INSTITUT HENRI POINCARE

SEMINAIRE DE CALCUL DES PROBABILITES

THEORIE DU CODAGE
ET DES EVENEMENTS RECURRENTS

Exposé de M.P. SCHUTZENBERGER

du 16 mars 1956

I. Rappel de la théorie de Feller :[1]

Soit $\xi(t)$ un processus stochastique. Un évènement $\mathcal{E}$ défini sur $\xi(t)$ est un "évènement récurrent" si et seulement si :

1) Il existe une règle permettant de décider si $\mathcal{E}$ s'est ou non produit au temps t_1 en connaissant seulement les valeurs de $\xi(t)$ pour $0 < t \leqslant t_1$.

2) Si $\mathcal{E}$ s'est produit en t_1 sur la suite de valeurs $\xi_1(t)$ $(0 < t \leqslant t_1)$ et si $\xi_2(t)$ est une autre suite $(0 < t \leqslant t_2)$, soit :

$$\xi_3(t) = \xi_1(t) \quad \text{pour} \quad 0 < t \leqslant t_1$$

$$= \xi_2(t - t_1) \quad \text{pour} \quad t_1 < t \leqslant t_2$$

$\mathcal{E}$ se produit au temps $t_1 + t_2$ pour ξ_3 si et seulement s'il se produit au temps t_2 pour ξ_2.

3) Dans ces conditions :

$$\Pr(\xi_3) = \Pr(\xi_1) \quad \Pr(\xi_2)$$

On dit aussi que $\xi(t)$ est un processus "régénératif"[2]

Exemples : 1) $\xi(t)$ est un processus continu à accroissements indépendants ξ défini par :

$$\int_0^t \xi(t)\, dt = 0.$$

---

(1) W. Feller. An Introduction to Probability Theory. Wiley N.Y. 1950), chap. 12.

(2) M.S. Bartlett. Stochastic Processes. Cambridge U. Press, 1955, chap. 3.

2) Le temps est discret. $\xi(t)$ est une suite de variables binomiales indépendantes ξ étant l'apparition de "0", suivi de ℓ "1", suivi d'un "0" ("run" de longueur ℓ). Nous supposerons toujours que le temps et l'ensemble des valeurs possibles de ξ sont discrets.

II. Rappel de la notion de code : (3) (4)

Un demi-groupe libre A engendré par les lettres $\{a_1 \ a_2 \ \dots \ a_n\}$ formant l'alphabet A_0

Une partie P_0 de A : le dictionnaire constitué par les "mots". P le sous demi-groupe de A engendré par P_0. Les séquences appartenant à ℓ étant appelées messages.

P_0 est un code si et seulement si toute suite de lettres $s \in P$ est d'une façon unique un produit de mots. Le code est unitaire (5) (sous-entendu "à gauche") si et seulement si $s\,s' \in P$ et $s \in P$ entrainent $s' \in P$. Le code est net (5) (sous-entendu: "à droite") si pour tout $s \in A$, il existe $s' \in A$ tel que $s\ s' \in P$.

Il est bien connu :

Une condition nécessaire et suffisante pour qu'un code soit unitaire est qu'aucun de ses mots ne soit le début d'un autre mot (en langage algébrique : ne soit diviseur à gauche d'un autre mot)..

Réciproquement, s'il en est bien ainsi : pour toutes les séquences de P_0, P_0 est un code.

Exemple :

$A_0 = \{a, b\}$ (alphabet "binaire")

$P_0 = \{a\ a;\ ab\ ;\ baa\ ;\ bab\ ;\ bb\}$ est l'ensemble des mots d'un code unitaire.

---

(3) Cf. l'exposé fait le 28/11 au Séminaire d'algèbre de M. P. Dubreil (on le référera par l'abréviation "M.P.S.")

(4) Ceci est un cas particulier d'une définition plus générale de J. Riguet.

(5) C'est la définition de P. Dubreil d'un sous groupe unitaire ou net. Mem. Acad. Sc. (1941) pp 1-52

Le dictionnaire opposé ({aa, ba, aab ; bab, bb} correspond bien à un code, mais celui-ci n'est pas unitaire.

Etant donnésun code et une suite "s" de lettres qui est un message, on appellera "décodage" l'opération qui consiste à décomposer s en un produit de p∈ P_o.

Exemple :

m = a a a b b a a se décode : /a a/a b/b a a

On remarque que si le code n'est pas unitaire, l'opération peut nécessiter la connaissance de tout le message avant que soient levées les ambiguités :

Dans le code opposé au précédent :

a a b b a a a =

soit/a a/ b b / a a / a ;

soit/a a b / b a / a a /

La deuxième possibilité est seule à retenir, puisque /a n'est pas un mot.

Si l'on attribue des probabilités fixes aux mots d'un code et que l'on effectue des tirages indépendants, on obtient un processus stochastique sur les suites de lettres :

A tout code unitaire peut être associé un évènement récurrent (l'évènement "fin de mot") et réciproquement (6). En effet, si un processus régénératif est donné, il suffit de considérer comme mots les suites $\xi_i(t)$ telles que ξ se produise en t et ne se soit produit pour aucun $t' < t$.(2) implique que le sous demi groupe est unitaire, donc correspond à un code ; (3) spécifie que les mots sont fournis par un processus indépendant à probabilités fixes.

III. Notion de Préfixe (7, cf. p. 4)

Soit $\approx$ la relation d'équivalence entre séquences de Λ définie par : $s_1 \approx s_2$ si et seulement si quelle que soit la séquence x :

$$Pr(s_1x/s_1) = Pr(s_2x/s_2)$$

---

(6) Ce résultat a été utilisé sous une forme un peu différente dès 1952 par B. Mandelbrot qui a le premier souligné les rapports entre codage et évènements récurrents.

Manifestement quel que soit $s' \in A$:

$$s_1 \approx s_2 \quad \text{entraîne} \quad s_1 s' \approx s_2 s'$$

car $\quad Pr(s_1 s' x / s_1 s') = Pr(s_1 s' x / s_1) : Pr(s_1 s'/s_1)$

$$= Pr(s_2 s' x / s_2) : Pr(s_2 s'/ s_2) = Pr(s_2 s' x/ s_2 s')$$

On appellera "préfixe" $\tilde{s}$ l'ensemble des séquences équivalentes à s selon $\approx$.

La notion de préfixe est une sorte de généralisation de celle de résumé exhaustif de G. Darmois : Si un message commence par s_1, la distribution de son "futur" à partir de sa dernière lettre ne dépend que de $\tilde{s}_1$.

Proposition III.1

Dans un code UN (unitaire à gauche et net à droite) l'ensemble des messages constitue un préfixe unique que l'on désignera toujours par α_1.

(immédiat d'après (I.(3)). La proposition admet une réciproque, mais nous n'aurons pas besoin de celle-ci). Dans les applications les plus fréquentes l'ensemble $\mathcal{A}$ des préfixes est fini - ou a une structure topologique très simple (Cf. M.P.S.).

Si $\xi(t)$ $(0 \leqslant t \leqslant t_1)$ représente les lettres successives d'une séquence aléatoire, on peut lui associer $\tilde{\xi}(t)$ qui est une variable aléatoire prenant ses valeurs dans $\mathcal{A}$: par construction, la suite des $\tilde{\xi}(t)$ est une chaîne de Markoff d'ordre 1 caractérisée par les probabilités de transition $p_{ij}(k)$ (que $\tilde{\xi}(t)$ passe du préfixe α_i au préfixe α_j par adjonction de la lettre $a_k \in A_0$).

Exemple :

(Code donné plus haut en exemple), les préfixes sont :

$$\alpha_1 \; ; \alpha_2 = \tilde{a} \; ; \alpha_3 = \tilde{b} \; ; \alpha_4 = \widetilde{ba} :$$

Les probabilités des mots étant :

---

(7) Les préfixes utilisés ici sont plus généraux que ceux introduits dans M.P.S. : en toute rigueur, on devrait les appeler préfixes stochastiques. On a d'ailleurs " $\approx$ plus fine que $\backsim$ " pour équivalence purement algébrique $\backsim$ définie plus bas (IV).

$Pr(a\ a) = p_1$; $Pr(a\ b)$ pr ; $Pr(b\ a\ a) = p_3$; $Pr(b\ a\ b) = p_4$

$Pr(b\ b) = p_5$ $(p_1 + p_2 + p_3 + p_4 + p_5 = 1)$, on botient la matrice :

	α_1	α_2	α_3	α_4
α_1	0	p_1+p_2	$p_3+p_4+p_5$	0
α_2	1	0	0	0
α_3	0	0	0	p_3+p_4 / $p_3+p_4+p_5$
α_4	1	0	0	0

On appellera Π cette matrice dans le cas général et on désignera par $h(\lambda)$ le déterminant $|\lambda\Pi - I|$.

Supposons en particulier que le code soit unitaire : s'il n'était pas net, il existerait un préfixe α_0 contenant toutes les séquences telles que $Pr(s\ s' \in P) = 0$ quel que soit s'. En outre deux cas peuvent se présenter :

<u>ou bien</u> il existe un préfixe au moins $\alpha_i \neq \alpha_0$ tel que l'on puisse trouver s et s' avec $\tilde{s} = \tilde{s}\tilde{s'} = \alpha_i$ sans que $\tilde{s}\ \tilde{s''} = \alpha_1$ pour aucun s'' diviseur à gauche de s' : dans ce cas il existe des mots de longueurs non bornées (toutes les séquences de la forme $s(s')^n\ s''$ où s'' est tel que $s\ s'' \in P$) même si α est fini ;

<u>ou bien</u> il n'en existe pas et alors l'ensemble α est fini si tous les mots ont une longueur bornée, car il existe en "ordre local" sur α (la matrice Π privée des lignes et des colonnes α_0 et α_1 est nulle en dessous de la diagonale principale).

Dans ce dernier cas qui est le plus important dans la Théorie des communications, on a :

<u>Proposition III.2</u>

$$h(\lambda) = 1 - \sum \Pi_i \lambda^2$$

où Π_i désigne la somme des probabilités des mots de longueur i.

et $H(\lambda) = 1 + \sum_{i=1}^{\infty} \Pi_i \lambda^i = 1/h(\lambda)$

où $\overline{\Pi}_i$ est la probabilité qu'une séquence de longueur i soit un message. En outre une condition nécessaire et suffisante pour que le code soit non est que $h(1) = 0$.

Dans ces conditions $\overline{\Pi}_i = \beta_1 + \sum_j \beta_j \rho_j^{-i} \ldots$

où les ρ_j sont les autres racines de $h(\lambda) = 0$ (dont les modules sont > 1, sauf si la longueur des mots admet un P.G.C.D. $\neq 1$) et où les β_j sont des constantes avec :

$$\beta_1^{-1} = \left[\frac{\partial}{\partial \lambda} h(\lambda) \right]_{\lambda=1} = \text{longueur moyenne des mots.}$$

Tous ces résultats sont des interprétations immédiates de résultats bien connus pour les chaînes de Markoff et nous n'en entreprendrons pas la démonstration : - l'identité $H(\lambda) = (h(\lambda))^{-1}$ est précisément le théorème fondamental de Feller. La preuve de sa validité et son extension au cas où les mots n'ont pas une longueur bornée ne demande que quelques précautions supplémentaires.

IV. Le décodage des messages altérés par le bruit

Supposons un code UN dont la longueur des mots est bornée, mais non constante.

Puisque le décodage s'effectue de proche en proche, il semblerait que la plus légère altération dût détruire , sauf chance exceptionnelle, la totalité de la signification. Par exemple (nous supposerons toujours pour simplifier qu'il s'agit d'un code binaire et que c'est la première lettre seule qui est altérée), dans le code que nous avons déjà utilisé

a a a b b a a se décode aa/ ab / aa /, mais b a a b b a a se décode
b a a/b b/a a/

Observons cependant que s'il se trouvait (comme dans le cas présent) que le décodage du message et du message altéré conduisent à placer la "fin de mot" à la même n-ième lettre, le reste de la séquence serait correctement déchiffrable - en supposant que de nouvelles altérations ne se produisent pas.

Donc, si ce phénomène est assez fréquent et si les altérations sont rares, une large partie du message sera encore utilisable.

Considérons algébriquement le problème : si s est le début du message, s' le début du message altéré, nous , nous voulons que les lettres suivantes s'' soient telles que :

$$s\, s'' \in P \quad \text{et} \quad s'\, s'' \in P$$

En particulier si $s \in P$, il faut que $s'' \in P$ (puisque le code est unitaire) et le problème qui se pose est de savoir s'il existe un $s' \notin P$ tel que $s' s'' \in P$.

Ceci est indépendant - dans une certaine mesure - de toute question de probabilités et nous redéfinissons des "préfixes-algébriques" par : $s_1 \sim s_2$ si et seulement si $s_1 x \in P$ entraine $s_2 x \in P$ pour tout $x \in A$.

Dans les cas étudiés $s_1 \approx s_2$ entraine $s_1 \sim s_2$ et il n'y aura pas d'inconvénient à parler désormais du préfixe $\tilde{s}$ pour signifier le préfixe algébrique formé par la classe d'équivalence de s par $\sim$.

Nous avons encore une représentation matricielle : et cette fois chacune des lettres $a_1 \in A_0$ correspond à une application dans lui-même de $\mathcal{C}$ c'est à dire à une matrice dont chaque ligne contient Un seul élément (égal à 1) différent de zéro.

α_1 et α_0 ont la même signification que précédemment (α_0 est vide puisque le code est supposé net). Ceci fournit une représentation du demi-groupe A engendré par A_0 qui est finie dans le cas étudié et par conséquent isomorphe à une image homomophe $\bar{A}$ de A que l'on a appelée "le demi groupe fondamental du code" (M.P.S. Section 3).

Naturellement, le code opposé aurait un autre système de préfixes, ce qui donnerait une autre représentation (toujours isomorphe, mais en général non équivalente à la précédente, d'ailleurs).

Définition

Un code sera dit absorbant s'il existe au moins une suite q finie, de probabilité non nulle telle que quel que soit $s \in A$ sq appartienne à P. Pour en terminer avec les considérations intuitives, indiquons que si un code est absorbant, quelle soit l'aléération unique produite au début du message, le décodage de celle-ci sera presque certainement correct à partir d'une certaine longueur : en effet, d'une part la dernière lettre de sq est toujours une fin de mot quel que soit s et par conséquent les décodages de $s\,q\,s''$ et $s'\,q\,s''$ concordent au moins à partir de cette lettre, d'autre part q appartient lui-même à P, donc les messages de m mots ne le contenant pas, ont une probabilité inférieure à :

$$(1 - Pr(q))^m \quad \text{qui tend vers zéro avec } 1/m.$$

On a donc comme une sorte de propriété ergodique en ce sens que le décodage d'un message dans un code absorbant tend à être indépendant du décodage correct ou non de ses premiers mots.

Proposition IV.1

Une condition nécessaire et suffisante pour qu'un code unitaire,

net, borné (UNB) soit absorbant est que quel soit $s \in A$, il existe $s' \in P$ tel que $ss' \in P$.

Considérons la représentation de A comme demi-groupe d'application de $\mathcal{O}_6$ dans lui-même : la condition signifie que quel que soit $\tilde{s}$ il existe $s' \in P$ avec $ss' \in P$: les matrices correspondant aux mots transforment le préfixe α_1 dans lui-même par hypothèse.

Soit donc $s_1' \in P$ tel que $ss_1' \in P$ pour au moins un $s_1 \notin P$; $\alpha s_1'$ est strictement plus petit que α, soit de nouveau $s_2 \notin P$, $s_2 \notin P$, $s_2 s_1' \notin P$, il existe $s_2' \in P$ tel que $s_2 s_2' \in P$ et $\alpha s_1' s_2'$ est encore de puissance strictement inférieure $\alpha s_1'$ etc. On finit ainsi par construire au moins une séquence $s_1', s_2' \ldots s_m' \in P$ qui est précisément la séquence q cherchée.

Observons d'ailleurs que "q" a une représentation très remarquable : "q" s'il existe correspond à la matrice dont les éléments de la colonne α_1 (et ceux-là seulement) sont égaux à 1.

Comme pour tous les problèmes ergodiques des conditions d'indécomposabilité jouent un rôle fondamental.

Proposition IV.2

Une condition nécessaire et suffisante pour que le code soit absorbant est que quelles que soient les suites s et s', on puisse trouver une suite s'' telle que $ss'' \sim s's''$ (c'est-à-dire $\widetilde{ss''} = \widetilde{s's''}$).

La condition est nécessaire, car si q existe $\widetilde{sq} = \alpha_1$ quel que soit q.

Elle est suffisante, car elle implique pour tout $s \notin P$ qu'il existe s' tel que $ss'' \sim s's''$ pour $s' \in P$. On en déduit :

Proposition IV.3

Les codes suivants ne sont pas absorbants :

a) les codes unitaires à droite

b) les codes dont le P.G.C.D. des longueurs des mots est différente de 1.

c) les codes "composés uniformes" dont l'ensemble des "mots" P_0 est l'ensemble P_0^{ν} de tous les messages formés de ν mots d'un autre code unitaire P_0'.

a) est immédiat : unitaire à droite signifie que $ss' \in P$ et $s' \in P$ impliquent $s \in P$ (l'existence de tels codes n'est pas évidente. Une famille infinie d'exemples est donnée dans M.P.S.)

b) : immédiat aussi : si la longueur $|s|$ de s différe de la longueur s' de s' par une quantité non équivalente à zéro (module ℓ) il en est de même de sx et de $s'x$ quel que soit $x \in A$.

c) Considérons s' formé de m' mots successifs de P'_o. Quel que soit $s \in P$, la séquence $s's$ est décodable dans P' et ^oy est formée de $m' + k_\nu$ mots (de P') donc elle ne peut jamais appartenir à P.

Malgré de longs efforts, je n'ai réussi ni à trouver d'autres codes (UNB) qui ne soient pas absorbants, ni à prouver que ce sont là les seuls codes qui ne jouissent pas de cette propriété ergodique.

En dehors de résultats encore fragmentaires, on ne connaît qu'un seul énoncé assez général que nous allons établir.

Préalablement, posons pour simplifier $|s|$ = la longueur de s quel que soit $s \in A$ et considérons le cas où, k étant le nombre des lettres de A_o, la probabilité d'un mot s quelconque est égale à $K^{-|s|}$. On vérifie que dans ce cas, si le code est UNB, les lettres successives du processus de codage apparaissent avec les mêmes fréquences que si elles étaient tirées indépendamment et avec des probabilités égales.

On a alors :

$\overline{\pi}_i = k^{-i}\, \overline{n}_i$ (respectivement $\pi_i = k_i^{-l}\, n_i$) où $\overline{n}_i$ (resp. n_i) est le nombre de messages (resp. mots) de longueur i. Soit alors h. la longueur moyenne des mots.

D'autre part v étant une séquence fixe, considérons l'ensemble $V \subset P_o^m$ (pour m assez grand fixe aussi) des séquences formées de m mots et admettant v comme diviseur à droite. Posons :

$$\gamma_v = \sum_{w \in V} k^{-|w|+|v|}$$

Proposition IV.4.

Une condition nécessaire et suffisante pour qu'un code UNB ne rentrant dans aucune des classes décrites en IV.3 soit absorbant est qu'il existe une séquence v telle que :

$$2k^{|v|}\, \gamma_v > h$$

Démonstration.

Appelons $D_\ell(u,v)$ l'ensemble des séquences de longueur $\ell \geq |u| + |v|$ admettant u et v respectivement comme diviseurs à gauche et à droite, où u est une séquence fixe.

1°) Le nombre des séquences distinctes de D (u, v) est

$$k^{-(\ell - |u| - |v|)} = d_\ell$$

Soit $D'_\ell(u,v)$ le sous-ensemble de $D_\ell(u, v)$ formé par les séquences de longueur ℓ de la forme ux avec $x \in P$.

2°) Le nombre des séquences distinctes de D' (u,v) est de la forme :

$$d'_\ell = k^{-(\ell-|u|)} \eta_N h^{-1} + \varepsilon(\ell) \quad \text{où} \quad k^{-\ell}\, \varepsilon(\ell) \longrightarrow 0 \quad \text{avec} \quad \ell .$$

En effet d'_ℓ est égal au nombre $V_{\ell-|u|}$ des messages de longueur $\ell - |u|$ admettant v comme diviseur à droite : c'est-à-dire si ν_i est le nombre des w de longueur i dans V :

$$V_{\ell-|u|} \sum \bar{n}_{\ell-|u|-i}\, \nu_i \quad \text{Or (III.2)} \quad n_i = k^{-i} h^{-1} + \varepsilon_1(i)$$

(car $h(\lambda)$ n'a qu'une seule racine de module égal à 1 après l'exclusion du cas b) de IV.2).

Donc :

$$V_{\ell|u|} = k^{-\ell+|u|} \cdot h^{-1} \left\{ \sum_{w \in V} k^{-|w|+v} \right\} + \varepsilon_2(\ell)$$

3°) Le nombre d'' des messages appartenant à $D_\ell(u,v)$ est de la forme :

$$d''_\ell = k^{-\ell+|u|} \eta_v h^{-1} + \varepsilon_3(\ell)$$

Soient u' le diviseur à gauche maximum de u qui appartienne à P et u''_1, u''_2 ... $u''_i \in U$ les messages minimum qui admettent u comme diviseur à droite

$$d''_\ell = \sum_{u''_i \in U} v_{\ell - |u''_i|} = \sum_{u'' \in U} k^{-\ell + |u''|} \; \eta_v^{h^{-1}} + \varepsilon_s(\ell)$$

$$= k^{\ell + |u'|} \; \eta_v^{h^{-1}} \left\{ \sum_{u'' \in U} k^{-|u'| + |u''|} \right\} + \varepsilon_s(\ell)$$

Or l'expression dans l'accolade est égale à $k^{|u| - |u'|}$ puisque le code est net.

4°) $d''_\ell + d'_\ell > d_\ell$ pour $\ell \to \infty$ est équivalent à =

$$2 \eta_v k^v > h$$

(le calcul est immédiat).

Donc, si l'inégalité est vérifiée, quel que soit u fini, il existe, pour ℓ assez grand, au moins un élément de $D_\ell(u,v)$ qui est un message, c'est-à-dire qu'il existe un $s \in P$ avec $u\,s \in P$, ce qui établit le résultat d'après IV. 1, et IV. 2.

Exemple : Nous reprenons le code déjà traité en exemple :

$k = 2$; $h = 2\ 1/4 + 2\ 1/4 + 2\ 1/4 + 3\ 1/8 = 18/8$

Soit $v = a\ a$; η_v $(1/4 + 1/8 = 3/8$. $|v| = 2$.

On a : $2^{-3/8} \times 4 > 18/8$. Donc ce code est absorbant. ($q = a\ a$ précisément;).

-:-:-:-:-:-:-:-