REVUE FRANÇAISE D'INFORMATIQUE ET DE RECHERCHE OPÉRATIONNELLE. SÉRIE ROUGE

ALAIN VIGUIER

Traitement d'un modèle de processus à stochasticités imbriquées (dynamic inference)

Revue française d'informatique et de recherche opérationnelle. Série rouge, tome 4, nº R1 (1970), p. 95-107

<http://www.numdam.org/item?id=M2AN_1970__4_1_95_0>

R.I.R.O.
(4e année, N° R-1, 1970, p. 95-107)

TRAITEMENT D'UN MODELE DE PROCESSUS A STOCHASTICITES IMBRIQUEES (DYNAMIC INFERENCE)

par Alain VIGUIER (1)

Résumé. — *Dans un article paru en 1965, Ronald A. Howard décrivait un modèle de processus à stochasticités imbriquées. On reprend ici la présentation de ce modèle et l'on donne une méthode itérative pour le traitement des observations passées en vue de la prévision des réalisations futures de processus.*

Dans un article paru en 1965 [1], Ronald A. Howard décrivait un modèle de processus à stochasticités imbriquées. Nous reprenons ici la présentation de ce modèle (I) et développons une méthode itérative pour le traitement des observations passées en vue de la prévision des réalisations futures du processus (II).

Cette méthode générale réduit considérablement le volume des calculs. Le traitement d'un cas concret par un ordinateur ne soulève plus aucune difficulté, même lorsque l'on opère sur un très grand nombre d'observations (III, § 1) ou pour un temps très long (III, § 2).

L'utilité de ce modèle réside dans la simulation et dans l'interprétation de phénomènes qui jusqu'alors étaient qualifiés d'étranges : en météorologie, la suite parfois déconcertante des jours pluvieux et des jours « secs », en marketing, celle des achats de tel ou tel produit, etc. C'est que ce modèle fonctionne avec ce que nous appellerons une mémoire : la suite des réalisations passées du processus est utilisée intégralement, avec sa chronologie, pour prévoir les réalisations futures.

La notation $\{ x \mid E \}$ désigne la loi ou la distribution de probabilité de la variable aléatoire x étant donnée l'information E. La même notation $\{ A \mid E \}$ où A est un événement désigne la probabilité de A conditionné par E.

(1) Departement d'Automatique et d'Informatique Économique.

Nous utiliserons aussi le symbole S symbole de « sommation généralisé », représentant l'opérateur de sommation discrète $(\sum)$ ou bien celui d'intégration $\left(\int\right)$, suivant le domaine de la variable sur laquelle il opère.

Nous utiliserons aussi le lemme suivant :

Lemme : *Si* $\{ A \mid B, C \} = \{ A \mid B \}$, *alors* $\{ C \mid B, A \} = \{ C \mid B \}$
en effet :

$$\{ A \mid B, C \} = \{ A, B, C \}/\{ B, C \} \qquad \text{et} \qquad \{ A \mid B \} = \{ A, B \}/\{ B \}$$

donc

$$\{ A, B, C \} = \{ A, B \} \{ B, C \}/\{ B \}$$

d'autre part

$$\{ C \mid B, A \} = \{ A, B, C \}/\{ B, A \}$$

donc

$$\{ C \mid B, A \} = \{ A, B \} \{ B, C \}/\{ A, B \} \{ B \} = \{ B, C \}/\{ B \} = \{ C \mid B \}$$

Ce lemme est une généralisation de l'indépendance simple :

$$\{ A \mid C \} = \{ A \} \Rightarrow \{ C \mid A \} = \{ C \}$$

Cela nous permettra de parler de l'indépendance d'événements ou d'informations conditionnées.

I. PRESENTATION DU MODELE

Le temps est supposé discret. Les réalisations du processus ont lieu aux instants $1, 2, 3 \ldots n \ldots$. Ces réalisations, que nous observons directement, sont notées $x(1)$, $x(2) \ldots x(n) \ldots$. L'origine des temps est le temps 1.

Les $x(i)$ sont tirées d'une loi statistique donnée qui dépend d'un ou plusieurs paramètres, eux-mêmes tirés d'une autre loi donnée, les tirages s'effectuent en des temps déterminés par les tirages d'une troisième loi. En pratique, cela fonctionne de la manière suivante :

— Une variable aléatoire discrète ρ, tirée d'une distribution statistique donnée, $\{ \rho \}$, décrit l'histoire des changements de paramètre :

— Au temps $C_1 = 1$, on tire une valeur ρ_1 de $\{ \rho \}$.

— Au temps $C_2 = C_1 + \rho_1$, on tire une valeur ρ_2 de $\{ \rho \}$, etc.

— Au temps $C_k = C_{k-1} + \rho_{k-1}$, on tire une valeur ρ_k de $\{ \rho \}$.

— En chaque point C_k qui sera dit « point du $k^{\text{ième}}$ changement », on décide de changer la valeur du paramètre et l'on fait un tirage de la loi $\{ \mu \}$, ce qui nous fournit la nouvelle valeur $\mu(C_k)$.

— En tout point i de l'intervalle $[C_k, C_k + \rho_k - 1]$, on conserve la même valeur $\mu(C_k)$ du paramètre et l'on fait un tirage de la loi $\{ x \mid \mu = \mu(C_k) \}$, ce qui nous donne la « sortie au temps i » $x(i)$.

On a les trois lois $\{ \rho \}$, $\{ \mu \}$, $\{ x \mid \mu \}$.

On observe les sorties $x(1)$, $x(2)$ ... $x(n - 1)$. Que peut-on dire de la future sortie $x(n)$ et des dates de changement de paramètre, connaissant seulement cette suite $X_{n-1} = [x(1), x(2) ... x(n - 1)]$?

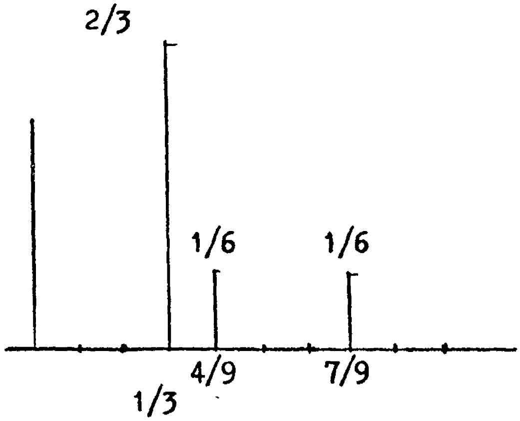
Prenons un exemple simple et pratique afin d'illustrer cela.

Exemple

Règle du jeu

1° Pour jouer à pile ou face, on dispose de trois pièces biaisées dont les probabilités de sorties « pile » sont respectivement :

pièce n° 1 : 1/3,
pièce n° 2 : 4/9,
pièce n° 3 : 7/9.



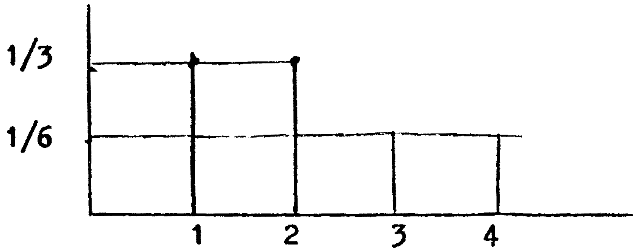
2° Avant chaque partie, on tire une pièce de telle sorte que la probabilité de choisir telle ou telle soit :

pièce n° 1 : 2/3,
pièce n° 2 : 1/6,
pièce n° 3 : 1/6.

3° Quand on a tiré une pièce, on la conserve jusqu'à la fin de la partie. Puis, on tire à nouveau une pièce, etc.

4° La longueur des parties, comptée en nombre de coups, est supposée obéir à la distribution suivante :

probabilité de jouer 1 coup : 1/3,
probabilité de jouer 2 coups : 1/3,
probabilité de jouer 3 coups : 1/6,
probabilité de jouer 4 coups : 1/6.



5° Nous sommes placés en observateur tenu à l'écart du déroulement du jeu. On ne nous laisse voir que le résultat (pile ou face) à chaque instant.

Formulation

Ce jeu peut se formuler de la manière suivante :

Au temps i, la « sortie » est $x(i) = 1$ (pile) ou 0 (face).

Le paramètre est $\mu(i) = \{ x(i) = 1 \}$, probabilité de sortir « pile ».

1° La loi de probabilité de la sortie, fonction du paramètre est donc :

$$\{ x \mid \mu \}$$

décrite par :

$$\{ x = 1 \mid \mu \} = \mu$$

et

$$\{ x = 0 \mid \mu \} = 1 - \mu$$

c'est-à-dire aussi que $\{ x \mid \mu \}$ est une loi de Bernoulli, de paramètre μ.

2° La loi de probabilité du paramètre μ, $\{ \mu \}$, est donnée par la figure 1.

3° La loi des intervalles inter-changements, $\{ \rho \}$ se voit sur la figure 2.

Mémoire

On observe la sortie (pile ou face) sur les dix premiers coups. Cette « histoire », ou mémoire, X_{10} pourra avoir la forme indiquée par le tableau de la figure 3. Elle constitue pour nous, dans l'attente du onzième coup la seule information. Bien sûr, nous connaissons aussi les lois et leurs imbrications, le fonctionnement du jeu. Mais nous n'avons pas accès aux divers tirages, nous ne savons pas, par exemple, quelle est la pièce dont le croupier va se servir pour ce onzième coup, ni si l'on est à la fin ou au début d'une partie, etc.

Nous noterons ainsi notre problème sous la forme :

$$\{ x(11) \mid X_{10} \}?$$

II. LA METHODE ITERATIVE

Revenons au cas général. Nous voulons attribuer une distribution de probabilité à $x(n)$ étant donné l'histoire X_{n-1} des observations aux temps $1, 2, \ldots n-1$. Pour cela, nous cherchons une méthode itérative permettant de ne pas refaire au temps n les calculs qui auront pu être déjà effectués lors des itérations précédentes. Écrivant ainsi $\{ x(n) \mid X_{n-1} \}$ sous la forme :

$$\{ x(n) \mid X_{n-1} \} = \{ x(n), X_{n-1} \} / \{ X_{n-1} \} \qquad (1)$$

on supposera que le dénominateur, $\{ X_{n-1} \}$, probabilité de la trajectoire particulière X_{n-1}, a déjà été calculé à l'itération précédente. Précisons tout de suite que cette hypothèse de travail est fondée, car la première itération, qui servira d'initiation à notre méthode, consiste seulement à calculer $\{ x(1) \}$, ce qui est possible en développant suivant les valeurs possibles du paramètre :

$$\{ x(1) \} = \mathop{S}_{m \in M} \{ x(1) \mid \mu(1) = m \} \{ \mu = m \}$$

où M est le domaine des valeurs du paramètre μ.

On développe, de la même façon, $\{ x(n), X_{n-1} \}$ suivant les valeurs du paramètre au temps n :

$$\{ x(n), X_{n-1} \} = \mathop{S}_{m \in M} \{ x(n), X_{n-1} \mid \mu(n) = m \} \{ \mu = m \} \qquad (2)$$

On sépare ensuite $x(n)$ de X_{n-1} par :

$$\begin{aligned} \{ x(n), X_{n-1} \mid \mu(n) = m \} &= \{ x(n) \mid X_{n-1}, \mu(n) = m \} \{ X_{n-1} \mid \mu(n) = m \} \\ &= \{ x(n) \mid \mu(n) = m \} \{ X_{n-1} \mid \mu(n) = m \} \qquad (3) \end{aligned}$$

puisque le tirage des $x(n)$, connaissant la valeur m du paramètre $\mu(n)$ à l'instant n, est parfaitement défini.

Dans cette formule (3), le premier terme est connu; nous développons le second suivant la date $C_p(n)$ du dernier changement de paramètre :

$$\begin{aligned} \{ X_{n-1} \mid \mu(n) = m \} = \sum_{k=1}^{n} & \{ C_p(n) = k \mid \mu(n) = m \} \\ & \{ X_{n-1} \mid C_p(n) = k, \mu(n) = m \} \quad (4) \end{aligned}$$

mais la valeur du paramètre au temps n, résultant d'un tirage [effectué au temps $C_p(n)$] de la loi $\{ \mu \}$, est indépendante de la date de ce tirage; on a donc :

$$\{ C_p(n) = k \mid \mu(n) = m \} = \{ C_p(n) = k \} \tag{5}$$

qui se calcule (voir p. 102) à partir de la loi $\{ \rho \}$ des intervalles inter-changements. Pour $\{ X_{n-1} \mid C_p(n) = k,\ \mu(n) = m \}$ nous distinguerons le cas d'un changement survenant à l'instant $n (k = n)$, des autres cas $(k < n)$:

— Si $C_p(n) = n$: on a

$$\{ X_{n-1} \mid C_p(n) = n,\ \mu(n) = m \} = \{ X_{n-1} \mid C_p(n) = n \} \tag{6}$$

Car la valeur du paramètre qui doit être choisie dans le futur, et qui résultera d'un nouveau tirage de la loi $\{ \mu \}$, ne peut affecter en rien la probabilité qu'on attribue à X_{n-1}. Par contre, le fait de devoir changer de paramètre en n limite les tirages possibles de la loi $\{ \rho \}$ dans la période $(1, n - 1)$ concernée, et affecte donc indirectement X_{n-1}. Pour expliciter cela, développons le second membre de (6) suivant les valeurs possibles de $C_p(n - 1)$:

$$\{ X_{n-1} \mid C_p(n) = n \} = \sum_{j=1}^{n-1} \{ C_p(n-1) = j \mid C_p(n) = n \} \cdot \{ X_{n-1} \mid C_p(n-1) = j,\ C_p(n) = n \} \tag{7}$$

et $$\{ X_{n-1} \mid C_p(n-1) = j,\ C_p(n) = n \} = \{ X_{n-1} \mid C_p(n-1) = j \} \tag{8}$$

Car dans X_{n-1}, quand on sait que l'on change pour la dernière fois en j, ni les sorties antérieures à j, effectuées par le seul fait de changer en j, ni les suivantes, dont le tirage ne dépendra que de la valeur du paramètre tiré en j, ne sont affectées par le fait de changer en n.

— Si $C_p(n) < n$: on a, pour $\{ X_{n-1} \mid C_p(n) = k,\ \mu(n) = m \}$ l'écriture équivalente :

$$\{ X_{n-1} \mid C_p(n-1) = k,\ \mu(k) = m \} \tag{9}$$

En reportant les résultats (3) à (9) dans l'équation (2), on obtient :

$$\{ x(n), X_{n-1} \} = \mathop{\mathrm{S}}_{m \in M} \{ \mu = m \} \{ x(n) \mid \mu(n) = m \} \Bigg[\sum_{k=1}^{n-1} \{ C_p(n) = k \} \cdot \{ X_{n-1} \mid C_p(n-1) = k,\ \mu(k) = m \} + \{ C_p(n) = n \} \sum_{j=1}^{n-1} \{ C_p(n-1) = j \mid C_p(n) = n \} \cdot \{ X_{n-1} \mid C_p(n-1) = j \} \Bigg]$$

Ce qui, amélioré quelque peu, s'écrit sous les trois formes opératoires :

a) en remarquant que

$$\{ C_p(n) = n \} \{ C_p(n-1) = j \mid C_p(n) = n \} = \{ C_p(n-1) = j, C_p(n) = n \} :$$

$$\begin{aligned} \{ x(n), X_{n-1} \} = {} & \{ x(n) \} \sum_{j=1}^{n-1} \{ C_p(n-1) = j, C_p(n) = n \} \\ & \cdot \{ X_{n-1} \mid C_p(n-1) = j \} \\ & + \mathop{\mathsf{S}}_{m \in M} \{ \mu = m \} \{ x(n) \mid \mu(n) = m \} \sum_{k=1}^{n-1} \{ C_p(n) = k \} \\ & \cdot \{ X_{n-1} \mid C_p(n-1) = k, \mu(k) = m \} \end{aligned} \qquad (10\,a)$$

où

$$\{ x(n) \} = \mathop{\mathsf{S}}_{m \in M} \{ \mu = m \} \{ x(n) \mid \mu(n) = m \}$$

b) en intervertissant les signes S et $\sum$ et en regroupant :

$$\begin{aligned} \{ x(n), X_{n-1} \} = {} & \sum_{k=1}^{n-1} (\{ C_p(n) = k \} \mathop{\mathsf{S}}_{m \in M} \{ \mu = m \} \{ x(n) \mid \mu(n) = m \} \\ & \cdot \{ X_{n-1} \mid C_p(n-1) = k, \mu(k) = m \} \\ & + \{ C_p(n-1) = k, C_p(n) = n \} \{ x(n) \} \{ X_{n-1} \mid C_p(n-1) = k \} \end{aligned} \qquad (10\,b)$$

c) en développant $\{ X_{n-1} \mid C_p(n-1) = k \}$ suivant les valeurs du paramètre $\mu(k)$ et en regroupant :

$$\begin{aligned} \{ x(n), X_{n-1} \} = {} & \mathop{\mathsf{S}}_{m \in M} \{ \mu = m \} \sum_{k=1}^{n-1} \{ X_{n-1} \mid C_p(n-1) = k, \mu(k) = m \} \\ & \cdot (\{ C_p(n-1) = k, C_p(n) = n \} \{ x(n) \} \\ & + \{ C_p(n) = k \} \{ x(n) \mid \mu(n) = m \}) \end{aligned} \qquad (10\,c)$$

où l'on pourra intervertir l'ordre des signes S et $\sum$.

La méthode consiste donc à calculer, à la $n^{\text{ième}}$ itération, $\{ x(n) X_{n-1} \}$ à l'aide de l'une des équations (10) et à déduire $\{ x(n) \mid X_{n-1} \}$ à l'aide de l'équation (1).

Pour l'itération suivante, on aura besoin de conserver en mémoire les quantités suivantes, $x(n)$ étant sortie :

$$\{ X_n \mid C_p(n) = k, \mu(k) = m \} = \{ x(n) \mid \mu(n) = m \}$$

$$x \{ X_{n-1} \mid C_p(n-1) = k, \mu(k) = m \} \qquad \text{si} \qquad k < n$$

$$\{ X_n \mid C_p(n) = n, \mu(n) = m \} = \{ x(n) \mid \mu(n) = m \}$$

$$X \sum_{k=1}^{n-1} \{ C_p(n-1) = k \mid C_p(n) = n \} \{ X_{n-1} \mid C_p(n-1) = k \}$$

$$\{ X_n \mid C_p(n) = k \} = \underset{m \in M}{\mathsf{S}} \{ \mu = m \} \{ X_n \mid C_p(n) = k, \mu(k) = m \}$$

et

$$\{ X_n \} = \{ x(n), X_{n-1} \}.$$

— La 1^re^ itération est immédiate :

$$\{ x(1) \} = \underset{m \in M}{\mathsf{S}} \{ \mu = m \} \{ x(1) \mid \mu(1) = m \}$$

$x(1)$ étant sortie, on conserve les quantités :

$$\{ X_1 \} = \{ x(1) \}$$

$$\{ X_1 \mid C_p(1) = 1, \mu(1) = m \} = \{ x(1) \mid \mu(1) = m \}$$

et

$$\{ X_1 \mid C_p(1) = 1 \} = \{ x(1) \}$$

— A la 2^e^ itération, on calcule $\{ x(2), X_1 \}$ à l'aide d'une équation (10) et $\{ x(2) \mid X_1 \}$ par (1). Lorsque $x(2)$ est sorti, on range les quantités :

$$\{ X_2 \} = \{ x(2), X_1 \}$$

$$\{ X_2 \mid C_p(2) = 1, \mu(1) = m \} = \{ x(2) \mid \mu(2) = m \}$$

$$x \{ X_1 \mid C_p(1) = 1, \mu(1) = m \}$$

$$\{ X_2 \mid C_p(2) = 2, \mu(2) = m \} = \{ x(2) \mid \mu(2) = m \}$$

$$x \{ C_p(1) = 1 \mid C_p(2) = 2 \} \{ X_1 \mid C_p(1) = 1 \}$$

etc.

Il nous reste à montrer comment on calcule les probabilités des dates de derniers changements.

— Pour $i = 1, 2, ..., n-2$, $\{ C_p(n) = i \}$ représente la probabilité qu'un changement de paramètre ait lieu en i et qu'il n'y en ait pas aux points $i + 1$, $i + 2, ..., n$. Pour que cette dernière condition soit réalisée, il faut et il suffit que l'on tire, de la loi $\{ \rho \}$, une valeur qui vérifie :

$$\rho > n - i$$

Le fait qu'un changement ait lieu en i, lui, dépend des précédents tirages de la loi $\{ \rho \}$. On a :

$$\{ C_p(n) = i \} = \{ \text{changement en } i \} \{ \rho > n - i \}$$

Connaissant la loi $\{ \rho \}$, on sait calculer $\{ \rho > n - i \}$.

— Pour $i = n$, on a simplement :

$$\{ C_p(n) = n \} = \{ \text{changement en } n \}$$

On est ainsi ramené au calcul des $\{ \text{changement en } j \}$.

Pour cela, on remarque que pour $j > 1$ (le cas $j = 1$ est évident), il y a forcément un changement qui a précédé celui de l'instant j, puisque l'instant initial 1 est considéré comme un point de changement. On a donc :

$$\{ \text{changement en } j \} = \sum_{i=1}^{j-1} \{ \text{changement en } i \} \{ \rho = j - i \}$$

ce qui permet, par récurrence, de calculer tous les $\{ \text{changements en } j \}$.

Appliquée à notre petit exemple, la méthode nous a fourni les résultats suivants :

SANS AUCUNE INFORMATION, LA PROB D'AVOIR UN 1 EST .4259

UN 0 VENANT DE SORTIR, LA PROB D'AVOIR UN 1 AU PROCHAIN COUP EST .3953
UN 0 VENANT DE SORTIR, LA PROB D'AVOIR UN 1 AU PROCHAIN COUP EST .3936
UN 1 VENANT DE SORTIR, LA PROB D'AVOIR UN 1 AU PROCHAIN COUP EST .4367
UN 0 VENANT DE SORTIR, LA PROB D'AVOIR UN 1 AU PROCHAIN COUP EST .4082
UN 0 VENANT DE SORTIR, LA PROB D'AVOIR UN 1 AU PROCHAIN COUP EST .3970
UN 1 VENANT DE SORTIR, LA PROB D'AVOIR UN 1 AU PROCHAIN COUP EST .4402
UN 1 VENANT DE SORTIR, LA PROB D'AVOIR UN 1 AU PROCHAIN COUP EST .4719
UN 1 VENANT DE SORTIR, LA PROB D'AVOIR UN 1 AU PROCHAIN COUP EST .4872
UN 1 VENANT DE SORTIR, LA PROB D'AVOIR UN 1 AU PROCHAIN COUP EST .4813
UN 0 VENANT DE SORTIR, LA PROB D'AVOIR UN 1 AU PROCHAIN COUP EST .4095

On peut trouver étrange que l'arrivée d'un 1 (pile) puisse faire décroître, comme c'est le cas au dixième coup, la probabilité d'avoir un 1 au coup suivant : là réside au contraire tout l'intérêt du modèle, capable de simuler pareils comportements. Exploitant des statistiques sur la vente d'un produit de consommation, ce modèle sera ainsi capable de tenir compte de « l'effet d'ennui », de lassitude que ressent le consommateur moyen à toujours acheter la même marque, ou à satisfaire trop souvent le même désir, à s'user sur la même satisfaction.

Cela tient, bien sûr, à la nature des lois choisies, et, pour un triplet différent de celui-ci, nous n'observerions pas forcément le même phénomène. L'absence de cette lassitude existe dans d'autres domaines de la consommation : il est certain qu'on ne se lasse point de manger, par exemple.

Une autre étrangeté est celle des « anomalies » mises en évidence par R. A. Howard dans le cas d'un triplet géométrique — bêta — Bernoulli. Le cas est justement un de ceux où l'arrivée d'un 1 ne peut faire que croître la probabilité d'avoir un 1 au coup suivant. Néanmoins, pour un γ (de loi géométrique $\{ \rho \}$) de .10, pour $r' = 1$ et $n' = 2$ (de la loi bêta $\{ \mu \}$), la probabilité de sortir un 1 après la séquence 0110000 est de 0.3023 quant à celle de sortir un 1 après, 1110000 est de 0.2898.

Cette « anomalie » trouve dans certains domaines une application pertinente : le consommateur qui avait choisi d'abord un autre produit (ou une autre marque) pour la satisfaction qu'il pensait en tirer, qui en a pris deux fois et puis s'est tourné à nouveau vers un autre produit, a montré qu'il était plus indécis que celui qui a attendu trois coups (les trois premiers) pour y renoncer, lui aussi. « Plus indécis » est une interprétation possible; on peut penser aussi qu'une perturbation est intervenue sur le marché, au temps $n = 4$, perturbation qui a forcé le jugement du 2e consommateur, du type fidèle, plus que celui du 1er, de type moins fidèle.

III. TRONQUAGE DES DONNEES. PHENOMENES D'OUBLI TRANSFORMATION DE LA NATURE DES INFORMATIONS

Bien que la méthode que nous venons d'exposer réduise autant qu'il est possible la longueur des calculs, la prévision de $x(n)$ connaissant

$$X_{n-1} = x(1, n-1)$$

peut devenir une tâche harassante si n est très grand. On pensera donc tout naturellement à se débarrasser des calculs inutiles, dont l'impact sur le résultat final est négligeable, à ignorer les données dont l'inférence au temps n est très réduite.

1er problème

On nous donne brutalement $n-1$ observations successives de $x(i)$, $i \in [1, n-1]$. On en supprime les $K-1$ premières. On cherche à déterminer le nombre K « en moyenne », c'est-à-dire dépendant seulement de la longueur $n-1$ des données et d'une limite η de l'erreur admise pour la prévision de la sortie au temps n.

Pour un $x(n)$ donné, on fait l'erreur

$$\begin{aligned}\Delta[x(1, n)] &= |\{ x(n) \mid x(1, n-1) \} - \{ x(n) \mid x(K, n-1)| \\ &= |\{ C_p(n) < K \mid x(1, n-1) \} \{ x(n) \mid x(1, n-1), C_p(n) < K \} \\ &\quad + \{ C_p(n) \geqslant K \mid x(1, n-1) \} \{ x(n) \mid x(1, n-1), C_p(n) \geqslant K \} \\ &\qquad - \{ x(n) \mid x(K, n-1) \}|\end{aligned}$$

mais $\{ x(n) \mid x(K, n-1) \} = \{ x(n) \mid x(K, n-1), C_p(n) \geqslant K \}$

et $\{ x(n) \mid x(1, n-1), C_p(n) \geqslant K \} = \{ x(n) \mid x(K, n-1), C_p(n) \geqslant K \}$

donc

$$\begin{aligned}\Delta[x(1, n)] &= |\{ C_p(n) < K \mid x(1, n-1) \} \{ x(n) \mid x(1, n-1), C_p(n) < K \} \\ &\quad - (1 - \{ C_p(n) \geqslant K \mid x(1, n-1), \varepsilon) \{ x(n) \mid x(K, n-1), C_p(n) \geqslant K \}|\end{aligned}$$

ou

$$\begin{aligned}\Delta[x(1, n)] &= \{ C_p(n) < K \mid x(1, n-1) \} \mid \{ x(n) \mid x(1, n-1), C_p(n) < K \} \\ &\qquad - \{ x(n) \mid x(K, n-1) C_p(n) \geqslant K \}|\end{aligned}$$

l'erreur commise « en moyenne » (au sens défini plus haut) est :

$$E = \underset{x(1,n)}{\mathsf{S}} \Delta[x(1,n)] \{ x(1, n-1) \}$$

ou
$$E = \underset{x(1,n-1)}{\mathsf{S}} \{ x(1, n-1) \} \{ C_p(n) < K \mid x(1, n-1) \}$$

$$\cdot \underset{x(n)}{\mathsf{S}} \mid \{ x(n) \mid x(1, n-1), C_p(n) < K \} - \{ x(n) \mid x(K, n-1), C_p(n) \geqslant K \} \mid$$

mais

$$\underset{x(n)}{\mathsf{S}} \mid \{ x(n) \mid x(1, n-1), C_p(n) < K \} - \{ x(n) \mid x(K, n-1), C_p(n) \geqslant K \} \mid$$

$$\leqslant \underset{x(n)}{\mathsf{S}} \{ x(n) \mid x(1, n-1), C_p(n) < K \} + \underset{x(n)}{\mathsf{S}} \{ x(n) \mid x(K, n-1), C_p(n) \geqslant K \}$$

$$\leqslant 2$$

et comme

$$\underset{x(1,n-1)}{\mathsf{S}} \{ C_p(n) < K \mid x(1, n-1) \} \{ x(1, n-1) \} = \{ C_p(n) < K \}$$

on a finalement

$$\boxed{E \leqslant 2 \{ C_p(n) < K \}}$$

Si donc la loi sur les intervalles de changements de paramètre $\{ \rho \}$ est telle que :

$$\exists\, K \in [1, n-1] : \{ C_p(n) < K \} < \eta/2 \qquad \text{et} \qquad \{ C_p(n) \geqslant K \geqslant \eta/2$$

On décidera de supprimer les $K - 1$ premières observations

$$x(i), i \in [1, K-1],$$

en étant conscient de faire sur la prévision de $x(n)$ une erreur « en moyenne » de :

$$\boxed{E \leqslant \eta}$$

On dispose donc d'un moyen de tronquer les données inutiles. Inversement, le modèle sera capable de simuler des phénomènes d'oubli progressif : si un événement $x(1)$ s'éloigne dans le temps, il arrivera un moment K à partir duquel cet événement « disparaîtra » à η près; si η est la mesure de la précision souhaitée, on pourra considérer que l'événement $x(1)$ a disparu à l'instant K de la « mémoire » du phénomène.

2^e^ problème

On sait que l'on va traiter le problème sur un grand nombre d'intervalle de temps, soit I.

On veut déterminer le temps $N(N < I)$ à partir duquel on pourra omettre de faire les calculs se rapportant à la période précédant N, et ne « charrier » ainsi que N « blocs de calculs », au lieu de $N + i(i \in [1, I - N])$; bien entendu, on voudra que l'erreur produite par cette approximation soit inférieure à un η donné.

On « effacera » donc des mémoires tous les $\{ X_n \mid C_p(n) = l \}$ et les

$$\{ X_n \mid C_p(n) = l, \mu(n) = m \}$$

pour $l \leqslant n - N - 1$, au fur et à mesure de leur caractère d'inutilité. C'est-à-dire que la $n^{\text{ième}}$ itération verra disparaître les quantités

$$\{ X_{n-1} \mid C_p(n) = n - N - 1 \}$$

calculées et utilisées pourtant encore à l'itération précédente.

A la $n^{\text{ième}}$ itération, on fait l'erreur :

$$\Delta[X_n] = \mathop{\mathsf{S}}_{l=1}^{n-N-1} (\{ C_p(n) = l \} \{ X_n \mid C_p(n) = l \}$$
$$+ \{ C_p(n-1) = l, C_p(n) = n \} \{ x(n) \} \{ X_{n-1} \mid C_p(n-1) = l \})$$

Nous chercherons, là encore, une règle de tronquage « en moyenne », et donc, à majorer la quantité :

$$\Delta_n = \mathop{\mathsf{S}}_{X_n} \Delta[X_n]$$
$$= \mathop{\mathsf{S}}_{l=1}^{n-N-1} \{ C_p(n) = l \} \mathop{\mathsf{S}}_{X_n} \{ X_n \mid C_p(n) = l \}$$
$$+ \sum_{l=1}^{n-N-1} \{ C_p(n-1) = l, C_p(n) = n \} \mathop{\mathsf{S}}_{X_{n-1}} \{ X_{n-1} \mid C_p(n-1) = l \} \mathop{\mathsf{S}}_{x(n)} \{ x(n) \}$$

$$\Delta_n = \sum_{l=1}^{n-N-1} (\{ C_p(n) = l \} + \{ C_p(n-1) = l, C_p(n) = n \})$$

$$\Delta_n = \sum_{l=1}^{n-N-1} \{ \text{changement en } l \} (\{ \rho > n - l \} + \{ \rho = n - l \})$$

mais $$\{ \rho > n - l \} + \{ \rho = n - l \} = \{ \rho > (n-1) - l \}$$

et $$\{ \text{changement en } l \} \{ \rho > (n-1) - l \} = \{ C_p(n-1) = l \}$$

donc $$\Delta_n \leqslant \sum_{l=1}^{n-N-1} \{ C_p(n-1) = l \}$$

ou

$$\boxed{\Delta_n \leqslant \{ C_p(n-1) < n - N \}}$$

si donc la loi sur les intervalles de changements $\{ \rho \}$ est telle que :

$$\forall \eta > 0, \exists N : \forall n > N \quad , \quad \{ C_p(n-1) < n - N \} < \eta$$