MATHÉMATIQUES ET SCIENCES HUMAINES

H. ROUANET

Les modèles stochastiques d'apprentissage

Mathématiques et sciences humaines, tome 5 (1964), p. 3-10

<http://www.numdam.org/item?id=MSH_1964__5__3_0>

H. ROUANET

LES MODELES STOCHASTIQUES D'APPRENTISSAGE*

INTRODUCTION

Depuis une quinzaine d'années, on assiste à un développement considérable des modèles mathématiques dans les sciences sociales. En Psychologie, en particulier, on a vu apparaitre à la suite de Bush et Mosteller (1955), Estes (1950) etc..., un nombre considérable de modèles dits "Modèles stochastiques d'apprentissage". Le présent exposé a pour but de fournir non pas une revue, mais une introduction à cette littérature. Il sera divisé en quatre parties: nous tenterons d'abord de situer le domaine d'application de ces modèles, puis nous présenterons un modèle général; en particularisant ce modèle général, nous retrouverons divers modèles effectivement utilisés; nous donnerons enfin, à titre d'illustration, des exemples d'applications.

En se reportant aux publications signalées en bibliographie, le lecteur pourra se faire une idée plus complète des questions abordées par ces modèles, aussi bien sur le plan mathématique que sur le plan expérimental.

I - Champ d'application des modèles stochastiques d'apprentissage

Il faut situer les modèles stochastiques d'apprentissage dans un contexte psychologique, et plus précisément dans le courant de la psychologie appelé behaviorisme. On sait que celui-ci, né il y a un demi-siècle en s'opposant aux méthodes traditionnelles fondées sur l'introspection, s'est donné pour programme de rendre compte du comportement ("behavior") des organismes (animaux ou humains) à partir des seules données observables. Développé parallèlement aux recherches sur le conditionnement de Pavlov et son école, ce courant a conduit à définir des notions précises: stimulus, réponse, renforcement, pour l'analyse objective des situations expérimentales.

Ce programme a conduit à étudier des situations très simples, où il est possible de contrôler avec précision tous les paramètres. D'où l'intérêt porté en particulier aux expériences de psychologie animale. C'est à celle-ci que nous emprunterons, pour préciser les notions précédentes sur un exemple (sans alourdir l'exposé en donnant des définitions techniques) le schéma d'une expérience classique, celle du labyrinthe en T "avec procédure de correction". A chaque essai, l'animal, préalablement affamé, est placé à l'entrée d'un couloir qui se ramifie en deux autres couloirs, droit et gauche. A l'extrémité du couloir droit, par exemple, on place de la nourriture. On observe quel est le comportement de l'animal, et en particulier quel est le premier couloir emprunté par le rat (si le premier couloir n'est pas celui de droite, on laisse l'animal revenir sur ses pas vers l'embranchement, jusqu'à ce qu'il ait trouvé la nourriture). On constate qu'au bout d'un certain nombre d'essais (disons une cinquantaine), l'animal emprunte toujours en pre-

* Ce texte reprend pour l'essentiel deux exposés faits au groupe d'Etude sur l'automatique non-numérique, les 23 et 30 Avril 1963.

mier le couloir de droite; on dit qu'il y a eu apprentissage de la réponse "droite". Dans une telle expérience, on peut caractériser le $n^{ième}$ essai par la succession d'une réponse a_n (le rat va à droite ou à gauche) et d'un renforcement e_n (la nourriture est à droite). Dans l'exemple précédent, le renforcement est constant, puisque la nourriture est toujours placée à droite. On utilise souvent des règles de renforcement plus compliquées. Une règle fréquemment employée est la règle dite aléatoire indépendante, où l'on place la nourriture à droite avec une probabilité π et à gauche avec une probabilité $1 - \pi$, π étant un paramètre fixé une fois pour toutes par l'expérimentateur. Si π est différent de 0 ou 1, on observe, après un grand nombre d'essais, non plus une tendance du rat à aller toujours du même côté, mais une tendance à donner à la longue la réponse "droite" avec une fréquence voisine de π. Ce résultat, retrouvé dans diverses expériences de même structure, a été parfois appelé "loi de l'ajustement". Il faut noter que dans ce cas les psychologues parlent encore d'apprentissage, car il y a eu une modification stable du taux d'apparition d'une réponse déterminée.

Chez l'homme, beaucoup d'expériences d'esprit analogue ont été également effectuées. A titre d'exemple, nous décrirons brièvement une procédure classique de l'expérience dite d'"association par paires". Au premier essai, on présente au sujet une liste comportant un nombre déterminé de paires d'items (syllabes sans signification ou chiffres). Le premier élement de chaque paire est considéré comme le stimulus, le deuxième élément constitue la réponse correcte à ce stimulus. Au deuxième essai, on présente successivement chaque stimulus, en demandant au sujet de fournir la réponse correspondante; après la réponse du sujet, l'expérimentateur fait connaître la réponse correcte (renforcement). Chacun des essais suivants s'effectue de façon semblable au deuxième essai (naturellement, l'ordre de présentation des stimuli est changé à chaque essai). L'expérience est poursuivie jusqu'à ce qu'un critère d'apprentissage (par ex.: aucune erreur lors de trois présentations consécutives) soit atteint. Malgré son apparence très simple, ce genre d'expérience a donné lieu à de nombreux débats au niveau des théories psychologiques de l'apprentissage. L'un des points les plus controversés concerne le rôle de la répétition dans la formation des associations. Selon la théorie continuiste classique, chaque répétition doit entraîner l'augmentation graduelle de la "force de la liaison" (variable latente hypothétique) entre le stimulus et la réponse correspondante. Cependant une expérience récente de Rock* semble prouver au contraire que la répétition ne joue aucun rôle dans la formation de l'association, qui s'établirait selon un processus discontinu. Rock modifie ainsi la procédure classique: dans le groupe expérimental, après chaque essai, les paires non apprises sont éliminées et remplacées par des paires nouvelles. De cette façon, les associations apprises sont toutes acquises à la première présentation. Or, si on soumet un groupe-contrôle parallèle à la procédure classique, on constate que le critère d'apprentissage est atteint aussi vite par le groupe expérimental que par le groupe-contrôle.

Ce genre de découverte a amené certains chercheurs à examiner plus en détail les données empiriques accumulées au cours d'innombrables expériences, et dont on n'avait pas toujours extrait toute l'information quantitative possible. Dès lors, l'intérêt presque exclusif que l'on avait jusque là porté aux courbes moyennes d'apprentissage a progressivement fait place à une analyse de plus en plus serrée des phénomènes séquentiels (étude d'essai en essai), voire à l'examen des séries

---

* Rock, I. The role of repetition in associative learning Amer. J. Psychol., 70, 1957, 186-193.

individuelles. Des modèles mathématiques sont apparus peu à peu servant de guide à ces recherches et suscitant en même temps, l'étude de nouvelles situations expérimentales. Ce caractère d'adaptation empirique progressive explique d'une part la variété de ces modèles, d'autre part le fait qu'ils n'ont été rigoureusement formalisés qu'assez tard et de façon encore incomplète. En dépit de cette diversité, ces modèles présentent un certain nombre de traits communs:

- leur but est toujours descriptif et jamais normatif; ils cherchent à rendre compte du comportement des organismes, et non pas à leur enseigner ce qu'ils ont à faire - ils ne cherchent pas non plus à fabriquer des mécanismes capables de performances déterminées.

- ce sont tous des modèles probabilistes. Ce dernier caractère fait de ces modèles des outils plus souples que les modèles déterministes (certains modèles déterministes avaient été proposés dans le passé). Il est clair qu'il suffit d'une exception pour invalider un modèle déterministe, alors qu'un modèle probabiliste laisse à l'organisme étudié une part plus ou moins grande de liberté. C'est le caractère probabiliste de ces modèles qui leur a valu le nom plus répandu de modèles stochastiques d'apprentissage.

Comme premier exemple simple de ces modèles, nous dirons quelques mots du modèle utilisé par Bower (1961) pour formaliser, dans la situation d'association par paires, une interprétation discontinue de l'acquisition voisine de celle avancée par Rock. Bower utilise la procédure traditionnelle mentionnée plus haut, avec dix stimuli, à associer à deux réponses seulement: le chiffre 1 pour cinq des stimuli, et le chiffre 2 pour les cinq autres. Pour chaque stimulus, on obtient une suite de réponse a_n (où a_n est, soit la bonne réponse a, soit la mauvaise réponse a') et de renforcements e_n (comme on a toujours $e_n = a$, le renforcement ici est en fait constant). Le modèle utilisé par Bower consiste à décrire la structure de ces suites obtenues pour chaque stimulus*. Selon ce modèle, le sujet peut se trouver dans l'un ou l'autre de deux états z (il connaît la bonne réponse) et z' (il ne connaît pas la bonne réponse). Si le sujet est dans l'état z, il donne la bonne réponse avec une probabilité égale à 1; s'il se trouve dans l'état z', il donne la bonne réponse avec une probabilité γ (en général $\gamma = 1/2$). A partir de l'état z', le sujet passe à l'état z avec la probabilité c (où c est un paramètre, $0 < c \leqslant 1$); si le sujet est dans l'état z, il reste dans cet état.

On peut donner à ce modèle de Bower une présentation plus formelle et qui se prêtera à la généralisation. Pour chaque stimulus, nous définissons un ensemble $A = \{a, a'\}$ (ensemble des réponses).

L'ensemble $E = \{a\}$ est l'ensemble (réduit à un élément) des renforcements. Le modèle postule qu'il existe un ensemble $Z = \{z, z'\}$ dit ensemble des états du sujet. Les axiomes du modèle stipulent (on pose $X_n = (z_n, a_n, e_n)$).

$$(R_0) \quad P\left[z_0 = z\right] = p_0 \quad \text{(où } p_0 \text{ est un paramètre)}$$

*Il est clair qu'en se plaçant ainsi "à stimulus constant", on suppose implicitement que les apprentissages des différentes suites se font de façon indépendante. La validité de cette supposition est plus ou moins acceptable selon la procédure expérimentale. Les modèles sont en tous cas des premières approximations, et d'autre part, dans la discussion qui va suivre, l'accent doit être mis sur la structure du modèle plutôt que sur sa validité intrinsèque.

$$(R_1) \quad P\left[a_n = a \mid z_n;\ X_0, X_1, \ldots, X_{n-1}\right] = P_a(z_n)$$
$$P\left[a_0 = a \mid z_0\right] = P_a(z_0)$$
$$\text{où} \quad P_a(z) = 1$$
$$P_a(z') = \gamma$$
$$(\text{avec } 0 \leqslant \gamma \leqslant 1, \text{ en fait ici } \gamma = \tfrac{1}{2})$$

$$(R_2) \quad P(z_{n+1} = z \mid z_n, a_n, e_n;\ X_0, X_1, \ldots X_{n-1}) = P_z(z_n)$$
$$P(z_1 = z \mid z_0, a_0, e_0) = P_z(z_0)$$
$$\text{où} \quad P_z(z) = 1$$
$$P_z(z') = c > 0$$

Aux axiomes du modèle proprement dits, il faut ajouter la condition

$$(R_3) \quad P(e_n = a \mid z_n, a_n;\ X_0, X_1, \ldots X_{n-1}) = 1$$

qui exprime que le renforcement est constant et égal à a.

Mathématiquement il est facile de voir que ces axiomes entraînent la convergence presque sûre de a_n vers a. D'une façon plus générale on peut étudier les propriétés du processus stochastique (a_n). Bower a calculé de nombreuses prédictions théoriques de ce modèle (nombre moyen d'erreurs, essai correspondant à la dernière erreur, etc.), ce qui lui a permis de le confronter en grand détail avec les données expérimentales.

II - Un modèle général d'apprentissage

Le lecteur qui désirerait se familiariser avec l'ensemble des modèles stochastiques d'apprentissage risque d'être rebuté par la diversité, considérable en apparence, de ces modèles (diversité corrélative des situations psychologiques) et par la rareté de systèmes formels généraux, englobant des familles de modèles adaptés à des situations particulières. C'est pourquoi il peut être utile de présenter un modèle général synthétique qui peut être regardé comme la généralisation de la plupart des modèles actuellement utilisés. Les propriétés de ce modèle sont actuellement en cours d'étude (P. Courrège et H. Rouanet, 1964, en préparation). Nous présenterons ici les axiomes de ce modèle et un théorème d'existence fondamental.

Considérons:

a) b) c) d) Quatre espaces mesurables* $(S, \mathcal{S})$, $(A, \mathcal{A})$, $(E, \mathcal{E})$ et $(Z, \mathcal{Z})$ dits respectivement des stimuli, des réponses, des renforcements** et des états.

e) Une probabilité Q_0 sur $(Z, \mathcal{Z})$.

---

* Un espace mesurable est la donnée d'un ensemble (A..) et d'une famille particulière de ses parties-tribu - $(\mathcal{A}\ldots)$.

**Le mot "renforcement" sera pris ici dans le sens de "renforcement observable", contrôlé par l'expérimentateur; il peut être préférable dans certains contextes, de lui substituer le terme plus neutre d'"issue" ou de "conclusion" de l'essai.

f) une probabilité de passage $P_a(\mathfrak{z}, s, .)$ de $Z \times S$ dans A

g) " " $P_{\mathfrak{z}}(\mathfrak{z}, s, a, e, .)$ de $Z \times S \times A \times E$ dans Z.

h) une probabilité de passage P_e (s, a, .) de $S \times A$ dans E dite règle de renforcement.

i) une probabilité de passage P_s (s, a, e, .) de $S \times A \times E$ dans S dite règle de présentation des stimuli.

Nous chercherons maintenant à déterminer un processus stochastique*.

$$X = (\Omega, \mathcal{F}, P, (X_n))$$

à valeurs dans $(\mathfrak{Z} \times \mathfrak{S} \times \mathfrak{A} \times \mathfrak{E}, \mathfrak{z} \times \mathfrak{s} \times \mathfrak{a} \times \mathfrak{e})$

et satisfaisant aux relations (R_i) suivantes**.

<u>Relations</u> (R_i)***

(R_o) $P(\mathfrak{z}_o \in \Delta) = Q_o(\Delta), \quad \forall \Delta \in \mathfrak{z}$

$P(s_o \in \Lambda) = R_o(\Lambda) \quad \forall \Lambda \in \mathfrak{s}$

(R_1) $P(a_n \in \Gamma | \mathfrak{z}_n, s_n; X_o, X_1, \ldots X_{n-1}) = P_a(\mathfrak{z}_n, s_n, \Gamma), \forall \Gamma \in \mathfrak{a}$

(R_2) $P(\mathfrak{z}_{n+1} \in \Delta | \mathfrak{z}_n, s_n, a_n, e_n; X_o, X_1, \ldots X_{n-1}) = P_z(\mathfrak{z}_n, s_n, a_n, e_n, \Delta), \forall \Delta \in \mathfrak{z}$

(R_3) $P(e_n \in H | s_n, a_n; X_o, X_1 \ldots X_{n-1}) = P_e(s_n, a_n, H), \quad \forall H \in \mathfrak{e}$

(R_4) $P(s_{n+1} \in G | s_n, a_n, e_n; X_o, X_1, \ldots X_{n-1}) = P_o(s_n, a_n, e_n, G), \forall G \in \mathfrak{s}$

Concrétement, les données a) b) c) correspondent aux <u>données observables</u> qui définissent la situation expérimentale:

S ensemble des stimuli

A ensemble des réponses

E ensemble des renforcements.

A chaque essai n se succédent les trois événements observables: s_n, a_n, e_n.

Les données d), e), f), g) sont des <u>constructions hypothétiques</u> qui caractérisent le modèle.

Z est dit ensemble des états du sujet.

---

* On sait qu'un processus stochastique $(\Omega, \mathcal{F}, P, (X_n))$ à valeur dans $(U, \mathcal{U})$ consiste en la donnée d'un espace de probabilité $(\Omega, \mathcal{F}, P)$ et d'une suite X_n d'applications mesurables de Ω dans U.

** En toute rigueur, on suppose que ces relations sont vérifiées seulement presque sûrement.

***A ces relations il faut adjoindre des relations analogues pour le cas

n = o

La relation R_0 précise la probabilité initiale.

La relation R_1 exprime que la réponse a_n à l'essai n ne dépend des événements antérieurs que par l'intermédiaire de l'état $\mathfrak{z}_n$ du sujet à cet essai et du stimulus présente à cet essai.

La relation R_2 exprime que l'état z_{n+1} du sujet à l'essai n+1 ne dépend des événements antérieurs que par l'intermédiaire de $\mathfrak{z}_n$, s_n, e_n, a_n.

La donnée h) et la relation (R_3) expriment la règle de renforcement choisie par l'expérimentateur - règle de renforcement très particulière (appelée souvent "simplement dépendante") qui consiste à choisir à l'essai n le renforcement de sorte qu'il ne dépende des événements antérieurs que par l'intermédiaire de la réponse a_n qui vient d'être donnée par le sujet. De même, la donnée i) et la relation (R_4) expriment la règle de présentation des stimuli, règle de présentation particulière, qui consiste à choisir à l'essai n + 1 le stimulus s_{n+1} de sorte qu'il ne dépende des événements antérieurs que par l'intermédiaire de s_n, a_n, e_n.

En faisant des hypothèses simplificatrices sur les fonctions P_e et P_0, on obtient des cas particuliers souvent utilisés pratiquement:

- Si P_s est une fonction constante en s, a, e, la règle de présentation des stimuli est dite indépendante.
- Si P_s est une mesure de Dirac pour chaque s, a, e, la règle de présentation est déterministe.
- La conjonction des deux cas précédents correspond au cas où le stimulus s_n est constant d'essai à essai.
- Si P_e est une fonction constante en s_a, la règle de renforcement est dite "aléatoire indépendante".
- Si P_e est une mesure de Dirac pour chaque s_a, la règle de renforcement est déterministe.
- La conjonction des deux cas précédents correspond au cas du renforcement constant dont un exemple a été donné dans la première partie (situation d'association par paires).

Ce qui suit tend à montrer qu'il existe un processus stochastique et un seul satisfaisant à ces diverses données et relations.

Construisons les fonctions Q et π_0 suivantes:

$$Q(z,s,a,e,M) = \int_{\mathbf{Z}} P_z(z,s,a,e,dz') \int_S P_s(s, a,e,ds') \int_A P_a(z', a', da') \int_E P_e(s', a', de')\, 1_M(z', s', a', e')$$

où

$$M \in \mathfrak{Z} \times \mathfrak{S} \times \mathfrak{A} \times \mathfrak{E}$$

et 1_M est la fonction indicatrice de M.

$$\pi_0(M) = \int_{\mathbf{Z}} Q_0(d\mathfrak{z}) \int_S R_0(ds) \int_A P_a(\mathfrak{z},s,da) \int_E P_e(s, a, de)\, 1_M(\mathfrak{z},s,a,e)$$

On vérifie immédiatement que

- Q (z,s,a,e,M) est une probabilité de passage de Z x S x A x E dans Z x S x A x E
- π_o (M) est une probabilité sur Z x S x A x E.

Il est alors possible de prouver le théorème suivant, que nous énoncerons sans démonstration.

Lemme

Soit $X = \left[\Omega, \mathfrak{F}, P, (X_n)_{n \in N}\right]$ un processus à valeurs dans Z x S x A x E; les deux propositions suivantes sont équivalentes:

a) X est satisfait aux conditions (R_i)

b) X est un processus de Markoff stationnaire, admettant Q pour fonction de transition et π_o pour distribution de probabilité initiale.

(en d'autres termes,

$$P\left[X_{n+1} \in M \mid X_n; X_o, X_1, \dots X_{n-1}\right] = P\left[X_{n+1} \in M \mid X_n\right] = Q(X_n, M)$$

et

$$P(X_o \in M) = \pi_o(M)$$

$$\forall M \in Z \times S \times A \times E$$

En permettant de remplacer la recherche d'un processus satisfaisant aux conditions (R_i) par la recherche d'un processus de Markoff, le lemme entraine immédiatement le théorème fondamental d'existence, grâce à l'application du théorème classique de Tulcea*.

Théorème fondamental

Soit $\Omega = (ZxSxAxE)^N$ (ensemble des suites de points de ZxSxAxE) et, pour tout n, $n \in N$ soit X_n la projection de Ω sur ZxSxAxE. Soit $\mathfrak{F}$ la tribu sur Ω engendrée par les x_n $(n \in N)$. Alors:

a) il existe <u>une probabilité P et une seule sur l'espace</u> mesurable $(\Omega, \mathfrak{F})$ telle que le processus

$$X = (\Omega, \mathfrak{F}, P, (X_n)_{n \in N})$$

satisfasse aux relations (R_i)

b) la probabilité P satisfait aux relations (caractère <u>markovien stationnaire</u>)

$$P\left[X_{n+1} \in M \mid X_n; X_o, X_1, \dots X_{n-1}\right] = P\left[X_{n+1} \mid X_n\right] = Q\left[X_n, M\right]$$

$$P\left[X_o \in M\right] = \pi_o(M)$$

pour tout $M \in Z \times S \times A \times E$

où Q et π_o sont les fonctions définies plus haut.

---

*Tulcea, I. Mesures dans les espaces produits. *Atti Accad. Naz. Lincei Rend.* 7. (1949) - 208-211.

Processus dérivés

Les conditions (R_i) étant toujours supposées vérifiées, on peut, à partir du processus markovien fondamental de terme

$$X_n = (\mathfrak{z}_n, s_n, a_n, e_n)$$

étudier les propriétés des processus dérivés. Les processus dérivés, de termes $(\mathfrak{z}_n, s_n, a_n)$, $(\mathfrak{z}_n, s_n, e_n)$, $(\mathfrak{z}_n, s_n)$ sont encore markoviens. Par contre, les processus qui ne font pas intervenir le couple $(\mathfrak{z}_n, s_n)$ comme le processus des événements observables (s_n, a_n, e_n) ne sont pas en général markoviens. Or c'est à ce processus qu'on s'intéresse. La démarche habituelle dans l'étude d'un modèle particulier, est la suivante. On étudie les propriétés des processus markoviens, et on en déduit les propriétés des processus observables, propriétés qui peuvent être comparées aux résultats expérimentaux. Nous montrerons des exemples de cette démarche dans la troisième partie.

Rappelons, avant de clore cette partie, que nous n'avons prouvé l'existence et l'unicité, d'une part, et le caractère markovien, d'autre part, du processus X_n, qu'en faisant des hypothèses particulières (caractère simplement dépendant) sur les règles de renforcement et de présentation des stimuli. Dans le cas de règles plus générales, on peut encore démontrer l'existence et l'unicité d'un processus X_n (Courrège et Rouanet, 1964), mais ce processus n'est plus markovien en général, ce qui en limite certainement l'intérêt pratique.

(Suite au prochain numéro)