$$

uniformément en $(k, l) \in Z \times Z$, $H \in \mathcal{A}$, $(\omega, \omega') \in \underline{\Omega} \times \underline{\Omega}$.

Démonstration : $\forall k \in Z$ et $\omega \in \underline{\Omega}$, on considère le générateur de probabilité

$$\mathcal{P}_\omega^k = \{ P_{k+1}(X^k(\omega); \cdot), (P_{k+n}(X^k(\omega), \cdot; \cdot), n \geq 2) \}$$

sur la suite d'espaces mesurables $(A_n, n \geq 1)$. On a bien $P_{k+1}(X^k(\omega); \cdot)$, probabilité sur A_1 , et $\forall n \geq 2$, $P_{k+n}(X^k(\omega), \cdot; \cdot)$, probabilité de transition

$$\prod_{i=1}^{n-1} (A_i) \prec A_n.$$

Soient alors, $\forall (k, l) \in Z \times Z$, et $(\omega, \omega') \in \underline{\Omega} \times \underline{\Omega}$,

$$\begin{cases} Y = (\underline{T}, \mathcal{T}, Q_\omega^k, (Y_n)_{n \geq 1}) \text{ et} \\ Z = (\underline{U}, \mathcal{U}, Q_{\omega'}^{k+l}, (Z_n)_{n \geq 1}) \end{cases}$$

les processus canoniques engendrés respectivement par les générateurs de probabilités \mathcal{P}_ω^k et $\mathcal{P}_{\omega'}^{k+l}$. On a

$$(\underline{T}, \mathcal{T}) = (\underline{U}, \mathcal{U}) = \prod_{i=1}^{\infty} (A_i, \mathcal{A}_i),$$

et $\forall n \geq 1$,

$$\begin{cases} Q_\omega^k \{ Y_n \in H \} = P_{k+n}^k(X^k(\omega); H) \\ Q_{\omega'}^{k+l} \{ Z_n \in H \} = P_{k+l+n}^{k+l}(X^{k+l}(\omega'); H) \end{cases}$$

d'où

$$\Delta_{k,l,H}^n(\omega, \omega') = |Q_\omega^k \{ Y_n \in H \} - Q_{\omega'}^{k+l} \{ Z_n \in H \}|.$$

On considère, dans l'espace produit $T \times U$, les pavés mesurables $G_h^{(m)}$ ($m \geq 1$, $h \geq m$), définis par

$$G_h^{(m)} = \{ (t, u) \in T \times U : (Y_{h-j}(t), Z_{h-j}(u)) \in H_0 \times H_0, \forall 0 \leq j \leq m-1 \},$$

et l'on désigne par $Q_{\omega, \omega'}^{k,l}$, la probabilité produit $Q_\omega^k \otimes Q_{\omega'}^{k+l}$ sur $T \times U$. On notera $Q_{\omega, \omega'}^{k,l} = Q$, lorsque les éléments k, l, ω, ω' sont fixés.

D'après la définition des processus Y et Z , on voit que, $\forall n > h \geq m$,

$$|Q_\omega^k \{ Y_n \in H | Y_1, \dots, Y_h \}(t) - Q_{\omega'}^{k+l} \{ Z_n \in H | Z_1, \dots, Z_h \}(u)| \leq \varepsilon_m^{(n-h)}$$

pour Q -presque tout $(t, u) \in G_h^{(m)}$.

En effet, le premier membre de cette inégalité s'identifie p. s. à

$$|P_{k+n}^{k+h}(X^k(\omega), Y_1(t), \dots, Y_h(t); H) - P_{k+l+n}^{k+l+h}(X^{k+l}(\omega'), Z_1(u), \dots, Z_h(u); H)|$$

et, d'après la définition des $\{\varepsilon_m^{(p)}, p \geq 1\}$ cette expression est bornée par $\varepsilon_m^{(n-h)}$, $\forall H \in \mathcal{A}$, $(k, l) \in Z \times Z$, $(\omega, \omega') \in \Omega \times \Omega$ dès que $(t, u) \in G_h^{(m)}$.

Or, $\forall m < n$, on peut majorer $|\mathcal{Q}_\omega^k \{Y_n \in H\} - \mathcal{Q}_{\omega'}^{k+l} \{Z_n \in H\}|$ de la façon suivante :

$$\begin{aligned} & |\mathcal{Q}_\omega^k \{Y_n \in H\} - \mathcal{Q}_{\omega'}^{k+l} \{Z_n \in H\}| \\ & \leq \int_{G_m^{(m)}} |\mathcal{Q}_\omega^k \{Y_n \in H | Y_1, \dots, Y_m\}(t) - \mathcal{Q}_{\omega'}^{k+l} \{Z_n \in H | Z_1, \dots, Z_m\}(u)| dQ(t, u) \\ & + \int_{T \times U - G_m^{(m)}} |\mathcal{Q}_\omega^k \{Y_n \in H | Y_1, \dots, Y_m\}(t) \\ & \quad - \mathcal{Q}_{\omega'}^{k+l} \{Z_n \in H | Z_1, \dots, Z_m\}(u)| dQ(t, u) \end{aligned}$$

Le premier terme de cette somme est majoré par $\varepsilon_m^{(n-m)} Q(G_m^{(m)})$ et le second terme peut s'écrire :

$$\begin{aligned} & \int_{(T \times U - G_m^{(m)}) \cdot G_{m+1}^{(m)}} |\mathcal{Q}_\omega^k \{Y_n \in H | Y_1, \dots, Y_{m+1}\}(t) \\ & \quad - \mathcal{Q}_{\omega'}^{k+l} \{Z_n \in H | Z_1, \dots, Z_{m+1}\}(u)| dQ(t, u) \\ & + \int_{T \times U - (G_m^{(m)} \cup G_{m+1}^{(m)})} |\mathcal{Q}_\omega^k \{Y_n \in H | Y_1, \dots, Y_{m+1}\}(t) \\ & \quad - \mathcal{Q}_{\omega'}^{k+l} \{Z_n \in H | Z_1, \dots, Z_{m+1}\}(u)| dQ(t, u) \end{aligned}$$

La première intégrale est majorée par $\varepsilon_m^{n-(m+1)} Q(G_{m+1}^{(m)} - G_m^{(m)})$, et la seconde se décompose encore.

$\forall m < n$, on obtient finalement, en poursuivant la décomposition :

$$\begin{aligned} & |\mathcal{Q}_\omega^k \{Y_n \in H\} - \mathcal{Q}_{\omega'}^{k+l} \{Z_n \in H\}| \\ & \leq \varepsilon_m^{(n-m)} Q(G_m^{(m)}) + \dots + \varepsilon_m^{(n-h)} Q\left(G_h^{(m)} - \bigcup_{i=m}^{h-1} G_i^{(m)}\right) + \dots \\ & \quad + \varepsilon_m^{(n-1)} Q\left(G_{n-1}^{(m)} - \bigcup_{i=m}^{n-2} G_i^{(m)}\right) \\ & + \int_{T \times U - \bigcup_{i=m}^{n-1} G_i^{(m)}} |\mathcal{Q}_\omega^k \{Y_n \in H | Y_1, \dots, Y_{n-1}\}(t) \\ & \quad - \mathcal{Q}_{\omega'}^{k+l} \{Z_n \in H | Z_1, \dots, Z_{n-1}\}(u)| dQ(t, u) \end{aligned}$$

Mais, d'après le lemme 1, $\varepsilon_m^{(p)} \rightarrow 0$, quand $m \rightarrow \infty$, uniformément en p . Donc, $\forall \varepsilon > 0$, $\exists M(\varepsilon) : m > M(\varepsilon) \Rightarrow \varepsilon_m^{(p)} \leq \varepsilon$, $\forall p$. On fixe arbitrairement un entier $m > M\left(\frac{\varepsilon}{2}\right)$. On a alors, $\forall n > m$,

$$\Delta_{k,l,H}^n(\omega, \omega') = |\mathcal{Q}_\omega^k \{Y_n \in H\} - \mathcal{Q}_{\omega'}^{k+l} \{Z_n \in H\}| \leq \frac{\varepsilon}{2} + Q\left(T \times U - \bigcup_{i=m}^{n-1} G_i^{(m)}\right)$$

Il reste donc, pour démontrer que $\Delta_{k,l,H}^n(\omega, \omega') \rightarrow 0$ ($n \rightarrow \infty$) uniformément en k, l, H, ω, ω' , à montrer que, pour n assez grand, et indépendant de k, l, ω, ω' ,

$$Q_{\omega, \omega'}^{k,l} \left(\bigcup_{i=m}^{n-1} G_i^{(m)} \right) \geq 1 - \frac{\varepsilon}{2}.$$

▼ l'entier $q \geq 1$, on désigne par $B_q^{(m)}$ l'ensemble $G_{qm}^{(m)}$, et par $\mathcal{B}_{q-1}^{(m)}$ la sous-tribu de $\mathcal{F} \otimes \mathcal{U}$ engendrée par les v. a. (Y_j, Z_j) où $j \leq (q-1)m$.

Pour chacun des ensembles $B_q^{(m)}$ ($q \geq 1$), et ▼ les éléments $(k, l) \in Z \times Z$ et $(\omega, \omega') \in \underline{\Omega} \times \underline{\Omega}$ à partir desquels on a construit la probabilité $Q_{\omega, \omega'}^{k,l}$ sur $T \times U$, la condition de positivité imposée à l'ensemble H_0 (hypothèse 1) entraîne que l'on a

$$Q_{\omega, \omega'}^{k,l} \{ B_q^{(m)} \mid \mathcal{B}_{q-1}^{(m)} \} \geq \delta^{2m} \quad \text{p. s.,}$$

d'où

$$Q_{\omega, \omega'}^{k,l} \left(\bigcup_{q=1}^n B_q^{(m)} \right) \geq 1 - (1 - \delta^{2m})^n,$$

et enfin, puisque

$$\bigcup_{i=m}^{nm} G_i^{(m)} \supset \bigcup_{q=1}^n B_q^{(m)}$$

$$Q_{\omega, \omega'}^{k,l} \left(\bigcup_{i=m}^{nm} G_i^{(m)} \right) \geq 1 - (1 - \delta^{2m})^n.$$

On pourra donc trouver, ▼ $\varepsilon > 0$, un entier N dépendant uniquement de ε (l'entier m ne dépendant lui-même que de ε), tel que

$$n > N(\varepsilon) \Rightarrow \Delta_{k,l,H}^n(\omega, \omega') \leq \varepsilon,$$

▼ $(k, l) \in Z \times Z, H \in \mathcal{A}, (\omega, \omega') \in \underline{\Omega} \times \underline{\Omega}$. Le lemme 2 est ainsi démontré.

Pour démontrer le théorème, on montre d'abord que ▼ $H \in \mathcal{A}$, ∃ un nombre réel $\Pi(H)$:

$$\forall k \in Z, \quad P_{k+n}^k(X^k; H) \rightarrow \Pi(H) \quad \text{P. p. s.,}$$

d'où

$$P \{ X_{k+n} \in H \mid \mathcal{F}_k \} \rightarrow \Pi(H) \quad \text{P. p. s.}$$

On montre ensuite que l'application $H \rightarrow \Pi(H)$ ainsi définie sur \mathcal{A} est une probabilité.

$H \in \mathcal{A}$ étant arbitrairement fixé, on désigne, ▼ $k \in Z$ et $n \in \mathbb{N}$, par $\Pi_n^{(k)}$ l'application $\omega \rightarrow P_{k+n}^k(X^k(\omega); H)$, de Ω dans $[0,1]$, et on démontre successivement les trois propositions suivantes :

1. $\forall k \in \mathbb{Z}$, \exists une application $\Pi^{(k)}$ p. s. définie sur $\underline{\Omega}$ t. q.

$$\Pi_n^{(k)} \rightarrow \Pi^{(k)} \quad \text{p. s.} \quad (n \rightarrow \infty)$$

2. Les applications $\Pi^{(k)}$ ($k \in \mathbb{Z}$) sont p. s. égales à une même application Π

3. L'application Π est p. s. égale à une constante, que l'on désigne par $\Pi(H)$.

1. $k \in \mathbb{Z}$ étant arbitrairement fixé, il suffit de montrer que la suite $(\Pi_n^{(k)}, n \in \mathbb{N})$ est p. s. une suite de Cauchy. Or, on peut trouver un ensemble P-négligeable \mathcal{N}^k , en dehors duquel on a :

$$\forall n \in \mathbb{N}, \quad \Pi_n^{(k)} = \mathbb{P} \{ X_{k+n} \in H \mid \mathcal{F}_k \}$$

soit

$$\Pi_n^{(k)} = E^{\mathcal{F}_k}[1_H(X_{k+n})]$$

d'où, $\forall n$ et $l \in \mathbb{N}$,

$$\begin{aligned} \Pi_n^{(k)} - \Pi_{n+l}^{(k)} &= E^{\mathcal{F}_k}[1_H(X_{k+n})] - E^{\mathcal{F}_k} E^{\mathcal{F}_{k+l}}[1_H(X_{k+l+n})] \\ &= E^{\mathcal{F}_k}[\mathbb{P} \{ X_{k+n} \in H \mid \mathcal{F}_k \} - \mathbb{P} \{ X_{k+l+n} \in H \mid \mathcal{F}_{k+l} \}] \end{aligned}$$

d'où

$$|\Pi_n^{(k)} - \Pi_{n+l}^{(k)}| \leq E^{\mathcal{F}_k} |\mathbb{P} \{ X_{k+n} \in H \mid \mathcal{F}_k \} - \mathbb{P} \{ X_{k+l+n} \in H \mid \mathcal{F}_{k+l} \}|$$

soit

$$|\Pi_n^{(k)} - \Pi_{n+l}^{(k)}| \leq E^{\mathcal{F}_k}(\bar{\Delta}_{k,l,H}^n),$$

en désignant par $\bar{\Delta}_{k,l,H}^n$ l'application $\omega \rightarrow \Delta_{k,l,H}^n(\omega, \omega)$, de $\underline{\Omega}$ dans $[0,1]$, c'est-à-dire la restriction de $\Delta_{k,l,H}^n$ à la diagonale de $\underline{\Omega} \times \underline{\Omega}$.

Or, le lemme 2 assure, en particulier, que la convergence vers 0 de la suite de v. a. $(\bar{\Delta}_{k,l,H}^n, n \in \mathbb{N})$ est uniforme en l , et, comme $|\bar{\Delta}_{k,l,H}^n| \leq 1 \forall n \in \mathbb{N}$, on voit (généralisation du théorème de Lebesgue à l'espérance conditionnelle), que la suite $(\Pi_n^{(k)} - \Pi_{n+l}^{(k)}; n \in \mathbb{N})$ converge vers 0, uniformément en l , quand $n \rightarrow \infty$, c'est-à-dire que la suite $(\Pi_n^{(k)}, n \in \mathbb{N})$ est une suite de Cauchy. En dehors de l'ensemble négligeable \mathcal{N}^k , la limite de cette suite, que l'on désigne par $\Pi^{(k)}$, est donc bien définie.

2. On va montrer maintenant, qu'en dehors de l'ensemble négligeable

$$\mathcal{N} = \bigcup_{k \in \mathbb{Z}} \mathcal{N}^k, \quad \Pi^{(k)} = \Pi^{(k+h)}, \quad \forall k \text{ et } h \in \mathbb{Z}.$$

Soient donc k et $h \in \mathbb{Z}$ arbitrairement fixés, on a, $\forall n \in \mathbb{N}$, en tout $\omega \notin \mathcal{N}$,

$$|\Pi^{(k)} - \Pi^{(k+h)}| \leq |\Pi^{(k)} - \Pi_n^{(k)}| + |\Pi_n^{(k)} - \Pi_n^{(k+h)}| + |\Pi_n^{(k+h)} - \Pi^{(k+h)}|$$

et, quand $n \rightarrow \infty$,

$$\Pi_n^{(k)} \rightarrow \Pi^{(k)} \quad \text{et} \quad \Pi_n^{(k+h)} \rightarrow \Pi^{(k+h)},$$

d'après la proposition 1,

$$|\Pi_n^{(k)} - \Pi_n^{(k+h)}| = \bar{\Delta}_{k,h,H}^n \rightarrow 0,$$

d'après le lemme 2.

$\forall \varepsilon > 0$, il existe donc un entier $N(\varepsilon)$, dépendant éventuellement de $k, h \in \mathbb{Z}$ et de $\omega \in \Omega - \mathcal{N}$, tel que :

$$n > N(\varepsilon) \Rightarrow |\Pi^{(k)} - \Pi^{(k+h)}| \leq \varepsilon,$$

ce qui démontre la proposition.

On désigne par Π la classe des v. a. $\Pi^{(k)}$ ($k \in \mathbb{Z}$), p. s. égales sur $\underline{\Omega}$.

3. On montre enfin que $\forall \omega$ et $\omega' \notin \mathcal{N}$, $\Pi(\omega) - \Pi(\omega') = 0$. $\forall \omega$ et $\omega' \in \underline{\Omega} - \mathcal{N}$, on a en effet, $\forall n \in \mathbb{N}$ et $k \in \mathbb{Z}$,

$$|\Pi(\omega) - \Pi(\omega')| \leq |\Pi(\omega) - \Pi_n^{(k)}(\omega)| \\ + |\Pi_n^{(k)}(\omega) - \Pi_n^{(k)}(\omega')| + |\Pi_n^{(k)}(\omega') - \Pi(\omega')|$$

et, quand $n \rightarrow \infty$,

$$\Pi_n^{(k)}(\omega) \rightarrow \Pi(\omega), \quad \Pi_n^{(k)}(\omega') \rightarrow \Pi(\omega'),$$

d'après la proposition 1,

$$|\Pi_n^{(k)}(\omega) - \Pi_n^{(k)}(\omega')| = \Delta_{k,0,H}^n(\omega, \omega') \rightarrow 0,$$

d'après le lemme 2.

$\forall \varepsilon > 0$, il existe donc, $\forall k \in \mathbb{Z}$, un entier $N(\varepsilon)$, dépendant éventuellement de $k \in \mathbb{Z}$ et de ω et $\omega' \in \underline{\Omega} - \mathcal{N}$, tel que :

$$n > N(\varepsilon) \Rightarrow |\Pi(\omega) - \Pi(\omega')| \leq \varepsilon.$$

L'application Π a donc p. s. sur $\underline{\Omega}$ une valeur constante que l'on désigne par $\Pi(H)$.

Ainsi, pour tout $H \in \mathcal{A}$, on a défini un nombre réel $\Pi(H)$, tel que, $\forall k \in \mathbb{Z}$,

$$\lim_{n \rightarrow \infty} \Pi_n^{(k)} = \Pi(H),$$

en tout $\omega \in \Omega - \mathcal{N}$, d'où

$$\lim_{n \rightarrow \infty} P \{ X_{k+n} \in H | \mathcal{F}_k \} = \Pi(H) \quad \text{p. s. sur } \Omega,$$

puisque, en dehors de l'ensemble négligeable \mathcal{N} , on a, $\forall k \in \mathbb{Z}$ et $n \in \mathbb{N}$,

$$\Pi_n^{(k)} = P \{ X_{k+n} \in H | \mathcal{F}_k \}.$$

De ce résultat, on déduit immédiatement le suivant :

$$\lim_{n \rightarrow \infty} P \{ X_n \in H \} = \Pi(H),$$

puisque, $\forall n \in \mathbb{N}$, et $k < n$,

$$P \{ X_n \in H \mid \mathcal{F}_k \} \in [0,1] \quad \text{p. s.},$$

et

$$P \{ X_n \in H \} = \int_{\Omega} P \{ X_n \in H \mid \mathcal{F}_k \} dP.$$

$H \in \mathcal{A}$ étant fixé, $\forall k \in \mathbb{Z}$ et $n \in \mathbb{N}$, $\Pi_n^{(k)}$ désignait l'application

$$\omega \rightarrow P_{k+n}^k(X^k(\omega); H),$$

de Ω dans $[0,1]$. Mais, pour tout $\omega \in \Omega$ fixé, l'application

$$H \rightarrow P_{k+n}^k(X^k(\omega); H)$$

est une probabilité sur A . En fixant arbitrairement $\omega \in \Omega - \mathcal{N}$, et $k \in \mathbb{Z}$, on obtient donc une suite de probabilité sur A , $\{ P_{k+n}^k(X^k(\omega); \cdot), n \in \mathbb{N} \}$, qui a la propriété suivante :

$$\Pi(H) = \lim_{n \rightarrow \infty} P_{k+n}^k(X^k(\omega); H)$$

existe $\forall H \in \mathcal{A}$. D'après le théorème de Vitali-Hahn-Saks ([4], IV-2) l'application $H \rightarrow \Pi(H)$ est donc une probabilité sur A .

II

Soient $A = (\underline{A}, \mathcal{A})$ et $E = (\underline{E}, \mathcal{E})$ deux espaces mesurables quelconques, et $(\Delta) = ((\delta), (\beta))$, un ensemble de données, relatives aux espaces A et E , constitué des éléments suivants :

- (δ) $\left\{ \begin{array}{l} \theta, \text{ nombre réel } \in]0,1] \\ G(\cdot; \cdot), \text{ probabilité de transition } E \prec A, \\ v, \text{ probabilité sur } A. \end{array} \right.$
- (β) $\left\{ \begin{array}{l} (\beta_n(\cdot; \cdot), n \in \mathbb{N}), \text{ où } \forall n \in \mathbb{N}, \beta_n(\cdot; \cdot) \text{ est une probabilité de transi-} \\ \text{tion } (A \times E)^{n-1} \times A \prec E. \end{array} \right.$

On appellera (δ) , les données paramétriques, et (β) la règle d'expérimentation.

DÉFINITION. — Un processus $(\Omega, \mathcal{F}, P, (a_n, e_n)_{n \in \mathbb{N}})$, défini sur \mathbb{N} , à valeurs dans $A \times E$, est appelé *processus d'apprentissage linéaire associé*

aux données paramétriques (δ) , et compatible avec la règle d'expérimentation (β) , s'il satisfait aux axiomes suivants (\mathcal{F}_n désignant la sous-tribu de \mathcal{F} engendrée par les (a_i, e_i) , $0 \leq i \leq n$) :

1. Axiome de réponse.

$$(R_0) \quad P\{a_0 \in H\} = v(H)$$

$$(R) \quad P\{a_{n+1} \in H | \mathcal{F}_n\} = (1 - \theta)P\{a_n \in H | \mathcal{F}_{n-1}\} + \theta G(e_n; H)$$

P. p. s. sur Ω

$$\forall n \geq 1 \quad \text{et} \quad \forall H \in \mathcal{A}$$

2. Axiome d'expérimentation.

$$(E) \quad P\{e_n \in K | \mathcal{F}_{n-1}, a_n\} = \beta_n(a_0, e_0, \dots, a_{n-1}, e_{n-1}, a_n; K)$$

P. p. s. sur Ω

$$\forall n \geq 0 \quad \text{et} \quad \forall K \in \mathcal{E}.$$

Existence et unicité.

PROPOSITION. — Tout ensemble de données $(\Delta) = ((\delta), (\beta))$ définit, de façon unique à une équivalence près, un processus d'apprentissage linéaire, associé aux données paramétriques (δ) , et compatible avec la règle d'expérimentation (β) ; ce processus est régulier.

En effet :

Soient $A = (\underline{A}, \mathcal{A})$ et $E = (\underline{E}, \mathcal{E})$ deux espaces mesurables, et $(\Delta) = ((\delta), (\beta))$ un ensemble de données relatives à A et E .

$\forall k \in \mathbb{N}$, on pose $A_k = A$, $E_k = E$, et l'on considère la suite des espaces mesurables $B_0, B_1, \dots, B_{2k}, B_{2k+1}, \dots$ où $\forall k \in \mathbb{N}$, $B_{2k} = A_k$, $B_{2k+1} = E_k$.

On construit alors, à partir des données (Δ) , le générateur de probabilité $\mathcal{P} = \{v_0, (v_n(\cdot; \cdot), n \geq 1)\}$ sur la suite d'espaces $(B_n, n \in \mathbb{N})$, en posant :

$$(5) \left\{ \begin{array}{l} v_0 = v, \text{ probabilité sur } B_0, \\ \forall n \geq 0, v_{2n+1}(\cdot; \cdot) = \beta_n(\cdot; \cdot), \text{ probabilité de transition} \\ \prod_{k=0}^{2n} B_k \prec B_{2n+1}, \\ \text{et } \forall n \geq 1, v_{2n}(\cdot; \cdot) = \alpha_n(\cdot; \cdot), \text{ probabilité de transition} \\ \prod_{k=0}^{2n-1} B_k \prec B_{2n}, \end{array} \right.$$

où la suite $(\alpha_n(\cdot; \cdot), n \geq 1)$ est définie, à partir des données (Δ) , par les relations de récurrence suivantes :

$$(6) \quad \left\{ \begin{array}{l} \alpha_1(a_0, e_0; H) = (1 - \theta)v_0(H) + \theta G(e_0; H) \quad \text{et} \quad \forall n \geq 1, \\ \alpha_{n+1}(a_0, e_0, \dots, a_n, e_n; H) \\ \quad = (1 - \theta)\alpha_n(a_0, e_0, \dots, a_{n-1}, e_{n-1}; H) + \theta G(e_n; H) \\ \quad \forall (a_i, e_i) \in A_i \times E_i, \quad \text{et} \quad \forall H \in \mathcal{A}. \end{array} \right.$$

Le générateur \mathcal{P} ainsi construit est univoquement déterminé par les données (Δ) .

Soit $Y = (\mathbf{B}, \mathcal{B}, P, (Y_n)_{n \in \mathbb{N}})$, le processus canonique engendré par ce générateur \mathcal{P} . Il satisfait aux axiomes 1 et 2 ; en effet, \mathcal{P} est une version régulière de Y (Proposition I) ; d'où, en notant, $\forall n \in \mathbb{N}, Y_{2n} = a_n$, et $Y_{2n+1} = e_n$, les relations suivantes :

$$\forall H \in \mathcal{A} \quad \text{et} \quad \forall K \in \mathcal{E},$$

$$\begin{array}{ll} (R_0) & P \{ a_0 \in H \} = v_0(H) \\ (R) & P \{ a_n \in H | \mathcal{F}_{n-1} \} = \alpha_n(a_0, e_0, \dots, a_{n-1}, e_{n-1}; H) \\ & \text{P. p. s.} \quad \forall n \geq 1 \\ (E) & P \{ e_n \in H | \mathcal{F}_{n-1}, a_n \} = \beta_n(a_0, e_0, \dots, a_n; K) \\ & \text{P. p. s.} \quad \forall n \geq 0 \end{array}$$

Mais, en tenant compte de la définition des $\alpha_n(\cdot; \cdot)$, ($n \geq 1$) (voir (6)), les relations (R) donnent :

$$(R_1) \quad P \{ a_n \in H | \mathcal{F}_{n-1} \} = (1 - \theta)P \{ a_{n-1} \in H | \mathcal{F}_{n-2} \} + \theta G(e_{n-1}; H) \\ \text{P. p. s.} \quad \forall n \geq 1$$

Les relations (R_0) , (R_1) et (E) expriment bien que le processus canonique Y , défini par le générateur \mathcal{P} , donc par les données (Δ) , satisfait aux axiomes 1 et 2.

On a donc montré, en construisant Y , qu'il existe un processus d'apprentissage linéaire, associé aux données paramétriques (δ) , et compatible avec la règle d'expérimentation (β) . Ce processus est régulier, puisqu'il admet \mathcal{P} comme version régulière.

Si un processus canonique Y' , construit sur le même espace de base $(\mathbf{B}, \mathcal{B})$, satisfait aux axiomes 1 et 2, il satisfait aussi aux relations (R_0) , (R), (E), admet donc \mathcal{P} comme version régulière, et par suite coïncide avec le processus canonique Y . D'où l'unicité de Y .

INTERPRÉTATION. — Dans de nombreuses expériences d'apprentissage, deux espaces mesurables définissent la situation expérimentale : l'espace des réponses, A , et l'espace des conclusions, E . L'expérience consiste en une suite d'essais : à l'essai n , le sujet donne une réponse a_n , à valeur dans A , puis l'expérimentateur donne une conclusion e_n , à valeur dans E .

La règle d'expérimentation, suivant laquelle l'expérimentateur choisit la conclusion e_n , en fonction des résultats des essais précédents, est définie par la donnée d'une suite $(\beta) = (\beta_n(\cdot; \cdot), n \in \mathbb{N})$, où $\forall n \in \mathbb{N}$, $\beta_n(\cdot; \cdot)$ est une probabilité de transition $(A \times E)^{n-1} \times A \rightarrow E$. On observe un processus expérimental $(a_n, e_n) n \in \mathbb{N}$.

A partir des espaces mesurables A et E , et de la règle d'expérimentation (β) , définissant les conditions expérimentales d'une expérience d'apprentissage de ce type, on construit, pour tout ensemble de données paramétriques (δ) relatives à A et E , le processus d'apprentissage linéaire associé aux données paramétriques (δ) , et compatible avec la règle d'expérimentation (β) (On a montré l'existence et l'unicité de ce processus). Ce processus théorique, $Y = (\Omega, \mathcal{F}, P, (a_n, e_n) n \in \mathbb{N})$, construit sur la base des axiomes 1 et 2, peut alors être interprété comme une prédiction sur le processus expérimental $(a_n, e_n) n \in \mathbb{N}$ obtenu lorsque l'expérience est réalisée avec la règle d'expérimentation (β) .

L'objet des paragraphes suivants est l'étude des propriétés asymptotiques de cette prédiction, lorsque les données paramétriques (δ) , et la règle d'expérimentation (β) satisfont à certaines hypothèses. Ce sont donc les propriétés d'un processus théorique qui seront envisagées.

La situation de cette étude théorique, du point de vue de son articulation avec une étude expérimentale des phénomènes d'apprentissage, peut être précisée en disant que, l'application qui à toute règle d'expérimentation (β) , associe une telle prédiction, correspond exactement à la notion de *modèle*, introduite par P. Courrege et H. Rouanet dans [5]. La rédaction de cette seconde partie étant écrite en toute cohérence avec le langage et les notions introduites dans [5], on pourra facilement s'y référer.

Convergence.

On se propose d'établir la convergence de ces processus d'apprentissage linéaires, en imposant des hypothèses aussi faibles que possible à l'ensemble de données $(\Delta) = ((\delta), (\beta))$ qui les définit.

Plus précisément, si $Y = (\Omega, \mathcal{F}, P, (a_n, e_n) n \in \mathbb{N})$ à valeurs dans $A \times E$, est le processus d'apprentissage linéaire associé aux données paramétriques (δ) , et compatible avec la règle d'expérimentation (β) (défini à une

équivalence près), on cherche à montrer qu'il existe une probabilité γ sur A , telle que,

$$\forall k \in \mathbb{N}, \quad \lim_{n \rightarrow \infty} P \{ a_n \in H \mid \mathcal{F}_k \} = \gamma(H) \quad \text{P. p. s.}$$

d'où
$$\lim_{n \rightarrow \infty} P \{ a_n \in H \} = \gamma(H), \quad \forall H \in \mathcal{A}.$$

Un cas particulièrement simple est celui où la règle d'expérimentation (β) est définie par la donnée d'une unique probabilité ρ sur E , telle que :

$$\beta_n(a_0, e_0, \dots, a_n; \cdot) = \rho(\cdot) \quad \forall n \in \mathbb{N}$$

et
$$\forall (a_0, e_0, \dots, a_n) \in (A \times E)^{n-1} \times A.$$

La convergence du processus Y , dans ce cas, est établi directement, et la probabilité limite γ est définie par les relations :

$$\gamma(H) = \int_E G(e; H) \rho(de), \quad \forall H \in \mathcal{A}.$$

Si l'on interprète $G(e; H)$ comme une densité de renforcement attachée par l'élément $e \in E$, à l'ensemble $H \in \mathcal{A}$, on voit que $\gamma(H)$ peut s'interpréter comme le renforcement moyen attaché à l'ensemble H , et par suite, le résultat :

$$\lim_{n \rightarrow \infty} P \{ a_n \in H \} = \gamma(H), \quad \forall H \in \mathcal{A},$$

peut s'interpréter comme la prédiction d'une loi d'ajustement.

L'étude d'un cas plus général fait l'objet du théorème suivant. Les hypothèses faites sur la règle d'expérimentation sont essentiellement une hypothèse de positivité, et une hypothèse de dépendance du passé de longueur m quelconque, mais fixée.

THÉORÈME II.

Hypothèses : Soient A et E deux espaces mesurables quelconques, et $(\Delta) = ((\delta), (\beta))$, un ensemble de données, relatives à A et E , satisfaisant aux hypothèses suivantes :

1. La probabilité de transition $G(\cdot; \cdot)$ est définie à partir d'une partition dénombrable $\sum_{i \in \mathbb{N}} D_i$ de E , et d'une famille de probabilité $(G_{D_i}, i \in \mathbb{N})$ sur A , par

$$G(e_n; \cdot) = \sum_{i \in \mathbb{N}} 1_{D_i}(e_n) G_{D_i}(\cdot).$$

2. La règle d'expérimentation $(\beta) = (\beta_n(\cdot; \cdot), n \in \mathbb{N})$ est définie par la donnée de $(m + 1)$ probabilité de transition $\beta_h (0 \leq h \leq m)$,

$$(A \times E)^h \times A \prec E,$$

m étant un entier positif quelconque fixé, et par la règle suivante,

$$\forall n > m, \quad \beta_n(a_0, e_0, \dots, a_{n-m}, e_{n-m}, \dots, a_n; \cdot) = \beta_m(a_{n-m}, e_{n-m}, \dots, a_n; \cdot)$$

3. Il existe D , élément de la partition

$$\sum_{i \in \mathbb{N}} D_i, \quad \text{et} \quad 0 < \eta < 1 :$$

$$\forall h \leq m, \quad \forall u \in (A \times E)^h \times A, \quad \beta_h(u; D) \geq \eta.$$

Soit alors $C \in \mathcal{A}$:

$$G_D(C) = \alpha > 0, \quad \text{et soit} \quad \sum_{j \in \mathbb{N}} C_j$$

une partition dénombrable de A , dont C est élément.

4. La probabilité de transition $\beta_m(\cdot; \cdot)$ est définie à partir de la partition $\sum_k \Lambda_k$ de $(A \times E)^m \times A$ induite par les partitions $\sum_i D_i$ et $\sum_j C_j$ de E et de A respectivement, et d'une famille de probabilités $(Q_{\Lambda_k}, k \in \mathbb{N})$ sur E , par

$$\beta_m(a_{n-m}, e_{n-m}, \dots, a_n; \cdot) = \sum_k 1_{\Lambda_k}(a_{n-m}, e_{n-m}, \dots, a_n) Q_{\Lambda_k}(\cdot)$$

Les probabilités de transition $\beta_h(\cdot; \cdot)$ ($0 \leq h < m$) sont définies suivant un schéma analogue.

Conclusions : Soit $Y = (\underline{B}, \mathcal{B}, P, (a_n, e_n) n \in \mathbb{N})$, le processus linéaire d'apprentissage associé aux données paramétriques (δ) , et compatible avec la règle d'expérimentation (β) .

Alors, il existe une probabilité γ sur A , t. q.

$$k \in \mathbb{N}, \quad \lim_{n \rightarrow \infty} P \{ a_{k+n} \in H \mid \mathcal{B}_k \} = \gamma(H) \quad \text{P. p. s.}$$

$$\text{et} \quad \lim_{n \rightarrow \infty} P \{ a_n \in H \} = \gamma(H)$$

$$\forall H \in \mathcal{A}.$$

Démonstration : L'ensemble de données $(\Delta) = ((\delta), (\beta))$ définit (voir (5)) un générateur de probabilité $\mathcal{P} = \{ \nu_0, (\nu_n(\cdot; \cdot), n \geq 1) \}$ sur la suite d'es-

paces $(B_n, n \in \mathbb{N})$ (où $B_{2k} = A_k = A$, $B_{2k+1} = E_k = E$, $\forall k \in \mathbb{N}$). Le processus d'apprentissage linéaire associé aux données paramétriques (δ) , et compatible avec la règle d'expérimentation (β) , est alors le processus canonique $Y = (B, \mathcal{B}, P, (a_n, e_n) n \in \mathbb{N})$ à valeurs dans $A \times E$, engendré par \mathcal{P} .

Pour établir la convergence de ce processus, on introduit un processus auxiliaire de la façon suivante : on pose $B_{-1} = E_{-1} = E$, et l'on considère la suite d'espaces mesurables $C_0, C_1, \dots, C_k, \dots$ où $\forall k \in \mathbb{N}$,

$$C_k = B_{2k-1} \times B_{2k}$$

soit encore

$$C_k = E_{k-1} \times A_k$$

A partir du générateur $\mathcal{P} = \{v_0, (v_n(\cdot; \cdot), n \geq 1)\}$ sur $(B_n, n \in \mathbb{N})$, associé aux données (Δ) , on construit le générateur de probabilité

$$\mathcal{S} = \{Q_0, Q_n(\cdot, \cdot), n \geq 1\}$$

sur la suite d'espaces $(C_n, n \in \mathbb{N})$, en définissant

$$(7) \left\{ \begin{array}{l} Q_0, \text{ probabilité sur } C_0, \text{ par les relations } Q_0(K \times H) = \mu(K) \cdot \nu(H), \\ \forall K \times H \text{ pavé mesurable de } C_0 = E_{-1} \times A_0 \text{ où } \mu \text{ est une probabilité quelconque sur } E, \text{ et } \forall n \geq 1, \\ Q_n(e_{-1}, a_0, \dots, e_{n-2}, a_{n-1}; K \times H) \\ \qquad \qquad \qquad = \int_K \beta_{n-1}(a_0, e_{n-2}, a_{n-1}; de_{n-1}) \alpha_n(a_0, a_{n-1}, e_{n-1}; H), \\ \forall (e_{i-1}, a_i) \in C_i \ (i = 0, \dots, n-1) \text{ et } \forall K \times H \text{ pavé mesurable de} \\ C_n = E_{n-1} \times A_n. \end{array} \right.$$

Soit alors $X = (G, \mathcal{G}, Q, (X_n) n \in \mathbb{N})$ le processus canonique engendré par le générateur $\mathcal{S} = \{Q_0, (Q_n(\cdot; \cdot), n \geq 1)\}$.

L'espace de base est

$$\underline{G}, \mathcal{G} = \prod_{i=0}^{\infty} C_i = \prod_{i=0}^{\infty} (E_{i-1} \times A_i), \quad \text{et} \quad \forall n \in \mathbb{N},$$

X_n est la coordonnée d'indice n de $(\underline{G}, \mathcal{G})$ dans $C_n = E_{n-1} \times A_n$. On notera éventuellement $X_n = (e_{n-1}, a_n)$, e_{n-1} (resp. a_n) désignant la composante de X_n sur E_{n-1} (resp. A_n).

On a les relations suivantes :

$$Q \{X_0 \in K \times H\} = Q_0(K \times H)$$

$$Q \{X_n \in K \times H \mid \mathcal{G}_{n-1}\} = Q_n(X^{n-1}; K \times H) \quad Q. \text{ p. s. sur } G.$$

$\forall K \times H$ pavé mesurable de $E \times A$ puisque le générateur \mathcal{S} est une version régulière du processus X (Proposition I), et, en introduisant les proba-

bilités de transition $Q_n^q(\cdot; \cdot)$ ($0 \leq q < n$) définies à partir des $Q_n(\cdot; \cdot)$ par les relations (1),

$$Q \{ X_n \in K \times H \mid \mathcal{G}_q \} = Q_n^q(X^q; K \times H) \quad Q. \text{ p. s. sur } G$$

Mais on peut d'après la définition de \mathcal{S} (relations (7)), faire les remarques suivantes : $Q_0(\underline{E} \times H) = \nu(H)$ (indépendant de μ), et

$$\text{donc} \quad \left. \begin{array}{l} Q_n(e_{-1}, a_0, \dots, e_{n-2}, a_{n-1}; K \times H) \\ Q_n^q(e_{-1}, a_0, \dots, e_{q-1}, a_q; K \times H) \end{array} \right\} \forall 0 \leq q < n,$$

sont indépendants de e_{-1} .

Par suite, les relations précédentes montrent que l'évolution du processus $(a_0, (X_n)_{n \geq 1})$ est indépendante de e_{-1} , donc de μ .

En particulier, $P \{ X_n \in K \times H \mid \mathcal{G}_k \}$ ($\forall 0 \leq k < n$) est indépendant de e_{-1} et de μ , et la limite (quand $n \rightarrow \infty$) de $P \{ X_n \in K \times H \mid \mathcal{G}_k \}$, si elle existe, est indépendante de μ .

C'est à ce processus $X = (G, \mathcal{G}, Q, (X_n)_{n \in N})$, à valeurs dans $E \times A$, que nous allons appliquer le théorème I, après avoir montré que la version régulière $\mathcal{S} = \{ Q_0, (Q_n(\cdot; \cdot), n \geq 1) \}$ de ce processus, et l'ensemble $H_0 = D \times C$, satisfont aux hypothèses de ce théorème.

1. — HYPOTHÈSE DE POSITIVITÉ.

Elle s'exprime ainsi :

$$\exists 0 < \delta < 1 \quad : \quad \forall k \in N, \quad Q_{k+1}(X^k; D \times C) \geq \delta.$$

D'après la définition des $Q_k(\cdot; \cdot)$ et des $\alpha_k(\cdot; \cdot)$ ($k \geq 1$), (voir les relations (7) et (6)),

$$Q_{k+1}(X^k; D \times C) = \int_D \beta_k(a_0, \dots, a_k; de_k) \alpha_{k+1}(a_0, \dots, e_k; C)$$

$$\text{et} \quad \alpha_{k+1}(a_0, \dots, e_k; C) \geq \theta G(e_k; C).$$

Mais, si $e_k \in D$, $G(e_k, C) = G_D(C) = \alpha$ (d'après les hypothèses 1 et 3), et $\beta_k(a_0, \dots, a_k; D) \geq \eta$ (d'après les hypothèses 2 et 3). En prenant $\delta = \theta \alpha \eta$, on a bien $0 < \delta < 1$ et $\forall k \in N, Q_{k+1}(X^k; D \times C) \geq \delta$.

2. — HYPOTHÈSE DE FAIBLE DÉPENDANCE DU PASSÉ LOINTAIN.

Elle s'exprime ainsi, compte tenu de ce que le processus X est défini sur N , et non sur Z :

$$\sum_{\mu=1}^{\infty} \varepsilon_{\mu}^{(1)}(D \times C) < \infty,$$

où, \forall l'entier $\mu \geq 1$,

$$E_{\mu}^{(1)}(\mathbf{D} \times \mathbf{C}) = \sup_{\substack{(k, l) \in \mathbb{N} \times \mathbb{N} \\ L \in \mathcal{E} \otimes \mathcal{A}}} \sup_{(g, g') \in E_{k, l}^{\mu}(\mathbf{D} \times \mathbf{C})} |Q_{k+1}(X^k(g); L) - Q_{k+l+1}(X^{k+l}(g'), L)|,$$

$E_{k, l}^{\mu}(\mathbf{D} \times \mathbf{C}) = \{(g, g') \in \underline{\mathbf{G}} \times \underline{\mathbf{G}} \text{ satisfaisant aux conditions } C_{k, l}^{\mu}(\mathbf{D} \times \mathbf{C})\}$
 et $(g, g') \in \underline{\mathbf{G}} \times \underline{\mathbf{G}}$ satisfont aux conditions $C_{k, l}^{\mu}$ si : \exists un ensemble de μ indices, soit $M = \{i_1, \dots, i_{\mu}\} \in \mathbb{N}$, tel que :

$$\begin{cases} k \geq i_1 > \dots > i_{\mu} \geq 0 \\ (X_{k-i}(g), X_{k+l-i}(g')) \in (\mathbf{D} \times \mathbf{C}) \times (\mathbf{D} \times \mathbf{C}) & \forall i \in M \\ X_{k-i}(g) = X_{k+l-i}(g') & \forall i \in \mathbb{N} - M, \quad 0 \leq i < i, \end{cases}$$

$\forall (k, l) \in \mathbb{N} \times \mathbb{N}$, $L \in \mathcal{E} \otimes \mathcal{A}$ et $(g, g') \in \underline{\mathbf{G}} \times \underline{\mathbf{G}}$, on pose

$$\Delta_{k, l, L}(g, g') = |Q_{k+1}(X^k(g), L) - Q_{k+l+1}(X^{k+l}(g'), L)|.$$

On cherche d'abord, $\forall (k, l) \in \mathbb{N} \times \mathbb{N}$, et $\forall L = K \times M$, pavé mesurable de $E \times A$, à majorer $\Delta_{k, l, K \times H}(\cdot, \cdot)$ sur $E_{k, l}^{\mu}(\mathbf{D} \times \mathbf{C})$.

D'après la définition des $Q_k(\cdot, \cdot)$, ($k \geq 1$) (relations (7)), on a, $\forall g$ et $g' \in \underline{\mathbf{G}}$

$$\left\{ \begin{array}{l} Q_{k+1}(X^k(g); K \times H) = \int_K \beta_k(a_0(g), \dots, a_k(g); de_{k+1}) \\ \qquad \qquad \qquad \qquad \qquad \qquad \qquad \alpha_{k+1}(a_0(g), \dots, a_k(g), e_k; H) \\ Q_{k+l+1}(X^{k+l}(g'); K \times H) = \int_K \beta_{k+l}(a_0(g'), \dots, a_{k+l}(g'); de_{k+l+1}) \\ \qquad \qquad \qquad \qquad \qquad \qquad \qquad \alpha_{k+l+1}(a_0(g'), \dots, a_{k+l}(g'), e_{k+l}; H). \end{array} \right.$$

On voit, d'après les hypothèses faites sur la règle d'expérimentation β , que si $\mu \geq m + 1$, et $(g, g') \in E_{k, l}^{\mu}(\mathbf{D} \times \mathbf{C})$, les probabilités

et
$$\begin{cases} \beta_k(a_0(g), \dots, e_{k-1}(g), a_k(g); \cdot) \\ \beta_{k+l}(a_0(g'), \dots, e_{k+l-1}(g'), a_{k+l}(g'); \cdot) \end{cases}$$

s'identifient à une même probabilité $Q_{\Lambda}(\cdot)$ sur E , Λ étant un élément de la partition $\sum_k \Lambda_k$ de $(A \times E)^m \times A$, introduite dans l'hypothèse 4. On a alors :

$$\Delta_{k, l, K \times H}(g, g') \leq \int Q_{\Lambda}(de) |\alpha_{k+1}(a_0(g), \dots, a_k(g), e; H) - \alpha_{k+l+1}(a_0(g'), \dots, a_{k+l}(g'), e; H)|$$

D'autre part, en appliquant la relation de récurrence

$$\alpha_{k+1}(a_0, \dots, a_n, e_n; \cdot) = (1 - \theta)\alpha_k(a_0, \dots, a_{n-1}, e_{n-1}; \cdot) + \theta G(e_n; \cdot)$$

à laquelle satisfont les $\alpha_n(\cdot; \cdot)$ ($n \geq 1$) (voir (6)), on obtient, $\forall e \in E$:

$$\left\{ \begin{array}{l} \alpha_{k+1}(a_0(g), \dots, a_k(g), e; H) \\ \qquad \qquad \qquad = (1 - \theta)\alpha_k(a_0(g), \dots, a_{k-1}(g), e_{k-1}(g); H) + \theta G(e, H) \\ \alpha_{k+l+1}(a_0(g'), \dots, a_{k+l}(g'), e; H) \\ \qquad \qquad \qquad = (1 - \theta)\alpha_{k+l}(a_0(g'), \dots, a_{k+l-1}(g'), e_{k+l-1}(g'); H) + \theta G(e, H) \end{array} \right.$$

Dans la différence, le terme $\theta G(e, H)$ disparaît, et il reste :

$$\Delta_{k,l,K \times H}(g, g') \leq (1 - \theta) |\alpha_k(a_0(g), \dots, e_{k-1}(g); H) - \alpha_{k+l}(a_0(g'), \dots, e_{k+l-1}(g'); H)|$$

Mais, si $(g, g') \in E_{k,l}^\mu(D \times C)$, $\forall i = 1, \dots, \mu$, $e_{k-i}(g)$ et $e_{k+l-i}(g')$ coïncident, ou appartiennent l'un et l'autre à l'ensemble D, et par suite

$$G(e_{k-i}(g); \cdot) = G(e_{k+l-i}(g'); \cdot)$$

En appliquant $(\mu - 1)$ fois la relation de récurrence sur les α_n , aux expressions

$$\text{et} \quad \left\{ \begin{array}{l} \alpha_k(a_0(g), \dots, e_{k-1}(g); H) \\ \alpha_{k+l}(a_0(g'), \dots, e_{k+l-1}(g'); H) \end{array} \right.$$

on obtient alors

$$\Delta_{k,l,K \times H}(g, g') \leq (1 - \theta)^\mu$$

On a ainsi montré que $\forall (k, l) \in \mathbb{N} \times \mathbb{N}$, la relation

$$(8) \quad \text{Sup}_{(g, g') \in E_{k,l}^\mu(D \times C)} |Q_{k+1}(X^k(g); L) - Q_{k+l+1}(X^{k+l}(g'); L)| \leq (1 - \theta)^\mu$$

est vérifiée $\forall L$, pavé mesurable de $E \times A$.

La classe des ensembles $L \in \mathcal{E} \otimes \mathcal{A}$ pour lesquels la relation (8) est vérifiée, contient donc la semi-algèbre des pavés mesurables. On voit d'autre part que cette classe est une classe monotone, et qu'elle est stable par complémentation. Elle contient donc tout ensemble $L \in \mathcal{E} \otimes \mathcal{A}$. Et par suite, on a

$$\varepsilon_\mu^{(1)}(D \times C) \leq (1 - \theta)^\mu, \quad \forall \mu \geq m + 1,$$

d'où

$$\sum_{\mu=1}^{\infty} \varepsilon_\mu^{(1)}(D \times C) < \infty.$$

La seconde hypothèse est vérifiée.

Au processus $X = (\underline{G}, \mathcal{G}, Q, (X_n)_{n \in \mathbb{N}})$, à valeurs dans $E \times A$, on peut alors appliquer la conclusion du théorème I :

Il existe une probabilité Π sur $E \times A$, telle que $\forall L \in \mathcal{L} \otimes \mathcal{A}$, et $\forall k \in \mathbb{N}$,

$$\lim_{n \rightarrow \infty} Q \{ X_n \in L \mid \mathcal{G}_k \} = \Pi(L) \quad Q. \text{ p. s. sur } G$$

d'où, en particulier, $\forall H \in \mathcal{A}$

$$(9) \quad \lim_{n \rightarrow \infty} Q \{ X_n \in \underline{E} \times H \mid \mathcal{G}_k \} = \Pi(\underline{E} \times H) \quad Q. \text{ p. s. sur } G$$

De la convergence de ce processus auxiliaire X , on va alors déduire la convergence du processus $Y = (\underline{B}, \mathcal{B}, P, (a_n, e_n)_{n \in \mathbb{N}})$, en montrant que, pour la probabilité γ sur A , définie par les relations

$$(10) \quad \gamma(H) = \Pi(\underline{E} \times H), \quad \forall H \in \mathcal{A},$$

on a, $\forall k \in \mathbb{N}$, et $\forall H \in \mathcal{A}$:

$$\lim_{n \rightarrow \infty} P \{ a_n \in H \mid \mathcal{B}_{k-1} \} = \gamma(H) \quad P. \text{ p. s. sur } \underline{B}$$

On a en effet :

$$P \{ a_n \in H \mid \mathcal{B}_{k-1} \} = v_n^{k-1}(a_0, e_0, \dots, a_{k-1}, e_{k-1}; H) \quad P. \text{ p. s.}$$

et

$$v_n^{k-1}(a_0, e_0, \dots, a_{k-1}, e_{k-1}; H) = \int \alpha_k(a_0, e_0, \dots, e_{k-1}; da_k) Q_n^k(e_{-1}, a_0, \dots, e_{k-1}, a_k; \underline{E} \times H)$$

(voir les définitions des $v_n^q(\cdot; \cdot)$ et des $Q_n^q(\cdot; \cdot)$ ($q < n$) à partir des relations (1)).

D'autre part,

$$Q_n^k(e_{-1}, a_0, \dots, a_k; \underline{E} \times H) = Q \{ X_n \in \underline{E} \times H \mid \mathcal{G}_k \} \quad Q. \text{ p. s.}$$

donc

$$\lim_{n \rightarrow \infty} Q_n^k(e_{-1}, a_0, \dots, a_k; \underline{E} \times H) = \gamma(H)$$

d'après (9) et (10).

On a donc, $\forall H \in \mathcal{A}$, et $\forall k \in \mathbb{N}$,

$$\lim_{n \rightarrow \infty} P \{ a_n \in H \mid \mathcal{B}_{k-1} \} = \gamma(H) \quad P. \text{ p. s.,}$$

et donc aussi

$$\lim_{n \rightarrow \infty} P \{ a_n \in H \} = \gamma(H)$$

comme il fallait le démontrer.

INTERPRÉTATION. — Si l'on interprète le processus Y construit à partir des données $(\Delta) = ((\delta), (\beta))$ satisfaisant aux hypothèses du théorème II, comme une prédiction sur le processus expérimental $(a_n, e_n)_{n \in \mathbb{N}}$ obtenu lorsque l'expérience est réalisée avec la règle d'expérimentation (β) , la propriété de convergence que l'on vient de mettre en évidence pour le processus Y , exprime la prédiction suivante : d'une part, lorsque $n \rightarrow \infty$, la réponse du sujet à l'essai n devient indépendante des résultats des k premiers essais, $\forall k$ — l'influence du passé devient donc négligeable quand ce passé devient très lointain ; d'autre part, la probabilité suivant laquelle le sujet choisit sa réponse dans A , à l'essai n , tend vers une probabilité limite γ sur A , indépendante de n ; il y a donc acquisition, par le sujet, d'un comportement asymptotique stable — ce que l'on peut exprimer encore en disant qu'il y a apprentissage.

BIBLIOGRAPHIE

- [1] W. DOEBLIN et R. FORTET, Sur des chaînes à liaisons complètes. *Bull. Soc. Math. France*, **65**, 1937, p. 132-148.
- [2] J. LAMPERTI et P. SUPPES, Chains of infinite order and their application to learning theory. *Pacific. J. Math.*, **9**, n° 3, 1959, p. 739-754.
- [3] Mme M. C. BERT, Convergence d'une chaîne à liaisons complètes et application à un modèle d'apprentissage. *C. R. Acad. Sci. Paris*, **259**, 1964, p. 1938-1940.
- [4] J. NEVEU, Bases mathématiques du calcul des probabilités. Masson et Cie, éditeurs, Paris, 1964.
- [5] P. COURREGE et H. ROUANET, Sur les fondements des modèles stochastiques d'adaptation. Chaire de Calcul des probabilités, I. H. P., Paris, 1964.

Manuscrit reçu le 4 septembre 1967.