

ANNALES SCIENTIFIQUES  
DE L'UNIVERSITÉ DE CLERMONT-FERRAND 2  
*Série Mathématiques*

GÉRARD LETAC

**Sur quelques aspects combinatoires du calcul des probabilités**

*Annales scientifiques de l'Université de Clermont-Ferrand 2*, tome 49, série *Mathématiques*, n° 8 (1972), exp. n° 1, p. 1-82

[http://www.numdam.org/item?id=ASCFM\\_1972\\_\\_49\\_8\\_A1\\_0](http://www.numdam.org/item?id=ASCFM_1972__49_8_A1_0)

© Université de Clermont-Ferrand 2, 1972, tous droits réservés.

L'accès aux archives de la revue « Annales scientifiques de l'Université de Clermont-Ferrand 2 » 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/>

Gérard LETAC

SUR QUELQUES ASPECTS COMBINATOIRES

DU CALCUL DES PROBABILITES



CHAPITRE I

---

"Reconstruction de probabilités à partir de  
processus simples"



## 1) Introduction

L'origine des questions traitées dans ce chapitre est une conversation que j'eus, il y a quelques années au Club du Département de Mathématiques Pures de l'Université de Montréal avec un étudiant en statistiques : Denis Labelle. Il s'intéressait à la meilleure manière de choisir entre plusieurs personnes, disons  $n$ , de façon équiprobable lorsqu'on dispose d'un dé, ou d'une pièce de monnaie, ou d'une manière générale d'un appareil aléatoire qui donne  $k$  résultats équiprobables. Il avait vu qu'il existait une stratégie pour que le nombre moyen de jets du dé "à  $k$  faces" soit  $\sum_{t=0}^{\infty} \frac{k^t \text{ modulo } n}{k^t}$ , et pensait, pour avoir essayé un certain nombre d'autres combinaisons, qu'il était impossible de faire mieux. Intrigué, je constatai quelques jours plus tard que tout reposait sur l'égalité :

$$E(T) = \sum_{t=0}^{\infty} P(T > t)$$

valable pour une variable aléatoire  $T$  à valeurs entières et positive, ici le nombre de jets, et que pour minimiser  $E(T)$ , il est suffisant de trouver un  $T$  qui rende  $P(T > t)$  minimum pour chaque  $t$ . En fait il y avait deux questions : trouver un algorithme, trouver le meilleur.

Diverses généralisations devenaient possibles :

- 1) Ne plus choisir les personnes de façon équiprobable mais avec des probabilités respectives  $p_1, p_2, \dots, p_k, \dots$  éventuellement en quantité dénombrable.

- 2) Changer d'appareil aléatoire à chaque jet en utilisant des dés avec des nombres de faces différents.
- 3) Prendre des appareils aléatoires qui ne donnent plus des résultats équiprobables, ni même discrétisés. Modifier de même les choix à faire.
- 4) Prendre des appareils aléatoires pour lesquels le résultat à l'instant  $t$  ne soit plus indépendant de ceux qui l'ont précédés.

Comment le problème se pose-t-il en termes mathématiques ? On dispose d'un espace de probabilité  $(\Omega, \mathcal{A}, P)$  -les résultats des appareils aléatoires- on cherche une sous-tribu  $\mathcal{B}$  de  $\mathcal{A}$  de sorte que  $(\Omega, \mathcal{B}, P)$  (en ce sens que les atomes  $y$  ont la même valeur) soit équivalent à un autre espace donné  $(E, \mathcal{C}, \pi)$  - par exemple les personnes parmi lesquelles il faut choisir. Pour fixer les idées, supposons que  $(\Omega, \mathcal{B}, P)$  soit l'espace associé à une suite infinie de pile ou face - équivalent au segment  $[0, 1]$  muni de la mesure de Lebesgue, et  $(E, \mathcal{C}, \pi)$  formé par trois points  $n^\circ 1, n^\circ 2, n^\circ 3$ , chargés chacun de la probabilité  $1/3$ . Bien sûr, on pourrait partager le segment  $[0, 1]$  en 3 ensembles de mesure  $1/3$ , mais si cette procédure correspondant à un nombre dénombrable de jets de pièces, elle est déraisonnable. Evidemment on procède ainsi : on jette la pièce deux fois et on fait le choix suivant :

pile pile  $\rightarrow n^\circ 1$

pile face  $\rightarrow n^\circ 2$

face pile  $\rightarrow n^\circ 3$

face face  $\rightarrow$  on recommence.

Ainsi notre espace de probabilité  $(\Omega, \mathcal{A}, P)$  est le produit d'espaces

$(\Omega_t, \mathcal{A}_t, P_t)_{t=1}^\infty$  et la tribu  $\mathcal{B}$  qu'on cherche à fabriquer doit être une sous-tribu de  $\mathcal{F}_T$  ; ici,  $T$  est un temps d'arrêt par rapport aux tribus  $(\mathcal{F}_t)_{t=1}^\infty$ ,  $\mathcal{F}_t$  étant la tribu "du passé à l'instant  $t$ ", produit de  $\mathcal{A}_1 \dots \mathcal{A}_t$  et des tribus triviales sur  $\Omega_t$ , si  $t' > t$ . Dans l'exemple ci-dessus de la pièce et des trois personnes,  $T$  est le nombre de jets conduisant au choix d'une personne.

Le programme est le suivant : nous donnons d'abord définitions et propriétés de l'outil essentiel que sera pour nous la signature d'un espace de probabilité (§ 2), en signalant chemin faisant quelques curiosités (§ 3). Nous établissons ensuite une inégalité très simple, dite "du déménageur" qui est en fait la réponse à la question que posait D. Labelle (§ 4). On passe ensuite au problème primitif plus délicat de la construction de la stratégie en déterminant, étant donnés  $(\Omega, \mathcal{A}, P)$  et une suite croissante  $(\mathcal{F}_t)_{t=1}^\infty$  de sous-tribus de  $\mathcal{A}$ , quels sont les espaces de probabilité équivalents à un  $(\Omega, \mathcal{B}, P)$ , où  $\mathcal{B}$  est sous-tribu de  $\mathcal{F}_T$  pour quelque temps d'arrêt  $T$ . Nous en déduisons au passage l'analogie pour les chaînes de Markov d'un célèbre théorème de Skohorod permettant de construire n'importe quelle marche aléatoire à partir du mouvement brownien (§ 5). Le meilleur temps d'arrêt pour notre problème n'est pas sans relation avec l'entropie de la signature (§ 6), ce qui nous amène à étudier la relation des questions traitées ici avec les problèmes de codage (§ 7).

Un résumé particulièrement significatif des résultats est fourni par l'ensemble des énoncés des théorèmes 1, 6, 7, 9, 11 et 15.

### Définition 1

Une signature est une application  $m$  du segment  $(0, 1]$  dans l'ensemble des entiers non négatifs telle que  $m(x)$  soit nulle sauf sur une partie dénombrable de  $(0, 1]$  et telle que

$$||m|| = \sum_x x m(x) \leq 1 .$$

Le nombre  $||m||$  est appelé masse de la signature et le nombre  $x(m) = \sup \{x ; m(x) \neq 0\}$  atome maximum de  $m$  si  $||m|| > 0$ .

Un atome  $A$  d'un espace de probabilité  $(\Omega, \mathcal{A}, P)$  est un élément de  $\mathcal{A}$  tel que  $P(A) > 0$  et tel que  $B \in \mathcal{A}$  et  $B \subset A$  entraîne  $P(B) = 0$  ou  $P(A)$ . Deux atomes sont dits équivalents si  $P(A_1 \setminus A_2) = P(A_2 \setminus A_1) = 0$ . L'ensemble des atomes équivalents à un même atome  $A$  est appelé classe d'atome et  $P(A)$  est appelé probabilité de la classe.

### Définition 2

La signature  $m$  de l'espace de probabilité  $(\Omega, \mathcal{A}, P)$  est définie par  $m(x) =$  nombre de classes d'atomes de probabilité  $x$ .  $P$  étant fixé et  $\mathcal{B}$  étant une sous-tribu de  $\mathcal{A}$ , la signature de  $\mathcal{B}$  est la signature de  $(\Omega, \mathcal{B}, P_1)$ , où  $P_1$  est la restriction de  $P$  à  $\mathcal{B}$ .

Il est clair qu'une signature d'espace de probabilité est bien une signature.  $||m||$  est la masse de la partie atomique,  $1 - ||m||$  celle de la partie continue,  $x(m)$  la masse de l'atome de plus forte

probabilité. Une signature quelconque est évidemment la signature d'un espace de probabilité. La définition 2 s'étend au cas où  $P$  est une mesure sur  $(\Omega, \mathcal{A})$  de masse  $< 1$ , la partie continue étant alors de masse  $P(\Omega) - ||m||$ .

La définition suivante de l'équivalence est celle de Röhrlin [ 11 ] qui parle d'espace "de même type".

### Définition 3

Deux espaces de probabilité  $(\Omega_1, \mathcal{A}_1, P_1)$  et  $(\Omega_2, \mathcal{A}_2, P_2)$  sont dits équivalents s'ils ont même signature. Le premier est dit plus fin que le second si ce dernier est équivalent à  $(\Omega_1, \mathcal{B}, P_1)$ , avec  $\mathcal{B} \subset \mathcal{A}_1$ . Dans ce cas la signature du premier est dite plus fine que celle du second.

### Remarques

- 1) Cette définition de l'équivalence suffit pour nos besoins : il est bien connu [ 11 ] que si deux espaces à bases dénombrables sont équivalents, il existe une application mesurable  $f$  de l'un dans l'autre transportant la mesure et telle que  $f$  soit bijective à des ensembles de mesure nulle près.
- 2) Il serait possible, quoique laborieux et peu utile de donner une expression analytique du fait que la signature  $m_1$  est plus fine que  $m_2$ .
- 3) La signature est un outil commode. Si on parle en effet d'espace dont les atomes ont pour masse  $\pi_1, \pi_2, \dots, \pi_n, \dots$  on introduit dans les atomes un ordre artificiel qui alourdit les énoncés et les

définitions par l'emploi d'indices multiples.

Les deux propositions suivantes - ou plutôt leurs corollaires - et le théorème 1 donnent les propriétés fondamentales des signatures.

### Proposition 1

$(\Omega, \mathcal{A}, P)$  étant un espace de probabilité et  $\mathcal{B}$  une sous-tribu de  $\mathcal{A}$ , tout atome de  $\mathcal{A}$  est contenu dans un atome de  $\mathcal{B}$ .

### Démonstration

A étant un atome de  $\mathcal{A}$ , soit  $\mathcal{C}$  l'ensemble des éléments de  $\mathcal{B}$  qui contiennent A à un ensemble de mesure nulle près. Alors il existe  $B \in \mathcal{C}$  tel que  $P(B) = \inf \{P(C) ; C \in \mathcal{C}\}$  car on peut trouver une suite  $C_n$  de  $\mathcal{C}$  telle que  $P(C_n)$  ait cet inf. pour limite ; on prend alors

$B = \bigcap_{n=1}^{\infty} C_n$ . Si B n'était pas un atome de  $\mathcal{B}$ , il existerait  $B' \in \mathcal{B}$

tel que  $0 < P(B') < P(B)$ . Mais  $P(B' \cap A) = 0$  car sinon, A étant un atome de  $\mathcal{A}$ , on aurait  $P(B' \cap A) = P(A)$  et donc

$0 < P(A) < P(B') < P(B)$  ce qui contredit la définition de B. Considérons alors  $B \setminus B'$  : à un ensemble de mesure nulle près il contient A et est de masse inférieure à  $P(B)$ , d'où la contradiction. B est donc un atome auquel on peut ajouter un ensemble de mesure nulle pour qu'il contienne strictement A.

### Corollaire

Si  $m$  est plus fine que  $m'$  alors  $\|m\| \leq \|m'\|$  et  $x(m) \leq x(m')$

### Proposition 2

Soit  $(\Omega, \mathcal{A}, P)$  le produit des espaces  $(\Omega_1, \mathcal{A}_1, P_1)$  et  $(\Omega_2, \mathcal{A}_2, P_2)$ .  
Toute classe d'atome de  $\mathcal{A}$  contient un atome de la forme  $A_1 \times A_2$ , et  
tout produit de cette forme est un atome de  $\mathcal{A}$ .

### Démonstration

Pour montrer que si  $A_1$  et  $A_2$  sont des atomes,  $A_1 \times A_2$  est un atome,  
on peut, sans perte de généralité, supposer  $P_1(A_1) = P_2(A_2) = 1$ . Les  
probabilités  $P_1$  et  $P_2$  ne prenant que les valeurs 0 et 1, il en est de  
même de la probabilité produit, et  $A_1 \times A_2$  est un atome.

Inversement, prenons dans chaque classe d'atomes de  $\mathcal{A}_1$  un représen-  
tant, désignons par  $H_1$  le complémentaire dans  $\Omega_1$  de la réunion de tous  
ces représentants :  $H_1$  est sans atome. Formons  $H_2$  dans  $\Omega_2$  de la même ma-  
nière. Nous avons à montrer que  $(\Omega_1 \times H_2) \cup (H_1 \times \complement_{\Omega_2} H_2)$  est sans  
atome. Si  $\mathcal{B}$  désigne la tribu qui est le produit de la tribu triviale  
sur  $\Omega_1$  et de la restriction à  $H_2$  de la tribu  $\mathcal{A}_2$ , si  $\mathcal{A}'$  désigne la  
restriction à  $\Omega_1 \times H_2$  de  $\mathcal{A}$  et  $P'$  la restriction correspondante de  $P$ ,  
puisque  $(\Omega_1 \times H_2, \mathcal{B}, P')$  est sans atome, d'après la proposition 2 il  
en est de même pour  $(\Omega_1 \times H_2, \mathcal{A}', P')$ ; on procède de même pour  
 $H_1 \times \complement_{\Omega_2} H_2$ , et ceci achève la preuve.

### Corollaire

Si  $m_1$  et  $m_2$  sont les signatures des composantes et  $m$  celle du pro-  
duit, on a :

$$1) \quad m(x) = \sum_t m_1(x/t) m_2(t)$$

$$2) \quad ||m|| = ||m_1|| ||m_2||$$

$$3) \quad x(m) = x(m_1) \times x(m_2)$$

Nous munissons maintenant l'ensemble des signatures d'une bonne topologie.

Au vu du dernier corollaire, il est commode de considérer une signature comme une mesure discrète sur le groupe multiplicatif des nombres réels positifs. Le théorème suivant montrera l'utilité de munir l'ensemble des signatures de la topologie vague des mesures sur ce groupe.

#### Définition 4

La suite  $(m_t)_{t=1}^{\infty}$  de signatures est dite convergente vers la signature  $m$  si pour toute fonction  $f$  continue à support compact dans  $(0, 1]$ , on a

$$\lim_{t \rightarrow \infty} \int_{\mathbf{x}} f(\mathbf{x}) m_t(\mathbf{x}) = \int_{\mathbf{x}} f(\mathbf{x}) m(\mathbf{x})$$

#### Théorème 1

Soit  $(\Omega, \mathcal{A}, P)$  un espace de probabilité de signature  $m$ ,  $(\mathcal{F}_t)_{t=1}^{\infty}$  une suite croissante de sous tribus de  $\mathcal{A}$ , de signatures  $m_t$ , et telle que  $\bigcup_{t=1}^{\infty} \mathcal{F}_t$  engendre  $\mathcal{A}$ . Alors  $m_t$  converge vers  $m$ .

La preuve repose sur le lemme suivant :

Lemme:

Sous les hypothèses du théorème 1, soit  $(F_t)_{t=1}^{\infty}$  une suite décroissante telle que  $F_t$  soit un atome de  $\mathcal{F}_t$ . Alors  $\bigcap_{t=1}^{\infty} F_t$  est ou bien de mesure nulle, ou bien un atome de  $\mathcal{A}$ .

Démonstration

Si  $\bigcap_{t=1}^{\infty} F_t$  n'est ni de mesure nulle, ni un atome de  $\mathcal{A}$ , il existe  $A \in \mathcal{A}$  tel que  $A \subset \bigcap_{t=1}^{\infty} F_t$  et  $0 < P(A) < \lim_{t \rightarrow \infty} P(F_t)$ . Comme  $\bigcup_{t=1}^{\infty} \mathcal{F}_t$  engendre  $\mathcal{A}$ , pour tout  $\varepsilon > 0$ , il existe  $t_\varepsilon$  et  $B_\varepsilon \in \mathcal{F}_{t_\varepsilon}$  tel que  $P(A \Delta B_\varepsilon) \leq \varepsilon$ . Prenons enfin  $\varepsilon$  strictement inférieur à  $P(A)$  et  $-P(A) + \lim_{t \rightarrow \infty} P(F_t)$ . Alors  $P(B_\varepsilon) > 0$  car sinon on aurait  $P(A) < \varepsilon$ . Puisque  $F_{t_\varepsilon}$  est un atome,  $B_\varepsilon$  contient  $F_{t_\varepsilon}$ , donc  $A$ , puisque  $A \subset F_{t_\varepsilon}$ , à un ensemble de mesure nulle près. Donc :

$$\varepsilon < -P(A) + \lim_{t \rightarrow \infty} P(F_t) \leq P(F_{t_\varepsilon} \setminus A) \leq P(B_\varepsilon \Delta A) \leq \varepsilon$$

est la contradiction cherchée.

Démonstration du théorème

Soit  $f$  continue sur  $(0, 1]$  et nulle sur  $(0, \delta]$ , où  $\delta$  est un nombre positif. Comme  $\sum_x x m(x) \leq 1$ , l'ensemble  $\mathcal{C}$  des classes d'atomes  $A$  de  $\mathcal{A}$  tels que  $P(A) \geq \delta$  est fini. Soit  $A \in \mathcal{C}$ ; d'après la proposition 1, il existe un atome, unique à un ensemble de mesure nulle près tel que  $F_t(A) \supset A$ . D'après le lemme  $\bigcap_{t=1}^{\infty} F_t(A)$  est de la classe de  $A$  et, l'ensemble  $\mathcal{C}$  étant fini, il existe un  $t_0$  tel que  $t \geq t_0$  entraîne

$F_t(A) \leq P(A) + \delta$ . Ceci assure que l'application de  $\mathcal{C}$  dans les classes d'atomes de  $\mathcal{F}_t$  définie par  $A \rightarrow F_t(A)$  est injective si  $t \geq t_0$ . En effet sinon il existerait un atome  $F$  de  $\mathcal{F}_t$  tel que  $F = F_t(A)$  et  $F = F_t(A')$  avec  $A$  et  $A'$  distincts dans  $\mathcal{C}$ ; on aurait  $F \supset A \cup A'$  et donc :

$$P(A) + \delta \leq P(A) + P(A') \leq P(F) < P(A) + \delta, \text{ puisque } t \geq t_0.$$

Il est ensuite trivial de constater que

$$\sum_x f(x) m(x) = \sum_{A \in \mathcal{C}} f(P(A)) = \lim_{t \rightarrow \infty} \sum_{A \in \mathcal{C}} f(P(F_t(A)))$$

Posant  $\mathcal{R}_t$  l'ensemble des classes d'atomes de  $\mathcal{F}_t$  qui ne sont pas de la forme  $F_t(A)$  pour  $A \in \mathcal{C}$ , et tels que  $P(F) \geq \delta$ , on remarque que si  $t \geq t_0$ , on a :

$$\sum_{F \in \mathcal{R}_t} f(P(F)) + \sum_{A \in \mathcal{C}} f(P(F_t(A))) = \sum_x f(x) m_t(x)$$

et il nous suffit de vérifier qu'il existe  $t_1$  tel que  $t \geq t_1$  entraîne  $\mathcal{R}_t = \emptyset$  pour achever la preuve du théorème.

D'après la proposition 1 et la définition de  $\mathcal{R}_t$ , si  $F \in \mathcal{R}_t$  il existe  $F' \in \mathcal{R}_{t-1}$  tel que  $F \subset F'$ . Si  $\mathcal{R}_t$  n'est jamais vide on peut donc trouver une suite décroissante  $(F_t)_{t=1}^{\infty}$  telle que  $F_t \in \mathcal{R}_t$ . D'après le lemme  $\bigcap_{t=1}^{\infty} F_t \in \mathcal{C}$ , ce qui fournit la contradiction cherchée.

### 3) Signature d'un espace produit

#### Définition 5

Si  $m_1$  et  $m_2$  sont deux signatures, on note  $m_1 * m_2$  la signature définie par :

$$m_1 * m_2 (x) = \sum_t m_1 (t) m_2 (x/t)$$

et si  $s$  est un nombre complexe de partie réelle  $> 1$ , on pose

$$\widehat{m}_1 (s) = \sum_x x^s m_1 (x).$$

D'après le corollaire de la proposition 2,  $m_1 * m_2$  est la signature de l'espace, produit des espaces de probabilité associés à  $m_1$  et  $m_2$ ,  $m_1$  et  $m_2$  pouvant être considérées comme des mesures discrètes sur le groupe multiplicatif des réels positifs,  $m_1 * m_2$  est le produit de convolution habituel dans un groupe. La transformée de Dirichlet  $\widehat{m}_1$  de  $m_1$  correspond d'ailleurs à la transformée de Fourier-Laplace. Son introduction n'est pas neuve, puisque la  $\alpha$ -entropie de A. Renyi de la signature  $m_1$  peut se définir comme  $H_s = \frac{1}{1-s} \text{Log } \widehat{m}_1 (s)$  si  $s$  est réel.

Nous passons aux produits infinis d'espaces de probabilité au moyen du théorème 2.

#### Théorème 2

Soient  $(\Omega_t, \mathcal{A}_t, P_t)_{t=1}^{\infty}$  une suite infinie d'espaces de probabilité et  $(m_t)_{t=1}^{\infty}$  leurs signatures. La signature du produit  $(\Omega, \mathcal{A}, P)$  de ces espaces est  $\prod_{t=1}^{\infty} m_t$ .

$$\text{De plus } \left\| \prod_{t=1}^{\infty} m_t \right\| = \prod_{t=1}^{\infty} \|m_t\| > 0 \quad \text{si } \prod_{t=1}^{\infty} x(m_t) > 0$$

$$\left\| \prod_{t=1}^{\infty} m_t \right\| = 0 \quad \text{si } \prod_{t=1}^{\infty} x(m_t) = 0$$

### Remarques

La convergence du produit infini de convolution  $\prod_{t=1}^{\infty} m_t$  est naturellement celle de la définition 4. Sauf si  $x(m_t)$ , et donc  $m_t$ , est nul pour quelque  $t$ , le théorème montre que  $(\Omega, \mathcal{A}, P)$  est sans atomes si et seulement si  $\sum_t (1 - x(m_t)) = +\infty$ . Si cette condition n'est pas vérifiée, le lemme de Borel Cantelli montre que le processus devient déterministe à l'infini. Un cas particulier du théorème mérite d'être mentionné, celui où  $\|m_t\| = 1$  pour tout  $t$ . Le théorème montre que le produit infini d'espaces de probabilité purement atomiques est ou bien purement atomique, ou bien continu (c'est-à-dire de signature nulle). C'est l'analogue de la célèbre "loi des types purs" de Jessen et Wintner.

### Démonstration du théorème

Désignons par  $\mathcal{F}_t$  la sous-tribu de  $\mathcal{A}$  produit des tribus  $\mathcal{A}_1, \dots, \mathcal{A}_t$  et des tribus triviales sur  $\Omega_{t+1}, \Omega_{t+2}, \dots$ . La signature de  $\mathcal{F}_t$  est  $m_1 * m_2 * \dots * m_t$  d'après la proposition 2. L'application du théorème 1 donne la première partie du théorème 2.

D'après le corollaire de la proposition 2 on a :

$$x \left( \prod_{t=1}^{\infty} m_t \right) = x(m_1) \dots x(m_s) x \left( \prod_{t=s+1}^{\infty} m_t \right) \leq x(m_1) \dots x(m_1)$$

$$\text{Donc } \prod_{t=1}^{\infty} x(m_t) = 0 \text{ entraîne } \prod_{t=1}^{\infty} m_t = 0 .$$

Supposons maintenant que  $\prod_{t=1}^{\infty} x(m_t) > 0$  . Puisqu'alors

$$x(m_t) \xrightarrow[t \rightarrow \infty]{} 1 , \text{ soit } t_0 \text{ tel que } t \geq t_0 \text{ entraîne } x(m_{t_0}) > \frac{1}{2} .$$

Pour  $t \geq t_0$  , il existe un atome  $A_t$  de  $\mathcal{O}_t$  , unique à un ensemble de mesure nulle près, tel que  $P(A_t) = x(m_t)$  . Soit  $A$  un atome de  $\mathcal{O}$  ,  $F_t(A)$  l'atome de  $\mathcal{F}_t$  contenant  $A$  . D'après la proposition 2, il existe des  $B_1, B_2, \dots, B_t, \dots$  tels que  $B_t$  soit un atome de  $\mathcal{O}_t$  et que

$$F_t(A) = B_1 \times B_2 \times \dots \times B_t \times \Omega_{t+1} \times \Omega_{t+2} \times \dots$$

à un ensemble de mesure nulle près.

Si l'ensemble  $T = \{t ; B_t \text{ n'est pas de la classe de } A_t\}$  était infini, il est clair qu'on aurait  $\lim_{t \in T} P_t(B_t) = 0$  et que  $\lim_{t \rightarrow \infty} P(F_t(A)) = 0$

ce qui contredirait le lemme du théorème 1. Définissons alors

$t(A) = 1 + \sup P$  . On peut alors écrire :

$$\left| \prod_{t=1}^{\infty} m_t \right| = \sum_{t=1}^{\infty} \sum_{\{A: t(A)=t\}} P(A)$$

où la dernière somme est prise pour tous les atomes  $A$  de  $\mathcal{O}$  , tels que  $t(A) = t$  . Introduisons enfin les quantités

$$a_t = \prod_{s>t} x(m_s)$$

$$b_t = \prod_{s<t} ||m_s|| \quad (b_1 = 1)$$

alors  $\sum_{\{A:t(A)=t\}} P(A)$ , si  $t = 1$  est  $a_1$ , et si  $t > 1$ , égale

$$(||m_{t-1}|| - x(m_{t-1})) a_t b_{t-1} = a_t b_t - a_{t-1} b_{t-1}$$

et il est clair que

$$\prod_{t=1}^{\infty} ||m_t|| = \lim_{t \rightarrow \infty} (a_t b_t).$$

la convergence du produit infini  $\prod_t x(m_t)$  entraîne que  $a_t \xrightarrow[t \rightarrow \infty]{} 1$ , ce qui achève la preuve du théorème.

Nous allons maintenant donner une démonstration combinatoire d'un résultat de Paul Lévy parallèle au théorème 2. Cette démonstration nous paraît plus naturelle, et peut être plus simple, que celle de Van Kampen [ 4 ] . Elle a de plus l'avantage de s'écrire directement dans un groupe abélien et non dans  $R$  . Nous notons, si  $X$  est une variable aléatoire à valeurs dans un espace mesurable  $(\Omega, \mathcal{L})$ , par  $m_X$  la signature de la probabilité induite par  $X$  sur  $(\Omega, \mathcal{L})$ .

### Théorème 3

Soient  $X_1, \dots, X_n, \dots$  des variables aléatoires indépendantes à valeurs dans un groupe topologique abélien muni de sa tribu borélienne, telles que le sous-groupe engendré par les atomes de  $X_1, \dots, X_n, \dots$  soit sans torsion. Si  $S_n = X_1 + \dots + X_n$ , alors :

$$x(m_{S_n}) \leq \left[ \sum_{t=1}^n (1 - x(m_{X_t})) \right]^{-1/4} .$$

Remarque

Naturellement, en général

$$m_{S_n} \neq m_{X_1} * \dots * m_{X_n} .$$

Corollaire (Paul Lévy)

Les hypothèses étant celles du théorème 3, supposons que  $S = \lim_n S_n$  existe presque-sûrement. Alors

$$x(m_S) = 0 \text{ si et seulement si } \prod_{t=1}^{\infty} x(m_{X_t}) = 0$$

Nous donnons d'abord la

Démonstration du corollaire

Si  $\prod_{t=1}^{\infty} x(m_{X_t}) = 0$  la série  $\sum_{t=1}^{\infty} (1 - x(m_{X_t}))$  diverge, et comme  $S = S_n + (\sum_{t>n} X_t)$  le théorème 3 entraîne immédiatement que  $x(m_S) = 0$ . La réciproque est triviale, compte tenu du fait que la signature d'une probabilité transportée par une application est toujours moins fine que la signature de la probabilité initiale. Ici la probabilité initiale est la probabilité produit sur  $G \times \dots \times G$  (n fois) induite par les  $X_1, \dots, X_n$ , et l'application est

$$(x_1, \dots, x_n) \longrightarrow x_1 + \dots + x_n$$

$$G \times \dots \times G \rightarrow G .$$

Pour démontrer le théorème 3, nous aurons besoin du lemme suivant.

Lemme

Soient  $Y_1, \dots, Y_n$  des variables aléatoires indépendantes et de même loi définie par  $P(Y_i = \pm 1) = \frac{1}{2}$  et  $u_1, u_2, \dots, u_n$  des éléments non nuls d'un groupe abélien sans torsion. Alors

$$P(Y_1 u_1 + Y_2 u_2 + \dots + Y_n u_n = 0) \leq \frac{1}{\sqrt{n+1}} .$$

Démonstration du lemme

Le groupe étant sans torsion, le sous-groupe engendré par  $\{u_1, u_2, \dots, u_n\}$  est isomorphe à  $Z^k$  et sans perte de généralité nous supposons ce groupe identique à  $Z^k$ ; ses éléments sont notés  $x = (x_1, \dots, x_k)$ , où les  $x_i$  sont des entiers. Désignons par  $S$  l'ensemble des  $(x_1, \dots, x_k)$  tels que si  $j = \inf \{i ; x_i \neq 0\}$ , alors  $x_j > 0$ . Il est clair que  $S$  est un semi groupe de  $Z^k$  et que  $S \cap -S = \{0\}$ . Aussi pouvons nous supposer tous les  $u_i$  dans  $S$ , puisque sinon  $-u_i \in S$ . Donc si pour  $T \subset \{1, 2, \dots, n\}$  on a

$\sum_{i \in T} u_i = 0$ , nécessairement  $T$  est vide. On en déduit immédiatement

que l'ensemble  $\mathcal{C}$  des  $T \subset \{1, 2, \dots, n\}$  tels que  $\sum_{i \in T} u_i - \sum_{i \in T'} u_i = 0$

( $T'$  est le complémentaire de  $T$ ) est une famille de Sperner, c'est-à-dire que si  $T_1$  et  $T_2 \in \mathcal{C}$  et  $T_1 \subset T_2$ , alors  $T_1 = T_2$ . Il est bien connu [7] qu'une telle famille de Sperner a au plus  $C_n^{\lfloor n/2 \rfloor}$  éléments. Il est facile de voir que

$$c_n^{\lfloor n/2 \rfloor} \leq \frac{2^n}{\sqrt{n+1}},$$

ce qui achève la preuve du lemme.

### Démonstration du théorème 3

Considérons des variables aléatoires  $X'_1, \dots, X'_n$  indépendantes entre elles et des  $X_1, \dots, X_n$ , telles que  $X_i$  et  $X'_i$  soient de même loi. Posons  $U_i = X_i - X'_i$ . Il est clair que

$$\left[ \chi(m_{S_n}) \right]^2 \leq P(U_1 + \dots + U_n = 0)$$

Désignons par  $T$  la variable aléatoire égale au nombre de  $U_1, \dots, U_n$  différents de zéro.  $H$  désignant le sous-groupe de  $G$  engendré par les atomes de  $X_1, X_2, \dots, X_n$ , il est clair que

$$\{U_1 + \dots + U_n = 0\} \subset \{U_1, \dots, U_n \in H\}$$

à un ensemble de mesure nulle près.

Mais les variables aléatoires  $U_1, \dots, U_n$  étant symétriques, et les variables aléatoires  $Y_1, \dots, Y_n$  étant définies comme dans le lemme (indépendamment des  $U_1, \dots, U_n$ ) on a

$$\begin{aligned} P(U_1 + \dots + U_n = 0) &= P(\{U_1 + \dots + U_n = 0\} \cap \{U_i \in H, \forall i\}) \\ &= P(\{Y_1 U_1 + \dots + Y_n U_n = 0\} \cap \{U_i \in H, \forall i\}) \\ &= E(P(\{Y_1 U_1 + \dots + Y_n U_n = 0\} \mid U_1, \dots, U_n; U_i \in H \forall i)) \end{aligned}$$

Le lemme permet de majorer la dernière probabilité conditionnelle, en ne tenant compte que des  $U_i \neq 0$  :

$$P(U_1 + \dots + U_n = 0) \leq \sum_{t=0}^n \frac{1}{\sqrt{1+t}} P(T = t) = E\left(\frac{1}{\sqrt{1+T}}\right)$$

Puisque  $E\left(\frac{1}{\sqrt{1+T}}\right) \leq \sqrt{E\left(\frac{1}{1+T}\right)}$ , nous majorons cette dernière quantité.

Posons  $p_t = P(U_t \neq 0)$ . Alors :

$$E(s^T) = \prod_{t=1}^n (1 - p_t + s p_t) \quad \text{et donc}$$

$$E\left(\frac{1}{1+T}\right) = \int_0^1 \prod_{t=1}^n (1 - p_t + s p_t) ds = \int_0^1 \prod_{t=1}^n (1 - p_t s) ds$$

Ensuite  $1 - p_t s \leq \exp - p_t s$  et on obtient

$$E\left(\frac{1}{1+T}\right) \leq \frac{1 - \exp - \left[ \sum_{t=1}^n p_t \right]}{\sum_{t=1}^n p_t} \leq \frac{1}{\sum_{t=1}^n p_t}$$

$$1 - p_t = P(U_t = 0) \leq x(m_t).$$

Donc, en résumant :

$$\begin{aligned} x(m_{S_n}) &\leq (P(U_1 + \dots + U_n = 0))^{1/2} \leq \left(E\left(\frac{1}{1+T}\right)\right)^{1/4} \\ &\leq \left(\sum_{t=1}^n (1 - x(m_t))\right)^{-1/4} \end{aligned}$$

### Remarques

1) Kolmogorov [5] a suivi une démarche analogue pour démontrer une inégalité sur la fonction de concentration de Paul Lévy.

2) Parallèlement au théorème 3, mentionnons ici un résultat de Guy Fourt : les hypothèses étant celles du théorème 3, on suppose de plus que le groupe sous-jacent est  $Z$  et que  $P(X_i \in \{0, 1\}) = 1$  pour tout  $i$ . Alors :

$$x(m_{S_{2n}}) \geq C_{2n}^n \frac{n^n (n+1)^n}{(2n+1)^{2n}}$$

$$x(m_{S_{2n+1}}) \geq C_{2n+1}^n \frac{1}{2^{2n+1}}$$

le résultat est naturellement le meilleur possible. Il est facile à vérifier si on suppose les  $X_i$  de même loi, mais c'est précisément le point délicat que de montrer que le minimum est atteint pour des variables de même loi.

Pour terminer ce paragraphe, nous allons montrer un curieux résultat concernant la possibilité de considérer un espace de probabilité comme un espace produit. Donnons d'abord la

### Définition 6

Un espace de probabilité  $(\Omega, \mathcal{A}, P)$  est dit  $n$ -divisible s'il existe  $n$  espaces  $(\Omega_t, \mathcal{A}_t, P_t)_{t=1}^n$  tous identiques tel que  $(\Omega, \mathcal{A}, P)$  soit équivalent au produit de ceux-ci. Il est dit indéfiniment divisible s'il est  $n$ -divisible pour tout  $n$ .

Naturellement, si  $m$  est la signature d'un espace  $n$ -divisible, il existe une signature  $m_1$  telle que,  $m = m_1^{*n}$  et c'est une condition nécessaire et suffisante de  $n$ -divisibilité.

Il existe évidemment des espaces indéfiniment divisibles : il suffit de prendre un espace formé d'un seul atome au plus et d'une partie continue. De tels espaces seront dits triviaux. Cependant :

#### Théorème 4

Il n'existe pas d'espaces indéfiniment divisibles non triviaux.

#### Démonstration

Désignons par  $\mathcal{C}$  la classe des mesures de probabilité  $\nu$  définies sur la demi-droite  $R^+ = \{x > 0\}$  telles que  $\nu$  soit atomique et telles qu'il existe un nombre  $\alpha_\nu > 0$  tel que  $e^{+\alpha_\nu y} \nu(\{y\})$  soit un nombre entier pour tout  $y > 0$ .

L'application  $m \rightarrow \nu_m$  de l'ensemble des signatures non nulles dans  $\mathcal{C}$  est définie par :

$$\nu_m(\{y\}) = m(e^{-y}) \exp(-\alpha y)$$

où  $\alpha$  est l'unique solution  $\geq 1$  de l'équation  $\hat{m}(\alpha) = 1$ . Cette correspondance n'est pas injective, mais il est facile de constater que :

$$\nu_{m_1 * m_2} = \nu_{m_1} * \nu_{m_2},$$

la convolution dans  $\mathcal{C}$  ayant le sens usuel. De plus  $m$  est la signature non nulle d'un espace trivial seulement si  $\nu_m$  est concentrée en un point.

Supposons donc qu'il existe un espace non trivial indéfiniment divisible. Soit  $m$  sa signature, soient  $m_n$  la signature telle que  $m_n^{*n} = m$ . On pose  $v_1 = v_m$  et  $v_n = v_{m_n}$ . Désignons par  $\alpha$  l'unique solution  $\geq 1$  de l'équation  $\hat{m}(\alpha) = 1$ . Il est utile de remarquer qu'on peut prendre  $\alpha_{v_n} = \alpha$ .

La mesure  $v_1$  étant atomique et concentrée sur  $(0, +\infty)$ , il existe des constantes  $\lambda$  et  $r$  positives, et une mesure atomique  $\mu$  sur  $[0, \infty)$  telles que

$$v_n = \varepsilon_{\lambda/nr} * \exp \frac{\lambda}{n} (\mu - \varepsilon_0) \quad \forall n = 1, 2, \dots$$

( $\varepsilon_a$  désigne la masse de Dirac en  $a$ ).

Soit  $y$  un atome fixé de  $\exp \lambda (\mu - \varepsilon_0)$  qui soit différent de zéro. C'est aussi un atome de  $\exp \frac{\lambda}{n} (\mu - \varepsilon_0)$  et de plus :

$$\exp \frac{\lambda}{n} (\mu - \varepsilon_0) (\{y\}) \xrightarrow[n \rightarrow \infty]{} 0 \quad (1)$$

Comme  $y + \frac{1}{nr}$  est un atome de  $v_n$  de masse  $\exp \frac{\lambda}{n} (\mu - \varepsilon_0) \{y\}$ , on en déduit que :

$$\exp \frac{\lambda}{n} (\mu - \varepsilon_0) (\{y\}) \exp - \alpha \left( y + \frac{\lambda}{nr} \right)$$

est un entier positif pour tout  $n$ , ce qui est évidemment incompatible avec (1).

#### 4) Image d'une probabilité - L'inégalité du déménageur

##### Définition 7

Soit  $(\Omega, \mathcal{A}, P)$  un espace de probabilité de signature  $m$ . L'ensemble  $F_m = \{P(A) ; A \in \mathcal{A}\}$  est appelé image de la probabilité  $P$  ou de la signature  $m$ . On définit également les fonctions sur  $[0, 1]$  suivantes

$$g_m(x) = \sup \{y \in F_m ; y \leq x\}$$

$$f_m(x) = x - g_m(x) .$$

Il est évident que  $F_m$  est invariant lorsque  $(\Omega, \mathcal{A}, P)$  est remplacé par un ensemble équivalent, d'où la notation.  $F_m$  est un fermé, aussi la fonction  $f_m(x)$ , nulle sur  $F_m$ , se construit-elle facilement sur chaque intervalle ouvert contigu à  $F_m$ , à la manière de [8] page 134. Le cas où  $F_m = [0, 1]$  est facile à caractériser : c'est le théorème de Kakeya dans le cas où  $||m|| = 1$ , et qu'il suffit de modifier légèrement dans le cas  $||m|| \leq 1$  :

##### Proposition 3

$F_m = [0, 1]$  si et seulement si

$$1 \leq \inf_{t \leq x(m)} \sum_{x < t} (x/t) m(x)$$

A. Renyi, dans son ouvrage sur le Calcul des Probabilités ([9], problème n° 48) "relègue au rang des exercices" ce théorème et propose la

généralisation suivante (ici modifiée pour avoir le cas  $||m|| \leq 1$ ) :

Proposition 4

La signature  $m$  est plus fine que toute signature  $m'$  telle que  $||m|| = ||m'||$  et

$$\sum_x m'(x) \leq \inf_{t \leq x(m)} \sum_{x \leq t} (x/t) m(x) .$$

Nous allons maintenant énoncer une inégalité simple et extrêmement utile puisqu'elle donne la réponse à la question initiale posée par D. Labelle. Nous l'appellerons l'inégalité du déménageur ; elle intervient en effet dans la question suivante : le déménageur dispose d'un camion de contenance 1 dans lequel se trouvent des boîtes de contenances respectives  $b_1, b_2, \dots, b_n, \dots$ , avec  $\sum_{n=1}^{\infty} b_n \leq 1$  . Il a à transporter un volume 1 d'objets, composés d'objets indémontables de volumes  $a_1, a_2, \dots, a_n, \dots$  plus une certaine quantité "continue" -du grain en vrac par exemple- de volume  $1 - \sum_{n=1}^{\infty} a_n$  . Il ne pourrait tout emmener que si les  $b_n$  et  $a_n$  satisfaisaient à certaines relations. Nous nous proposons de donner une minoration du volume d'objets qu'il doit laisser. Plus précisément :

### Théorème 5

Soit  $(\Omega, \mathcal{A}, P)$  un espace de probabilité  $\mathcal{F}$  et  $\mathcal{B}$  deux sous-tribus de  $\mathcal{A}$  telles que  $\mathcal{F}$  soit de signature  $\mu$  et  $\mathcal{B}$  ait une signature plus fine qu'une signature donnée  $m$ . Soit  $F$  la réunion des atomes de  $\mathcal{F}$  qui ne sont pas entièrement contenus dans un atome de  $\mathcal{B}$  (à un ensemble de mesure nulle près).

Alors, si la trace de  $\mathcal{B}$  sur  $\int_{\mathcal{F}}$  est contenue dans  $\mathcal{F}$ , on a l'inégalité :

$$P(F) \geq \sum_{\mathbf{x}} m(\mathbf{x}) f_{\mu}(\mathbf{x}).$$

### Démonstration

Soit  $\mathcal{B}'$  une tribu moins fine que  $\mathcal{B}$  et de signature  $m$ . Si l'inégalité annoncée est fautive, alors

$$P(\int_{\mathcal{F}}) > 1 - \sum_{\mathbf{x}} m(\mathbf{x}) f_{\mu}(\mathbf{x}) = 1 - ||m|| + \sum_{\mathbf{x}} m(\mathbf{x}) g_{\mu}(\mathbf{x}) \quad (1)$$

Comme tout atome de  $\mathcal{F}$  contenu dans  $\int_{\mathcal{F}}$  est contenu dans un atome de  $\mathcal{B}$ , et à fortiori de  $\mathcal{B}'$ , il existe un atome  $B_0$  de  $\mathcal{B}'$  tel que :

$$g_{\mu}(P(B_0)) < P(\int_{\mathcal{F}} \cap B_0) \quad (2)$$

En effet s'il n'en était pas ainsi, on aurait

$$\sum_{\mathbf{x}} m(\mathbf{x}) g_{\mu}(\mathbf{x}) = \sum_{\mathcal{B}} g_{\mu}(P(B)) \geq \sum_{\mathcal{B}} P(\int_{\mathcal{F}} \cap B) = P(\int_{\mathcal{F}} \cap A) \quad (3)$$

où les sommes  $\sum_{\mathcal{B}}$  sont prises sur toutes les classes d'atomes de  $\mathcal{B}'$  et où  $A$  est la réunion des atomes de  $\mathcal{B}'$ . En comparant les inégalités (1) et (3) on obtient :

$$P ( \mathcal{C}_F \cap A ) < P ( \mathcal{C}_F ) - P ( \mathcal{C}_A )$$

qui est impossible. On est donc assuré de l'existence d'un  $B_0$  satisfaisant (2).

Alors, la trace de  $\mathcal{B}$  sur  $\mathcal{C}_F$  étant contenue dans  $\mathcal{F}$ , donc  $\mathcal{C}_F \cap B_0 \in \mathcal{F}$ . L'inégalité :

$$g_\mu ( P ( B_0 ) ) < P ( \mathcal{C}_F \cap B_0 ) \leq P ( B_0 )$$

contredit alors la définition de  $g_\mu (x)$  pour

$$x = P ( B_0 ) .$$

Le théorème suivant est un cas typique d'application du théorème 5 :

### Théorème 6

Soit  $(\mathcal{F}_t)_{t=1}^\infty$  une famille croissante de sous-tribus de  $\mathcal{A}$ ,  $T$  un temps d'arrêt par rapport à cette famille, avec  $P ( T = +\infty ) \geq 0$ . Si  $\mathcal{F}_T$  a une signature plus fine que  $m$ , alors

$$P ( T > t ) \geq \sum_x m (x) f_{\mu_t} (x) ,$$

où  $\mu_t$  est la signature de  $\mathcal{F}_t$ . En particulier :

$$E ( T ) \geq 1 + \sum_x m (x) \sum_{t=1}^\infty f_{\mu_t} (x)$$

### Démonstration

Il suffit de montrer qu'on est dans les conditions d'application du théorème 5 avec  $F \subset \{ T > t \}$   $\mathcal{F} = \mathcal{F}_t$  et  $\mathcal{B} = \mathcal{F}_T$ .

Par définition la trace de  $\mathcal{F}_T$  sur  $\{T \leq t\}$  est contenue dans  $\mathcal{F}_t$ .  
 Soit  $A$  un atome de  $\mathcal{F}_t$  non contenu dans un atome de  $\mathcal{F}_T$ . Je dis  
 qu'à un ensemble de mesure nulle près  $A \subset \{T < t\}$ . En effet  
 $A \cap \{T \leq t\} \in \mathcal{F}_t$  et est contenu dans  $A$ .  $A$  étant un atome, ou bien  
 $P(A \cap \{T \leq t\}) = 0$ , ce qu'on veut démontrer, ou bien

$$P(A \cap \{T \leq t\}) = P(A).$$

Nous montrons que cette deuxième hypothèse est impossible ; elle entraîne  
 qu'il existe un entier  $s$  entre  $l$  et  $t$  tel que :

$$P(A \cap \{T = s\}) > 0.$$

D'après la proposition 1, il existe un atome  $B$  de  $\mathcal{F}_s$  tel que  $B \supset A$ .  
 Je dis que  $B$  est un atome de  $\mathcal{F}_T$ , ce qui contredira l'hypothèse faite  
 sur  $A$ . On a en effet  $P(B \cap \{T = s\}) > 0$  ; comme  $B \cap \{T = s\} \in \mathcal{F}_s$   
 et que  $B$  est un atome,  $B \cap \{T = s\} = B$ . Si  $F \in \mathcal{F}_T$  et  $F \subset B$ ,  
 alors  $F \cap \{T = s\} \in \mathcal{F}_s$  et donc  $P(F) = 0$  ou  $P(B)$ , ce qui mon-  
 tre que  $B$  est un atome de  $\mathcal{F}_T$  et achève la démonstration du théorème 6.

Pour donner quelques applications du théorème 6, la remarque sui-  
 vante nous sera utile :

### Proposition 5

Pour toutes signatures  $m_1$  et  $m_2$  on a

$$\sup_x f_{m_1 \# m_2}(x) \leq \sup_x f_{m_1}(x) \cdot \sup_x f_{m_2}(x)$$

### Démonstration

Posant, pour  $i = 1$  ou  $2$ ,  $s_i = \sup_x f_{m_i}(x)$ . Soit  $x \in [0, 1]$ ; nous avons à montrer qu'il existe un élément  $d$  de  $F_{m_1} * m_2$  dans  $(x - s_1, s_2, x]$ . Si  $x \in F_{m_1}$  c'est évident. Si  $x \notin F_{m_1}$  soit  $(a, b)$  l'intervalle contigu à  $F_{m_1}$  tel que  $a < x < b$ . Il existe  $c$  dans  $F_{m_2}$  tel que

$$\frac{x - a}{b - a} - s_2 < c \leq \frac{x - a}{b - a}$$

Comme  $b - a \leq s_1$ , le nombre  $d = a + (b-a)c$  sera le nombre cherché si on montre que  $d \in F_{m_1} * m_2$ .

Or  $c$  s'écrit  $\alpha + \sum_x n(x) x$

où  $\alpha$  est un nombre de  $[0, 1 - ||m_2||]$

et  $n(x)$  un entier de  $[0, m_2(x)]$ . Donc :

$$d = (1 - ||m_2|| - \alpha) a + \alpha b + \sum_x (m_2(x) - n(x)) a x + \sum_x n(x) b x$$

ce qui, puisque  $a$  et  $b$  sont dans  $F_{m_1}$ , achève la démonstration.

### Corollaire

Si  $A_1$  et  $A_2$  sont des entiers, si  $F_{m_i} = \left\{ \frac{k}{A_i} ; k = 0, 1, \dots, A_i \right\}$  pour  $i = 1$  ou  $2$ , alors :

$$F_{m_1} * m_2 = \left\{ \frac{k}{A_1 A_2} ; k = 0, 1, \dots, A_1 A_2 \right\}$$

Remarque

La forme particulière de  $F_{m_i}$  n'entraîne pas que  $m_i \left(\frac{1}{A_i}\right) = A_i$ , cas où le corollaire serait évident.

Considérons maintenant des espaces  $(\Omega_t, \mathcal{A}_t, P_t)_{t=1}^{\infty}$  d'images respectives

$$F_{m_t} = \left\{ \frac{k}{D_t} ; k = 0, 1, \dots, D_t \right\} ,$$

où  $D_t$  est un entier. Par exemple, on peut avoir  $m_t \left(\frac{1}{D_t}\right) = D_t$ , mais aussi  $n$  atomes de masses respectives  $2^k / 2^n - 1$ , avec  $D_t = 2^n - 1$ . Formons alors l'espace produit  $(\Omega, \mathcal{A}, P)$  et introduisons les tribus  $\mathcal{F}_t$ , produit des tribus  $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_t$  et des tribus triviales sur  $\Omega_{t+1}, \Omega_{t+2}, \dots$ . Posons  $q_t = D_1 \dots D_t$  et désignons par  $\mu_t$  la signature de  $\mathcal{F}_t$ . Alors on a, d'après le corollaire précédent

$$F_{\mu_t} = \left\{ \frac{k}{q_t} ; k = 0, 1, \dots, q_t \right\}$$

La fonction  $f_{\mu_t}(x)$  a une expression très simple :

$$f_{\mu_t}(x) = \{x q_t\} / q_t ,$$

où  $\{a\} = a - [a]$  et  $[a]$  est la partie entière du nombre positif  $a$ .

Le théorème 6 devient alors

$$E(T) \geq 1 + \sum_x m(x) \sum_{t=1}^{\infty} \{x q_t\} / q_t .$$

Particularisons encore en supposant qu'il existe un entier  $k$  tel que  $m(\frac{1}{k}) = k$  (cas où on a choisi entre  $k$  personnes de façon équiprobable). Alors :

$$E(T) \geq \sum_{t=0}^{\infty} \frac{q_t \text{ modulo } k}{q_t}$$

ce qui est la formule demandée par D. Labelle lorsque les  $D_t$  sont tous égaux.

Une dernière application du théorème 6 considère le cas où  $\Omega_t = \{0, 1\}$  avec

$$P_t(\{0\}) = q \quad P_t(\{1\}) = p = 1 - q$$

Supposons qu'on veuille "fabriquer", à l'aide d'un temps d'arrêt une probabilité ayant même signature qu'une loi géométrique de moyenne  $1/p$ . La méthode naturelle consiste à poser, avec des notations évidentes :

$$T = \inf \{t ; \omega_t = 1\}$$

Le théorème 6 montre que si  $q < p$  cette méthode naturelle est la plus rapide. En effet, ici

$$f_{\mu_t}(q^{k-1} p) = 0 \quad \text{si } t \geq k$$

$$f_{\mu_t}(q^{k-1} p) = q^{k-1} p \quad \text{si } t < k$$

et on a 
$$\frac{1}{p} = 1 + \sum_x m(x) \sum_{t=1}^{\infty} f_{\mu_t}(x) .$$

Nous verrons cependant que les cas où la borne de  $E(T)$  du théorème 6 est atteinte sont rares, les exemples traités ci-dessus étant l'exception. En général, il est possible de trouver un temps d'arrêt  $T$  tel que  $P(T > t)$  atteigne le minimum pour un  $t$  donné, mais ce  $T$  optimal varie avec chaque  $t$ .

## 5) Les signatures des $\mathcal{F}_T$

Nous démontrons dans ce paragraphe le théorème (n° 9) fondamental d'existence d'un temps d'arrêt permettant de reconstruire une probabilité donnée. Dans tout ce paragraphe, on se donne une fois pour toutes un espace de probabilité  $(\Omega, \mathcal{A}, P)$  et une suite croissante  $(\mathcal{F}_t)_{t=1}^{\infty}$  de sous-tribus de  $\mathcal{A}$ , de signatures respectives  $\mu_t$ . Si  $(\Omega, \mathcal{A}, P)$  est le produit d'espaces  $(\Omega_t, \mathcal{A}_t, P_t)_{t=1}^{\infty}$ , de signatures  $m_t$ , les  $\mathcal{F}_t$  seront les tribus produit des tribus  $\mathcal{A}_1, \dots, \mathcal{A}_t$  et des tribus triviales sur  $\Omega_{t+1}, \Omega_{t+2}, \dots$ . On se réfèrera brièvement à ce cas particulier en disant qu'on est dans le cas indépendant. Les temps d'arrêt  $T$  s'entendent relativement aux  $(\mathcal{F}_t)_{t=1}^{\infty}$ , et nous imposerons toujours à  $T$  de vérifier  $P(T < \infty) = 1$ .

Le théorème 7 précise l'idée intuitive qu'on ne peut "fabriquer du continu avec du discret" suivant une procédure finie.

### Théorème 7

Si  $T$  est un temps d'arrêt tel que la signature de  $\mathcal{F}_T$  soit plus fine que la signature  $m$ , alors

$$||m|| \geq \lim_{t \rightarrow \infty} ||\mu_t|| = \lambda .$$

De plus, dans le cas indépendant, l'égalité n'est atteinte que si  $\lambda = 0$  ou s'il existe  $t_0$  tel que  $||\mu_{t_0}|| = \lambda$ .

### Démonstration

Sans perte de généralité, on peut supposer que  $m$  est la signature de  $\mathcal{F}_T$ . Les atomes de  $\mathcal{F}_T$  sont caractérisés par le lemme suivant :

### Lemme

$A$  est un atome de  $\mathcal{F}_T$  si et seulement si il existe  $t$  tel que  $A$  soit un atome de  $\mathcal{F}_t$  contenu dans  $\{T = t\}$ .

La démonstration du lemme se fait de la même manière qu'au théorème 6, et nous ne la détaillerons pas.

Soit  $B_t$  la réunion des atomes de  $\mathcal{F}_t$  (à un ensemble de mesure nulle près). On pose  $B = \bigcap_{t=1}^{\infty} B_t$ . Il est clair que  $P(B_t) = \|\mu_t\|$ , que  $B_t \supset B_{t+1}$  et que  $P(B) = \ell$ .

D'après le lemme on peut écrire :

$$\|\mu\| = \sum_{t=1}^{\infty} P(\{T = t\} \cap B_t) \geq \sum_{t=1}^{\infty} P(\{T = t\} \cap B) = \ell$$

puisque  $P(T < \infty) = 1$ .

Pour étudier séparément le cas indépendant, on introduit  $A_t$  la réunion dans  $\Omega_t$  des atomes de  $\mathcal{A}_t$ ,  $A'_t = \Omega_1 \times \dots \times \Omega_{t-1} \times A_t \times \Omega_{t+1} \times \dots$  et enfin  $C_t = \bigcap_{i=t+1}^{\infty} A'_i$ .

Comme on a  $B_t = \bigcap_{i=1}^t A'_i$  à un ensemble de mesure nulle près, on a donc

$B_t \cap C_t = B$  et  $B_t$  et  $C_t$  sont indépendants. Ceci permet d'écrire

$$\sum_{t=1}^{\infty} P(\{T = t\} \cap B_t) = \ell + \sum_{t=1}^{\infty} P(\{T = t\} \cap B_t) (1 - P(C_t)).$$

Si  $\ell > 0$ , il existe donc  $t_0$  tel que

$$0 < P(B \cap \{T = t_0\}) \leq P(\{T = t_0\} \cap B_{t_0})$$

Si de plus on n'a pas  $\|\mu_{t_0}\| = \ell$ , alors il existe  $t \geq t_0$  tel que  $\|\mu_t\| < 1$  et donc

$$1 - P(C_{t_0}) > 0.$$

On voit que sous ces hypothèses on a  $\|m\| > \ell$ .

### Remarque

L'exemple suivant montre que en dehors des cas cités, l'égalité est possible dans le cas non-indépendant : Prenons  $\Omega = [0, 1/2]$  muni de la mesure de Lebesgue plus une masse 1/2 en zéro.  $\mathcal{F}_t$  est la tribu engendrée par les atomes 0 et  $(0, \frac{1}{2^t}]$ , et les boréliens de  $(\frac{1}{2^t}, \frac{1}{2}]$ . On pose  $T(0) = 1$  et  $T(\omega) = \inf \{t ; \frac{1}{2^t} < \omega\}$ . Il est facile de constater que l'égalité est atteinte dans ce cas.

### Corollaire

Si  $\mathcal{F}_T$  est sans atome,  $\lim_{t \rightarrow \infty} \|\mu_t\| = 0$   
 et  $E(T) \geq 1 + \sum_{t=1}^{\infty} \|\mu_t\|$ .

### Démonstration

Le fait que  $\|\mu_t\| \xrightarrow{t \rightarrow \infty} 0$  est une conséquence immédiate du théorème 6. Pour le reste, il suffit de vérifier que  $P(T > t) \geq \|\mu_t\|$  pour tout  $t$ . S'il n'en était pas ainsi, il existerait  $t$  tel que

$P (T \leq t) > 1 - ||\mu_t||$  et,  $B_t$  désignant la réunion des atomes de  $\mathcal{F}_t$ , on aurait  $P (\{T \leq t\} \cap B_t) > 0$ , ce qui d'après le lemme, entraînerait que  $\mathcal{F}_T$  possède un atome.

Remarque

Il est immédiat que si  $\lim_{t \rightarrow \infty} ||\mu_t|| = 0$  il existe  $T$  tel que  $\mathcal{F}_T$  soit sans atome, il suffit de définir  $T = \inf \{t ; \omega \notin B_t\}$ . Et on a  $E (T) = 1 + \sum_{t=1}^{\infty} ||\mu_t||$ .

Le théorème suivant ne serait qu'un cas particulier du théorème 9 s'il n'avait son intérêt propre, celui d'exhiber, dans un cas assez proche de celui envisagé par D. Labelle, un temps d'arrêt qui atteint la borne définie par le théorème 6.

Théorème 8

Dans le cas indépendant, supposons que  $m_t (\frac{1}{D_t}) = D_t$ , où  $(D_t)_{t=1}^{\infty}$  est une suite d'entiers  $\geq 2$ . Alors, pour toute signature  $m$  telle que  $||m|| = 1$ , il existe un temps d'arrêt  $T$  tel que la signature de  $\mathcal{F}_T$  soit plus fine que  $m$  et tel que

$$E (T) = 1 + \sum_x m (x) \sum_{t=1}^{\infty} \{x q_t\} / q_t$$

(Ici  $q_t = D_1 \dots D_t$ ).

### Démonstration

Pour construire  $T_0$ , on pose :  $R = \{(x, i) ; m(x) > 0 ; i = 1, 2, \dots, m(x)\}$ . On construit alors les  $C_1(x, i)$  dans  $\mathcal{F}_1$ , disjoints et tels que  $P(C_1(x, i)) = [x q_1] / q_1$ . Supposons qu'on ait construit les  $C_n(x, i)$  dans  $\mathcal{F}_n$ , avec  $n \leq t$  et  $(x, i) \in R$ , de sorte que pour  $n$  fixés les  $C_n(x, i)$  soient disjoints,  $C_n(x, i) \subset C_{n+1}(x, i)$  et  $P(C_t(x, i)) = [x q_t] / q_t$ . Alors on construit les  $C_{t+1}(x, i)$  pour que les  $C_n(x, i)$ ,  $n \leq t+1$  aient les mêmes propriétés. Pour cela on choisit les  $B(x, i)$  dans  $\mathcal{F}_{t+1}$  disjoints entre eux, disjoints de  $\bigcup \{C_t(x, i) ; (x, i) \in R\}$  et tels que :

$$P(B(x, i)) = [x q_{t+1}] / q_{t+1} - [x q_t] / q_t .$$

Cela est toujours possible, en vertu de l'inégalité

$D_{t+1}[a] \leq [D_{t+1} a]$ . Les  $C_{t+1}(x, i) = C_t(x, i) \cup B(x, i)$  ont les propriétés requises. Et on définit  $T_0$  par :

$$\{T_0 \leq t\} = \bigcup \{C_t(x, i) ; (x, i) \in R\} .$$

Voici maintenant le théorème principal du chapitre ;

### Théorème 9

Supposons  $\lim \mu_t = 0$ . Soit  $m$  une signature satisfaisant à :

$$||m|| \geq \ell = \lim_{t \rightarrow \infty} ||\mu_t|| ,$$

l'inégalité étant stricte lorsque :

$$0 < \ell < ||\mu_t|| \quad \text{pour tout } t .$$

Alors il existe un temps d'arrêt  $T$  tel que la signature de  $\mathcal{F}_T$  soit plus fine que  $m$ .

### Lemme

$(\Omega, \mathcal{B}, P)$  étant un espace de signature  $m$ , soit  $E \in \mathcal{B}$ . Pour toute suite  $(a_n)_{n=1}^{\infty}$  de nombres positifs ou nuls, tels que

$\sum_{n=1}^{\infty} a_n \leq P(E)$ , il existe une suite  $(A_n)_{n=1}^{\infty}$  telle que  $A_n \in \mathcal{B}$ ,

$A_n \subset E$ , les  $A_n$  soient deux à deux disjoints, et

$$a_n - x(m) < P(A_n) \leq a_n \quad \text{pour tout } n$$

### Démonstration

Nous construisons les  $a_n$  par récurrence. Soit  $F_1$  le fermé image par  $P$  de la trace de  $\mathcal{B}$  sur  $E$ , il est clair que tout intervalle contigu à  $F_1$  est de longueur inférieure à  $x(m)$ . Il existe donc  $A_1 \subset E$  et  $A_1 \in \mathcal{B}$  tel que

$$a_1 - x(m) < P(A_1) \leq a_1.$$

Supposons qu'on ait construit  $A_1 \dots A_n$ . On définit alors  $F_{n+1}$  comme le fermé image par  $P$  de la trace de  $\mathcal{B}$  sur  $E \setminus \bigcup_{i=1}^n A_i$ , et l'argument analogue permet de conclure.

### Démonstration du théorème

Nous commençons par le cas  $||m|| = 1$ . On pose comme plus haut  $R = \{(x, i) ; m(x) > 0 ; i = 1, \dots, m(x)\}$ . Si  $r = (x, i)$  est un élément de  $R$ , on définit  $x_r$  par  $r = (x_r, i)$ .

Nous allons construire, par récurrence sur  $t$  des ensembles deux à deux disjoints  $D_t^{(r)}$ , avec  $D_t^{(r)} \in \mathcal{F}_t$  et  $r \in R$ , tels que

$$0 \leq x_r - \sum_{i=1}^t P(D_i^{(r)}) \leq x(\mu_t) \quad (1)$$

On construit les  $D_1^{(r)}$  en appliquant le lemme au cas où  $\mathcal{B}$  est  $\mathcal{F}_1$ ,  $E$  est  $\Omega$  et les  $(a_n)_{n=1}^\infty$  sont les  $(x_r)_{r \in R}$ . Supposant les  $D_i^{(r)}$  construits pour tout  $r \in R$  et pour tout  $i < t$  on pose, si  $t > 1$

$$E_t = \Omega \setminus \bigcup_{i=1}^{t-1} \bigcup_{r \in R} D_i^{(r)}.$$

Appliquons le lemme au cas où  $\mathcal{B}$  est  $\mathcal{F}_t$ ,  $E$  est  $E_t$  et les  $a_n$  sont

les  $x_r - \sum_{i=1}^{t-1} P(D_i^{(r)})$ . Les  $A_n$  correspondront aux  $D_t^{(r)}$ .

Les  $\{D_t^{(r)} ; r \in R \quad t = 1, 2, \dots\}$  satisfont bien à (1). Ensuite, comme  $\lim_{t \rightarrow \infty} \mu_t = 0$ , alors  $x(\mu_t) \rightarrow 0$ , et donc  $P(A_r) = x_r$ , où

$A_r = \bigcup_{t=1}^\infty D_t^{(r)}$ . Pour terminer, définissons le temps d'arrêt  $T$  par

$$\{T = t\} = \bigcup_{r \in R} D_t^{(r)}.$$

Il est clair que les  $A_r$  sont  $\mathcal{F}_T$  mesurables.

Reste à envisager le cas  $0 < ||m|| < 1$ , le cas  $||m|| = 0$  étant réglé par la remarque qui suit le corollaire du théorème 7. On commence par fabriquer la partie continue ; désignons par  $B_t$  la réunion des

atomes de  $\mathcal{J}_t$  (à un ensemble de mesure nulle près). L'hypothèse sur  $m$  entraîne qu'il existe  $t_0$  tel que :

$$||m|| \geq ||\mu_{t_0}||$$

On définit alors  $T$  sur  $D = \Omega \setminus B_{t_0}$  par

$$T = \inf \{t ; \omega \notin B_t\}$$

Il est clair que  $D$ , muni de la tribu  $\mathcal{F}_T$  restreinte à  $D$  est sans atome. On peut donc trouver  $C \subset D$ ,  $C \in \mathcal{F}_T$  tel que

$$P(C) = 1 - ||m||.$$

Définissant enfin

$$R = \{(x, i) ; m(x) > 0, i = 1 \dots m(x)\}$$

il existe une partition  $\{C^{(r)} ; r \in R\}$  de  $D \setminus C$  telle que  $C(x, i) \in \mathcal{F}_T$  et  $P(C^{(r)}) \leq x_r$ . On aura alors

$$\sum_{r \in R} P(C^{(r)}) = ||m|| - ||\mu_{t_0}||.$$

La partie de  $m$  restant à construire, c'est-à-dire la signature  $m'$  définie par

$$m'(x) = \text{nombre de } \{(y, i) \in R ; y - P(C(y, i)) = x\}$$

satisfait à  $||m'|| = ||\mu_{t_0}|| = P(B_{t_0})$  ; en posant  $\{T > t_0\} = B_{t_0}$ , on est ramené au cas précédent  $||m|| = 1$ .

Remarque :

On aimerait avoir un énoncé analogue en enlevant la condition  $\lim_{t \rightarrow \infty} \mu_t = 0$  et en la remplaçant par "m est une signature moins fine que la signature de la tribu engendrée par les  $\mathcal{F}_t$ "

Un tel énoncé est impossible, comme le montre l'exemple suivant :  $\Omega$  est l'ensemble des rationnels,  $\mathcal{O}$  la tribu la plus fine, P une probabilité qui charge positivement chaque rationnel.  $\mathcal{F}_t$  est la tribu engendrée par les intervalles  $\left[ \frac{k}{2^t}, \frac{k+1}{2^t} \right)$   $k \in \mathbb{Z}$ . Si m est égale à la signature de  $(\Omega, \mathcal{O}, P)$ , tout temps d'arrêt T tel que  $\mathcal{F}_T$  ait une signature plus fine que m doit nécessairement satisfaire à  $\mathcal{F}_T = \mathcal{O}$  ce qui est évidemment impossible.

En application du théorème 9, nous allons voir un énoncé analogue au théorème suivant de Skohorod [12] .

Théorème

Soit B (t) un mouvement brownien standard sur R avec B (0) = 0, et  $X_1, \dots, X_n, \dots$  des variables aléatoires indépendantes entre elles et de B (t). On suppose  $E (X_n) = 0$  et  $\sigma^2 (X_n) < \infty$  .

Soit  $S_n = X_1 + \dots + X_n$  .

Alors il existe une suite croissante de temps d'arrêt

$T_1 \leq T_2 \leq \dots \leq T_n \leq \dots$  pour B (t) tels que

1°) Les processus  $(B (T_n))_{n=1}^{\infty}$  et  $(S_n)_{n=1}^{\infty}$  sont identiques.

2°)  $E (T_n) = \sigma^2 (S_n)$

3°) Il existe  $L_m$  tel que  $E ( T_n - T_{n-1} )^m \leq L_m E ( X_n^m )$

4°) Si les  $|X_n| \leq h$ , alors

$$|B(s) - B(T_n)| \leq h \text{ si } s \in [T_n, T_{n+1}].$$

Rappelons la méthode de construction de  $T_1$  en prenant pour simplifier  $X_1$  symétrique.

1°) On tire  $X_1$

2°) On place des bornes  $-|X_1|$  et  $|X_1|$  de part et d'autre de l'origine.

3°) On laisse filer le mouvement brownien  $B(t)$  jusqu'à ce qu'il frappe - au bout du temps  $T_1$  - l'une des bornes  $-|X_1|$  et  $|X_1|$ .

Quelques calculs standards montrent alors que  $T_1$  a les propriétés annoncées. Si  $X_1$  n'est plus symétrique, on remplace à l'étape n° 2 les bornes par  $X_1$  et  $G(X_1)$ , où  $G(x)$  est une fonction convenablement choisie.

Remarquons que les temps d'arrêt sont ici définis par rapport à des tribus plus riches que celles  $\mathcal{F}_t$  engendrées par  $\{B(s); 0 \leq s \leq t\}$  puisqu'il faut en effet choisir  $X_1$ . Si on veut s'en tenir strictement aux  $\mathcal{F}_t$ , puisque  $\mathcal{F}_t$  est sans atome pour tout  $t > 0$  (voir [ ], page ) on peut utiliser  $\mathcal{F}_\varepsilon$  avec  $\varepsilon > 0$  pour choisir  $X_1$ : ceci retardera  $T_1$  de  $\varepsilon$  et modifie légèrement l'énoncé ci-dessus.

### Théorème 10

Soit  $(B(t))_{t=0}^{\infty}$  une chaîne de Markov irréductible récurrente et homogène sur l'ensemble d'états dénombrable  $E$ ,  $\mathcal{F}_t$  la tribu engendrée par les  $\{B(0), \dots, B(t)\}$  et  $(S_n)_{n=1}^{\infty}$  une chaîne de Markov quelconque sur  $E$  indépendante de  $(B(t))_{t=0}^{\infty}$ . Alors il existe une suite croissante  $T_1 \leq \dots \leq T_n \leq \dots$  de temps d'arrêt par rapport aux  $(\mathcal{F}_t)_{t=0}^{\infty}$  telle que les processus  $(B(T_n))_{n=1}^{\infty}$  et  $(S_n)_{n=1}^{\infty}$  soient identiques.

Nous esquissons la preuve : il est clair que la construction de  $T_1$  suffira.

On vérifie d'abord que  $\lim_{t \rightarrow \infty} \mu_t = 0$ . En effet, si on partage  $E$  en deux ensembles disjoints  $A_1$  et  $A_2$ , soit

$$p_{ij} = P(B(t+1) \in A_j \mid B(t) \in A_i) \quad i, j = 1, 2$$

$B(t)$  étant récurrente et irréductible, aucun des  $p_{ij}$  n'est égal à 1. Comme  $m = \max_{i,j} p_{ij} < 1$ , on peut affirmer que  $x(\mu_t) \xrightarrow{t \rightarrow \infty} 0$  car  $x(\mu_t) \leq m^t$ .

On peut alors appliquer le théorème 9 en fabriquant un temps d'arrêt  $T$  tel que la signature de  $\mathcal{F}_T$  soit plus fine que la probabilité induite sur  $E$  par  $\mathcal{S}_1$ . ( $\mathcal{F}_T$  joue alors le rôle de  $\mathcal{F}_\varepsilon$  dans l'exemple du mouvement brownien). Il existe donc une partition de  $\Omega$   $(A_e)_{e \in E}$  telle que  $A_e \in \mathcal{F}_T$  et  $P(A_e) = P(S_1 = e)$ . Définissons  $e(\omega)$  par  $\omega \in A_{e(\omega)}$ . Puisque le processus  $B(t)$  est markovien et récurrent, la variable aléatoire suivante :

$$T_1(\omega) = \inf \{t > T ; B(t) = e(\omega)\}$$

est bien un temps d'arrêt par rapport aux  $\mathcal{F}_t$  tel que  $P(T_1 < \infty) = 1$ .

Ainsi par construction  $B(T_1)$  et  $S_1$  sont-ils de même loi.

### Remarques

1. Le problème posé par le fait que la solution de Skohorod est insatisfaisante puisque faisant appel à des tribus un peu trop riches a été résolu en 1968 par Dubins [3]. Dubins considère des parties dénombrables  $E_n$  de  $R$  qui portent des mesures  $\mu_n$  convergeant vaguement vers la mesure  $\mu$  à construire. On définit  $T_n$  comme le temps de frappe de  $E_n$  par le mouvement brownien  $B(t)$  et on a choisi de façon habile  $E_n$  de sorte que  $B(T_n)$  soit de loi  $\mu_n$ . On vérifie enfin que  $T = \lim_n T_n < \infty$  et que  $B(T)$  est de loi  $\mu$ . Cette méthode par temps de frappe est en principe inapplicable à notre problème sur une chaîne de Markov, l'espace d'états n'étant pas assez grand : imaginer le cas où la chaîne a deux états. Il faut donc se résigner à construire un peu lentement un appareil aléatoire comme nous l'avons fait.

2. Bien entendu, si on a des renseignements sur la chaîne, les résultats peuvent être beaucoup plus précis. Ceci arrive en particulier lorsqu'on considère des variables aléatoires  $(X_t)_{t=1}^{\infty}$  indépendantes et de même loi  $\mu$  à valeurs dans un espace mesurable  $(\Omega, \mathcal{F})$ , et qu'on cherche à construire un temps d'arrêt  $T$  tel que  $X_T$  soit de loi donnée. Nous donnerons les détails dans une autre publication.

## 6) L'optimisation des temps d'arrêt et l'entropie de la signature

Dans tout ce paragraphe, nous nous placerons dans ce que nous avons décrit au § 5 comme le cas indépendant, et les notations du §5 restent en vigueur.

### Définition 8

L'entropie  $e(m)$  d'une signature est le nombre, fini ou non :

$$e(m) = \sum_x m(x) \times \text{Log} \frac{1}{x}$$

Nous nous proposons maintenant de montrer que le temps moyen nécessaire  $E(T)$  pour construire  $m$  est de l'ordre de l'entropie de  $m$  en général - ou plutôt, sous des hypothèses raisonnables en ce qui concerne les  $m_t$  - Pour simplifier, nous supposerons toujours que  $m$  est purement atomique, mais ce point pourrait être amélioré.

Commençons par une proposition concernant le cas où  $m$  et  $m_t$  correspondent à des cas équiprobables.

### Proposition 6

Si  $m(1/k) = k$  et  $m_t(1/D_t) = D_t$ , où  $k$  et les  $D_t$  sont des entiers, tels que  $D_t \geq k$  pour tout  $t$  alors il existe un temps d'arrêt  $T$  tel que la signature de  $\mathcal{F}_T$  soit plus fine que  $m$  et  $E(T)$  satisfait à :

$$E(T) \leq \frac{43}{28}$$

l'égalité n'étant atteinte que pour  $k = 4$ ,  $D_1 = 7$  et  $D_t = 5$  pour  $t \geq 2$ .

Si de plus les  $D_t$  sont tous égaux entre eux, on a :

$$E(T) \leq \frac{3}{2}$$

l'égalité n'étant atteinte que pour  $k = 2$  et  $D_t = 3$ .

### Démonstration

On a vu dans les applications du théorème 6, compte tenu du théorème 8 qu'il existe un  $T$  tel que :

$$E(T) = \sum_{t=0}^{\infty} (q_t \bmod k) / q_t$$

où  $q_t = D_1 \dots D_t$  et  $0 \leq q_t \bmod k < k$ . Posons  $q_1 \bmod k = r$ .

Alors  $r / q_1 \leq r / (r + k)$ . Si  $q_t \not\equiv 0 \pmod k$ , alors  $D_2, D_3, \dots, D_t > k$ .

Donc  $(q_t \bmod k) / q_t \leq (k - 1) / (r + k) (k + 1)^{t-1}$  si  $t \geq 2$ .

$$\text{Donc } E(T_0) \leq 1 + \frac{r}{r+k} + \frac{k-1}{k(r+k)}.$$

$$\text{Or } \sup_{0 \leq r < k} \frac{r}{r+k} + \frac{k-1}{k(r+k)} = \frac{k^2-1}{2k^2-k} \quad (r = k-1)$$

$$\sup_{k=2,3,\dots} \frac{k^2-1}{2k^2-k} = \frac{15}{28} \quad (k = 4)$$

On a donc toujours  $E(T_0) \leq 43/28$ . Le cas de l'égalité se traite facilement.

Le cas où les  $D_t$  sont égaux est traité en [6] page 58 et nous ne détaillerons pas.

### Théorème 11

Si  $||m|| = ||m_t|| = 1$  et si il existe un nombre  $a > 0$  tel que  $m_t(x) \neq 0$  entraîne  $x \geq a$ , et ceci pour tout  $t$ , alors tout temps d'arrêt  $T$  tel que la signature de  $\mathcal{F}_T$  soit plus fine que  $m$  satisfait à

$$E(T) \geq \frac{e(m)}{|\text{Log } a|}$$

### Démonstration

Elle s'appuie naturellement sur "l'inégalité du déménageur", que les conditions du problème rendent non triviales.

Posons  $t(x) = \sup \{t ; x < a^t\}$ . Alors, si  $t \leq t(x)$  on a  $f_{\mu_t}(x) = x$ , et donc

$$\sum_{t=1}^{\infty} f_{\mu_t}(x) \geq x \cdot t(x)$$

Comme  $t(x) \geq \frac{\text{Log } 1/x}{|\text{Log } a|} - 1$ , d'après le théorème 6,  $T$  satisfait à

$$E(T) \geq 1 + \sum_x x \cdot t(x) \cdot m(x) \geq \frac{e(m)}{|\text{Log } a|} .$$

### Théorème 12

Si  $\sup_t x(m_t) = q < 1$ , il existe des constantes  $A$  et  $B$  ne dépendant que de  $q$ , telles que pour toute signature  $m$ , avec  $||m|| = 1$ , il existe un temps d'arrêt  $T$  tel que la signature de  $\mathcal{F}_T$  soit plus fine que  $m$  et satisfaisant à

$$E(T) \leq A + B e(m) .$$

On commence par démontrer le théorème dans un cas particulier, par le lemme suivant :

Lemme

Le théorème 11 est vrai dans le cas où  $m_t (1/D_t) = D_t$  pour tout  $t$ , les  $D_t$  étant entiers  $\geq 2$ . Dans ce cas :

$$E(T) \leq 3 + \frac{e(m)}{\text{Log } d},$$

avec  $d = \inf_t D_t$ .

Démonstration

Posant  $q_t = D_1 \dots D_t$  on a vu au théorème 8 qu'il existait un  $T$  satisfaisant à

$$E(T) = 1 + \sum_x m(x) \sum_{t=1}^{\infty} \{x q_t\} / q_t.$$

Posons  $t(x) = \sup \{t ; x q_t < 1\}$ . Alors :

$$\begin{aligned} \sum_{t=1}^{\infty} \{x q_t\} / q_t &\leq x t(x) + \sum_{t>t(x)} q_t^{-1} \\ &\leq x t(x) + q_{t(x)+1}^{-1} \left[ 1 + D_{t(x)+2}^{-1} + D_{t(x)+2}^{-1} D_{t(x)+3}^{-1} + \dots \right] \\ &\leq x t(x) + 2 q_{t(x)+1}^{-1} \leq x t(x) + 2 x. \end{aligned}$$

Comme  $t(x) \leq (\text{Log } d) (\text{Log } 1/x)$ , on a bien le résultat annoncé.

### Démonstration du théorème

L'art consiste à adapter la démonstration du théorème 9 pour fabriquer un jeu de pile ou face avec le processus, ce qui est possible dans le cas indépendant où nous nous sommes placés. Le temps moyen pour fabriquer un seul jet de pièce sera uniformément borné grâce à la condition  $\sup_t x(\mu_t) \leq q < 1$ . Le jeu de pile ou face, grâce au théorème 8 et au lemme, nous servira ensuite à construire  $m$ , en un temps moyen dépendant linéairement de l'entropie.

Si nous reprenons la démonstration du théorème 9, nous constatons que le procédé de construction de  $T$  fournit l'inégalité

$$P(T > t) \leq x(\mu_t) \sum_x m(x)$$

puisque pour tout  $r$  on avait :

$$P\left(\bigcup_{i>t} D_i^r\right) \leq x(\mu_t)$$

Appliquons ici le théorème 9 au cas où on construit un temps d'arrêt  $T_1$  tel qu'il existe  $A \in \mathcal{F}_{T_1}$  tel que  $P(A) = 1/2$ . Puisque  $x(\mu_t) \leq q^t \xrightarrow[t \rightarrow \infty]{} 0$  on est bien -d'après le théorème 1- dans les conditions d'application du théorème 9. La remarque qui précède montre que l'on peut prendre  $T_1$  tel que

$$E(T_1) \leq 1 + 2 \sum_{t=1}^{\infty} q^t = \frac{1+q}{1-q}$$

Puisqu'on est dans le cas indépendant, nous pouvons répéter le procédé et construire ainsi une suite  $T_1 < T_2 < \dots < T_s < \dots$  de temps

d'arrêt satisfaisant aux conditions suivantes :

1°) Les variables aléatoires  $I_s = T_s - T_{s-1}$ ,  $I_1 = T_1$  sont indépendantes et satisfont à

$$E(I_s) \leq \frac{1+q}{1-q} \text{ pour tout } s$$

2°) Il existe des ensembles  $(A_s)_{s=1}^{\infty}$  indépendants tels que  $A_s \in \mathcal{F}_{T_s}$  et  $P(A_s) = \frac{1}{2}$

Soit  $\mathcal{G}_s$  la tribu engendrée par  $A_1, A_2, \dots, A_s$ . D'après le lemme, il existe un temps d'arrêt  $S$  par rapport à  $(\mathcal{G}_s)_{s=1}^{\infty}$  tel que la signature de  $\mathcal{G}_S$  soit plus fine que  $m$  et tel que :

$$E(S) \leq 3 + \frac{e(m)}{\text{Log } 2}$$

Posant alors  $T = T_S$ , il est immédiat que  $\mathcal{G}_S$  est une sous-tribu de  $\mathcal{F}_T$ . Enfin puisque

$$E(T) \leq \frac{1+q}{1-q} E(S)$$

le théorème se trouve démontré, avec les constantes

$$A = 3 \frac{1+q}{1-q} \text{ et } B = \frac{1}{\text{Log } 2} \cdot \frac{1+q}{1-q} .$$

La proposition suivante peut donner une meilleure estimation que le théorème 11 :

Proposition 7

Si  $\sum_x m(x) = N$  ,  $s = \sup_{t,x} f_{m_t}(x) < 1$  et si le temps d'arrêt  $T$  réalise l'égalité du théorème 6, alors

$$E(T) \leq \frac{1 + (N-1) s}{1-s} .$$

Démonstration

D'après la proposition 4,  $f_{m_t}(x) \leq s^t$  . Le résultat est alors immédiat.

## 7) Codage

Nous allons maintenant éclaircir les analogies entre la théorie précédente et le problème du codage instantané. Afin de justifier les définitions, une description informelle du codage d'un alphabet est nécessaire.

On dispose d'un ensemble fini ou dénombrable  $R$  qui est l'alphabet à coder.  $R$  est muni d'une probabilité  $p$  et  $m$  est la signature de cette probabilité. D'autre part, on dispose d'alphabets  $\Omega_t = \{1, \dots, D_t\}$  de  $D_t$  lettres  $t = 1, 2, \dots$  et on forme  $\Omega = \prod_{t=1}^{\infty} \Omega_t$ . Le mot  $s = (\omega_1, \omega_2, \dots, \omega_t)$ , avec  $\omega_i \in \Omega_i$  est identifié au sous-ensemble de  $\Omega$  défini par  $\{\omega_1\} \times \dots \times \{\omega_t\} \times \Omega_{t+1} \times \dots$ ; on note par  $t = t(s)$  la longueur du mot  $s$ , et l'ensemble des mots de longueur  $t$  engendre la tribu  $\mathcal{F}_t$  dans  $\Omega$ .  $\mathcal{S}$  désigne l'ensemble des mots. Un codage de  $R$  est une application  $C$  de  $R$  dans  $\mathcal{S}$ . Le codage est dit instantané si  $C(r) \subset C(r')$  entraîne  $r = r'$ . La longueur moyenne du codage est le nombre

$$\bar{C} = \sum_t t (C(r)) p(r).$$

### Définition 9

$(\Omega, \mathcal{A})$  étant un espace mesurable, un atome-strict de  $\mathcal{A}$  est un élément  $A$  non vide de  $\mathcal{A}$  tel que  $B \in \mathcal{A}$  et  $B \subset A$  entraîne  $B = A$  ou  $B = \emptyset$ . La tribu  $\mathcal{A}$  est strictement-atomique si elle est engendrée par une famille dénombrable d'atomes tous disjoints.

On se donne pour le reste du paragraphe un espace mesurable  $(\Omega, \mathcal{A})$  et  $(\mathcal{F}_t)_{t=1}^{\infty}$  une famille croissante de sous-tribus de  $\mathcal{A}$  strictement-atomiques qui engendrent  $\mathcal{A}$ . Si  $(\Omega, \mathcal{A})$  est formé par le produit d'espaces  $(\Omega_t, \mathcal{A}_t)_{t=1}^{\infty}$ , les  $\mathcal{F}_t$  seront construits de la manière habituelle à partir des  $\mathcal{A}_t$ ,  $D_t$  ( $\leq \infty$ ) désignera le nombre d'atomes stricts de  $\mathcal{A}_t$ ,  $q_t = D_1 \dots D_t$  celui de  $\mathcal{F}_t$ , et nous nous référerons brièvement à ces hypothèses en parlant du cas indépendant.

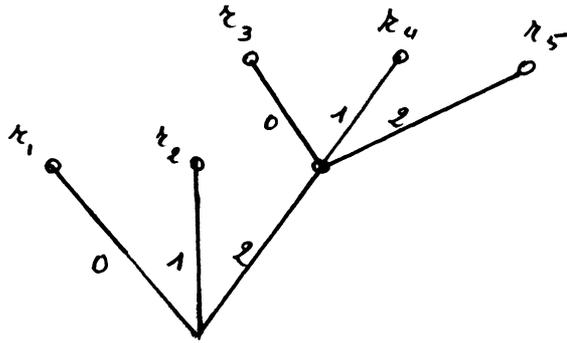
#### Définition 10

Soit  $m$  une signature telle que  $\|m\| = 1$ . On appelle code instantané de  $m$  la donnée

- 1°) d'un temps d'arrêt  $T$  par rapport aux  $(\mathcal{F}_t)_{t=1}^{\infty}$ ,  $T = \infty$  n'étant pas exclu
- 2°) d'une probabilité  $P$  sur  $\mathcal{F}_T$  de signature  $m$  telle que  $P(T < \infty) = 1$ .  $E(T)$  est la longueur moyenne du code.

Quelques commentaires sont nécessaires :

- 1) Les  $\mathcal{F}_t$  ne sont pas probabilisés
- 2) La définition réserve la possibilité d'avoir des nombres de "branches" différents, mais infinis à chaque noeud de l'arbre.
- 3) Dans le cas indépendant, les "sommets" non utilisés sont simplement affectés de la probabilité zéro. Ainsi si  $D_t = 3$ ,  $t = 1, 2, \dots$   $m\left(\frac{1}{4}\right) = 4$ , le codage est décrit par l'arbre suivant :



Ici  $P(r_i) = \frac{1}{4}$  si  $i = 1, 2, 3, 4$  et  $P(r_5) = 0$

$T(\omega) = 1$  si  $\omega \in \{0, 1\} \times \Omega_2 \times \Omega_3 \times \dots$

$T(\omega) = 2$  si  $\omega \in \{2\} \times \Omega_2 \times \Omega_3 \times \dots$

4) Nous excluons la possibilité de codage continu. En effet si  $\|m\| < 1$ , ou bien on permet à un  $\mathcal{F}_t$  de ne pas être strictement atomique, ce qui permet de construire la partie continue, ou bien on aura  $T = \infty$  avec une probabilité positive.

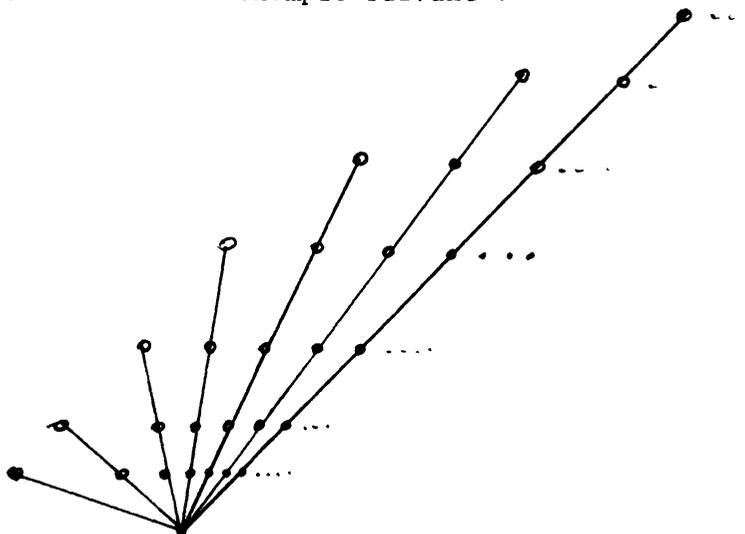
Nous énonçons une proposition qui permet de mieux comprendre le rôle joué par  $\{T = \infty\}$  dans la définition.

Proposition 8

On suppose que les  $\mathcal{F}_t$  n'ont qu'un nombre fini d'atomes. Soit  $T$  un temps d'arrêt par rapport aux  $(\mathcal{F}_t)_{t=1}^{\infty}$  tel que  $\{T = +\infty\} = \emptyset$ . Alors il existe une constante  $a$  telle que  $\{T > a\} = +\emptyset$  et  $\mathcal{F}_T$  n'a qu'un nombre fini d'atomes.

Remarque

La restriction concernant le nombre fini d'atomes des  $\mathcal{F}_t$  est essentielle, comme le montre l'exemple suivant :



Cette proposition se démontre commodément en munissant  $\Omega$  de la topologie la moins fine telle que les éléments de  $\bigcup_{t=1}^{\infty} \mathcal{F}_t$  soient ouverts. L'hypothèse de finitude des  $\mathcal{F}_t$  entraîne que  $\Omega$  est quasi-compact. Reste à observer que tout temps d'arrêt  $T$  tel que  $\{T = \infty\} = \emptyset$  est une fonction continue sur  $\Omega$ , donc bornée si  $\Omega$  est quasi-compact.

Théorème 13

Si  $(\Omega, \mathcal{A})$  est muni d'une probabilité  $P_1$ , et si  $T$  est un temps d'arrêt tel que la signature de  $\mathcal{F}_T$  soit plus fine que  $m$ , alors il existe une probabilité  $P$  sur  $\mathcal{F}_T$  telle que  $(T, P)$  forme un code instantané de  $m$ .

### Démonstration

Soit  $R = \{(x, i) ; m(x) \neq 0, i = 1 \dots m(x)\}$ .  $x_r$  est défini sur  $R$  par  $r = (x_r, i)$ . Soit  $(A(r))_{r \in R}$  une partition de  $\Omega$   $\mathcal{F}_T$  mesurable avec  $P_1(A(r)) = x_r$ . Pour tout  $r$ , définissons :

$$t(r) = \inf \{t ; A(r) \cap \{T_1 = t\} \neq \emptyset\}.$$

Choisissons  $C(r)$  comme un atome de  $\mathcal{F}_{t(r)}$  contenu dans  $A(r) \cap \{T = t(r)\}$

Il est clair que  $T(\omega)$  est un temps d'arrêt. On munit alors  $\mathcal{F}_T$  de la probabilité  $P$  suivante :

$$P(C(r)) = x_r, P\left(\bigcup_r C(r)\right) = 0$$

### Corollaire 1

Dans le cas indépendant  $(D_t)_{t=1}^{\infty}$ , il existe un code instantané tel que

$$E(T) = \sum_x t(x) x m(x)$$

où  $t(x) = \inf \{t ; x q_t \geq 1\}$

### Démonstration

Il suffit en effet de probabiliser  $(\Omega, \mathcal{A})$  en prenant sur  $(\Omega_t, \mathcal{A}_t)$  la probabilité de signature  $m_t(1/D_t) = D_t$ . D'après le théorème 8, il

existe un temps d'arrêt  $T$  permettant de construire  $m$  tel que les  $A(r) \cap (T \leq t)$  sont de probabilité  $\frac{[x_r q_t]}{q_t}$ . Donc  $A(r) \cap (T = t)$  est non vide si  $t = t(x_r)$ , ce qui démontre le corollaire.

### Corollaire 2

Dans le cas indépendant, il existe un code instantané tel que

$$\frac{e(m)}{\text{Log } D} \leq E(T) < \frac{e(m)}{\text{Log } D} + 1$$

lorsque  $D_t = D$  pour tout  $t$ .

### Démonstration

Il suffit d'appliquer le corollaire précédent, en remarquant que

$$t(x) \geq \frac{\text{Log } \frac{1}{x}}{\text{Log } D} > t(x) - 1$$

Le corollaire 2 montre que l'approche "probabilité" que nous donnons du problème du codage instantané est bonne, puisque un théorème de Shannon [1] montre que le meilleur codage est situé dans les bornes du corollaire 2. Dans le cas d'une signature finie et de  $D_t$  constants. Le meilleur codage, celui qui minimise  $E(T)$  n'est pas donné par notre algorithme mais celui d'Huffmann. Cependant celui-ci est inopérant

1°) dans le cas de  $D_t$  distincts

2°) dans le cas où  $\sum_x m(x) = +\infty$ . D'ailleurs, comme dans le cas du

théorème 9, il n'est pas clair qu'il existe un  $T$  qui atteigne la borne inférieure des  $E(T)$ .

8) Complément : la version dénombrable d'un théorème de Banach

Etant donnés des nombres  $(p_t)_{t=1}^{\infty}$  de  $(0, 1)$ , il est connu qu'on peut trouver une probabilité  $P$  sur  $N$  et des parties  $A_n \subset N$  telles que  $p(A_t) = p$  et telles que les évènements  $(A_t)_{t=1}^{\infty}$  soient indépendants si et seulement si  $\sum_{t=1}^{\infty} \min(p_t, 1 - p_t) < \infty$ . Le théorème 3 nous permet de généraliser sans difficulté ce résultat. Plus précisément :

Théorème 14

Soient  $(m_t)_{t=1}^{\infty}$  des signatures de masse unité. Il existe une probabilité  $P$  sur  $N$  et des tribus  $(\mathcal{A}_t)_{n=1}^{\infty}$  sur  $N$  indépendantes et de signatures respectives  $(m_t)_{t=1}^{\infty}$  si et seulement si  $\prod_{t=1}^{\infty} x(m_t) > 0$ .

Démonstration

Supposons  $\prod_{t=1}^{\infty} x(m_t) > 0$ . On construit les espaces  $(\Omega_t, \mathcal{A}_t, P_t)_{t=1}^{\infty}$  de signatures respectives  $m_t$  et le produit  $(\Omega, \mathcal{A}, P)$  de ceux-ci. La signature de ce dernier est  $m = \sum_{t=1}^{\infty} m_t$  d'après le théorème 3 et est de masse unité. Il suffit ensuite d'énumérer les atomes de  $\mathcal{A}$  pour terminer.

Inversement, supposons que  $N$  soit probabilisé et que les  $\mathcal{A}_t$  soient des tribus indépendantes sur  $N$  de signature  $m_t$ . Soit  $n \in N$  fixé de masse positive et  $A_t$  l'atome de  $\mathcal{A}_t$  qui le contient. Donc

$$\{n\} \subset \bigcap_{t=1}^N A_t$$

Donc  $P(\{n\}) \leq \prod_{t=1}^N P(A_t) \leq \prod_{t=1}^N x(m_t)$  ce qui montre que  $\prod_{t=1}^{\infty} x(m_t) > 0$ .

Ce théorème suggère d'étendre à  $\mathbb{N}$  le théorème suivant de S. Banach [2] :

### Théorème

Soient  $(\Omega, \mathcal{A}_t, P_t)_{t=1}^{\infty}$  des espaces de probabilité sur un même  $\Omega$ ,  $\mathcal{A}$  la tribu engendrée par les  $\mathcal{A}_t$ . Si les  $\mathcal{A}_t$  sont tels que  $\emptyset \neq A_t \in \mathcal{A}_t$  entraîne  $\bigcap_t A_t \neq \emptyset$ , alors il existe une probabilité  $P$  unique sur  $(\Omega, \mathcal{A})$  telle que  $P|_{\mathcal{A}_t} = P_t$  et telle que les  $\mathcal{A}_t$  soient indépendantes.

Il n'est pas question d'appliquer ce théorème dans le cas où  $\Omega = \mathbb{N}$  si les  $\mathcal{A}_t$  ne sont pas triviaux sauf pour un nombre fini. En effet la condition  $\bigcap_t A_t \neq \emptyset$  si  $\emptyset \neq A_t \in \mathcal{A}_t$  contredirait la dénombrabilité. Aussi avons nous le théorème suivant, dont le corollaire sera la version dénombrable cherchée.

### Théorème 15

Soient  $(\mathbb{N}, \mathcal{F}_t, P_t)_{t=1}^{\infty}$  des espaces de probabilité sur  $\mathbb{N}$  tels que  $(\mathcal{F}_t)_{t=1}^{\infty}$  forme une suite croissante de tribus qui engendrent la tribu  $\mathcal{A}$ , et tels que  $P_{t+1}|_{\mathcal{F}_t} = P_t$ . Soient  $(\mu_t)_{t=1}^{\infty}$  leurs signatures respectives. Alors il existe une probabilité  $P$  sur  $(\mathbb{N}, \mathcal{A})$  telle que  $P|_{\mathcal{F}_t} = P_t$  si et seulement si  $\mu_t$  converge vers une signature  $\mu$  telle que  $\|\mu\| = 1$ ;  $\mu$  est alors la signature de  $P$ .

### Corollaire

Soient  $(\mathbb{N}, \mathcal{A}_t, P_t)_{t=1}^{\infty}$  des espaces de probabilité sur  $\mathbb{N}$  de signatures respectives  $(\mu_t)_{t=1}^{\infty}$  et  $\mathcal{A}$  la tribu engendrée par les  $\mathcal{A}_t$ . On suppose que les  $\mathcal{A}_t$  sont tels que  $\bigcap_{t=1}^n A_t \neq \emptyset$  pour tout  $n$  si

$\emptyset \neq A_t \in \mathcal{A}_t$ . Alors il existe une probabilité  $P$  sur  $(N, \mathcal{A})$  telle que  $P|_{\mathcal{A}_t} = P_t$  et telle que les  $\mathcal{A}_t$  soient indépendantes, si et seulement si  $\prod_{t=1}^{\infty} \mu_t > 0$ .

### Démonstration du théorème

Si  $P$  existe, la conclusion du théorème est la conséquence immédiate du théorème 1. Inversement supposons que  $\mu_t$  converge vers une signature  $\mu$  de masse unité.

On considère alors l'algèbre  $\mathcal{F} = \bigcup_{t=1}^{\infty} \mathcal{F}_t$  et on définit sur  $\mathcal{F}$  la fonction simplement additive  $P(A) = P_t(A)$  si  $A \in \mathcal{F}_t$ . Nous avons donc à démontrer que si  $A_t \in \mathcal{F}$  et  $A_t \supset A_{t+1}$  pour tout  $t = 1, 2, \dots$ , alors  $\bigcap_t A_t = \emptyset$  entraîne  $\lim_t P(A_t) = 0$ . Le théorème d'extension de Kolmogoroff permettra alors d'étendre  $P$  à  $\mathcal{A}$  et il est clair que cette extension aura les propriétés requises.

Supposons donc que  $\bigcap_t A_t = \emptyset$  et qu'il existe une constante  $a$  telle que :

$$P(A_t) \geq a > 0.$$

Sans perte de généralité, on peut supposer  $A_t \in \mathcal{F}_t$ . Soit  $B_t$  l'atome de grande masse contenu dans  $A_t$ ; alors  $P_t(B_t) \xrightarrow[t \rightarrow \infty]{} 0$  car sinon il existerait une constante  $b$  telle que :

$$P(B_t) \geq b > 0 \text{ pour tout } t$$

Comme  $\mathcal{F}_t$  n'a qu'un nombre fini d'atomes  $\geq b > 0$ , un argument de compacité montre qu'alors  $\bigcap_{t=1}^{\infty} A_t$  ne serait pas vide. Maintenant, le

fait que  $||\lim_t \mu_t|| = 1$  entraîne que pour tout  $\varepsilon > 0$  il existe  $\eta(\varepsilon)$  tel que :

$$\sum_{x < \eta(\varepsilon)} x \mu_t(x) < \varepsilon$$

pour tout  $t = 1, 2, \dots$

Prenons alors  $\varepsilon = a$  et  $t$  assez grand pour que  $P(B_t) \leq \eta(a)$ .

Alors l'inégalité :

$$P(A_t) \leq \sum_{x < \eta(a)} x \mu_t(x) < a$$

fournit la contradiction cherchée.

#### Démonstration du corollaire

Soit  $\mathcal{F}_t$  la tribu engendrée par  $\mathcal{A}_1, \dots, \mathcal{A}_t$ . D'après le théorème de Banach cité ci-dessus et appliqué au cas fini, il existe une probabilité  $P'_t$  sur  $\mathcal{F}_t$  telle que  $P'_t | \mathcal{A}_s = P_s$  si  $s \leq t$  et telle que  $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_t$  soient indépendantes ; et en vertu de l'unicité on a  $P'_{t+1} | \mathcal{F}_t = P'_t$ . Soit  $\mu_t$  la signature de  $\mathcal{F}_t$ . Alors

$$\mu_t = m_1 * \dots * m_t$$

(on adapte en effet facilement la démonstration du lemme du théorème pour montrer que si  $\mathcal{A}$  et  $\mathcal{B}$  sont deux tribus indépendantes engendrant  $\mathcal{C}$ , les atomes de  $\mathcal{C}$  sont de la forme  $A \cap B$ , où  $A$  et  $B$  sont des atomes de  $\mathcal{A}$  et  $\mathcal{B}$ , et la signature de  $\mathcal{C}$  est la convolution des signatures de  $\mathcal{A}$  et  $\mathcal{B}$ ). Le théorème 15 donne alors le corollaire .

### Remarque

Si la condition  $||\lim_t \mu_t|| = 1$  n'est pas remplie, que se passe-t-il ?  
Pour récupérer la masse perdue, il est nécessaire de compléter  $N$  par une frontière. Plus précisément on désigne par  $\hat{N}$  l'ensemble de toutes les suites décroissantes d'atomes  $(A_t)_{t=1}^\infty$ , avec  $A_t \in \mathcal{F}_t$ .  $\mathcal{F}_t$  et  $P_t$  s'étendent canoniquement à  $\hat{N}$ , ainsi que  $N$ , et on peut démontrer l'existence d'une probabilité  $P$  sur  $\hat{N}$  ayant les propriétés annoncées au théorème 15. Les atomes de  $P$  sont situés dans  $N$  et la signature de  $P$  reste  $\lim_t \mu_t$ .

## Bibliographie du Chapitre I

- [1] ASH Robert  
"Information theory" chapitre 2  
Interscience Wiley (New-York) 1965
  
- [2] BANACH Stefan  
"On measures of independent fields"  
Studia Mathematica, vol. 10 (1948) pp. 159-181
  
- [3] DUBINS Lester E.  
"On a theorem of Skorohod"  
Annals of Mathematical Statistics, vol. 39 (1968) pp. 2094-2097
  
- [4] VAN KAMPEN  
"Infinite products measures and infinite convolutions"  
American Journal of Mathematics, vol. 62 (1940) pp. 417-448
  
- [5] KOLMOGOROV A.  
"Sur les propriétés des fonctions de concentration de M. Paul Lévy"  
Annales de l'Institut Henri Poincaré, XVI, vol. 1, pp. 27-34 (1958)
  
- [6] LETAC Gérard  
"Problèmes de probabilité". Problème n° 19 - P.U.F. (1970)
  
- [7] LUBELL David  
"A short proof of Sperner lemma"  
Journal of combinatorial theory, vol. 1 (1966) pp. 299
  
- [8] MEYER Paul-André  
"Séminaire de probabilités IV de l'Université de Strasbourg" (pp.134)  
Lectures Notes in Mathematics n° 124 - SPRINGER VERLAG, Berlin (1970)

- [9] RENYI Alfred  
"Calcul des Probabilités" page 71, DUNOD (Paris) 1966
- [10] RENYI Alfred  
"Foundations of probability" page 167  
Holden Day, San Francisco (1970)
- [11] ROHRLIN V.A.  
"On the fundamental ideas of measure theory"  
In "Translations, série 1, vol. 10, Functional analysis and measure  
theory"  
A.M.S. Rhode Island (1962)
- [12] SKOROKHOD A.V.  
"Studies in the theory of random processes" page 163  
Addison-Wesley (Reading) (1965)

## CHAPITRE II

---

"Reconstruction de probabilités à partir de processus  
mal connus ; problème de Von Neumann, polynomes et  
coefficients multinomiaux"

Ce chapitre reprend les thèmes et les énoncés (mais aucune des démonstrations) de l'article "Construction d'évènements équiprobables et coefficients multinomiaux modulo  $p^n$ " de Jacques Bernard et l'auteur, à paraître à l'Illinois Journal of Mathematics.

Il complète l'article précité sur divers points ; les preuves des résultats nouveaux sont données. Les directions étant diverses, chaque paragraphe contient sa propre introduction.

## 1) Problème de Von Neumann

La démarche initiale de ce chapitre est très proche de celle du chapitre I : supposons qu'on dispose d'une pièce de monnaie assez irrégulière pour qu'on ignore la probabilité de ses faces. Comment choisir entre deux personnes de manière rigoureusement équiprobable lorsqu'on ne dispose que de cette pièce ? Une réponse a été donnée en 1951 par John Von Neumann : il suffit de jeter la pièce un nombre pair de fois, disons  $T$ , où  $T$  est le premier nombre pair tel que les jets aux instants  $T - 1$  et  $T$  donnent des résultats différents. Alors les résultats Pile - Face et Face - Pile aux instants  $T - 1$  et  $T$  sont exactement de probabilité  $1/2$ .

La connaissance parfaite de l'appareil aléatoire que nous étions censés avoir au chapitre I est ici remplacée par l'hypothèse : les expériences successives, bien que mal connues, sont rigoureusement de même loi et indépendantes.

On peut formuler le problème de Von Neumann de la façon suivante : soient  $(\Omega_t, \mathcal{A}_t, P_t)_{t=1}^{\infty}$  des espaces de probabilité tous identiques,  $(\Omega, \mathcal{A}, P)$  leur produit. On prend une sous-tribu  $\mathcal{J}$  de  $\mathcal{A}_1$  engendrée par une partition  $(A_i)_{i \in I}$  de  $\Omega_1$  où  $I$  est un ensemble fini ou dénombrable. Notre connaissance de  $P$  se résume à l'hypothèse suivante :

$$\pi_i = P_1(A_i) < 1 \text{ pour tout } i \text{ de } I \quad (H)$$

### Définition

Un évènement universel est un élément  $A$  de  $\mathcal{A}$  tel que  $P(A)$  soit indépendant de  $P$  lorsque celui-ci satisfait (H).  $\mathcal{A}^u$  désigne la tribu

des évènements universels,  $P$  la probabilité canonique sur  $\mathcal{A}^u$ .

(L'exemple de Von Neumann précédemment cité montre que ces définitions ne sont pas vides !).

On désigne par  $\mathcal{F}_t$  la tribu habituelle du passé à l'instant  $t$ , produit des tribus  $\mathcal{A}_1, \dots, \mathcal{A}_t$  et des tribus triviales sur  $\Omega_{t+1}, \Omega_{t+2}, \dots$ . Notre problème est le suivant : étant donnée une signature  $m$  quelconque de masse 1 (c'est-à-dire sans partie continue ; voir le chapitre I), trouver un temps d'arrêt  $T$  par rapport aux  $(\mathcal{F}_t)_{t=1}^\infty$  tel que  $P(T < \infty) = 1$  pour tout  $P$  satisfaisant à (H) et tel que la signature de la tribu  $\mathcal{F}_T \cap \mathcal{A}^u$  soit plus fine que  $m$ .

Sans perte de généralité, on voit qu'on peut identifier  $I$  avec  $\Omega_1 = \dots = \Omega_t = \dots$  et  $\mathcal{A}_t$  avec la tribu la plus fine, ce que nous supposerons désormais. Nous avons la proposition suivante, dont la démonstration est immédiate :

#### Proposition 1

$\mathcal{A}^u \cap \mathcal{F}_t$  ne contient que des évènements de probabilité 0 ou 1.

Introduisons quelques nouvelles définitions.

$\mathcal{G}$  est le groupe abélien libre engendré par  $I$ , identifié aux suites d'entiers  $x = (x_i)_{i \in I}$  telles que  $x_i \neq 0$  pour un nombre fini de  $i$  seulement ;  $\mathcal{M}$  est le monoïde abélien libre engendré par  $I$ , identifié aux suites d'entiers non négatifs  $x = (x_i)_{i \in I}$  de  $\mathcal{G}$ . On définit en particulier  $\varepsilon^j$  dans  $\mathcal{M}$  avec  $j \in I$  par  $(\varepsilon^j)_i = 1$  si  $i = j$  et  $(\varepsilon^j)_i = 0$  si  $i \neq j$ . Si  $X \in \mathcal{G}$ , on pose  $|x| = \sum_i x_i$  et  $c(x) = 0$

si  $x \notin \mathcal{M}$ ,  $c(x) = \frac{|x|!}{\prod_i (x_i!)}$  si  $x \in \mathcal{M}$ .

Si  $(X_i)_{i \in I}$  est un ensemble de variables formelles, on pose

$X^x = \prod_i X_i^{x_i}$  si  $x \in \mathcal{M}$ . En particulier si  $\pi_i = P_1(A_i)$ , on pose

$\pi^x = \prod_i \pi_i^{x_i}$ . Les atomes-stricts de  $\mathcal{F}_t$  (voir chapitre I, § 7) s'iden-

tifient aux mots non abéliens de longueur  $t$  sur  $I$ , et on désigne par

$\gamma$  l'application canonique des mots non abéliens dans  $\mathcal{M}$ .

### Proposition 2

Soient  $A$  et  $B$  dans  $\mathcal{F}_t$  tels que  $P(A) = P(B)$  pour tout  $P$ . Alors pour tout  $x \in \mathcal{M}$  le nombre d'éléments de  $\gamma^{-1}(x) \cap A$  et  $\gamma^{-1}(x) \cap B$  est le même.

### Démonstration

Soient  $n(x)$  et  $m(x)$  ces deux nombres. Ils sont inférieurs ou égaux à  $c(x)$  naturellement.  $P(A) = P(B)$  entraîne

$$\sum_{x: |x|=t} (n(x) - m(x)) \pi^x = 0 \quad \text{pour tout } \pi = (\pi_i)_{i \in I} \text{ tel}$$

que  $\pi_i \geq 0$  pour tout  $i$  et  $\sum_i \pi_i = 1$ .

Il est immédiat que si la proposition est vraie pour  $\sum \pi_i = 1$  elle est encore vraie pour  $\sum \pi_i < 1$ . Distinguons  $1$  dans  $I$ . On peut écrire :

$$0 = \sum_{x: |x|=t} (n(x) - m(x)) \pi^x = \sum_{x_1=0}^t \pi_1^{x_1} \sum (n(x'+x_1 \varepsilon^1) - m(x'+x_1 \varepsilon^1)) \pi^{x'}$$

où la dernière somme est prise pour les  $x'$  de  $\mathcal{M}$  tels que

$$x'_1 = 0 \quad \text{et} \quad |x' + x_1 \varepsilon^1| = t.$$

La formule (1), compte tenu de la remarque précédente permet sans difficulté de démontrer la proposition par récurrence sur  $t$ .

La proposition 1 montre qu'à priori les temps d'arrêt dont nous avons besoin ne peuvent être bornés. La proposition 2 nous met sur la voie d'une construction possible en utilisant le fait que deux "chemins" aboutissant au même  $x$  de  $\mathcal{M}$  ont la même probabilité. Plus précisément, on considère l'ensemble des variables aléatoires  $(y_t)_{t \geq 1}$  à valeurs dans  $\mathcal{M}$  définies sur  $\Omega$  par

$$y_t(\omega) = \varepsilon^i \quad \text{si la } t^{\text{ième}} \text{ composante de } \omega \text{ est } i \in I.$$
 Les  $y_t$  sont indépendantes et de même loi. On pose  $S_t = y_1 + \dots + y_t$ , avec  $S_0 = 0$ . Il est clair que

$$\begin{aligned} P(S_t = x) &= c(x) \pi^x && \text{si } |x| = t \\ &= 0 && \text{si } |x| \neq t. \end{aligned}$$

Tentons de trouver des  $T$  correspondant au problème de Von Neumann parmi les temps de frappe  $T(E)$  de  $E \subset \mathcal{M}$  par la promenade  $(S_t)_{t=1}^{\infty}$ . Il faudra d'abord que  $E$  soit "universellement frappé", c'est-à-dire que  $P(T(E) < \infty) = 1$  pour tout  $P$  satisfaisant à la condition (H). Désignons par  $S'_t$  le processus stoppé, c'est-à-dire

$$\begin{aligned} S'_t &= S_t && \text{si } t < T(E) \\ S'_t &= S_{T(E)} && \text{si } t \geq T(E), \end{aligned}$$

on définit  $m_E(x)$  dans  $\mathcal{M}$  par

$$P(S'_t = x) = m_E(x) \pi^x.$$

Il est clair que  $m_E$  est indépendant de  $\pi$ . Si on désigne par  $\mathcal{Y}(E)$  l'ensemble des points  $x$  de  $E$  tel que  $P(S'_t = x) > 0$ , on a la caractérisation suivante de l'image de la probabilité  $P^u$  sur  $\mathcal{F}_{T(E)} \cap \mathcal{A}^u$  :

Proposition 3

Les valeurs prises par la probabilité  $P^u$  sur  $\mathcal{F}_{T(E)} \cap \mathcal{A}^u$  sont les nombres de la forme  $\sum_{x \in \mathcal{Y}(E)} s(x) \pi^x$ , où  $s(x)$  est un entier défini pour tout  $x$  de  $\mathcal{Y}(E)$  tel que

1°)  $0 \leq s(x) \leq m_E(x)$

2°)  $\sum_{x \in \mathcal{Y}(E)} s(x) \pi^x$  soit indépendant de  $\pi$ .

Démonstration

Soit  $A \in \mathcal{F}_{T(E)} \cap \mathcal{A}^u$ . Alors  $A$  s'écrit comme une réunion finie ou dénombrable d'atomes  $(A_k)_{k \in K}$  de la tribu  $\mathcal{F}_{T(E)}$ . Or, comme on l'a vu au chapitre I,  $A_k$  est un atome de  $\mathcal{F}_t$  contenu dans  $(T = t)$  pour quelque  $t = t_k$ . Donc  $P(A_k) = \pi^x$  pour un certain  $x \in \mathcal{Y}(E)$  dépendant de  $k$ , avec  $|x| = t_k$ . Désignant par  $s(x)$  le nombre de  $A_k$  ainsi associés à  $x$  de  $\mathcal{Y}(E)$ , on voit que

$$P(A) = \sum_{x \in \mathcal{Y}(E)} s(x) \pi^x.$$

Les  $A_k$  étant disjoints, on a bien  $s(x) \leq m_E(x)$ . La deuxième condition résulte de l'appartenance de  $A$  à  $\mathcal{A}^u$ .

Cette proposition montre que les signatures possibles de  $\mathcal{F}_{T(E)} \cap \mathcal{A}^u$  sont soumises à de sévères restrictions. Aussi pour

construire une signature quelconque prendrons nous une procédure mixte : construire un parfait jeu de pile ou face ou son équivalent avec  $D$  possibilités, à partir duquel on pourra appliquer les procédures du chapitre I.

Nous cherchons donc à trouver des ensembles  $E$  de  $\mathcal{M}$  tels que la signature de  $\mathcal{F}_{T(E)} \cap \sigma^u$  soit plus fine que la signature  $m_D$  définie par  $m(1/D) = D$ , où  $D$  est un entier, en cherchant à diminuer  $E$  ( $T(E)$ ) le plus possible.

On doit à Wassily Hoeffding et Gordon Simons [12] l'idée extraordinaire de chercher des ensembles  $E$  de  $N^2$  tels que l'entier  $m_E(x)$  correspondant soit toujours divisible par 2, et la découverte d'un ensemble  $E$  maximal ayant cette propriété. J'ai eu la chance d'intéresser Jacques Bernard à ce problème et cette fructueuse collaboration nous a permis de mieux comprendre et de généraliser les méthodes de [12].

Considérons en particulier le théorème suivant :

#### Théorème 1

Si  $E \subset \mathcal{M}$ , alors les trois affirmations sont équivalentes :

- a)  $m_E(e) = 0 \pmod D$  si  $x \in E$
- b)  $c(x) = 0 \pmod D$  si  $x \in E$
- c)  $m_E(x) = c(x) \pmod D$  pour tout  $x$  de  $\mathcal{M}$ .

La démonstration est une simple généralisation de [1].

Avons nous ainsi résolu, au vu de la proposition 3, le problème de la construction de  $m_D$ ? Pas tout à fait, car le point délicat est de

trouver un tel ensemble  $E$  tel que  $P(T(E) < \infty)$  pour tout  $P$  satisfaisant à l'hypothèse  $H$ . Si cela est fait, définissant

$$H_D = \{x ; c(x) \equiv 0 \pmod{D}\},$$

puisque  $E \subset H_D$ , on a  $P(T(H_D) < \infty)$  pour tout  $P$ , et  $H_D$  est le meilleur ensemble possible.

Pour trouver cet  $E$ , nous allons au § 2 découvrir ou redécouvrir des propriétés des coefficients multinomiaux  $c(x)$ , et chemin faisant des polynomes multinomiaux. Nous reviendrons à l'espérance de  $T(H_D)$  au § 3.

## 2) Congruences des nombres et polynomes multinomiaux

Nous rassemblons dans ce paragraphe quelques propriétés des  $c(x)$  et de leurs analogues polynomiaux. Seule, une partie d'entre elles sera réellement utilisée dans la suite, en particulier, pour la solution au problème de Von Neumann au § 3. Nous conservons les notations du § 1. De plus, on se fixe un nombre premier  $p$ . Dans les notations suivantes, qui s'y rapportent, on se dispense de faire figurer l'indice  $p$ .

Si  $n$  est un entier  $\geq 0$ , on pose :

$$n = \sum_{\alpha \geq 0} h_{\alpha}(n) p^{\alpha}, \text{ avec } 0 \leq h_{\alpha}(n) < p \text{ et } h_{\alpha}(n) \text{ entier.}$$

$$v(n) = \inf \{ \alpha : h_{\alpha}(n) > 0 \}, \text{ avec } v(0) = +\infty$$

$$d(n) = \sup \{ \alpha : h_{\alpha}(n) > 0 \}, \text{ avec } d(0) = -\infty$$

$$\text{et } ||n|| = \sum_{\alpha \geq 0} h_{\alpha}(n).$$

Recensons, dans la proposition suivante, quelques faits importants :

### Proposition 4

$$a) \quad ||pn|| = ||n||$$

$$b) \quad ||n|| \leq (d(n) - v(n) + 1)(p - 1) \quad \text{si } n \neq 0$$

$$c) \quad ||n - 1|| = (p - 1)v(n) + ||n - p^{v(n)}||$$

$$d) \quad (p - 1)v(n!) = n - ||n||$$

$$e) \quad ||n + m|| \leq ||n|| + ||m||$$

La démonstration de cette proposition est facile et peut être trouvée dans [1].

Si  $x \in \mathcal{M}$ , on définit  $h_\alpha(x)$  dans  $\mathcal{M}$  par  $(h_\alpha(x))_i = h_\alpha(x_i)$

$$v(x) = \inf_i v(x_i)$$

$$d(x) = \sup_i d(x_i)$$

$$\text{et } ||x|| = || |x| ||$$

D'après la proposition 4, d) on a :

$$(p - 1) v(c(x)) = - ||x|| + \sum_i ||x_i|| \quad (1)$$

Introduisons les entiers  $q_0(x), \dots, q_\alpha(x), \dots$  définis par

$$p q_0(x) = |h_0(x)| - h_0(|x|)$$

$$p q_\alpha(x) = q_{\alpha-1}(x) + |h_\alpha(x)| - h_\alpha(|x|) \text{ si } \alpha \geq 1.$$

Un peu de réflexion montre que les  $q_\alpha(x)$  sont des entiers  $\geq 0$  et qu'ils constituent la "retenue" effectuée dans la colonne  $\alpha$  lorsqu'on effectue l'addition des  $x_i$  en base  $p$ .

Voici maintenant l'énoncé du théorème de Kummer :

### Théorème 2

$$v(c(x)) = \sum_{\alpha \geq 0} q_\alpha(x) \quad (\text{voir [1] pour la démonstration}).$$

### Corollaire

$c(x) \equiv 0 \pmod p$  si et seulement si  $|h_\alpha(x)| < p$   
pour tout  $\alpha \geq 0$ .

Nous allons maintenant démontrer l'analogue du théorème de Kummer pour les polynômes multinomiaux. On introduit, si  $x \in \mathcal{M}$  le polynôme en  $X$  défini par :

$$C_X(x) = \frac{(X^{|x|} - 1) (X^{|x|-1} - 1) \dots (X - 1)}{\prod_{i \in I} (X^{x_i} - 1) (X^{x_i-1} - 1) \dots (X - 1)}$$

appelé polynôme multinomial lié à  $x$ . On convient  $C_X(0) = 1$  ;  
il est immédiat que  $C_1(x) = c(x)$ . Il s'agit effectivement d'un  
polynôme, ce qu'on pourrait voir par examen des racines des numé-  
rateurs et dénominateurs. Mais il est plus intéressant de le voir  
- exactement comme pour les coefficients multinomiaux, dont il n'est  
pas évident qu'ils sont entiers - en donnant une interprétation com-  
binatoire ou probabiliste de  $C_X(x)$ . Cette interprétation combinatoire  
montre de plus que les coefficients de  $C_X(x)$  sont entiers non négat-  
tifs. Pour  $|I| = 2$ , ceci a été fait par de nombreux auteurs, depuis  
Sylvester [11] jusqu'à Georges Polya très récemment [10] . Pour  
 $|I| > 2$  , ceci est fait par P. Mac Mahon [ 8 ] et par l'auteur [7] ,  
et presque simultanément par L. Carlitz [2] et Louis Comtet [3] .  
Nous reviendrons sur cette interprétation combinatoire au § 4.

Le nombre premier  $p$  étant fixé, nous considérons également les  
polynômes  $R_0(X) = 1 + X + X^2 + \dots + X^{p-1}$  et  $R_\alpha(X) = R_0(X^{p^\alpha})$ .  
Nous avons le théorème suivant :

Théorème 3

Le polynôme  $C_X(x)$  est divisible par le polynôme  $\prod_{\alpha \geq 0} [R_\alpha(X)]^{q_\alpha(x)}$  .

Le quotient est à coefficients entiers et n'est divisible par aucun  
polynôme  $R_\alpha(X)$ .

### Démonstration

On commence par vérifier que  $\frac{X^k - 1}{X - 1}$  est divisible par

$\prod_{\alpha < v(k)} R_\alpha(X) = \frac{X^{p^{v(n)}} - 1}{X - 1}$ . On observe ensuite que le nombre

de multiples de  $p^\alpha$  inférieurs ou égaux à  $n$  est exactement

$$a_\alpha(n) = \frac{1}{p^\alpha} \left[ n - \sum_{\beta < \alpha} h_\beta(n) p^\beta \right], \text{ ce qui montre que}$$

$\frac{X^n - 1}{X - 1} \cdot \frac{X^{n-1} - 1}{X - 1} \cdots \frac{X^2 - 1}{X - 1}$  est divisible par  $\prod_{\alpha} R_\alpha^{a_\alpha(n)}(X)$ .

Reste à vérifier que  $q_\alpha(x) + \sum_i a_\alpha(x_i) = a_\alpha(|x|)$ . On procède

pour cela par récurrence sur  $\alpha$ . C'est évident pour  $\alpha = 0$ .

Supposant la propriété vraie pour  $\alpha - 1$ , on écrit :

$$pq_\alpha(x) = q_{\alpha-1}(x) + |h_\alpha(x)| - h_\alpha(|x|) =$$

$$a_{\alpha-1}(|x|) - h_\alpha(|x|) - \sum_i (a_{\alpha-1}(x_i) - h_\alpha(x_i))$$

comme  $a_\alpha(n) = \sum_{\beta \geq \alpha} h_\beta(n) p^{\beta-\alpha}$ , on voit immédiatement que

$a_{\alpha-1}(n) - h_\alpha(n) = p a_\alpha(n)$ , ce qui achève la preuve de la

première partie du théorème. Le fait que le quotient  $Q$  soit

à coefficients entiers est garanti par le fait que les  $R_\alpha$  sont

des polynomes moniques. D'après le théorème de Kummer, l'entier

obtenu en remplaçant  $X$  par  $1$  dans ce quotient est premier avec  $p$ .

Si  $Q(X)$  était divisible par un  $R_\alpha(X)$ ,  $Q / R_\alpha$  serait à coefficients

entiers et comme  $R_\alpha(1) = p$ , le théorème 2 serait contredit.

### Remarques

Je ne connais pas d'exemple où ce quotient

$Q = C_X(x) / \prod_{\alpha} R_\alpha^{q_\alpha(x)}(X)$  ne soit pas à coefficients positifs.

Pour le démontrer dans le cas général, il paraîtrait raisonnable de trouver une interprétation combinatoire de  $Q$ . Posons

$A_p(X) = \prod_{\alpha} R_{\alpha}^{q_{\alpha}(x)}(X)$ ; si  $p$  varie, les  $A_p$  sont premiers entre eux, mais le polynôme  $C_X(x) / \prod_p A_p$  n'est plus en général à coefficients positifs. Le problème de la factorisation de la variable aléatoire à valeurs entières admettant  $C_X(x)/c(x)$  comme fonction génératrice en  $X$  semble donc délicat.

Nous tirons du théorème 2 l'énoncé suivant :

#### Théorème 4

Si  $x \in \mathcal{M}_b$  et  $x \neq 0$ , alors

a)  $v(|x|) - v(x) \leq v(c(x))$

b)  $d(|x|) - d(x) \leq v(c(x))$

c)  $v(c(x)) \leq (d(|x|) - v(x)) (a(x) - 1)$

où  $a(x)$  est le nombre de  $i$  tels que  $x_i \neq 0$ .

(voir [ ] pour la démonstration de ces trois propriétés ; signalons que a) est connu depuis Charles Hermite [6]). Les conséquences de ce théorème 4 sont nombreuses, et un grand nombre de propriétés des coefficients multinomiaux souvent laborieusement démontrées, en sont des applications presque immédiates ; nous allons en détailler quelques-unes.

#### Corollaire 1

Si le p.g.c.d. des  $x_i$  est 1, alors  $|x|$  divise  $c(x)$ .

#### Démonstration

Par hypothèse  $v(x) = 0$  pour tout  $p$ , donc d'après le théorème 4a)  $v(|x|) \leq v(c(x))$  pour tout  $p$ .

### Corollaire 2

$$v(c(x)) \leq (a(x) - 1) d(|x| + 1) - v(|x| + 1) - (a(x)-2)v(x')$$

où  $x'$  dans  $\mathbb{M}$  est défini par  $x'_i = x_i + 2i$  si  $x_i > 0$

$$x'_i = 0 \text{ si } x_i = 0.$$

### Démonstration

Soit  $i_0$  l'indice tel que  $v(x'_{i_0})$  soit minimum. Comme

$a(x) = a(x + \varepsilon_{i_0})$ , on pose  $x'' = x + \varepsilon_{i_0}$  et on applique le théorème

$$\begin{aligned} 4c) \text{ à } x'' : v(c(x'')) &\leq (a(x) - 1) [d(|x| + 1) - v(x'')] \\ &= (a(x) - 1) d(|x| + 1) - (a(x) - 2) v(x') - v(x'_{i_0} + 1). \end{aligned}$$

or  $(x'_{i_0} + 1) c(x'') = (|x| + 1) c(x)$ , d'où le résultat.

### Corollaire 3

Si  $a(x) = 2$

$$v(c(x)) \leq d(|x| + 1) - v(|x| + 1)$$

Il est facile de voir que cette borne supérieure est atteinte pour toute valeur de  $|x|$ . Je dois cet énoncé à Robert Meynieux auteur d'un remarquable théorème sur les nombres binomiaux. Pour tout nombre premier  $p$ , on désigne par  $\rho_p(n)$  le plus grand entier  $\rho \leq n/2$  tel que :

$$v(c(\rho, n-\rho)) = d(n+1) - v(n+1)$$

et  $q(n) = 1 + \inf_p \rho_p(n)$ . Le théorème de R. Meynieux s'énonce ainsi :

l'entier  $q(n)$  est la plus petite puissance de nombre premier supérieure ou égale à  $(n + 2)/3$ .

#### Corollaire 4

Si  $x$  de  $\mathcal{M}$  n'est pas de la forme  $x = (p^n - 1, 1)$ ,  $c(x)$  n'est jamais une puissance de nombre premier.

#### Démonstration

Puisque  $c(x) = c(x_1, |x| - x_1) c(x - x_1 \varepsilon_1)$  il suffit de le démontrer pour  $|I| = 2$ , c'est-à-dire pour des coefficients binomiaux. Si  $c(x_1, x_2) = p^m$ ,  $v(c(x)) = m$ ; d'après le théorème 4.c), on a donc  $m \leq d(|x|)$  et donc  $c(x) = p^m \leq |x|$ . Or, il est facile de vérifier que pour tout  $x$ ,  $c(x) \geq |x|$ , et que l'égalité entraîne que  $x$  est de la forme  $x = (|x| - 1, 1)$ , ce qui démontre le théorème.

#### Remarques

Ce corollaire est dû à Herring [ 5]. Si on appelle nombre-binomial un nombre de la forme  $c(x)$  avec  $x = (x_1, x_2)$  et  $x_1$  et  $x_2 > 1$ , si on appelle support d'un entier l'ensemble des nombres premiers qui le divise, le résultat de Herring conduit à conjecturer : il n'y a qu'un nombre fini de nombres-binomiaux à support fini donné. Ce problème semble fort difficile.

### 3) Le temps de frappe de $H_D$

Nous revenons au problème de Von Neumann et considérons

$H_D = \{x \in \mathcal{M} ; c(x) \equiv 0 \pmod{D}\}$  où  $D$  est un entier  $> 1$ .

#### Théorème 5

Le temps de frappe  $T(H_D)$  est presque sûrement fini pour tout  $D$  satisfaisant à l'hypothèse (H) et  $E(T(H_D)) \ll K$ , où  $K$  est le minimum pour tous les couples  $(i, j)$  avec  $i \neq j$  de la quantité

$$D - \frac{(\pi_i + \pi_j)^{D-1}}{\pi_i^D + \pi_j^D}$$

#### Démonstration

Désignons par  $E$  l'ensemble des points  $x$  de  $\mathcal{M}$  tels qu'il existe  $i$  et  $j$  tels que :

$$x_i \equiv 1 \pmod{D} \quad \text{et} \quad x_j \equiv -1 \pmod{D}. \text{ Nous montrons d'abord}$$

que  $E \subset H_D$ . En effet, si  $x \in E$ , alors  $c(x)$  est divisible par

$c(x_i, x_j)$ , dont nous allons démontrer qu'il est nul modulo  $D$ .

Pour cela, on pose  $x_i = aD + 1$  et  $x_j = bD - 1$  et on considère

un nombre premier  $p$  fixé. Posons  $D = hp^k$  avec  $v(\mathbf{h}) = 0$ . Alors :

$$(p-1) v(c(aD + 1, bD - 1)) = ||ahp^{k+1}|| + ||bhp^{k-1}|| - ||ah+bh||$$

Mais il est facile de constater, d'après la proposition 4, que :

$$||ahp^k + 1|| = 1 + ||ah|| \quad \text{et}$$

$$||bhp^k - 1|| = (p-1)k + ||bh - 1||.$$

$$v(c(aD+1, bD-1)) = k + \frac{1}{p-1} (||ah|| + 1 + ||bh-1|| - ||ah+bh||).$$

D'après la proposition 4,e)

$$||ah + bh|| \ll ||ah|| + ||bh - 1|| + ||1||$$

Par conséquent

$$v(c(x_i, x_j)) \geq v(D)$$

et ceci pour tout  $p$ , ce qui montre que  $E \subset H_D$ . Nous allons maintenant voir que  $T(E) < \infty$  presque-sûrement si  $P$  satisfait à l'hypothèse (H) et que  $E(T(E)) \leq K$ . Comme  $T(H_D) \leq T(E)$  cela entraînera le théorème.

Il existe nécessairement  $i$  et  $j$  distincts dans  $I$  tels que  $\pi_i$  et  $\pi_j$  soient positifs. Considérons l'homomorphisme de  $\mathcal{G}$  dans  $[Z / DZ]^2$  défini par  $h(x) = (x_i \bmod D, x_j \bmod D)$ . La promenade aléatoire  $S_t$  est alors transportée par  $h$  sur  $[Z / DZ]^2$ , on a :

$$h(S_t) = h(y_1) + \dots + h(y_t),$$

avec  $P(h(y_t) = (1, 0)) = \pi_i$

$$P(h(y_t) = (0, 1)) = \pi_j$$

$$P(h(y_t) = (0, 0)) = 1 - \pi_i - \pi_j$$

Désignons alors par  $G$  le sous-groupe de  $[Z / DZ]^2$  des points  $(x, y)$  tels que  $x + y = 0$  et soit  $T_1, T_2, \dots, T_n, \dots$  les instants successifs de visite de  $G$  par la promenade  $h(S_t)$ . Les variables  $I_1 = T_1, \dots, I_n = T_n - T_{n-1}$  sont indépendantes et de même loi, et il est clair que  $I_1$  suit une loi de fonction génératrice

$$E(s^{I_1}) = \left[ \frac{(\pi_i + \pi_j) s}{1 - (1 + \pi_i + \pi_j) s} \right]^D \left[ \frac{(\pi_i + \pi_j) s}{1 - (1 - \pi_i - \pi_j) s} \right]^D$$

donc de moyenne  $\frac{D}{\pi_i + \pi_j}$ .

Les  $(h(S_{T_n}))_{n=0}$  définissent dans  $G$  une promenade aléatoire

gouvernée par la loi de  $h(S_{T_1})$ , qui est :

$$P(h(S_{T_1}) = 0) = p^D + q^D$$

$$P(h(S_{T_1}) = k) = c(k, D-k) p^k q^{D-k}$$

pour  $0 < k < D$ .

(Ici, nous identifions  $G$  à  $\mathbb{Z}/D\mathbb{Z}$  au moyen de l'application

$(x, y) \longrightarrow x$  et nous posons  $p = \pi_i / (\pi_i + \pi_j)$ ,  $q = \pi_j / (\pi_i + \pi_j)$ .

Si on pose enfin

$$N = \inf \{n ; h(S_{T_n}) = \pm 1\}$$

on aura  $T(E) = T_N$ , et d'après l'identité de Wald :

$$E(T(E)) = E(T_1) E(N).$$

Procédons maintenant à la majoration de  $E(N)$ .

Si  $(S_n)_{n=1}^\infty$  est une promenade sur  $G$  gouvernée par une probabilité  $\mu$

et si  $F \subset G$ , on note par  $T_X = \inf \{n ; S_n \notin F\}$  si  $S_0 = x$  et

$f(x) = E(T_X)$ . Il est clair qu'on a

$$f(x) = 0 \quad \text{si } x \notin F$$

$$f(x) = 1 + \sum_{y \in G} f(x+y) \mu(\{y\}) \quad \text{si } x \in F$$

Introduisons l'opérateur  $A$  sur les fonctions  $g$  définies sur  $G$  par

$$Ag(x) = 1_F(x) \cdot \sum_{y \in G} g(x+y) \mu(\{y\})$$

il est clair que

$$f = 1_F + Af$$

Si, de plus,  $F \neq G$  et si  $\mu$  n'est pas concentrée dans un sous-groupe de  $G$ , on voit immédiatement que  $f$  est l'unique solution de

$$(I - A)f = 1_F, \text{ soit}$$

$$f = \left( \sum_{n=0}^{\infty} A^n \right) 1_F$$

Particularisons en prenant  $E \setminus F = \{-1, +1\}$  et  $\mu$  comme loi de  $h(S_{T_1})$ . Alors  $N = f(0)$ . Ici :

$$\max_{x \in G} A 1_F(x) = m < 1 \text{ et donc}$$

$$f(x) \leq \frac{1}{1-m} \quad \text{si } x \in F.$$

$1-m$  est le minimum de la quantité  $\mu(\{x+1\}) + \mu(\{x-1\})$  quand  $x$  décrit  $\mathbb{Z}/D\mathbb{Z}$ . Or, on sait que la suite  $c(k, D-k)p^k q^{D-k}$ , avec  $k = 0, 1, \dots, D$  est croissante puis décroissante.

Donc  $1-m = \inf_{x=0, \pm 1, \pm 2} \mu(\{x+1\}) + \mu(\{x-1\})$ . Il est clair alors

que  $1-m \geq p^D + q^D$ . Donc :

$$E(N) \leq (\pi_i + \pi_j)^D / (\pi_i^D + \pi_j^D)$$

$$E(T_D) \leq D(\pi_i + \pi_j)^{D-1} / (\pi_i^D + \pi_j^D), \text{ ce qui achève}$$

la démonstration.

Nous allons voir (corollaire 2 du théorème 6) que si  $D$  est un nombre premier  $p$ , on a un résultat beaucoup plus précis.

En effet, l'examen d'un tableau des restes modulo  $p$  des  $c(x)$ ,

facile à construire pour  $|I| = 2$  (puisque c'est le triangle de Pascal associé au corps  $F_p = \mathbb{Z}/\mathbb{Z}_p$ ), rend plausible l'énoncé suivant :

Théorème 6

Si  $x \notin H_p$ , alors :

$$m_{H_p}(x) = \prod_{\alpha \geq 0} c(h_\alpha(x)).$$

$(m_{H_p}(x))$  a la signification donnée au § 1).

La démonstration assez longue peut être trouvée en [1].

Les corollaires de ce théorème sont fort intéressants.

En voici deux :

Corollaire 1

$$c(x) \equiv \prod_{\alpha \geq 0} c(h_\alpha(x)) \text{ pour tout } x \text{ de } \mathcal{M}_0.$$

Ceci est la conséquence des théorèmes 1 et 6 ; ce corollaire est du à Lucas, et les théorèmes 1 et 6 en fournissent l'interprétation combinatoire.

Corollaire 2

$$E(T(H_p)) = p \prod_{n=1}^{\infty} \frac{1 - \left[ \sum_i \pi_i p^n \right]^n}{1 - \sum_i \pi_i p^n}$$

Ceci découle de la formule

$$E(T(H_p)) = \sum_{t=0}^{\infty} P(T(H_p) > t) = \sum_{x \notin H_p} m_{H_p}(x) \pi^x$$

4) La formule de Van der Monde

On désigne sous le nom de formule de Van der Monde la relation suivante entre coefficients binomiaux :

$$c(r, s) = \sum_k c(k, t-k) c(r-k, s+k-t)$$

où  $r, s$  et  $t$  sont des entiers positifs et où la somme est prise pour les entiers  $k$  tels que  $\max(0, t-s) \leq k \leq \min(s, t)$ . Elle prend un aspect plus agréable si on l'écrit sous la forme :

$$c(x) c(y) = \sum_z c(z)$$

où  $x = (r, s)$        $y = (t, r + s - t)$

et où  $z$  décrit l'ensemble des matrices  $(2, 2)$  à coefficients entiers de marges  $x$  et  $y$ . Sous cet aspect, cette formule se généralise aisément :

Proposition 5

Soient  $I_1, \dots, I_d$  des ensembles quelconques

$\mathcal{M}_{I_1}, \dots, \mathcal{M}_{I_d}, \mathcal{M}_{I_1 \times \dots \times I_d}$  les monoïdes abéliens libres

engendrés. Si les  $x^{(k)} \in \mathcal{M}_{I_k}$  sont tels que  $|x^{(1)}| = \dots = |x^{(d)}|$ ,

alors :

$$c(x^{(1)}) \dots c(x^{(d)}) = \sum_z c(z)$$

où la somme est prise pour tous les  $z = (z_{i_1}, \dots, z_{i_d})$  de

$\mathcal{M}_{I_1 \times \dots \times I_d}$  de marges  $x^{(1)}, \dots, x^{(d)}$ , c'est-à-dire tels que

$$i_2 \cdot \dots \cdot i_d \cdot z_{i_1, \dots, i_d} = x_{i_1}^{(1)} \quad \text{etc...}$$

La démonstration est faite dans [7] pour  $d = 2$  et se généralise trivialement.

De nombreux cas particuliers de cette formule sont intéressants. Nous n'en signalerons cependant qu'un seul, relié au chapitre 3 ; c'est celui où les  $x^{(k)}$  sont de la forme  $(1, 1, \dots, 1)$ . Alors  $c(x^{(k)}) = N!$   $c(z)$  pour  $z$  de marges convenables, et la proposition montre de façon très simple qu'il y a  $(N!)^{d-1}$  façons de placer  $N$  points dans  $\{1, \dots, N\}^d$  de sorte que la projection de ces  $N$  points sur un quelconque des  $d$  axes soit bijective. Il est facile de voir que ceci entraîne que l'ensemble convexe des probabilités sur  $\{1, \dots, N\}^d$  de marges uniformes à  $(N!)^{d-1}$  points ; c'est ainsi qu'on généralise le célèbre théorème de Birkhoff sur les matrices bistochastiques.

Notre propos est ici de démontrer dans le cas  $d = 2$ , la formule de Van der Monde pour les polynômes multinomiaux. Pour cela, on pose  $I_1 = I$  et  $I_2 = J$  et il est nécessaire de supposer  $I$  et  $J$  totalement ordonnés ; on prend donc  $I = J = \{1, 2, \dots\}$ .

Si  $z \in \mathcal{M}_{I \times J}$ , on pose :

$$\sigma(z) = \sum \{z_{ij} z_{i'j'} ; i < i' \quad j < j'\}.$$

Le théorème s'énonce :

### Théorème 7

Si  $x \in \mathcal{M}_I$  et  $y \in \mathcal{M}_J$  sont tels que  $|x| = |y|$ , alors

$$C_X(x) C_X(y) = \sum_z X^{\sigma(z)} C_X(z)$$

où la somme est prise pour tous les  $z \in \mathcal{M}_{I \times J}$  de marges  $x$  et  $y$ ,

c'est-à-dire tels que

$$\sum_j z_{ij} = x_i \quad \text{et} \quad \sum_i z_{ij} = y_j$$

### Démonstration

Quelques notations nouvelles nous sont nécessaires. Si

$|x| = |y|$  et  $x \in \mathcal{M}_I$  et  $y \in \mathcal{M}_J$ , on note

$M(x, y)$  l'ensemble des  $z$  de  $\mathcal{M}_{I \times J}$  de marges  $x$  et  $y$ .

Si  $y$  est de la forme  $(1, 1, \dots; 1, 0, \dots)$  on note

$M(x, y) = M(x)$ . Si  $z \in M(x, y)$  on note  $z_{.j}$  l'élément de  $\mathcal{M}_X$

défini par  $i \longrightarrow z_{ij}$ .

Pour démontrer la formule, on commence par remarquer que le cas particulier  $y = (1, 1, \dots, 1, 0, 0, \dots)$  a déjà été démontré dans la note [7], et je ne reprends pas cette preuve ici. Si  $|x| = N$ ,  $y = (1, \dots, 1, 0, 0, \dots)$  et  $z \in M(x)$ , alors  $C_X(z) = C_X(y)$ . On peut donc affirmer que

$$C_X(x) = \sum_{z \in M(x)} X^{\sigma(z)} \quad (1)$$

Prenons maintenant  $y \in \mathcal{M}_J$  quelconque. Il est clair que

si  $z \in M(x, y)$  alors  $C_X(y)$  divise  $C_X(z)$ . Nous avons donc à montrer

$$\text{que :} \quad C_X(x) = \sum_{z \in M(x, y)} X^{\sigma(z)} \prod_j C_X(z_{.j}) \quad (2)$$

D'après la formule (1), le second membre de (2) s'écrit :

$$\sum_{z \in M(x, y)} X^{\sigma(z)} \prod_{j \in J} \sum_{z'_j \in M(z, .j)} X^{\sigma(z'_j)} = \sum_{z \in M(x, y)} \sum_{z'_j \in M(z, .j)} X^{\sigma(z)} + \sum_{j \in J} \sigma(z'_j)$$

Nous avons donc maintenant à établir une correspondance bijective entre  $M(x)$  et  $\bigcup_{z \in M(x, y)} \prod_{j \in J} M(z, .j)$  qui respecte de façon convenable les exposants, et (1) nous donnera (2).

On considère donc un élément  $z'$  de  $M(x)$  soit  $(z'_{i,k})$ , où  $i \in I$  et  $k = 1, 2, \dots, |x|$ . On pose  $s_0 = 0$  et  $s_j = y_1 + y_2 + \dots + y_j$ ,

et on définit  $z'_{ij} = \sum_{k=s_{j-1}+1}^{s_j} z'_{ik}$ , et  $z'_{ij} = z'_{ik}$  si

$s_{j-1} < k \leq s_j$ . Il est clair que  $z \in M(x, y)$  et que  $z'_j \in M(z, .j)$ .

Cette correspondance est naturellement bijective. Enfin,

$\sigma(z') = \sigma(z) + \sum_{j \in J} \sigma(z'_j)$  se vérifie facilement, ce qui achève

la démonstration.

Comme exemple d'application, on peut considérer le cas  $I = J = \{1, 2\}$  et  $x = y = (n, n)$ . Alors, si  $z \in M(x, y)$   $z$  est de la forme :

$$z = \begin{bmatrix} k & n-k \\ n-k & k \end{bmatrix}$$

et on a :

$$\left[ C_X(n, n) \right]^2 = \sum_{k=0}^n X^{k^2} \left[ C_X(k, n-k, k, n-k) \right]$$

ou encore :

$$C_X(n, n) = \sum_{k=0}^n X^{k^2} \left[ C_X(k, n-k) \right]^2.$$

### 5) Comportement p-adique des coefficients multinomiaux

Au § 3, nous avons considéré l'ensemble  $H_p$  des  $x$  de  $\mathcal{M}$  tels que  $c(x) \equiv 0 \pmod{p}$ . L'examen d'un tableau de  $H_p$  (voir ) conduit à découvrir de nouvelles propriétés aux  $c(x)$ . En voici trois, dont on trouvera les démonstrations en [1]. Seule la seconde semblait être connue auparavant, et ceci dans le cas binomial [4]. Nous développerons ensuite des applications de la troisième.

#### Théorème 8

Soient  $x$  et  $y$  dans  $\mathcal{M}$ , avec  $v(x_i) > d(y_i)$  pour tout  $i$ , et  $\eta_i = \pm 1$ . Notons  $(x + \eta y)_i = x_i + \eta_i y_i$ . Alors si  $\sum \eta_i y_i \geq 0$ , nécessairement :

$$v(c(x + \eta y)) \geq v(c(x))$$

#### Théorème 9

Soit  $x$  dans  $\mathcal{M}$ ; pour tout  $i$ , on a :

$$v [c(x) - c(x + p^n \varepsilon^i)] \geq n - d(x - x_i \varepsilon^i)$$

#### Corollaire

$c(x + p^n \varepsilon^i) \xrightarrow[n \rightarrow \infty]{} c(x)$  dans l'anneau  $\mathbb{Z}_p$  des entiers p-adiques.

#### Théorème 10

Soit  $x$  dans  $\mathcal{M}$ ; on a :

$$v [c(px) - c(x)] \geq v(x) + 1 + v(c(x))$$

#### Corollaire

$L(x) = \lim_{n \rightarrow \infty} c(p^n x)$  existe dans  $\mathbb{Z}_p$ .

Les applications de ce théorème dépendent de la proposition technique suivante :

Proposition 6

Soit  $U = \{x \in \mathcal{M} ; v(x) = 0\}$  et  $(M_n)_{n=0}^{\infty}$  une famille de parties finies de  $\mathcal{M}$  telle que  $x \in M_n$  entraîne  $v(|x|) \geq n$  et telle que les  $p^{n-k}(M_k \cap U)$ , avec  $k = 0, 1, \dots, n$  forment une partition de  $M_n$ . Soit  $b(x)$  une fonction définie sur  $\mathcal{M}$  à valeurs dans  $\mathbb{Z}_p$  telle que  $b(x) = b(px)$ , et soit  $f$  définie sur  $\mathbb{Z}_p$  à valeurs dans  $\mathbb{Z}_p$  telle que  $v(f(u)) \geq v(u)$ . Alors :

$$\lim_n \sum_{x \in M_n} b(x) f(c(x)) = \sum \{b(x) f(L(x)) ; x \in U \cap \bigcup_{n=1}^{\infty} M_n\}$$

Démonstration

$$\sum_{x \in M_n} b(x) f(c(x)) = \sum_{k=0}^n \sum_{x \in M_k} b(x) f(c(p^{n-k}x)).$$

Soit  $x \in M_k \cap U$  ; alors  $v(|x|) \geq k$  et  $v(x) = 0$ . D'après le théorème , on a donc  $v(c(x)) \geq k$ .

Donc si  $n < k$ , on a  $L(x) \equiv 0 \pmod{p^n}$ , ce qui garantit la convergence de la série  $\sum \{b(x) f(L(x)) ; x \in \bigcup_{k=0}^{\infty} \{M_k \cap U\}\}$

Si  $n \geq k$ ,  $L(x) \equiv c(p^{n-k}x) \pmod{p^\alpha}$ , où  $\alpha = n - k + 1 + v(c(x)) > n$ .

Par conséquent :

$$\sum_{x \in M_n} b(x) f(c(x)) \equiv \sum_{k=0}^{\infty} \sum_{x \in M_k \cap U} b(x) f(L(x)) \pmod{p^n}, \text{ ce qui}$$

démontre la proposition.

La première application est une version p-adique de la formule de Van der Monde.

Proposition 7

Soient  $x^{(i)} \in \mathcal{M}_{I_i}$   $i = 1, 2, \dots, d$  avec

$$|x^{(1)}| = |x^{(2)}| = \dots = |x^{(d)}| \quad \text{et} \quad \inf_i v(x^{(i)}) = 0$$

$$L(x^{(1)}) L(x^{(2)}) \dots L(x^{(d)}) = \sum_x L(x)$$

où la somme est prise pour tous les  $x$  de  $\mathcal{M}_{I_1} \times \dots \times I_d$

de marges  $p^k x_i$  pour un certain  $k$  et tels que  $v(x) = 0$ .

Démonstration

Il suffit de considérer  $M_n \subset \mathcal{M}_{I_1} \times \dots \times I_d$  ensemble des  $x$  de marges  $p^n x^{(1)}, \dots, p^n x^{(d)}$  et d'appliquer le théorème 10 et la proposition 6, avec  $b = 1$  et  $f = \text{identité}$  !

Proposition 8

Soient  $(X_i)_{i \in I}$  des racines (p-1)ième de l'unité dans  $\mathbb{Z}_p$  telles que  $\sum_i X_i \equiv 0 \pmod{p}$ , et  $a$  un entier avec  $v(a) = 0$ . Alors :

$$\sum_x L(x) X^x = 0$$

où la somme est prise pour tous les  $x$  de  $\mathcal{M}$  tels que  $|x| = a p^{v(|x|)}$  et  $v(x) = 0$ .

Démonstration

On remarque d'abord que  $(\sum_i X_i)^{p^n a} \equiv 0 \pmod{p^n}$ , ensuite que

$X_i^{p-1} = 1$  entraîne que  $b(x) = X^x$  satisfait à  $b(x) = b(px)$ , enfin que  $M_n = \{x ; |x| = a p^n\}$  a les propriétés requises à la proposition 6 en prenant  $f$  égale à l'identité. Ceci achève la preuve.

La dernière application est la version  $p$ -adique de la célèbre formule de Dixon :

$$\frac{(-1)^n 3n!}{(n!)^3} = \sum_{k=0}^{2n} (-1)^k \left[ \frac{2n!}{k! (2n-k)!} \right]^3 .$$

### Proposition 9

Si  $p > 2$  et si  $a$  est un entier tel que  $v(a) = 0$ , alors

$$L(a, a, a) = \sum_x (-1)^{x_1} (L(x))^3 \text{ où la somme est prise pour tous}$$

les  $x = (x_1, x_2)$  tels que  $x_1 + x_2 = 2a p^{v(x_1 + x_2)}$  et  $v(x) = 0$ .

### Démonstration

Il suffit d'appliquer la proposition 6 du cas où  $f(u) = u^3$ ,  $M_n = \{x = (x_1, x_2) ; x_1 + x_2 = 2ap^n\}$  et  $b(x) = (-1)^{x_1}$ . Le fait que  $p$  soit impair entraîne que  $b(x) = b(px)$ . La formule de Dixon entraîne alors le résultat.

### Remarque

Cette formule est à rapprocher de  $L(a, a) = \sum_x (L(x))^2$ , où la somme est prise pour les  $x = (x_1, x_2)$  tels que  $x_1 + x_2 = ap^{v(x_1+x_2)}$  et  $v(x) = 0$  qu'on obtient à partir de la proposition ou encore directement à partir de

$$c(n, n) = \sum_{k=0}^n (c(k, n-k))^2 .$$

## Bibliographie du Chapitre II

- [1] BERNARD Jacques et LETAC Gérard  
"Construction d'évènements équiprobables et coefficients multinomiaux modulo  $p^n$ "  
A paraître à l'Illinois Journal of Mathematics
- [2] CARLITZ Léonard  
"Sequences and Inversions"  
Duke Mathematical Journal, vol. 37 n° 1 (1970) pp. 193-198
- [3] COMTET Louis  
"Analyse combinatoire" Tome 2, page 108 - P.U.F. (1970)
- [4] FRAY Robert D.  
"Congruences properties of ordinary and  $q$  - binomial coefficients"  
Duke Mathematical Journal, vol. 34 (1967)
- [5] HERING  
"Eine Beziehung zwischen Binomialkoeffizienten und Primzahlpotenzen"  
Archiv für Mathematik, vol. 19 (1968) pp. 411-412
- [6] HERMITE Charles  
cité par L.E. DICKSON in "History of the theory of numbers"  
vol. 1, chapitre 6  
G.E. STECHERT (New-York) (1934)
- [7] LETAC Gérard  
"Sur certains polynômes à coefficients positifs"  
Comptes rendus de l'Académie des Sciences, vol. 269 (3.12.69)  
pp. 1066-1068
- [8] MAC MAHON P.A.  
"The indices of permutations and the derivation there from of functions of a single variable associated with the Permutations of any Assemblage of Objects"  
American Journal of Mathematics, vol. 48, tome 4 (1913) pp. 282-321

- [9] MEYNIEUX Robert  
"Sur le plus petit commun multiple des coefficients du polynôme  $(1+z)^n$  et celui de certains de ces coefficients"  
Comptes rendus de l'Académie des Sciences, vol. 271 (2.10.1970)  
pp. 861-864
- [10] POLYA Georges  
"On the number of certain lattice polygons"  
Journal of Combinatorial Theory, vol. 6 (1969) pp. 102-105.
- [11] SYLVESTER J.J.  
"A constructive theory of partitions, arranged in three acts, an interact and an exodion"  
American Journal of Mathematics, vol. 5 (1882) pp. 251-330

### CHAPITRE III

---

#### LE PROBLEME DES MARGES



L'expression "problème des marges" évoque d'abord à l'esprit un couple de lois de variables aléatoires réelles  $X$  et  $Y$  dont on se demande quelle est leur loi jointe, puis ensuite l'idée de trois lois de variables aléatoires dans le plan notées  $(X_1, X'_1)$   $(X_2, X'_2)$   $(X_3, X'_3)$ , telles que les lois de  $X_i$  et  $X'_i$  soient les mêmes pour tout  $i$  et dont on se demande s'il existe une loi dans  $\mathbb{R}^3$  dont ces trois lois seraient les projections sur des plans convenables [3]. Ces deux exemples sont typiques. Dans le deuxième cas, on a un problème d'existence, dans le premier un problème sinon d'unicité, du moins de caractérisation de l'ensemble des solutions.

Commençons par donner une formulation générale du "problème des marges" par le biais du "problème des moments". Soit  $(\Omega, \mathcal{A})$  un espace mesurable,  $E$  un espace vectoriel de fonctions  $\mathcal{A}$  mesurables sur  $F$ ,  $f$  une forme linéaire sur  $E$ . Soit  $\mathcal{M}$  l'ensemble convexe des probabilités  $P$  sur  $(\Omega, \mathcal{A})$  telles que

$$f(e) = \int e(\omega) dP(\omega) \text{ pour tout } e \text{ de } E.$$

Le "problème des moments" est une expression pour désigner l'ensemble des questions suivantes :

- 1°)  $\mathcal{M}$  est-il non vide ?
- 2°)  $\mathcal{M}$  a-t-il des points extrémaux ? Quels sont-ils ?
- 3°)  $\mathcal{M}$  est-il pourvu d'agréables propriétés permettant sa représentation par extrémales ?

Ainsi dans les problèmes des moments classiques  $(\Omega, \mathcal{O})$  est  $\mathbb{R}$  ou  $[0, 1]$  munis de leurs boréliens et  $E$  est l'espace des polynomes. Un autre exemple remplace les polynomes par des polynomes trigonométriques à support donné ; l'exemple fondamental est donné par le cas où on se donne une famille  $(\mathcal{O}_\alpha)_{\alpha \in I}$  de sous-tribus de  $\mathcal{O}$ , où  $E$  est l'espace engendré par les espaces  $E_\alpha$  des fonctions bornées  $\mathcal{O}_\alpha$ -mesurables et où  $f$  est induit sur  $E_\alpha$  par une probabilité  $P_\alpha$  sur  $(\Omega, \mathcal{O}_\alpha)$  ; Le "problème des moments" est alors appelé "problème des marges" pour  $(\Omega, \mathcal{O}_\alpha, P_\alpha)_{\alpha \in I}$ .

On s'attend à ce qu'une question posée en des termes aussi vagues reste sans réponse générale.

Cependant des cas particuliers fort intéressants n'ont pas encore prouvé de solution :

1. Bien qu'une réponse fonctionnelle au problème d'existence posé par le second exemple soit possible, il n'y a pas de test maniable.
2. On ignore toujours quelles sont les extrémales de l'ensemble des probabilités sur un carré  $[0, 1] \times [0, 1]$  dont les projections sur les côtés sont la mesure de Lebesgue (voir § 3 et 7).
3. Etant données  $n$  lois sur  $Z$  ensemble des entiers, les lois extrémales sur  $Z^n$  les admettant pour lois marginales sont inconnues pour  $n \geq 3$ , en dehors des extrémales de Fréchet (voir § 8). Ce problème est le moins désespéré des trois, nous le formulerons en une conjecture précise au § 10.

Fort heureusement le problème du 3ème pour  $n = 2$  a un agréable parfum combinatoire, et c'est sur lui que nous allons nous concentrer. Nous saluerons le long de ce fil conducteur les généralisations lorsque celles-ci seront possibles.

Nous commençons (§ 2) par étudier la représentation du convexe  $\mathcal{M}$  des mesures de marges données au moyen des extrémales. Le cadre raisonnable auquel on se limite - celui d'espaces localement compacts équipés de leurs tribus boréliennes - permet de munir  $\mathcal{M}$  d'une structure compacte renvoyant ainsi le problème à la théorie de Choquet bien connue.

Le §3 donne une caractérisation facile des extrémales dans le cas le plus général et le § 4 donne une intéressante condition nécessaire d'extrémalité dans le cas où  $\Omega$  est dénombrable. Le § 5 contient l'important théorème des flots convergents qui a été grandement simplifié par une remarque de P.L. Hennequin. Un corollaire intéressant de ce théorème consiste en une formule de représentation des fonctions de classe A sur le tore à 2 dimensions qui sont nulles le long de deux droites. L'application du théorème des flots convergents au problème des marges est en § 6 , l'emploi des flots convergents ayant beaucoup simplifié la démonstration du théorème 9 (comparer avec [18]). Le § 7 recense ce qu'on peut dire des extrémales du cas continu (en dehors du théorème de Joram Lindenstrauss [19], qui affirme que si  $m$  est une extrémale du convexe des mesures sur le carré dont les projections sur les côtés sont la mesure de Lebesgue, alors  $m$  est singulière par rapport à la mesure de Lebesgue du carré). Le § 7 exhibe en particulier une extrémale dont le

support est le carré en entier. Le § 8 s'appuie sur les travaux de Maurice Fréchet, et étudie le rôle particulier joué par les extrémales qui portent son nom. Le § 9 reprend d'une manière plus simple les idées de [18]. J'ai été conduit à cette refonte par les commentaires pertinents de Pierre Cartier du sujet de [18] dans le Zentralblatt [7]. Les définitions utiles concernant les arbres sont extraites de l'article de Jacques Tits [28]. C'est Nicolas Varopoulos qui m'a signalé le théorème de Rado qui est la clé du § 10, et la lecture de [29] m'a fait simplifier la démonstration du théorème 14. Le § 11 rassemble des résultats épars ; j'ai profité de nombreuses conversations avec Dominique Foata. Le § 12 rassemble le peu qui m'a paru nouveau sur le difficile problème de l'existence, pour lequel l'article de V. Strassen [27], le dernier chapitre du livre de P.A. Meyer [21] et l'exposé de G. Choquet [8] apportent beaucoup d'information.

Un échantillon des résultats est fourni par l'ensemble des énoncés des théorèmes 1, 9, 13, 14 et 15.

## 2) Représentation par extrémales

Soient  $(\Omega_n, \mathcal{B}_n, P_n)_{n \in I}$  une famille finie ou dénombrable d'espaces de probabilité, où  $(\Omega_n, \mathcal{B}_n)$  est un espace localement compact muni de sa tribu borélienne. L'espace  $\prod_{n \in I} \Omega_n$ , étant muni de sa topologie produit, désignons par  $\mathcal{F}$  la famille des parties  $\Omega$  fermées et localement compactes de  $\prod_{n \in I} \Omega_n$  telles que la projection canonique de  $\Omega$  sur  $\Omega_n$  soit  $\Omega_n$  en entier. Si  $\Omega$  est dans  $\mathcal{F}$ , on désigne par  $\mathcal{O}_n$  la trace sur  $\Omega$  de la tribu qui est produit de  $\mathcal{B}_n$  et des tribus triviales sur  $\Omega_{n'}$ , quand  $n \neq n'$ . Evidemment  $P_n$  induit sur  $(\Omega, \mathcal{O}_n)$  une probabilité  $P_n^*$ . La tribu  $\mathcal{O}$  sur  $\Omega$  engendrée par les  $\mathcal{O}_n$  est aussi la tribu borélienne sur  $\Omega$  et on désigne par  $\mathcal{M}(\Omega)$  l'ensemble convexe des probabilités  $P$  sur  $(\Omega, \mathcal{O})$  dont la restriction à  $\mathcal{O}_n$  soit  $P_n^*$ .

Donnons maintenant quelques exemples :

1.  $\Omega_n = \mathbb{R}^2$   $n = 1, 2, 3$ ,  $\Omega$  sous espace vectoriel de  $\mathbb{R}^6$  des points de la forme  $(x_1, x_2, x_2, x_3, x_3, x_1)$
2.  $\Omega_1 = \Omega_2 = [0, 1]$   $\Omega = \Omega_1 \times \Omega_2$ ,  $P_1 = P_2 =$  mesure de Lebesgue
3.  $\Omega_1, \dots, \Omega_n$  dénombrables,  $\Omega = \Omega_1 \times \dots \times \Omega_n$   
 $P_1, P_2, \dots, P_n$  quelconques
4.  $\Omega_n = \mathbb{Z}$   $n = 1, 2, \dots$ ,  $\Omega$  espace des suites  $(x_n)_{n=1}^{\infty}$  nulles à partir d'un certain rang,  $(P_n)_{n=1}^{\infty}$  suite de lois sur  $\mathbb{Z}$  tels que

$$\sum_{n=1}^{\infty} \left[ 1 - P_n(\{0\}) \right] < \infty$$

5. Soit  $A$  un ensemble fini,  $F_1, F_2, \dots, F_n$  une famille de parties de  $A$  telle que  $F_i \subset F_j$  entraîne  $i = j$ . Soient  $(X_\alpha, \mathcal{C}_\alpha)_{\alpha \in A}$  une famille d'espaces localement compacts munis de leurs boréliens. On définit pour tout  $i = 1, \dots, n$   $(\Omega_i, \mathcal{B}_i)$  comme le produit des espaces mesurables  $(X_\alpha, \mathcal{C}_\alpha)_{\alpha \in F_i}$ . L'ensemble  $\Omega$  sera la partie de  $\Omega_1 \times \dots \times \Omega_n$  formée des suites de la forme :

$(x_\alpha^i)_{i, \alpha}$  où  $i = 1 \dots n$  et  $\alpha$  décrit  $F_i$ , avec  $x_\alpha^i \in X_\alpha$  et telles que  $x_\alpha^i = x_\alpha^j$  si  $\alpha \in F_i \cap F_j$ .

On aura reconnu les exemples 1, 2 et 3 de l'introduction. Le 5 est naturellement la généralisation de 1.

Signalons au passage que l'exemple 5 a été étudié par Vorobev [30]. Si  $F_1, \dots, F_n$  sont disjoints, il est clair que  $\mathcal{M}(\Omega)$  n'est jamais vide quelles que soient les probabilités marginales. Si les  $F_i$  ne sont pas disjoints, disons que les probabilités marginales sont compatibles si les projections canoniques de  $P_i$  et  $P_j$  sur  $\prod_{\alpha \in F_i \cap F_j} X_\alpha$  sont les mêmes.

Vorobev donne une condition nécessaire et suffisante (trop longue pour être décrite ici) pour que  $\mathcal{M}(\Omega)$  soit non vide quelles que soient les probabilités marginales compatibles.

Le mode de présentation a un avantage immédiat, celui de donner un aspect intrinsèque aux extrémales de  $\mathcal{M}_b(\Omega)$ , c'est-à-dire lié aux  $P_n$  et non à  $\Omega$ . Plus précisément, nous avons la proposition très simple suivante :

Proposition 1

Soient  $\Omega$  et  $\Omega'$  dans  $\mathcal{F}$  tels que  $\Omega \cap \Omega'$  soit aussi dans  $\mathcal{F}$  et soit  $P$  dans  $\mathcal{M}_b(\Omega \cap \Omega')$ . Si  $P$  est une extrémale de  $\mathcal{M}_b(\Omega)$ , c'est une extrémale de  $\mathcal{M}_b(\Omega')$ .

Démonstration

En effet si il existe  $P'$  et  $P''$  différents dans  $\mathcal{M}_b(\Omega')$  tels que  $P = \frac{P' + P''}{2}$ , nécessairement les supports de  $P'$  et  $P''$  relativement à  $\Omega'$  sont contenus dans le support de  $P$ . Donc  $P'$  et  $P''$  appartiennent à  $\mathcal{M}_b(\Omega \cap \Omega')$  et donc à  $\mathcal{M}_b(\Omega)$ . Comme  $P$  est extrémale dans  $\mathcal{M}_b(\Omega)$  on a la contradiction.

Pour mieux voir l'intérêt de cette proposition, reprenons l'exemple 5 ci-dessus dans le cas où les espaces  $X_\alpha$  sont dénombrables. Supposons qu'on ait pu caractériser toutes les extrémales de  $\mathcal{M}_b(\Omega_1 \times \dots \times \Omega_n)$ . Alors les extrémales de  $\mathcal{M}_b(\Omega)$  seront simplement celles de  $\mathcal{M}_b(\Omega_1 \times \dots \times \Omega_n)$  qui ont leur support dans  $\Omega$ . La recherche des extrémales de  $\mathcal{M}_b(\Omega)$  est ainsi débarrassée de la complication représentée par les  $F_i$ , il suffit de travailler avec les espaces dénombrables  $\Omega_i$ .

Parallèlement à la proposition 1, nous pouvons remarquer également

### Proposition 2

Soit  $\Omega$  dans  $\mathcal{F}$ ,  $P$  une extrémale de  $\mathcal{M}(\Omega)$  et  $\Omega'$  une partie fermée de  $\Omega$ , telle que  $P(\Omega') > 0$ . Soit  $P'(\cdot) = P(\cdot | \Omega')$ . Alors, si les  $P'_n$  sont les projections de  $P'$  sur  $\Omega'_n$  et  $\mathcal{M}'(\Omega')$  est l'ensemble des mesures sur  $\Omega'$  de marges  $P'_n$ ,  $P'$  est une extrémale de  $\mathcal{M}'(\Omega')$ .

### Démonstration

Si  $P' = \frac{Q + R}{2}$  avec  $Q$  et  $R$  dans  $\mathcal{M}'(\Omega')$  distincts, nécessairement :

$$P = \frac{1}{2} (1_{\mathcal{C}_{\Omega'}}, P + P(\Omega') Q) + \frac{1}{2} (1_{\mathcal{C}_{\Omega'}}, P + P(\Omega') R)$$

où  $1_{\mathcal{C}_{\Omega'}}, P$  est la restriction de  $P$  au complémentaire de  $\Omega'$  relativement à  $\Omega$ . Il est clair qu'alors  $P$  ne peut être extrémale.

Passons maintenant au théorème qui permettra la représentation de  $\mathcal{M}(\Omega)$ . Sa simplicité est un peu inattendue :

### Théorème 1

$\mathcal{M}(\Omega)$  muni de la topologie vague des mesures est compact si  $\Omega$  est dans  $\mathcal{F}$ .

### Démonstration

On sait naturellement que  $\mathcal{M}$  est relativement compact pour cette topologie, étant contenu dans la boule unité de l'espace des mesures de Radon sur l'espace localement compact  $\Omega$ . Pour voir que  $\mathcal{M}$  est fermé,

on montre que  $\mathcal{M}_b$  est uniformément tendu, c'est-à-dire que pour tout  $\varepsilon > 0$  il existe un compact  $K(\varepsilon)$  de  $\Omega$  tel que  $P(K(\varepsilon)) \geq 1 - \varepsilon$  pour tout  $P \in \mathcal{M}_b$ . Pour cela on prend  $K_n(\varepsilon)$  un compact de  $\Omega_n$  tel que  $P_n(K_n(\varepsilon)) \geq 1 - \frac{\varepsilon}{2^n}$  et on désigne par  $K(\varepsilon)$  la trace sur  $\Omega$  du compact  $\prod_{n=1}^{\infty} K_n(\varepsilon)$ . Alors :

$$1 - P(K(\varepsilon)) \leq \sum_{n=1}^{\infty} (1 - P_n(K_n(\varepsilon))) \leq \varepsilon$$

(ici est intervenu le fait que  $\Omega$  soit fermé).

Soit enfin  $P'$  un point de la fermeture de  $\mathcal{M}_b$ . On a toujours  $P(K(\varepsilon)) \geq 1 - \varepsilon$ . Soit  $f$  une fonction continue à support compact sur  $\Omega_{n_0}$ ,  $P'_{n_0}$  la projection de  $P'$  sur  $\Omega_{n_0}$ . On a à démontrer que :

$$\int_{\Omega_{n_0}} f(\omega_0) dP_{n_0}(\omega) = \int_{\Omega_{n_0}} f(\omega) dP'_{n_0}(\omega)$$

Notons  $\tilde{f}$  le relèvement de  $f$  dans  $\prod_{n=1}^{\infty} \Omega_n$  défini par  $\tilde{f}(\omega) = f(\omega_{n_0})$  pour tout  $\omega$  dont la  $n_0$ -ième coordonnée est  $\omega_{n_0}$ . Alors pour tout  $P$  de  $\mathcal{M}_b$  on peut écrire :

$$a = \left| \int_{\Omega_{n_0}} f(\omega) d(P_{n_0} - P'_{n_0})(\omega) \right| = \left| \int_{\Omega} \tilde{f}(\omega) d(P - P')(\omega) \right|$$

Donc, pour tout  $\varepsilon > 0$  on a :

$$a \leq \left| \int_{K(\varepsilon)} \tilde{f}(\omega) d(P - P')(\omega) \right| + 2\varepsilon \max_{\omega \in \Omega} |f(\omega)|$$

La fonction  $\tilde{f} \cdot 1_{K(\varepsilon)}$  étant à support compact, il est possible maintenant de trouver  $P$  dans  $\mathcal{M}_b$  assez proche de  $P'$  pour que

$$\left| \int_{K(\varepsilon)} \tilde{f}(\omega) d(P - P')(\omega) \right| \leq \varepsilon$$

a pouvant être rendu aussi petit qu'on veut est donc nul.

### Corollaire

Si  $\Omega$  est dans  $\mathcal{F}$ , tout point de  $\mathcal{M}(\Omega)$  est barycentre d'une mesure de probabilité  $\mu$  sur  $\mathcal{M}(\Omega)$  muni de sa tribu de Baire, telle que tout élément de la tribu ne contenant aucun point extrême soit de  $\mu$  mesure nulle. (La tribu de Baire est engendrée par les parties vaguement compactes de  $\mathcal{M}(\Omega)$  qui sont intersections dénombrables d'ouverts).

### Démonstration

C'est l'application du théorème de Choquet-Bishop-de Leeuw [23] à  $\mathcal{M}(\Omega)$ .

### Remarque

Dans le cas où  $I = \{1, 2\}$  et où  $\Omega_1$  et  $\Omega_2$  sont dénombrables, on verra au § 6 que l'ensemble des extrémales est alors fermé : le corollaire ci-dessus ne nécessite plus alors que l'appel au théorème de Krein-Millman.

### 3) Caractérisation fonctionnelle des extrémales

Voici une caractérisation très simple des extrémales de  $\mathcal{M}_b$  dans le cas du problème des marges, généralisation d'un résultat de Douglas [11]. Les notations sont celles de l'introduction.

Soit  $\mathcal{L}_\alpha$  l'espace des fonctions  $\sigma_\alpha$ -mesurables et  $P_\alpha$  intégrables,  $\mathcal{L}$  le sous-espace vectoriel de  $R^\Omega$  engendré par les  $\mathcal{L}_\alpha$ . Si  $P \in \mathcal{M}_b$ ,  $L^1(P)$  et  $L^\infty(P)$  sont les espaces usuels de classes de fonctions et on identifiera par abus de langage  $\mathcal{L}$  à un sous espace de  $L^1(P)$ .

#### Théorème 2

Si  $P$  est dans  $\mathcal{M}_b$ , c'est une extrémale de  $\mathcal{M}_b$  si et seulement si  $\mathcal{L}$  est dense dans  $L^1(P)$ .

#### Démonstration

Si  $P$  n'est pas extrémale, il existe  $P_1$  et  $P_2$  dans  $\mathcal{M}_b$  distincts tels que :

$$P = \frac{P_1 + P_2}{2}$$

c'est un exercice (voir [17]) de constater que  $P_1$  et  $P_2$  sont absolument continus par rapport à  $P$  et que leurs dérivées  $f_1$  et  $f_2$  de Radon-Nikodym appartiennent à  $L^\infty(P)$ . Il est clair que si  $g \in \mathcal{L}$ , alors :

$$\int_{\Omega} (f_1(\omega) - f_2(\omega)) g(\omega) dP(\omega) = \int_{\Omega} g(\omega) dP_1(\omega) - \int_{\Omega} g(\omega) dP_2(\omega) = 0$$

comme  $f_1 - f_2 \in L^\infty(P)$  et que  $f_1 - f_2 \neq 0$ ,  $\mathcal{L}$  est non dense dans  $L^1(P)$ .

Inversement, si  $\mathcal{L}$  est non dense dans  $L^1(P)$  il existe  $f$  non nul

dans  $L(P)$  tel que

$$\int_{\Omega} f(\omega) g(\omega) dP(\omega) = 0 \text{ pour tout } g \text{ de } \mathcal{L}$$

Posant  $f_1 = 1 + \frac{f}{\|f\|_{\infty}}$  et  $f_2 = 1 - \frac{f}{\|f\|_{\infty}}$  où  $\|f\|_{\infty}$  est le sup essentiel de  $f$ ,  $f_1$  et  $f_2$  sont  $\geq 0$  et tels que

$\int_{\Omega} f_1 dP = \int_{\Omega} f_2 dP = 1$ . Il est clair que les mesures  $P_1 = f_1 P$  et  $P_2 = f_2 P$  sont dans  $\mathcal{L}$  et sont distinctes, et que  $P$  ne peut être extrême puisque  $P = \frac{P_1 + P_2}{2}$ .

Il est intéressant de donner ici une caractérisation de l'espace  $\mathcal{L}$ , dans le cas où les hypothèses sont proches de celles de l'exemple 5 du § 2. Plus précisément :

Soient  $X_1, X_2, \dots, X_m$  des ensembles quelconques,  $X$  leur produit cartésien. Si  $F \subset \{1, 2, \dots, m\}$  on note  $\mathcal{L}(F)$  l'espace des fonctions  $f(x)$  réelles définies sur  $X$  qui ne dépendent que des coordonnées  $x^{(i)}$  de  $x$ , telles que  $i \in F$ . En d'autres termes, si  $P_F$  désigne la projection canonique de  $X$  sur  $\prod_{i \in F} X_i = X_F$ ,  $\mathcal{L}(F)$  est l'espace des fonctions sur  $X$  de la forme  $f = g \circ P_F$ , où  $g$  est une fonction sur  $X_F$ .  $P_F^{\perp}$  notant  $P_{C_F}$ , tout élément de  $X$  peut s'écrire  $x = (P_F x, P_F^{\perp} x)$ . On définit pour tout  $x_0$  de  $X$  l'opérateur  $\Pi(F, x_0)$  sur  $\mathcal{L}[\{1, \dots, m\}]$  par

$$\Pi(F, x_0) f(x) = f(x) - f(P_F x, P_F^{\perp} x_0)$$

Nous avons alors la caractérisation suivante de  $\mathcal{L}(F)$  :

Proposition 3

$\mathcal{L}(F) = \text{Ker } \pi(F, x_0)$  pour tout  $x_0$  de  $X$ .

En effet, si  $f \in \text{Ker } \pi(F, x_0)$ , posons, pour  $u$  dans  $X_F$  :

$$g(u) = f(u, P_F^\perp x_0)$$

Alors  $f = g \circ P_F$ .

Inversement si  $f = g \circ P_F$  où  $g$  est une fonction sur  $F$ , on a :

$$\pi(F, x_0)(g \circ P_F)(x) = g \circ P_F(x) - g \circ P_F(P_F x, P_F^\perp x_0) = 0.$$

On peut remarquer également que

$$(1 - \pi(F, x_0))(1 - \pi(F', x_0)) = (1 - \pi(F \cap F', x_0))$$

ce qui entraîne en particulier que pour  $x_0$  fixé et  $F$  variable les

$\pi(F, x_0)$  commutent, et que  $\pi(F, x_0)$  est idempotent.

Soient maintenant  $F_1, F_2, \dots, F_n$  des parties de  $\{1, 2, \dots, m\}$  et  $\mathcal{L}$  l'espace des fonctions  $f$  définies sur  $X$  de la forme

$$f = \sum_{i=1}^n f_i, \text{ où } f_i \in \mathcal{L}(F_i)$$

Voici le théorème de caractérisation de  $\mathcal{L}$ , du à P.J. Fishburn [13] :

Théorème 3

$$\mathcal{L} = \text{Ker} \left[ \prod_{i=1}^n \pi(F_i, x_0) \right]$$

pour tout  $x_0$  de  $X$ .

### Démonstration

Si  $f \in \mathcal{L}$ , il existe d'après la proposition 3, pour tout  $i = 1, 2, \dots, n$ , des  $f_i$  dans  $\text{Ker } \pi (F_i, x_0)$  tels que  $f = \sum_{i=1}^n f_i$ .

Comme les  $\pi (F_i, x_0)$  commutent entre eux, il est immédiat que

$$\left[ \prod_{i=1}^n \pi (F_i, x_0) \right] f = 0$$

Pour démontrer l'implication inverse, nous procédons par récurrence sur  $n$ . La proposition 3 montre que c'est vrai pour  $n = 1$ . Supposons le théorème vrai pour  $n - 1$ . Alors si

$$\pi (F_n, x_0) f \in \text{Ker } \left[ \prod_{i=1}^{n-1} \pi (F_i, x_0) \right]$$

d'après l'hypothèse de récurrence, on peut écrire

$$f_n = \pi (F_n, x_0) f = \sum_{i=1}^{n-1} f_i$$

avec  $f_i \in \mathcal{L} (F_i)$  pour  $i = 1, 2, \dots, n-1$ .

Or  $\pi (F_n, x_0)$  est idempotente. Donc  $\pi (F_n, x_0) (f - f_n) = 0$  et  $f - f_n \in \mathcal{L} (F_n)$ .

On peut donc écrire :

$$f = \sum_{i=1}^{n-1} f_i + (f - f_n)$$

ce qui montre que  $f$  est dans  $\mathcal{L}$ .

### Remarque

Dans le cas où les  $X_i$  n'ont qu'un nombre fini d'éléments  $a_j$ , il est possible - et utile - de calculer la dimension de  $\mathcal{L}$ . Posons  $a_T = \prod a_j$ , où le produit est pris pour tous les  $j \in \bigcap_{i \in T} F_i$ , et où

$T \subset \{1, 2, \dots, n\}$ . On convient qu'un produit sur un ensemble vide d'indices est 1. Il est connu que si  $E_1, \dots, E_n$  sont des sous-espaces vectoriels d'un même espace  $E$  tels que  $(\sum_{i \in T} E_i) \cap E_j = \sum_{i \in T} E_i \cap E_j$  pour tout  $T \subset \{1, \dots, n\}$  et pour tout  $j$ , on a :

$$\dim(E_1 + \dots + E_n) = \sum_{\emptyset \neq T \subset \{1, \dots, n\}} (-1)^{|T|+1} \dim\left(\bigcap_{i \in T} E_i\right)$$

Prenant  $E_i = \mathcal{L}(F_i)$ , on constate que

$$\dim\left(\bigcap_{i \in T} \mathcal{L}(F_i)\right) = a_T \text{ et donc}$$

$$\dim \mathcal{L} = \sum_{\emptyset \neq T \subset \{1, \dots, n\}} (-1)^{|T|+1} a_T$$

En particulier, si  $F_i = \{1\}$  pour tout  $i$

$$\dim \mathcal{L} = a_1 + \dots + a_n - n + 1.$$

#### 4) Les extrémales dans le cas dénombrable

Dans ce paragraphe, nous considérons une famille finie ou dénombrable  $(X_n)_{n \in I}$  d'ensembles également finis ou dénombrables. Ils sont naturellement équipés chacun de la tribu la plus fine et d'une probabilité  $P_n$ . Si  $I$  est dénombrable, on distingue un élément  $0 = 0_n$  dans chaque  $X_n$  et on postule que

$$\sum_{n \in I} [1 - P_n(\{0\})] < \infty$$

On désigne par  $X$  l'ensemble des suites  $(x_n)_{n \in I}$  telles que  $x_n = 0$  sauf pour un nombre fini de  $n$ , par  $\mathcal{M}$  l'ensemble des probabilités sur l'ensemble fini ou dénombrable  $X$  de marges  $P_n$ . La convergence de la série ci-dessus garantit d'ailleurs, d'après Borel Cantelli, que si  $P$  est une probabilité sur  $\prod_{n \in I} X_n$  de marges  $P_n$ , nécessairement  $P(X) = 1$ .

#### Définition

Un cube  $C$  de  $X$  est une partie de  $X$ , finie et non vide de la forme  $C = \prod_{n \in I} A_n$  où  $A_n \subset X_n$  et où  $A_n = \{0\}$  sauf pour un nombre fini de  $n$ . La norme du cube est définie par

$$||C|| = \sum_{n \in I} [(\text{nombre d'éléments de } A_n) - 1].$$

Le théorème suivant donne une condition nécessaire pour que  $P$  soit extrémale, qui a un caractère local.

#### Théorème 4

Si  $P$  est une extrémale de  $\mathcal{M}$ , son support n'a pas plus de  $1 + ||C||$  points dans le cube  $C$ .

### Démonstration

Soit  $S(P) = \{x \in X ; P(x) > 0\}$  le support de  $P$ . Soit  $N$  le nombre d'éléments de  $S(P) \cap C$ . Si  $N > 0$ ,  $P(C) > 0$  et d'après la proposition 2 du § 2,  $P(\cdot | C)$  est une extrémale de  $\mathcal{M}$ , ensemble des probabilités sur  $C$  de marges  $P(\cdot | A_n)$  (avec  $C = \prod_{n=1}^{\infty} A_n$  et  $A_n \subset X_n$ ).

D'après le théorème 2 du § 3 l'espace  $\mathcal{L}$  des fonctions  $f$  sur  $C$  de la forme  $f = \sum_{n \in I} f_n$ , où  $f_n$  ne dépend de la coordonnée  $x_n$  de  $x$ , est dense dans  $L^1(P(\cdot | C))$  espace qui est isomorphe à  $\mathbb{R}^N$ . Il nous suffit de voir pour achever la démonstration que  $\mathcal{L}$  est de dimension  $1 + ||C||$  : c'est la remarque qui termine le § 3.

### Remarque

C'est un théorème fondamental que la réciproque soit vraie dans le cas où  $I$  a deux éléments : nous le démontrerons au § 6. Elle est fautive si  $I$  a plus de deux éléments :

Prenons  $X_1, X_2, X_3$  à deux éléments notés 0 et 1 et

$S(P) = \{(0, 0, 0), (0, 0, 1), (1, 1, 0), (1, 1, 1)\}$  il est évident

que  $P$  ne peut être extrémale et pourtant  $S(P)$  n'a pas plus de

$1 + ||C||$  points dans le cube  $C$ .

## 5) Flots convergents

Pour démontrer le résultat annoncé dans la remarque qui précède, nous devons faire un détour. Le théorème que nous démontrerons dans ce paragraphe a plusieurs formulations équivalentes. La plus simple est celle qui s'exprime en termes de séries doubles.

### Théorème 5

Soit  $(a_{p,q})_{p,q=1}^{\infty}$  une série double, non nulle, à termes réels, absolument convergente, telle que :

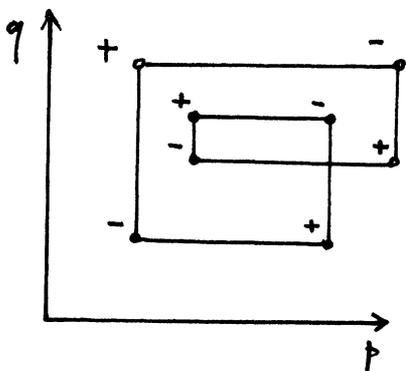
$$\sum_{q=1}^{\infty} a_{p_0,q} = \sum_{p=1}^{\infty} a_{p,q_0} = 0 ,$$

pour tous  $p_0$  et  $q_0$ . Alors on peut trouver deux suites finies d'entiers  $p_1, p_2, \dots, p_k$  et  $q_1, q_2, \dots, q_k$  telles que

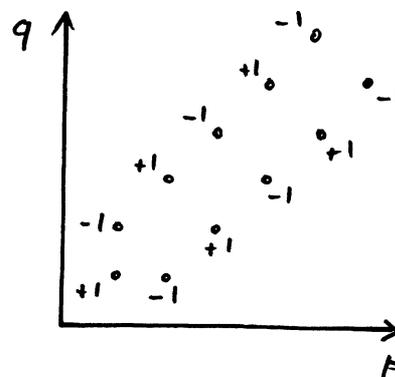
$$a_{p_i, q_i} > 0 \text{ pour tout } i = 1, 2, \dots, k$$

$$a_{p_{i+1}, q_i} < 0 \text{ pour tout } i = 1, 2, \dots, k ,$$

en convenant  $p_{k+1} = p_1$ .



Exemple n° 1



Exemple n° 2

L'exemple n° 2, où la série  $a_{m,n}$  est définie par :

$$a_{p,p+1} = a_{p+1,p} = (-1)^p$$

avec  $a_{1,1} = 1$  et  $a_{p,q} = 0$  dans les autres cas montre que la condition d'absolue convergence est indispensable.

Pour voir le théorème 5 sous un autre aspect, quelques définitions sont nécessaires. Soit  $X$  un ensemble fini ou dénombrable et  $m$  une fonction sur  $X \times X$  à valeurs  $\geq 0$  telle que

$$\sum_{x \in X} m(x, y) = \sum_{x \in X} m(y, x)$$

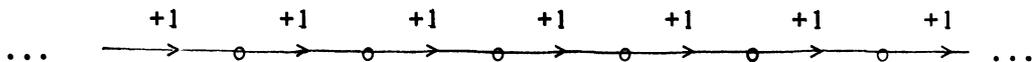
pour tout  $y$ . Un tel couple  $(X, m)$  est appelé un flot (nous nous écartons légèrement du vocabulaire habituel qui distingue deux sommets : entrée et sortie). Si de plus

$$\sum_{x,y \in X} m(x, y) < \infty$$

le flot est dit convergent. Un flot induit un graphe orienté ayant  $X$  pour ensemble des sommets et pour ensemble d'arêtes  $A$  l'ensemble des couples  $(x, y)$  tels que  $m(x, y) > 0$  : en particulier  $(x, x)$  peut être une arête. On appelle cycle dans le graphe orienté  $(X, A)$  une suite finie  $(x_1, x_2, \dots, x_k)$  de points distincts de  $X$  telle que  $(x_i, x_{i+1}) \in A$  pour tout  $i$ , en convenant  $x_{k+1} = x_1$ . Nous allons voir que le théorème 5 peut encore s'énoncer :

### Théorème 6

Si  $(X, m)$  est un flot convergent, le graphe  $(X, A)$  possède un cycle  
 Imaginons que  $X$  représente les carrefours (dénombrables !) d'une ville,  
 $A$  ses rues,  $m(x, y)$  une mesure de la capacité de la circulation de la  
 rue  $(x, y)$  (dans le sens  $x, y$ ). Le théorème dit simplement que si  
 $\sum m(x, y) < \infty$  il est possible de revenir au point de départ. C'est  
 faux sans cette condition :



Exemple n° 3

Nous allons d'abord démontrer que les énoncés des théorèmes 5 et 6  
 sont équivalents.

Supposons que le théorème 5 soit vrai ; sans perte de généralité  
 on peut supposer que  $m(x, x) = 0$  pour tout  $x$  car sinon  $\{x\}$  est un  
 cycle. Numérotons  $1, 2, \dots, n, \dots$  les éléments de  $X$  et définissons  
 la série double  $(a_{p,q})_{p,q=1}^{\infty}$  par

$$a_{p,q} = m(p, q) \quad \text{si } p \neq q$$

$$a_{p,p} = - \sum_{q=1}^{\infty} m(p, q)$$

La série satisfaisant aux hypothèses du théorème 5, il existe des points  
 $p_1, p_2, \dots, p_k$  de  $X$  tels que  $m(p_i, p_{i+1}) > 0$ , avec  $p_{k+1} = p_1$ .

Inversement supposons que le théorème 6 soit vrai. L'idée de la  
 démonstration suivante m'est venue en regardant la vue dite "éclatée"  
 d'un moteur en trois dimensions : il s'agit d'amener "les signes moins"

de la série  $a_{m,n}$  sur la diagonale. Pour cela on les isole dans chaque ligne en ajoutant artificiellement des lignes puis on recommence avec les colonnes. L'exemple suivant fait comprendre le procédé, que nous définissons ensuite rigoureusement :

p \ q	1	2	3
1	-	+	+
2	-	+	-
3	+	-	-

	(1,1)	(1,2)	(2,3)	(3,2)	(3,3)
1	-	0	+	+	+
2	0	-	+	-	0
3	+	+	-	0	-

	(1,1)	(1,2)	(2,3)	(3,2)	(3,3)
(1,1)	-	0	+	+	+
(1,2)	0	-	+	0	0
(2,3)	+	+	-	0	0
(3,2)	0	0	+	-	0
(3,3)	+	+	0	0	-

Exemple n° 4

Soit donc  $(a_{p,q})_{p,q=1}^{\infty}$  une série double satisfaisant aux hypothèses du théorème 5. On pose

$$X = \{(p, q) ; p = 1, 2, \dots \quad q = 1, 2, \dots\} \quad \text{et}$$

$$m((p_0, q_0), (p, q)) = a_{p_0, q}^+ \quad a_{p, q_0}^-$$

où  $a^+$  signifie  $\max(0, a)$  et  $a^- = (-a)^+$

On a alors

$$\sum_{p,q=1}^{\infty} m((p_0, q_0), (p, q)) = \left( \sum_{q=1}^{\infty} a_{p_0, q}^+ \right) \left( \sum_{p=1}^{\infty} a_{p, q_0}^- \right)$$

$$\sum_{p,q=1}^{\infty} m((p, q), (p_0, q_0)) = \left( \sum_{q=1}^{\infty} a_{p_0, q}^- \right) \left( \sum_{p=1}^{\infty} a_{p, q_0}^+ \right)$$

et deux nombres sont égaux puisque  $(a_{p,q})_{p=1, q=1}^{\infty}$  satisfait aux hypothèses du théorème 5.

Il est clair alors que  $m$  est un flot convergent et que le cycle associé fournit les suites  $p_1, \dots, p_k$  et  $q_1, \dots, q_k$  cherchées.

#### Démonstration du théorème 6

Une démonstration directe étant laborieuse, il est plus rapide et plus intéressant d'appliquer un résultat classique concernant les chaînes de Markov.  $(X, m)$  étant le flot convergent considéré, supposons, sans perte de généralité, que

$$\pi(y) = \sum_{x \in X} m(x, y)$$

est positif pour tout  $y$  dans  $X$ . Alors

$$p(y | x) = m(x, y) / \pi(x)$$

est tel que :

$$\sum_{y \in X} p(y | x) = 1$$

$$\sum_{x \in X} p(y | x) \pi(x) = \pi(y)$$

Par conséquent  $p(y | x)$  est la probabilité de transition d'une chaîne de Markov homogène sur l'espace d'états  $X$  et possédant une mesure invariante  $\pi$  (voir W. Feller [12] ou bien Hennequin et Tortrat [16] pour

les détails). Comme  $\sum \pi(x) < \infty$ , il existe des états récurrents positifs dans  $X$ ; en particulier il existe donc une suite  $(x_1, \dots, x_k)$  de points distincts de  $X$  tels que  $p(x_{i+1} | x_i) > 0$  pour  $i = 1, 2, \dots, k$  et  $x_{k+1} = x_1$ , ce qui est bien le résultat annoncé.

Voyons quelques conséquences des théorèmes 5 et 6.

#### A - Lois symétriques

$X$  étant dénombrable, soit  $P_s(X)$  l'ensemble des  $p$  tels que  $(X, p)$  soit un flot convergent et  $\sum_{x,y} p(x,y) = 1$ . (On peut considérer  $p$  comme la loi "jointe" d'un couple de variables aléatoires ayant les mêmes lois marginales).

$P_s(X)$  forme un convexe dont on cherche les extrémales. A toute suite  $C = (x_1, \dots, x_k)$  de points distincts de  $X$ , on associe  $p_C \in P_s(X)$  défini par :

$$p_C(x, y) = \frac{1}{k}$$

si  $(x, y)$  est de la forme  $(x_i, x_{i+1})$  ( $i = 1, \dots, k, x_{k+1} = x_1$ ) et  $p_C(x, y) = 0$  sinon.

#### Théorème 7

Les extrémales de  $P_s(X)$  sont les lois de la forme  $p_C$ . De plus, si  $p \in P_s(X)$  il existe une suite finie ou dénombrable de couples  $(\alpha_n, C_n)_{n=1}^{\infty}$  où  $\alpha_n > 0$ ,  $C_n$  est une suite finie de points distincts de  $X$  telle que

$$p(x, y) = \sum_{n=1}^{\infty} \alpha_n p_{C_n}(x, y)$$

### Démonstration

Soit  $p$  une extrémale de  $P_s(X)$ . D'après le théorème 1, il existe un  $C$  tel que  $p(x, y) \geq \varepsilon p_C(x, y)$  pour  $\varepsilon$  assez petit. Donc

$$p = \frac{1-\varepsilon}{2} \frac{p - \varepsilon p_C}{1 - \varepsilon} + \frac{1+\varepsilon}{2} \frac{p + \varepsilon p_C}{1 + \varepsilon}$$

Comme  $\frac{p \pm \varepsilon p_C}{1 \pm \varepsilon}$  sont dans  $P_s(X)$  ces deux lois sont égales et  $p = p_C$ . Réciproquement, il est aisé de voir que  $p_C$  est extrémale pour tout  $C$ .

Passons au problème de la représentation. A toute suite finie  $C = (x_1, \dots, x_k)$  d'éléments distincts de  $X$  et à tout flot convergent  $m$  on fait correspondre :

$$\varepsilon(m, C) = \inf_i \{m(x_i, x_{i+1}) ; i = 1, \dots, k\}$$

toujours avec la convention  $x_{k+1} = x_1$ . Puis on pose

$$\alpha(m) = \sup_C \varepsilon(m, C) |C| ,$$

$|C|$  désignant la longueur de la suite  $C$ . On a toujours

$$\alpha(m) \leq \sum_{x, y \in X} m(x, y) , \text{ et le théorème 6 s'énonce : } \alpha(m) > 0.$$

On construit les flots  $m_n$  sur  $X$  et les suites  $C_n$  de la manière suivante. On pose  $m_0 = p$ . Si  $m_n$  est construit, on choisit  $C_{n+1}$  tel que :

- i)  $\varepsilon(m_n, C_{n+1}) > 0$
- ii)  $\varepsilon(m_n, C_{n+1}) |C_{n+1}| > \alpha(m_n) - 1/n+1$

Posant  $\varepsilon(m_n, C_{n+1}) |C_{n+1}| p_{C_{n+1}}(x, y) = \mu_{n+1}$ , on définit enfin

$m_{n+1} = m_n - \mu_n$ . Alors :

$$\mu = p - \sum_{n=1}^{\infty} \mu_n = \lim_n m_n$$

existe (les limites sont prises au sens de la convergence simple). Il est facile de voir, par convergence monotone que  $(X, \mu)$  est, comme  $(X, m_n)$ , un flot convergent lorsque  $\mu$  n'est pas nul. D'autre part, il est clair que

$$\alpha(\mu) \leq \alpha(m_n) \leq \frac{1}{n+1} + \sum_{x,y \in X} \mu_{n+1} \text{ pour tout } n$$

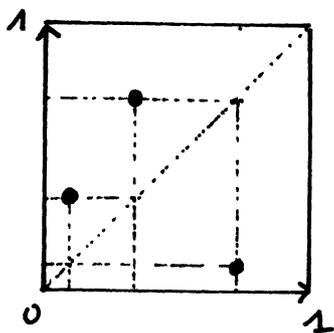
Donc, comme  $\sum_{n=1}^{\infty} \mu_n(x, y) \leq m(x, y)$

$$\alpha(m_n) \xrightarrow[n \rightarrow \infty]{} 0 \text{ et } \alpha(\mu) = 0$$

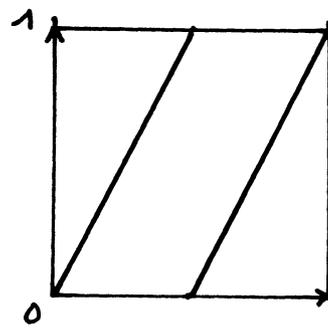
ce qui entraîne  $\mu = 0$  d'après le théorème 6. On a le résultat annoncé en posant  $\alpha_n = \varepsilon(m_{n-1}, C_n) |C_n|$ .

### Remarques

1.  $P_s[X]$  n'est nullement compact pour la topologie vague, ce qui nécessite de faire cette démonstration "à la main" de la représentation.
2. Si on remplace  $X$  par le segment  $[0, 1]$  et qu'on considère de la même façon  $P_s[0, 1]$ , cet ensemble possède des extrémales du même type (voir exemple 5)



Exemple n° 5



Exemple n° 6

Mais il en possède d'autres. Considérons la mesure  $p$  concentrée sur les droites  $y - 2x = 0$  et  $y - 2x + 1 = 0$  et se projetant sur l'axe des  $x$  suivant la mesure de Lebesgue. Son support est le graphe de la transformation  $y = T(x)$  du segment  $[0, 1]$  dans lui-même. Le segment  $[0, 1]$  étant muni de la mesure de Lebesgue  $m$ ,  $T$  conserve la mesure  $m$ . Il est bien connu (voir Billingsley [ 5 ] page 11) que  $T$  est ergodique pour cette mesure. Supposons que  $p$  ne soit pas extrême. Alors il existe  $p_1$  et  $p_2$  distincts dans  $P_s [0, 1]$  tels que  $p = \frac{p_1 + p_2}{2}$ . Les mesures  $p_i$  sont donc absolument continues par rapport à  $p$  (et même à dérivée de Radon-Nikodym essentiellement bornée par 2). Les projections  $m_1$  et  $m_2$  sont donc absolument continues par rapport à  $m$ . Les supports de  $p_1$  et  $p_2$  étant contenus dans le graphe de  $T$ , on en déduit que  $T$  conserve  $m_1$  et  $m_2$ . Soient  $f_1$  et  $f_2$  les dérivées de Radon Nikodym de  $m_1$  et  $m_2$  par rapport à  $m$ . D'après le théorème ergodique de Birkhoff, on sait que

$$\frac{1}{n} \sum_{k=1}^n f_i [T^k(x)] \xrightarrow[n \rightarrow \infty]{} C_i \quad \text{presque partout}$$

pour la mesure de Lebesgue,  $C_i$  étant une constante. En effet  $T$  est ergodique et  $f_i$  est intégrable. Or  $f_i [T^k(x)] = f_i(x)$  presque partout car la mesure  $m_i$  est invariante.  $f_i$  est donc équivalent à une constante et  $m = m_1 = m_2$ , ce qui entraîne  $p = p_1 = p_2$ .

## B. Mesures de marges nulles

X et Y étant des ensembles finis ou dénombrables, on note  $\mathcal{M}_0$  l'ensemble convexe des mesures réelles m sur X x Y telles que

$$\sum_{x \in X} m(x, y_0) = \sum_{y \in Y} m(x_0, y) = 0$$

pour tous  $(x_0, y_0)$  de X x Y et telles que

$$\|m\| = \sum_{(x,y) \in X \times Y} |m(x, y)| \leq 1.$$

On appelle cycle dans (X, Y) un couple  $\sigma = (\sigma^+, \sigma^-)$  de parties disjointes de X x Y de la forme

$$\sigma^+ = \{(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)\}$$

$$\sigma^- = \{(x_1, y_2), (x_2, y_3), \dots, (x_{k-1}, y_k), (x_k, y_1)\}.$$

où  $x_1, \dots, x_k$  et  $y_1, \dots, y_k$  sont des points distincts de X et Y respectivement. A tout cycle  $\sigma = (\sigma^+, \sigma^-)$ , on associe  $m_\sigma \in \mathcal{M}_0$  défini par :

$$m_\sigma(x, y) = \frac{1}{2k} \quad \text{si } (x, y) \in \sigma^+$$

$$m_\sigma(x, y) = -\frac{1}{2k} \quad \text{si } (x, y) \in \sigma^-$$

et  $m(x, y) = 0$  sinon. Ici k est le nombre d'éléments de  $\sigma^+$  et de  $\sigma^-$ .

### Théorème 8

Les extrémales de  $\mathcal{M}_0$  sont les mesures de la forme  $m_\sigma$ . Pour tout m de  $\mathcal{M}_0$ , il existe une suite  $(\alpha_n, \sigma_n)_{n=1}^\infty$  finie ou dénombrable telle que  $\alpha_n \geq 0$ , que  $\sigma_n$  soit un cycle, que  $\sum_{n=1}^\infty \alpha_n \leq 1$  et que

$$\sum_{n=1}^{\infty} \alpha_n m_{\sigma_n},$$

la convergence étant prise au sens de la convergence simple.

### Démonstration

Soit  $m$  une extrémale de  $\mathcal{M}_0$ . D'après le théorème 5, il existe  $\sigma$  tel que

$$0 \leq \frac{\varepsilon m_{\sigma}(x, y)}{m(x, y)} \leq 1$$

pour tout  $(x, y)$  tel que  $m(x, y) \neq 0$ , avec  $\varepsilon$  assez petit. Donc

$$m = \frac{1-\varepsilon}{2} \frac{m - \varepsilon m_{\sigma}}{1 - \varepsilon} + \frac{1+\varepsilon}{2} \frac{m + \varepsilon m_{\sigma}}{1 + \varepsilon}$$

où  $\frac{m \pm \varepsilon m_{\sigma}}{1 \pm \varepsilon} \in \mathcal{M}$ . Donc  $m = m_{\sigma}$ .

Réciproquement, si  $m_{\sigma}$  n'est pas extrémale, il existe  $m_1$  et  $m_2$  dans  $\mathcal{M}_0$  tels que :

$$m_{\sigma} = \frac{m_1 + m_2}{2}.$$

De plus,  $1 = ||m_{\sigma}|| \leq \frac{1}{2} (||m_1|| + ||m_2||)$  entraîne

$||m_1|| = ||m_2|| = 1$ . Vérifions que  $m_1(x, y) \neq 0$  entraîne  $m(x, y) \neq 0$ .

Sinon, supposons  $m_1(x_0, y_0) \neq 0$  et  $m(x_0, y_0) = 1$ . Alors on aurait :

$$\begin{aligned} 2 &= \sum_{\substack{x \neq x_0 \\ y \neq y_0}} |m_1(x, y) + m_2(x, y)| \\ &\leq \sum_{\substack{x \neq x_0 \\ y \neq y_0}} |m_1(x, y)| + |m_2(x, y)| = 2 - 2|m_1(x_0, y_0)| < 2 \end{aligned}$$

Comme  $\sigma$  est un cycle et que  $m_1 \in \mathcal{M}_0$ , il est clair que  $m_1 = m_2 = m_\sigma$  et  $m_\sigma$  est bien extrême.

Le reste de la démonstration, celle qui concerne la représentation par extrémales, se copie sans difficulté sur celle du théorème 7.

On peut donner un corollaire intéressant de ce théorème en considérant le cas où  $X = Y = Z$  où  $Z$  est le groupe des entiers relatifs. A tout cycle  $\sigma = (\sigma^+, \sigma^-)$  on associe la fonction définie sur  $T^2$  ( $T = [0, 2\pi)$ ) définie par

$$f_\sigma(\alpha, \beta) = \frac{1}{2k} \sum_{j=1}^k e^{i x_j \alpha} \left[ e^{i y_j \beta} - e^{i y_{j+1} \alpha} \right]$$

Soit  $A$  l'espace des fonctions définies sur  $T^2$  de série de Fourier absolument convergente, et  $M_0$  le sous espace de  $A$  formé des fonctions  $f(\alpha, \beta)$  telles que  $f(0, \beta) = f(\alpha, 0)$  pour tout  $(\alpha, \beta) \in T^2$ . Alors

### Corollaire

Si  $f \in M_0$  il existe une suite  $(c_n, \sigma_n)_{n=1}^\infty$  de nombres complexes et de cycles de  $Z^2$  tels que  $\sum_{n=1}^\infty |c_n| < \infty$  et :

$$f(\alpha, \beta) = \sum_{n=1}^\infty c_n f_{\sigma_n}(\alpha, \beta)$$

la série convergeant uniformément.

### Démonstration

Il suffit d'écrire

$$f(\alpha, \beta) = \sum_{p, q=-\infty}^{+\infty} a_{p, q} \exp i(p\alpha + q\beta)$$

et d'appliquer le théorème 8 aux deux éléments de  $\mathcal{M}_0$  définis par

$$b_{p,q} = \operatorname{Re} (a_{p,q}) \left( \sum_{p,q} |a_{p,q}| \right)^{-1}$$

$$b'_{p,q} = \operatorname{Im} (a_{p,q}) \left( \sum_{p,q} |a_{p,q}| \right)^{-1}$$

### Remarques

3.  $\mathcal{M}_0$  n'est nullement compact pour la convergence vague.

4. Comme pour  $P_S [0, 1]$ , on peut définir  $\mathcal{M}_0$  pour le carré

$[0, 1] \times [0, 1]$  ; mais les extrémales ne sont plus nécessairement

atomiques. Considérons en effet la mesure sur le carré  $[0,1] \times [0,1]$

qui est concentrée sur la diagonale  $x = y$  et dont les projections

sur les axes sont la mesure de Lebesgue. Soit  $p$  la mesure définie à

la remarque 2 qui suit le théorème 7, et, définissons  $m \in \mathcal{M}_0$  par

$m = \frac{p - p'}{2}$ . Alors  $m$  est une extrémale de  $\mathcal{M}_0$ . En effet, suppo-

sons que  $m$  s'écrive  $m = \frac{m_1 + m_2}{2}$  avec  $m_1$  et  $m_2$  distincts dans  $\mathcal{M}_0$ .

Nécessairement  $m_1$  est absolument continue par rapport à  $\frac{p + p'}{2} = |m|$

car  $m_1$  et  $m_2$  sont de norme inférieure ou égale à 1. De plus, si  $f_i$

est la dérivée de Radon-Nikodym de  $m_i$  par rapport à  $|m|$ ,  $f_i$  est né-

cessairement positive sur le support de  $p$  et négative sur celui de  $p'$ ,

ceci au moins presque partout relativement à  $|m|$ . Par conséquent  $m_1$

restreint au support de  $p$  est une mesure positive absolument continue par rapport à  $p$  et qui a mêmes projections sur les côtés du carré.

On a vu à la remarque 2 que ceci entraîne que  $m_1$  restreint au support

de  $p$  est proportionnelle à  $p$ . On voit alors qu'il est facile de ter-

miner le raisonnement et de voir que  $m = m_1 = m_2$ .

5. Il n'en reste pas moins que les  $m$  sont aussi des extrémales du  $\mathcal{M}_0$  précédent. Ce fait est utilisé (voir Buck [ 6 ], page 18, et Diliberto et Strauss [ 9 ] ) pour mesurer la qualité de l'approximation de la fonction  $f(x, y)$  (définie dans le carré) fournie par les fonctions de la forme  $g(x) + h(y)$ .

## 6) Ensembles d'unicité

Dans ce paragraphe et les suivants, on considère une fois pour toutes deux ensembles finis ou dénombrables  $X$  et  $Y$  équipés de probabilités  $r$  et  $s$ . On désigne par  $\mathcal{M}$  l'ensemble convexe des probabilités  $p$  sur  $X \times Y$  telles que :

$$\sum_{x \in X} p(x, y_0) = s(y_0) \quad \text{et} \quad \sum_{y \in Y} p(x_0, y) = r(x_0)$$

pour tout  $(x_0, y_0)$  de  $X \times Y$ . Si  $p$  est une probabilité sur  $X \times Y$ , on appelle support de  $p$  l'ensemble  $S(p) = \{(x, y) ; p(x, y) > 0\}$ .

Il est intéressant de voir le problème de la caractérisation de  $\mathcal{M}$  sous l'angle des chaînes de Markov. Supposons que sur un espace d'états  $E$  dénombrable, nous considérons une chaîne de Markov mal connue, mais dont nous savons que tous ses états sont récurrents et dont nous connaissons une probabilité stationnaire  $r(x)$ , c'est-à-dire, si  $p(y | x)$  est la probabilité de transition, telle que l'on ait

$$r(y) = \sum_x p(y | x) r(x)$$

(on peut estimer  $r(x)$  si un grand nombre de particules parcourent  $E$  de façon indépendante en suivant la loi  $p$ ).

Considérons alors la probabilité  $q(x, y)$  définie sur  $E \times E$  par  $q(x, y) = p(y | x) r(x)$ . Il est clair qu'elle a pour marges  $r$  et  $r$  et que c'est tout ce que nous savons. Connaître  $\mathcal{M}$  permettra de connaître toutes les chaînes possibles.

Rappelons quelques définitions classiques de théorie des graphes : Un graphe  $G = (S, A)$  (non-orienté) se définit comme un couple formé par

un ensemble  $S$  appelé ensemble de sommets et une famille de parties à deux éléments appelées arêtes (donc  $(x, x)$  n'est pas une arête). Un sous-graphe  $G' = (S', A')$  de  $G$  est un graphe tel que  $S' \subset S$  et  $A' \subset A$ . Un graphe  $(S, A)$  est connexe si pour tous  $x$  et  $y$  distincts dans  $S$  il existe une suite  $x = x_1, x_2, \dots, x_{k+1} = y$  de points de  $S$  telle que  $\{x_i, x_{i+1}\} \in A$  pour tout  $i = 1, 2, \dots, k$ . Un cycle dans  $(S, A)$  est une suite de points distincts  $x_1, x_2, \dots, x_k$  de  $S$  tels que  $(x_i, x_{i+1}) \in A$  pour tout  $i = 1, 2, \dots, k$ , en convenant  $x_{k+1} = x_1$ .

Un arbre est un graphe connexe et sans cycle. Un sous-arbre (parfois appelé forêt) est un sous-graphe d'un arbre.

Revenons à  $X$  et  $Y$ . Toute partie  $E$  de  $X \times Y$  définit un graphe  $G(E)$  ayant  $X \cup Y$  pour ensemble de sommets et  $E$  pour ensemble d'arêtes.

### Définition

On appelle ensemble d'unicité une partie  $U$  de  $X \times Y$  telle que quelles que soient  $r$  et  $s$ , il existe au plus un élément  $p$  de  $\mathcal{M}$  tel que  $S(p) \subset U$ .

Voici une conséquence importante du théorème 5, dont le corollaire nous donnera une caractérisation des extrémales de  $\mathcal{M}$  à l'aide de leur support.

### Théorème 9

$U$  est un ensemble d'unicité si et seulement le graphe  $G(U)$  est un sous-arbre.

### Démonstration

L'implication directe est presque immédiate ; en effet, si  $U$  est un ensemble d'unicité, fixons une mesure  $p$  telle que  $S(p) = U$ . Si  $U$  n'est pas un sous-arbre, il possède un cycle  $(z_1, z_2, \dots, z_k)$ . Ici  $k$  est nécessairement pair car il n'y a pas d'arêtes joignant deux points de  $X$  ou deux points de  $Y$ .

Posons  $k = 2n$  et sans perte de généralité, supposons  $z_1$  dans  $X$ . On note le cycle  $(x_1, y_1, x_2, y_2, \dots, x_n, y_n)$  avec  $x_i \in X$  et  $y_i \in Y$ . Comme  $(x_i, y_i) \in U$  et  $(x_{i+1}, y_i) \in U$  (en convenant  $x_{n+1} = x_1$ ) posons

$$\sigma^+ = \{(x_i, y_i) \quad i = 1, 2, \dots, n\}$$

$$\sigma^- = \{(x_{i+1}, y_i) \quad i = 1, 2, \dots, n\}$$

$$\text{et } \varepsilon = \inf \{p(x, y) ; (x, y) \in \sigma^+ \cup \sigma^-\}$$

Avec les notations du théorème 8, on peut écrire que  $p + \varepsilon m_\sigma$  et  $p - \varepsilon m_\sigma$  ont les mêmes marges et ont leur support dans  $U$ . L'ensemble  $U$  n'est donc pas un ensemble d'unicité.

Inversement, supposons que  $G(U)$  soit un sous-arbre et que  $U$  ne soit pas un ensemble d'unicité. Alors il existe deux probabilités  $p_1$  et  $p_2$  à support dans  $U$  de mêmes marges et distinctes. La mesure  $\frac{p_1 - p_2}{2}$  étant non nulle et de marges nulles, il existe, d'après le théorème 5, une partie  $C$  de  $U$  de la forme  $C = \sigma^+ \cup \sigma^-$  où

$$\sigma^+ = \{(x_i, y_i) ; i = 1, 2, \dots, n\}$$

$$\sigma^- = \{(x_{i+1}, y_i) ; i = 1, 2, \dots, n\}$$

ce qui entraîne évidemment que  $G(U)$  possède un cycle.

### Corollaire 1

Si  $p$  est dans  $\mathcal{M}$ ,  $p$  est une extrémale de  $\mathcal{M}$  si et seulement si  $S(p)$  est un ensemble d'unicité.

### Démonstration

Si  $S(p)$  est un ensemble d'unicité, il est trivial que  $p$  est extrémale. Inversement, si  $p$  est extrémale, si  $S(p)$  n'est pas un ensemble d'unicité, d'après le théorème 9  $G(S(p))$  possède un cycle, qu'on associe à  $\sigma = \sigma^+ \cup \sigma^-$ ; et il est clair que

$$p = \frac{p + \varepsilon m_\sigma}{2} + \frac{p - \varepsilon m_\sigma}{2} \quad \text{ou} \quad \frac{p \pm \varepsilon m_\sigma}{2}$$

appartiennent à  $\mathcal{M}$  pour  $\varepsilon$  assez petit.

### Corollaire 2

La famille des ensembles d'unicité de  $X \times Y$  est inductive pour l'inclusion. Un ensemble d'unicité est un arbre seulement si son graphe associé est un arbre.

La démonstration est immédiate.

Le théorème 9 admet également pour corollaire la réciproque du théorème 4, annoncée à la fin du § 4. Rappelons qu'un cube  $C$  dans  $X \times Y$  est un ensemble de la forme  $C_1 \times C_2$ , où  $C_1$  et  $C_2$  sont des parties finies de  $X$  et  $Y$ . Si  $C_1$  a  $p$  éléments et  $C_2$  a  $q$  éléments, on avait posé

$$||C|| = p + q - 2 .$$

### Corollaire 3

$U$  est un ensemble d'unicité si et seulement si pour tout cube  $C$ ,  $U \cap C$  a au plus  $1 + ||C||$  éléments.

### Démonstration

Si  $U$  est un ensemble d'unicité, soit  $p$  une probabilité dont le support est tout  $U$ ; le problème de marges correspondant admettant évidemment  $p$  pour extrémale, il suffit d'appliquer le théorème 4 pour avoir le résultat.

Inversement si pour tout cube  $C$   $U \cap C$  a au plus  $1 + ||C||$  éléments, supposons que  $U$  ne soit pas un ensemble d'unicité. D'après le théorème 8, il existe deux suites  $(x_1, \dots, x_k)$  et  $(y_1, \dots, y_k)$  de points distincts de  $X$  et  $Y$  respectivement tels que  $(x_i, y_i)$  et  $(x_{i+1}, y_i)$  appartiennent à  $U$  pour tout  $i = 1, 2, \dots, k$ . Considérant le cube

$$C = \{x_1, \dots, x_k\} \cap \{y_1, \dots, y_k\}$$

on voit que  $1 + ||C|| = 2k - 1$

et que  $U \cap C$  possède au moins  $2k$  éléments, ce qui est la contradiction cherchée.

Voici maintenant une légère généralisation du théorème 9 et de son corollaire, qui, dans le cas où  $X$  et  $Y$  sont finis est due à Djoković [10]. Il s'agit du problème des marges avec contraintes : sur  $X \times Y$  on se donne des fonctions  $a(x, y)$  et  $b(x, y)$  telles que

$$0 \leq a(x, y) \leq b(x, y), \quad \sum_y a(x, y) \leq r(x) \leq \sum_y b(x, y)$$

et  $\sum_x a(x, y) \leq s(y) \leq \sum_x b(x, y)$ , et on désigne par  $\mathcal{M}_c$

l'ensemble des probabilités sur  $X \times Y$  contenues dans  $\mathcal{M}$ , telles que de plus

$$a(x, y) \leq p(x, y) \leq b(x, y).$$

Si  $p \in \mathcal{M}_c$ , on définit

$$S'(p) = \{(x, y) ; a(x, y) < p(x, y) < b(x, y)\}$$

### Théorème 10

Si  $p \in \mathcal{M}_c$ ,  $p$  est une extrémale de  $\mathcal{M}_c$  si et seulement si  $S'(p)$  est un ensemble d'unicité.

### Démonstration

Si  $S'(p)$  est un ensemble d'unicité, supposons que  $p$  ne soit pas extrémale. Alors  $p = \frac{p_1 + p_2}{2}$ , avec  $p_1$  et  $p_2$  dans  $\mathcal{M}_c$  et distincts. Alors si  $p_1(x, y) \neq p_2(x, y)$ , comme  $p_1(x, y)$  et  $p_2(x, y)$  appartiennent à l'intervalle  $[a(x, y), b(x, y)]$  celui-ci est d'intérieur non vide,  $p(x, y)$  appartient à cet intérieur et  $(x, y)$  est dans  $S'(p)$ . Donc  $S'(p)$  supporte la mesure de marges nulles  $(p_1 - p_2) / 2$  et ne peut pas être un ensemble d'unicité.

Inversement si  $p$  est une extrémale,  $S'(p)$  est un ensemble d'unicité car sinon, d'après le théorème 9,  $G(S(p))$  possède un cycle qu'on associe à  $\sigma = \sigma^+ \cup \sigma^-$  posant

$$\varepsilon = \inf \{ p(x, y) - a(x, y), b(x, y) - p(x, y) ; (x, y) \in \sigma^+ \cup \sigma^- \}$$

il est clair que

$$p = \frac{p + \varepsilon m_\sigma}{2} + \frac{p - \varepsilon m_\sigma}{2}, \text{ où } \frac{p \pm \varepsilon m_\sigma}{2} \in \mathcal{M}_c$$

ce qui contredit le fait que  $p$  est une extrémale.

### Remarques

1. Dans le cas où  $X$  et  $Y$  ont un nombre fini d'éléments  $n$  et si  $r(x) = s(y) = 1/n$  pour tous  $x$  et  $y$ , le théorème 9 et son corollaire 1 redonne naturellement le théorème de Birkhoff-Von-Neumann, qui établit que les extrémales de l'ensemble convexe des matrices bistochastiques sont les matrices de permutation.
2. On voit, d'après le théorème 9, que l'ensemble des points extrémaux de  $\mathcal{M}$  est fermé pour la convergence vague. En effet, si  $p_n \rightarrow p$ , les supports  $S(p_n)$  des  $p_n$  étant d'unicité, supposons que  $S(p)$  ne soit pas un ensemble d'unicité. Notons par  $\sigma = (\sigma^+, \sigma^-)$  un cycle tel que  $\sigma^+ \cup \sigma^- \subset S(p)$   
si  $\varepsilon = \inf \{p(x, y) ; (x, y) \in \sigma^+ \cup \sigma^-\}$   
pour  $n$  assez grand  $p_n(x, y) \geq \varepsilon/2$  pour tout  $(x, y)$  de  $\sigma^+ \cup \sigma^-$ ,  
ce qui contredit l'extrémalité de  $p_n$ .

## 7) Exemples d'extrémales dans le cas continu

1. La généralisation au cas continu de ces théorèmes 9 et 10 n'est pas apparente. Considérons en effet le cas où  $X = Y = [0, 1]$  et où  $r$  et  $s$  sont la mesure  $m$  de Lebesgue.

Considérons une application  $T : [0, 1] \rightarrow [0, 1]$  mesurable et préservant la mesure de Lebesgue ; la probabilité  $p_T$  c'est-à-dire telle que  $m(T^{-1}(B)) = m(B)$  sur  $X \times Y$  est définie par

$$p_T(A \times B) = m(A \cap T^{-1}(B))$$

pour tout couple de parties mesurables  $A$  et  $B$  de  $[0, 1]$ . Il est évident que  $p_T \in \mathcal{M}$ . De plus  $p_T$  est une extrémale de  $\mathcal{M}$ , car si

$$p_T = \frac{p_1 + p_2}{2}$$

avec  $p_1 \neq p_2$  et  $p_1$  et  $p_2 \in \mathcal{M}$ , alors il existe un couple  $A$  et  $B$  de parties mesurables de  $[0, 1]$  telle que, disons :

$$p_2(A \times B) < p_T(A \times B) = m[A \cap T^{-1}(B)]$$

Donc, à fortiori

$$p_2[(A \cap T^{-1}(B)) \times B] < m[A \cap T^{-1}(B)]$$

et donc

$$p_1[(A \cap T^{-1}(B)) \times B] > m(A \cap T^{-1}(B))$$

comme

$$\begin{aligned} p_1[A \cap T^{-1}(B) \times B] &\leq p_1[A \cap T^{-1}(B) \times [0, 1]] \\ &= m(A \cap T^{-1}(B)) \end{aligned}$$

on a la contradiction cherchée.

Il devient possible à partir de cet exemple de construire un grand nombre d'extrémales du problème. Partageons les segments  $X = [0, 1]$  et  $Y = [0, 1]$  en une infinité dénombrable d'ensembles mesurables  $(X_m)_{m=1}^{\infty}$  et  $(Y_n)_{n=1}^{\infty}$  de mesures positives.

On pose également

$$A_m = \bigcup_{k>m} X_k$$

$$B_n = \bigcup_{k>n} Y_k$$

Si  $A$  et  $B$  sont des parties de  $[0, 1]$  de mesures positives, munies de la mesure de Lebesgue normalisée par  $1/m(A)$  et  $1/m(B)$ , il est toujours possible de définir une application mesurable  $T : A \rightarrow B$  qui préserve la mesure, et de construire ainsi sur  $A \times B$  une probabilité  $p_T$  comme on l'a fait plus haut avec  $A = B = [0, 1]$ . On considère alors de telles applications

$$T_i : A_{i-1} \rightarrow Y_i$$

$$S_i : B_i \rightarrow X_i$$

et enfin la probabilité :

$$p = \sum_{i=1}^{\infty} m(A_{i-1}) m(Y_i) p_{T_i} + \sum_{i=1}^{\infty} m(B_i) m(X_i) p_{S_i}$$

dont on voit que  $p \in \mathcal{M}$ , et par une méthode analogue à celle utilisée ci-dessus, que  $p$  est extrémale.

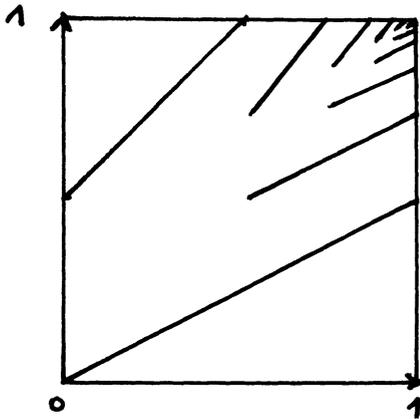


Figure n° 1

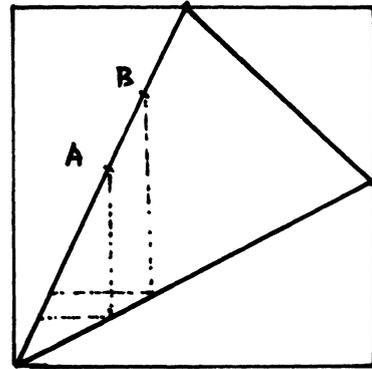


Figure n° 2

La figure n° 1 ci-dessus illustre la construction précédente ; ici les  $X_n$  et  $Y_n$  sont des intervalles de longueur  $2^{-n}$  et les  $T_n$  et  $S_n$  sont simplement des fonctions affines.

3. Mais la figure n° 2 montre que toutes les extrémales du problème ne sont pas de la forme ci-dessus. On considère les segments limités par le carré des trois droites  $y = 2x$  ,  $y = x/2$  ,  $x + y = 3/2$  . Soit  $F$  la réunion de ces trois segments,  $p$  une mesure portée par  $F$  ,  $r$  et  $s$  ses projections sur les côtés. (Pour avoir  $r = s =$  mesure de Lebesgue, il suffirait de charger chacun des trois segments suivant la mesure de Lebesgue de sorte que chacun d'eux soient de masse  $1/3$ ). Je dis que  $p$  est extrêmale de  $\mathcal{M}$  . En effet sinon il existe  $p_1$  et  $p_2$  dans  $\mathcal{M}$  distincts tels que  $p = \frac{p_1 + p_2}{2}$  .  $p_1 - p_2$  est donc une mesure bornée de marge nulle, portée par  $F$  .

Soient alors deux points  $A = (a, 2a)$  et  $B = (b, 2b)$  tels que le segment

$$I = \{(x, 2x) ; a \leq x \leq b\}$$

soit tel que :

$$(p_1 - p_2) (I) \neq 0$$

(Il est clair que A et B existent si  $p_1 \neq p_2$ ).

Alors

$$I' = \left\{ \left( x, \frac{x}{2} \right) ; a \leq x \leq b \right\}$$

satisfait à  $(p_1 - p_2) (I') + (p_1 - p_2) (I) = 0$

Donc posant

$$I_n = \left\{ (x, 2x) ; \frac{a}{2^n} \leq x \leq \frac{b}{2^n} \right\} \quad n = 0, 1, \dots$$

il est clair que les ensembles  $I_n$  sont disjoints et que

$$(p_1 - p_2) (I_n)$$

est une constante non nulle, ce qui est évidemment impossible

puisque  $p_1 - p_2$  est une mesure bornée.

4. On voit que l'exemple précédent construit un ensemble d'unicité dans le cas continu. Mais la difficulté à généraliser vient qu'on ignore quelle est la bonne notion de support à adopter. Ainsi, répondant à une question posée par Michael Keane, nous allons construire une extrêmale dont le support est le carré  $[0, 1] \times [0, 1]$ .

Partageons  $[0, 1]$  en une infinité dénombrable d'ensembles mesurables disjoints  $(X_n)_{n=1}^{\infty}$  tels que  $X_n$  soit dense dans  $[0, 1]$  et  $m(X_n)$  soit de mesure positive pour tout  $n$ . Soit  $(p_n/q_n)_{n=1}^{\infty}$  une énumération de tous les rationnels de  $[0, 1]$ ,  $f$  la fonction définie sur  $(0, 1]$  par

$$f(x) = p_n / q_n \quad \text{si} \quad x \in X_n$$

On définit ensuite  $g(x) = x + f(x)$  modulo 1. Soit  $\mu$  la probabilité transportée par  $g$  de la mesure de Lebesgue ; il n'y a pas d'atome à  $\mu$ , car si  $\mu(\{a\}) > 0$  il existe un ensemble  $A$  sur  $[0, 1]$  de mesure de Lebesgue non nulle telle que :

$$x + f(x) = a \quad \text{pour tout } x \in A .$$

Cela est impossible car alors il existe  $n$  tel que  $X_n \cap A$  soit de mesure positive et donc  $x = a - p_n / q_n$  pour tout  $x$  dans  $X_n \cap A$ . Donc la fonction  $F(x) = \mu([0, x])$  est continue en  $x$  et donc la fonction, définie sur  $[0, 1]$  :

$$T(x) = F \circ g(x) ,$$

conserve la mesure de Lebesgue.

Cependant l'ensemble des points  $\{(x, T(x)) ; x \in [0, 1]\}$  est dense dans le carré et donc la mesure  $p_T$ , définie par

$p_T(A \times B) = m(A \cap T^{-1}(B))$  dont nous savons par le 1 qu'elle est extrêmeale a pour support le carré tout entier.

5. Le problème de la caractérisation des extrêmeales dans le cas où  $r(dx)$  est discrète, c'est-à-dire ne comprend que des atomes et où  $s(dy)$  est sans atome est presque trivial.

Sans perte de généralité, on peut supposer que  $s(dy)$  est la mesure de Lebesgue concentrée sur le segment  $Y = [0, 1]$  et considérer  $X =$  ensemble des atomes de  $r(dx)$ . Si  $\mathcal{M}$  est l'ensemble des probabilités sur  $X \times Y$  de marges  $r$  et  $s$   $p$  dans  $\mathcal{M}$  sera une extrêmeale si et seulement si il existe une partition  $(A_x)_{x \in X}$  de  $Y$  en ensembles mesurables tels que pour tout  $x \in X$  on ait :

$$p(\{x\} \times A_x) = r(x) .$$

## 8) Extrêmes de Fréchet-Barbelés

Une contribution significative dans le problème des marges est due à Maurice Fréchet (voir [15] et ses références). L'expression "extrême de Fréchet" se réfère à une certaine inégalité (voir proposition ci-dessous) mais il se trouve que c'est aussi une extrême au sens convexe du mot. Soient  $r_1, r_2, \dots, r_n$  des probabilités sur  $\mathbb{R}$  et désignons par  $\mathcal{M}$  l'ensemble convexe des probabilités sur  $\mathbb{R}^n$  de marges  $r_1, r_2, \dots, r_n$ .

Soit  $p$  la probabilité sur  $\mathbb{R}^n$  définie par l'égalité :

$$p \left( \prod_{i=1}^n (-\infty, x_i] \right) = \inf_i r_i \left( (-\infty, x_i] \right) \text{ pour tous } x_i. \text{ Alors}$$

### Proposition 3

La probabilité  $p$  est une extrême de  $\mathcal{M}$  et tout élément  $p'$  de  $\mathcal{M}$  satisfait à  $p'(A) \leq p(A)$  pour tout  $A$  de la forme  $A = \prod_{i=1}^n (-\infty, x_i]$ .

### Démonstration

Il est immédiat que  $p$  appartient à  $\mathcal{M}$  et satisfait à l'inégalité annoncée. Pour vérifier que  $p$  est une extrême de  $\mathcal{M}$ , il suffit de voir que si  $p = \frac{p_1 + p_2}{2}$ , avec  $p_1$  et  $p_2 \in \mathcal{M}$  alors d'après l'inégalité, on a :

$$p_1(A) \leq p(A) \quad \text{et} \quad p_2(A) \leq p(A)$$

pour tout  $A = \prod_{i=1}^n (-\infty, x_i]$ .

Ceci entraîne  $p_1(A) = p_2(A) = p(A)$  pour tout  $A$  et donc  $p_1 = p_2$  ce qui montre que  $p$  est extrême.

On voit que pour la construction de  $p$ , on a utilisé la structure d'ordre de  $R$  là où elle joue un rôle peu naturel. Aussi, si on munit  $R$  d'autres structures d'ordre (total) de sorte que les segments  $\{x ; a \prec x \prec b\}$  restent mesurables, va-t-on obtenir d'autres extrémales. C'est d'ailleurs ainsi que procède M. Fréchet en inversant l'ordre de  $R$  pour obtenir l'autre "extrémale de Fréchet".

Obtient-on toutes les extrémales de  $\mathcal{M}$  ainsi ? Evidemment non : voir l'exemple n° 3 du paragraphe précédent, ou bien celui-ci :

$$r(\{i\}) = \frac{1}{3} \quad i = 0, 1, 2$$

$$s(\{j\}) = \frac{1}{4} \quad j = 0, 1, 2, 3$$

Définissons  $p$  dans  $R^2$  par

$$p(\{i, 0\}) = \frac{1}{12} \quad i = 0, 1, 2$$

$$p(\{i, i+1\}) = \frac{1}{4} \quad i = 0, 1, 2.$$

D'après le théorème 9,  $p$  est extrémale de  $\mathcal{M}$ , et un examen patient de tous les cas de figures possibles montre qu'il est impossible d'ordonner  $R$  pour que  $p$  soit une extrémale de Fréchet (le théorème en donnera une autre démonstration et montrera qu'il n'y a pas de meilleur contre-exemple).

Nous nous proposons dans ce paragraphe (dans le cas où  $X$  et  $Y$  sont dénombrables) de déterminer les cas où une extrémale est une extrémale de Fréchet.

Un ordre sur un ensemble dénombrable  $X$  est une application bijective des entiers relatifs  $Z$  dans  $X$  notée  $m \rightarrow x_m$ . On posera

$$X_m = \{x_k ; k \leq m\}$$

Si un ordre est défini sur l'ensemble dénombrable  $Y$  on définit de même  $Y_n$ .

### Définition

Une partie  $F$  de  $X \times Y$  est appelée ensemble de Fréchet si il existe des ordres sur  $X$  et  $Y$  tels que pour tous  $m$  et  $n$  on ait

$$\text{ou bien } (X_m \times Y) \cap F \subset X_m \times Y_n$$

$$\text{ou bien } (X \times Y_n) \cap F \subset X_m \times Y_n$$

Il est clair que si  $p$  est une probabilité ayant son support dans un ensemble de Fréchet,  $r$  et  $s$  désignant ses marges

$$p(X_m \times Y_n) = \inf(r(X_m), s(Y_n))$$

ce qui, grâce à la proposition 5 montre que  $p$  est extrême et que tout ensemble de Fréchet est d'unicité.

Il est amusant de vérifier directement ce résultat en montrant qu'un ensemble de Fréchet est sans cycle ; soit  $F$  un ensemble de Fréchet contenant une suite de points de la forme

$$(x_{m_i}, y_{m_i})_{i=1}^k \quad \text{et} \quad (x_{m_{i+1}}, y_{n_i})_{i=1}^k$$

(où les  $m_i$  et  $n_i$  sont tous distincts). Il faut montrer qu'il existe un couple de points de  $F$   $(x_{\alpha_1}, y_{\beta_1})$  et  $(x_{\alpha_2}, y_{\beta_2})$  tel que

$$(\alpha_1 - \alpha_2)(\beta_1 - \beta_2) < 0. \text{ Prenant } m = \min(\alpha_1, \alpha_2) \text{ et } n = \min(\beta_1, \beta_2)$$

on verra que la définition d'un ensemble de Fréchet est contredite.

Supposons donc que

$$(n_i - n_j)(m_i - m_j) > 0 \text{ pour tous } i \text{ et } j (i \neq j)$$

Alors  $\prod_{i < j} (n_i - n_j) \prod_{i < j} (m_i - m_j) > 0$

et nécessairement - en convenant  $m_{k+1} = m_i$

$$\prod_{i < j} (n_i - n_j) \prod_{i < j} (m_{i+1} - m_{j+1}) < 0$$

ce qui démontre le résultat.

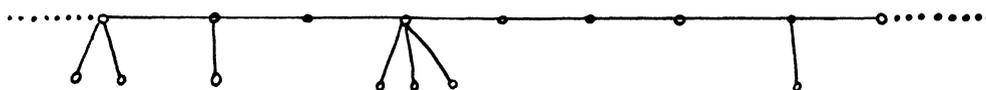
Pour caractériser les ensembles de Fréchet parmi les ensembles d'unicité de façon maniable, nous introduisons quelques définitions supplémentaires.

La distance  $d(x, y)$  entre deux sommets d'un arbre est le nombre d'arêtes de l'unique chemin joignant  $x$  et  $y$ .

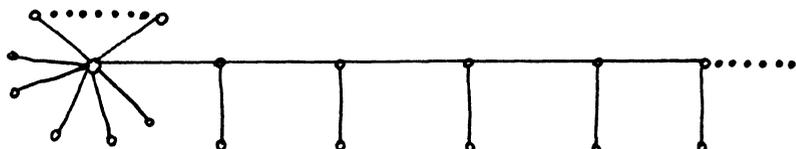
Définition

Un arbre est appelé un barbelé s'il contient un sous-graphe  $S$  isomorphe à un segment, à  $\mathbb{N}$  ou à  $\mathbb{Z}$  tel que

- 1) Tout sommet de l'arbre est dans  $S$  ou à une distance 1 d'un sommet de  $S$
- 2) Si  $s$  n'est pas une extrémité de  $S$   $\{x ; d(s, x) = 1\}$  est fini.

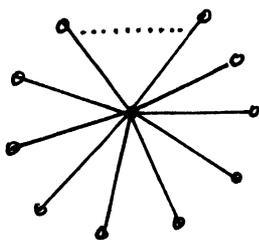


Exemple n° 1



Exemple n° 2

On voit tout de suite qu'il y a trois sortes de barbelés : les "doublement infinis" qui ont un sous-graphe isomorphe à  $\mathbb{Z}$ , les "simplement infinis" qui contiennent un sous-graphe isomorphe à  $\mathbb{N}$  mais ne sont pas doublement infinis, et les autres (qui ne sont pas nécessairement finis), comme le montre l'exemple de l'oursin ci-dessous.



Exemple n° 3

Un sous-barbelé est un sous-graphe d'un barbelé. Attention : à la différence des sous-arbres, deux barbelés ne font pas un sous-barbelé : deux exemples n° 1 ne peuvent être connectés pour former un barbelé.

Voici le théorème qui permet de diagnostiquer les extrémales de Fréchet :

### Théorème 11

L'ensemble  $F$  est un ensemble de Fréchet si et seulement si son graphe est un sous-barbelé.

Nous utilisons le lemme suivant :

### Lemme

Si  $F$  est un ensemble de Fréchet,  $X$  et  $Y$  étant munis de l'ordre correspondant, alors les entiers  $\alpha_n \geq -\infty$  et  $\beta_n \leq +\infty$

$$\alpha_n = \inf \{m ; (x_n, y_m) \in F\}$$

$$\beta_n = \sup \{m ; (x_n, y_m) \in F\}$$

s'ils sont définis, satisfont à :

$$1) (x_{n'}, y_m) \in F \quad \text{et} \quad n' < n \quad \text{implique} \quad m \leq \alpha_n$$

$$2) (x_{n'}, y_m) \in F \quad \text{et} \quad n' > n \quad \text{implique} \quad m \geq \beta_n$$

#### Démonstration du lemme

Si il existe  $n' < n$  et  $m > \alpha_n$  tels que  $(x_{n'}, y_m) \in F$ , alors  $(n' - n)(m - \alpha_n) < 0$ , ce qui contredit le fait que  $F$  soit un ensemble de Fréchet. On procède de même pour le 2).

#### Démonstration du théorème

Soit  $F$  un ensemble de Fréchet. Sans perte de généralité on peut supposer que l'ensemble des entiers  $n$  tels que  $\alpha_n$  et  $\beta_n$  soient définis est de la forme  $(-\infty, 0]$ ,  $[0, +\infty)$  ou  $(-\infty, +\infty)$ . On construit alors sur  $X \times Y$  l'ensemble  $F_1$  défini par

$$F_1 = \bigcup_n \{(x_n, y_m) ; \alpha_n \leq m \leq \beta_n\} .$$

Il est clair d'après le lemme que  $F_1$  est un ensemble de Fréchet convexe. Il suffit de voir que son graphe associé est un barbelé, ce qui est immédiat.

Réciproquement, supposons que le graphe de  $F$  soit un sous-barbelé. Alors il est toujours possible de construire une partie  $F_1$  de  $X \times Y$  contenant  $F$  telle que le graphe de  $F_1$  soit un barbelé. La seule difficulté, en complétant le graphe de  $F$ , est de vérifier qu'on peut ajouter de nouvelles arêtes qui connectent le graphe de  $F$  en ne joignant

toujours que des points de  $X$  à des points de  $Y$ . Les deux figures ci-dessous donnent les deux possibilités (typiques)

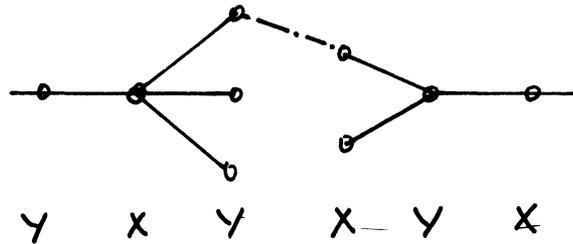


Figure n° 1

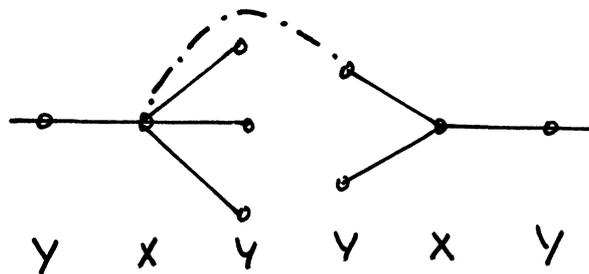


Figure n° 2

On ordonne ensuite les éléments de  $X \cup Y$  en fixant une origine  $0$  sur  $X \cup Y$  et en procédant comme l'indique la figure ci-dessous

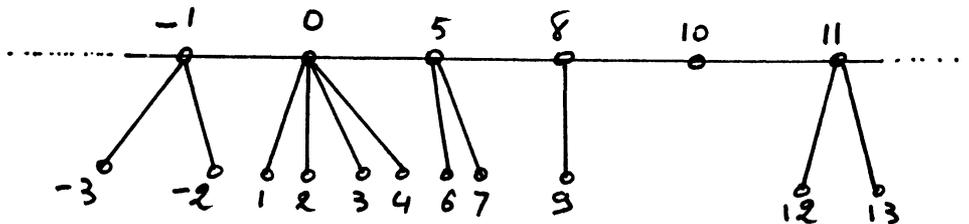


Figure n° 3

Cette numérotation induit sur  $X$  et  $Y$  des ordres convenables qui font de  $F_1$  et à fortiori de  $F$  des ensembles de Fréchet.

### Corollaire

La famille des ensembles de Fréchet de  $X \times Y$  est inductive pour l'inclusion, et un ensemble de Fréchet est maximal, seulement si son graphe est un barbelé.

## 9) La formule du relèvement

Dans ce paragraphe nous répondons - d'une manière théorique - à la question suivante :

Etant données des marges  $r$  et  $s$  sur les ensembles dénombrables  $X$  et  $Y$ , et un ensemble d'unicité  $U$  de  $X \times Y$  existe-t-il une probabilité  $p$  de  $\mathcal{M}$ , portée par  $U$  ? Si oui, nous savons qu'elle est unique ; comment la calculer ? La méthode a l'avantage de donner une autre démonstration du théorème 9 (voir [18] à ce sujet). Cette méthode s'expose de façon naturelle dans le cadre des arbres. Soit donc un arbre d'ensemble de sommets  $S$  et d'ensemble d'arêtes  $A$ . On se donne deux applications :

$$m : S \rightarrow R$$

$$p : A \rightarrow R$$

satisfaisant à :

$$\sum_{x \in S} |p(x, y)| < \infty \quad (1)$$

$$\text{et } \sum_y p(x, y) = m(x) \quad \forall x \in S \quad (2)$$

Naturellement la sommation n'est faite que pour les  $y$  de  $S$  tels que  $(x, y)$  soit une arête. Pour énoncer le prochain lemme, quelques définitions sont nécessaires. Ici  $d(x, y)$  est la distance entre les sommets définie au § 8. Soient  $x$  et  $y$  deux sommets voisins de l'arbre et  $z$  un sommet quelconque. On a toujours :

$$|d(x, z) - d(y, z)| = 1$$

Nous disons que le couple (ordonné)  $(y, x)$  est dirigé vers  $z$  si  $d(y, z) = d(x, z) + 1$ . On note  $B_n(x, y)$  l'ensemble des points  $z$  situés à une distance  $n$  de  $x$  et tels que le couple  $(y, x)$  soit dirigé

vers  $z$ . Enfin si  $B$  est un ensemble de sommets, on note

$$m(B) = \sum_{x \in B} m(x). \text{ Nous avons le lemme suivant :}$$

### Lemme 1

Si  $m$  et  $p$  satisfont (1) et (2), alors pour tout entier  $n \geq 0$

$$p(x, y) = \sum_{k=0}^n (-1)^k m(B_k(x, y)) \quad (3)$$

$$+ (-1)^{n+1} \sum \{ p(z_1, z_2) ; z_1 \in B_n(x, y), z_2 \in B_{n+1}(x, y) \}$$

### Démonstration

Nous procédons par récurrence sur  $n$ , le lemme se réduit en effet à la formule (2) pour  $n = 0$ . Supposons (3) vraie pour  $n-1$ . Alors, d'après (2) si  $z_1 \in B_{n-1}(x, y)$  et  $z_2 \in B_n(x, y)$  on a :

$$p(z_1, z_2) = - \sum_{z_3} p(z_3, z_2) + m(z_2) \quad (4)$$

où la somme sur  $z_3$  est prise pour  $z_3$  décrivant  $B_1(z_2, z_1)$ . Or, on remarque que :

$$z_1 \in B_{n-1}(x, y), z_2 \in B_n(x, y) \text{ et } z_3 \in B_1(z_2, z_1)$$

entraîne  $z_3 \in B_{n+1}(x, y)$ . Il suffit donc de remplacer dans l'expression (3) correspondant à  $n-1$   $p(z_1, z_2)$  par l'expression (4) pour avoir la formule (3) correspondant à  $n$ , et achever la preuve du lemme.

### Remarque

Le lemme reste valable en enlevant la condition (1) et en remplaçant par la condition que tout sommet n'appartient qu'à un nombre fini d'arêtes, cas fréquent.

### Théorème 12

Si  $\sum_{x \in S} |m(x)| < \infty$  il existe au plus une fonction  $p(x, y)$  définie sur l'ensemble des arêtes satisfaisant (1) et (2). Cette fonction  $p$  existe seulement si pour tous  $x$  et  $y$  voisins, on a

$$\sum_{k=0}^{\infty} (-1)^k m(B_k(x, y)) = \sum_{k=0}^{\infty} (-1)^k m(B_k(y, x)) \quad (5)$$

et  $p$  est alors égale à la valeur commune des deux membres de cette égalité.

### Démonstration

Supposons que  $p$  existe. Je dis que  $p(x, y)$  est nécessairement égal au premier membre de (5), ce qui démontre l'unicité. Pour cela, il suffit de remarquer que puisque le graphe  $A$  est un arbre

$$z_n \xrightarrow[n \rightarrow \infty]{} +\infty \quad \text{si } z_n \in B_n(x, y)$$

( $z_n \xrightarrow[n \rightarrow \infty]{} +\infty$  signifie que la suite sort de tout ensemble fini).

La condition (1) étant réalisée on a donc, d'après le lemme

$$p(x, y) - \sum_{k=0}^n (-1)^k m(B_k(x, y)) \xrightarrow[n \rightarrow \infty]{} 0$$

ce qui démontre l'égalité annoncée. Puisque  $p(x, y) = p(y, x)$  l'égalité (5) est nécessairement réalisée.

Arrêtons nous sur l'égalité (5).

Fixons un sommet quelconque  $x_0$  pour origine et désignons par  $X$  l'ensemble des sommets de l'arbre situés à une distance paire de  $x_0$  par  $Y$  le complémentaire de  $X$  dans  $S$ . Alors on a :

### Proposition

Si  $\sum_{x \in S} |m(x)| < \infty$ , l'égalité (5) est réalisée pour toute arête si et seulement si  $m(X) = m(Y)$ .

### Démonstration

On pose

$$p(x, y) = \sum_{k=0}^{\infty} (-1)^k m(B_k(x, y))$$

Remarquons que l'ensemble des sommets situés à la distance  $n$  du sommet est exactement  $B_n(x, y) \cup B_{n-1}(y, x)$  et que  $B_i(x, y)$  et  $B_j(y, x)$  sont disjoints pour tous  $i$  et  $j$

$$X = \bigcup_{k=0}^{+\infty} B_{2k}(x, y) \cup B_{2k+1}(y, x)$$

$$Y = \bigcup_{k=0}^{\infty} B_{2k+1}(x, y) \cup B_{2k}(y, x)$$

De ceci il découle que

$$m(X) - m(Y) = p(x, y) - p(y, x)$$

pour toute arête  $(x, y)$ , ce qui démontre la proposition.

### Théorème 13

Supposons  $m(x) \geq 0$  pour tout  $x$ ,  $\sum_{x \in S} m(x) < \infty$  et  $m(X) = m(Y)$ . Alors  $p(x, y)$  défini par

$$p(x, y) = \sum_{k=0}^{\infty} (-1)^k m(B_k(x, y))$$

pour toute arête  $(x, y)$ , tend vers zéro si  $(x, y)$  tend vers l'infini.

De plus si  $p(x, y) \geq 0$  pour toute arête, nécessairement  $p$  satisfait (1) et (2).

### Démonstration

Le fait que  $p(x, y) \xrightarrow{(x,y) \rightarrow \infty} 0$  a déjà été démontré (voir [18], théorème 8). Le reste du théorème 13 y est énoncé ([18], théorème 7) mais démontré avec insuffisamment de rigueur.

Montrons que si  $p(x, y) \geq 0$ ,  $p$  satisfait à (2), ce qui entraînera naturellement que (1) est vérifié.

On peut écrire :

$$p(x, y) + \sum_{z \in B_1(x, y)} p(x, z) = \sum_{k=0}^{\infty} (-1)^k m(B_k(x, y)) + \sum_{z \in B_1(x, y)} \left[ \sum_{k=0}^{\infty} (-1)^k m(B_k(z, x)) \right]$$

(Remarquer l'inversion  $B_k(z, x)$  dans la dernière somme, justifiée par  $m(X) = m(Y)$  et la proposition précédente)

$$\text{Or } \bigcup_{z \in B_1(x, y)} B_k(z, x) = B_{k+1}(x, y) .$$

Les ensembles  $\{B_k(z, x) ; k = 0, 1, \dots, z \in B_1(x, y)\}$  sont disjoints. Comme  $\sum_x m(x) < \infty$  nous inversons les sommations pour écrire :

$$\sum_y p(x, y) = m(B_0(x, y)) - \sum_{k=0}^{\infty} (-1)^k B_{k+1}(x, y) + \sum_{k=0}^{\infty} (-1)^k \sum_{z \in B_1(x, y)} m(B_k(z, x)) = m(x)$$

ce qui achève la démonstration du théorème.

L'application du théorème 13 au problème des marges est alors immédiate. Soit  $U$  un ensemble d'unicité ; les composantes connexes du graphe associé à  $U$  partagent  $S = X \cup Y$  en ensembles disjoints  $(S_\alpha)_{\alpha \in I}$  ( $I$  ensemble d'indices fini ou dénombrable) de la forme  $S_\alpha = X_\alpha \cup Y_\alpha$ .

U induit sur  $S_\alpha$  un arbre. Si  $r$  et  $s$  sont des mesures de probabilité, on définit  $m$  sur  $S_\alpha$  par  $m(x) = r(x)$  si  $x \in X_\alpha$  et  $m(y) = s(y)$  si  $y \in Y_\alpha$ . On peut alors énoncer

### Corollaire

Il existe sur U une probabilité de marges  $r$  et  $s$  si et seulement si

$$1^\circ) r(X_\alpha) = s(Y_\alpha) \quad \forall \alpha \in I$$

$$2^\circ) p(x, y) = \sum_{k=0}^{\infty} (-1)^k m_\alpha(B_k(x, y))$$

est positif ou nul pour tout  $(x, y)$  de  $X_\alpha \times Y_\alpha$  et pour tout  $\alpha$ , et  $p$  définit la probabilité cherchée.

### Remarque

On voit que si  $p$  est extrémale de  $\mathcal{M}$  alors pour tout  $(x, y)$  il existe  $A \subset X$  et  $A' \subset Y$  tels que  $p(x, y) = r(A) - s(A')$ , c'est-à-dire que  $p(x, y) \in F_r - F_s$ , où  $F_r$  et  $F_s$  sont les fermés images des probabilités  $r$  et  $s$ .

#### 10) Réunion d'un nombre fini d'ensembles d'unicité

Ce paragraphe est basé sur un étonnant résultat de Rado [24] .  
Soit  $E$  un espace vectoriel sur un corps  $K$  ; si  $A$  est une partie finie de  $E$  ,  $|A|$  est le nombre de ses éléments et  $r(A)$  est son rang.

#### Théorème

Une partie  $S$  de  $E$  est réunion de  $m$  ensembles indépendants si et seulement si pour tout  $A$  fini contenu dans  $S$  on a  $|A| \leq m r(A)$ .  
(Dans le cas où  $S$  est fini, une démonstration en est proposée en exercice dans N. Bourbaki, Algèbre (nouvelle édition 1970) chapitre 2, § 7 , exercice 7).

#### Théorème 14

Soient  $X$  et  $Y$  dénombrables et  $S$  une partie de  $X \times Y$  . Alors  $S$  est réunion de  $m$  ensembles d'unicité si et seulement si pour tout cube  $C$  de  $X \times Y$  ,  $S \cap C$  a au plus  $m(1 + ||C||)$  éléments.  
( $||C||$  a la signification habituelle des § 3 et § 6).

#### Démonstration

La partie directe du théorème découle immédiatement du corollaire n° 3 du théorème 10.

Nous passons à l'implication inverse, pour laquelle nous aurons besoin du théorème de Rado. Introduisons quelques notations.

Si  $A$  est un ensemble dénombrable, on note  $\ell_1(A)$  la partie de  $\mathbb{R}^A$  formée des suites  $(x_a)_{a \in A}$  sommables  $k(A) = \mathbb{R}^{(A)}$  est l'ensemble des suites nulles sauf sur un ensemble fini ; c'est un sous-espace de  $\ell_1(A)$ . Si  $B \subset A$  on identifie canoniquement  $\ell_1(B)$  à un sous-espace de  $\ell_1(A)$  ainsi que  $k(B)$  pour  $k(A)$ . On considère enfin l'application linéaire

$$h : \ell_1(X \times Y) \rightarrow \ell_1(X \cup Y)$$

définie par  $m \rightarrow h(m)$

où  $k(m)$  est la projection de  $m$  sur  $X$  et  $Y$  ; en d'autres termes si

$(e_{(x,y)})_{(x,y) \in X \times Y}$  est la base canonique de  $k(X \times Y)$  et  $(e_x)_{x \in X}$  et  $(e_y)_{y \in Y}$  celles de  $k(X)$  et  $k(Y)$ ,

$$m = \sum_{x,y} m(x,y) e_{(x,y)} \rightarrow h(m) = \sum_{x,y} m(x,y) (e_x + e_y)$$

Le noyau de  $h$  est donc constitué par les mesures de marges nulles.

Si  $S$  est une partie de  $X \times Y$ , on identifie  $S$  à une partie de  $k(X \times Y)$  par la correspondance  $(x, y) \rightarrow e_{(x,y)}$ , et  $h(S)$  désigne donc une partie de  $k(X \cup Y)$ . Naturellement  $h$  restreint à  $S$  est injectif.

#### Lemme 1

$r(h(C)) = 1 + ||C||$  si  $C$  est un cube.

La preuve est facile : c'est la conséquence du fait que si  $A_1, \dots, A_n$  est une partition d'une base d'un espace vectoriel, l'espace engendré par  $A_1 + A_2 + \dots + A_n$  est de codimension  $n-1$ . On peut aussi procéder comme au § 3.

Une partie  $A$  de  $h(S)$  est dite connexe si le graphe associé au sous-ensemble de  $X \times Y$  défini par  $A' = S \cap h^{-1}(A)$  est lui-même connexe. Soient  $X_1$  et  $Y_1$  les plus petites parties de  $X$  et  $Y$  satisfaisant à  $A' \subset X_1 \times Y_1$ . Si  $A$  est fini, on note par  $C(A)$  le cube  $X_1 \times Y_1$ . Il est immédiat que si  $A$  est une partie quelconque de  $h(S)$ ,  $A$  se décompose en composantes connexes  $A = \bigcup_{\alpha \in I} A_\alpha$ , et que si  $C(A_\alpha) = X_\alpha \times Y_\alpha$ , les ensembles  $(X_\alpha)_{\alpha \in I}$  sont disjoints, ainsi que les  $(Y_\alpha)_{\alpha \in I}$ .

### Lemme 2

$r(A) = 1 + ||C(A)||$  si  $A$  est une partie connexe de  $h(S)$ .

### Démonstration

Si  $C(A) = X_1 \times Y_1$  il nous suffit, compte tenu du lemme 1, de démontrer que l'espace engendré par  $A$  dans  $k(X_1 \cup Y_1)$  est de codimension 1. Soit  $f$  une forme linéaire sur  $k(X_1 \cup Y_1)$  qui s'annule en  $A$ . Alors si  $(x, y) \in S \cap h^{-1}(A) = A'$

on a  $f(e_{(x)}) + f(e_{(y)}) = 0$ .

On en déduit que  $f(e_x)$  est indépendant de  $x$  quand  $x$  décrit  $X_1$ . En effet si  $x$  et  $x'$  sont des points de  $X_1$ , par définition de  $C(A)$  il existe une suite  $(x_i, y_i)_{i=1}^k$  de points de  $A'$  telle que

1°)  $x_1 = x$  et  $x_k = x'$

2°) ou bien  $(x_{i+1}, y_i) \in A'$  pour tout  $i = 1, \dots, k-1$

ou bien  $(x_i, y_{i+1}) \in A'$  pour tout  $i = 1, \dots, k-1$ .

On en déduit que  $f(e_{(x)}) = f(e_{(x')})$  et on procède de même pour voir que  $f(e_{(y)}) = f(e_{(y')})$  pour tous  $y$  et  $y'$  dans  $Y_1$ , ce qui achève la preuve du lemme.

Soit maintenant  $S \subset X \times Y$  fixé. Supposons que pour tout cube  $C$ ,

$$|S \cap C| \leq m(1 + ||C||)$$

Prenons  $A$  une partie finie de  $h(S)$ , je dis que :

$$|A| \leq m r(A)$$

En effet, décomposons  $A$  en composantes connexes, soit  $A = \bigcup_{\alpha \in I} A_\alpha$ .

Alors

$$\begin{aligned} |A| &= \sum_{\alpha} |A_\alpha| \leq \sum_{\alpha} |S \cap C_\alpha| \leq \sum_{\alpha} m(1 + ||C_\alpha||) \\ &= m \sum_{\alpha} r(A_\alpha) = m r(A) \end{aligned}$$

La première égalité vient du fait que  $h$  restreint à  $S$  est bijective, l'égalité suivante est le lemme 2, la dernière vient du fait que les espaces engendrés par les  $A$  sont en somme directe, puisque les  $X_\alpha$  et  $Y_\alpha$  correspondants sont disjoints. D'après le théorème de Rado appelé au début, on en déduit que  $h(S)$  se partage en  $m$  ensembles indépendants et donc qu'il existe une partition  $S_1, S_2, \dots, S_m$  de  $S$  telle que  $h(S_i)$  soit indépendant pour  $i = 1, 2, \dots, m$ .

Maintenant il est clair que dire " $h(S_i)$  est un ensemble indépendant" est dire " $h$  restreint à  $k(S_i)$  est injectif". Or, et ceci est fondamental, nous avons vu au théorème 9 que :

$h$  restreint à  $k(S_i)$  est injectif entraîne que  $h$  restreint à  $l(S_i)$  est injectif.

## Remarques

1. Il est important de remarquer que si  $p$  est une mesure de marges  $r$  et  $s$  tel que le support  $S(p)$  soit réunion de  $m$  ensembles d'unicité,  $p$  n'est pas cependant nécessairement le barycentre de  $m$  extrémales, car ces ensembles d'unicité ne portent pas nécessairement une probabilité (de marges  $r$  et  $s$  (voir § 9).

Voici deux contre-exemples :

Si  $X = Y = \{1, \dots, n\}$ ,  $r(k) = s(k) = \frac{1}{n}$  pour tout  $k$  et  $p$  est la mesure produit, il est facile de constater que  $\frac{n}{2} < m < \frac{n}{2} + 1$ , alors que  $p$  est barycentre d'au moins  $n$  extrémales.

Si  $X = Y = \{1, 2, \dots, n, \dots\}$ ,  $r(k) = s(k) = \frac{1}{2^k}$  pour tout  $k$  et  $S = \{(x, y) ; |x - y| \leq 1\}$ , alors  $m = 2$  et il est facile de construire sur  $S$  une probabilité qui ne soit pas barycentre d'un nombre fini d'extrémales.

2. On adapte sans difficulté la première partie de cette démonstration pour démontrer le résultat suivant :

Soit  $(S, A)$  un graphe connexe quelconque. Si  $C$  est une partie finie de  $S$ ,  $a(C)$  est le nombre d'arêtes du sous-graphe plein associé, c'est-à-dire que  $a(C)$  est le nombre d'éléments  $(x, y)$  de  $A$  tels que  $\{x, y\} \subset C$ . Alors il existe des parties  $S_1, \dots, S_m$  de  $S$  telles

que  $\bigcup_{i=1}^m S_i = S$  et, tels que les sous-graphes pleins associés  $(S_i, A_i)$  soient des arbres et tels que  $\bigcup_{i=1}^{\infty} A_i = A$  si et seulement si  $a(C) \leq m(|C| - 1)$  pour tout  $C$  fini contenu dans  $S$ .

Le seul point nouveau est ici qu'il faut vérifier que  $a(C) \leq |C| - 1$  pour tout  $C$  fini est la condition nécessaire et suffisante pour qu'un graphe connexe  $(S, A)$  soit un arbre.

3. Il est facile de généraliser la première partie de la démonstration du théorème au problème des marges à  $n$  composantes et non plus 2. Prenons  $n = 3$  pour simplifier. Alors de la même façon on définit

$$h : \ell_1(X \times Y \times Z) \rightarrow \ell_1(X \cup Y \cup Z)$$

qui à  $e_{(x,y,z)}$  associe  $e_x + e_y + e_z$

Mais la phrase soulignée dans la deuxième partie du théorème nous fait nous heurter à la difficulté fondamentale : si  $S$  est tel que  $h$  est injectif sur  $k(S)$ , est-ce que  $h$  est injectif sur  $\ell_1(S)$  ?

Appelons mesure de marges nulles sur  $X \times Y \times Z$  un élément  $m$  de  $\ell_1(X \times Y \times Z)$  tel que

$$\sum_{y,z} m(x_0, y, z) = \sum_{x,z} m(x, y_0, z) = \sum_{x,y} m(x, y, z_0) = 0$$

pour tous  $(x_0, y_0, z_0)$ . La question précédente peut encore se formuler ainsi :

Question : Existe-t-il toujours une mesure bornée de marges nulles à support fini et absolument continue par rapport à une mesure de marges nulles donnée quelconque ?

Nous avons vu que c'est vrai pour  $n = 2$  et avons pu le démontrer parce que nous avons une idée précise des mesures de marges nulles à support fini minimales (voir théorème 8); ce sont les  $m_\sigma$  : on pouvait soit utiliser un résultat sur les chaînes de Markov (théorème 6), soit la formule de relèvement pour les arbres.

Pour  $n \geq 3$ , nous ne savons rien sur la structure combinatoire des ensembles ne pouvant pas porter de mesures de marges nulles finies, et la question reste sans réponse. Je conjecture néanmoins que la réponse à la question posée est encore oui pour  $n \geq 3$ .

## 11) Ensembles d'unicité finis

Dans ce paragraphe, nous étudions quelques procédés d'identification des ensembles d'unicité et nous faisons quelques remarques sur la cardinalité de l'ensemble des extrémales.  $X$  et  $Y$  seront ici des ensembles finis à  $p$  et  $q$  éléments respectivement.

Commençons par une définition :

### Définition

$E \subset X \times Y$  est appelé ensemble d'existence si quels que soient les fonctions  $r$  et  $s$  sur  $X$  et  $Y$  respectivement telles que

$$\sum_x r(x) = \sum_y s(y)$$

il existe une fonction  $p$  sur  $E$  telle que  $\sum_y p(x_0, y) = r(x_0)$  et

$$\sum_x p(x, y_0) = s(y_0) \text{ pour tous } x_0 \text{ et } y_0 .$$

Il n'y a aucune difficulté à caractériser les ensembles d'existence:

### Proposition 4

$E$  est un ensemble d'existence si et seulement si les deux conditions suivantes sont réalisées

- 1°) les projections canoniques de  $E$  sur  $X$  et  $Y$  sont  $X$  et  $Y$
- 2°) le graphe associé à  $E$  est connexe

### Démonstration

Si  $E$  est un ensemble d'existence, la première condition est évidente. Si le graphe associé à  $E$  n'est pas connexe, il existe  $X_1 \subset X$  et  $Y_1 \subset Y$  tel que  $E \subset (X_1 \times Y_1) \cup (X'_1 \times Y'_1)$  où  $X'_1$  et  $Y'_1$  sont les

complémentaires de  $X_1$  et  $Y_1$  dans  $X$  et  $Y$  respectivement. Prenant  $r$  et  $s$  tels que  $\sum_{x \in X_1} r(x) \neq \sum_{y \in Y_1} s(y)$ , il est clair que  $E$  n'est pas un ensemble d'existence.

Inversement, si les deux conditions sont réalisées, on peut trouver un sous graphe du graphe associé à  $E$  qui soit un arbre et qui passe par tous les sommets. D'après la formule de relèvement du § 9 (théorème 12), il est clair que l'ensemble  $E_1$  correspondant à ce sous graphe est un ensemble d'existence.

### Corollaire

Les ensembles d'existence minimaux sont les ensembles d'unicité maximaux.

Ces ensembles d'existence sont connus en analyse de la variance sous le nom de plans d'expérience à deux facteurs (voir Barra [2], chapitre IX). On trouve dans le travail de Tuan Pham Dinh [22] un atlas des ensembles d'existence minimaux, et donc, au vu du corollaire, d'ensembles d'unicité maximaux, pour  $p$  et  $q \leq 6$ .

### Proposition 5

Il y a exactement  $p^{q-1} q^{p-1}$  ensembles d'unicité maximaux.

Ceci est une conséquence d'une remarquable formule d'Austin [1].

Soient  $S_1, S_2, \dots, S_m$  une partition d'un ensemble fini  $S$ ; le nombre d'éléments de  $S_i$  est  $p_i$  et  $n = \sum_{i=1}^m p_i$ ; alors le nombre d'arbres  $(S, A)$  tels que quel que soit  $i$ , il soit impossible d'avoir  $\{x, y\} \subset S_i$  et  $\{x, y\} \in A$ , est égal à

$$n^{m-2} \prod_{i=1}^m (n - p_i)^{p_i - 1}$$

Le cas particulier de cette formule  $m = 2$  a été retrouvé deux ans plus tard par Scoins [26]. Il s'énonce : si  $X$  et  $Y$  ont  $p$  et  $q$  éléments, le nombre d'arbres  $(X \cup Y, A)$  tels que toute arête connecte un élément de  $X$  et un élément de  $Y$ , est  $p^{q-1} q^{p-1}$ . Cette formule est passée à la postérité sous le nom de formule de Scoins. Elle donne évidemment la preuve de notre proposition.

#### Remarque 1

En généralisant la formule exponentielle d'Hurwitz (voir Foata-Schutzenberger [14], chapitre 3) on pourrait montrer que le nombre  $a_{p,q}$  d'ensembles d'unicité de  $X \times Y$  est donné par la relation :

$$\exp \sum_{p,q \geq 1} a_{p,q} x^p y^q / p! q! = 1 + \sum_{p,q \geq 1} p^{q-1} q^{p-1} x^p y^q / p! q!$$

Nous donnerons les détails dans une autre publication.

#### Remarque 2

Bien entendu, le nombre  $p^{q-1} q^{p-1}$  donne une borne supérieure beaucoup trop grande du nombre d'extrémales de  $\mathcal{M}_b$ , puisque seulement certains ensembles d'unicité portent une probabilité de  $\mathcal{M}_b$ . Le nombre exact d'extrémales est une fonction probablement fort compliquée de  $r$  et  $s$ , comme dans tous les problèmes de dénombrement analogues dans le cas où  $r$  et  $s$  prennent des valeurs entières : calculs du nombre de matières à coefficients 0 ou 1, ou à coefficients entiers positifs, de marge donnée (voir Ryser [25] ou Berge [4] à ce sujet).

### Remarque 3

Si  $F_r$  et  $F_s$  sont les fermés, image des probabilités  $r$  et  $s$ , on a vu en fin du § 9 qu'une extrémale prend ses valeurs dans  $F_r - F_s$ . Si  $p = \lambda q$ , où  $\lambda$  est entier, et si  $r(x) = 1/p$  et  $s(y) = 1/q$  pour tous  $x$  et  $y$ , on voit que  $p$  ne peut prendre que les valeurs  $1/\lambda q$ . Ceci montre qu'on obtient toutes les extrêmes en découpant  $X$  en  $\lambda$  ensembles égaux, et que le nombre d'extrêmes est alors  $p! / \lambda!$ . Une méthode analogue permet de donner une numération compliquée du nombre d'extrêmes si  $r(x) = 1/p$  et  $s(y) = 1/q$ , avec  $p$  et  $q$  quelconque : si  $p \geq q$ , on pose :

$$p = a_1 q + r_1 \quad q = a_2 r_1 + r_2, \quad r_1 = a_3 r_2 + r_3, \quad \dots,$$

$$r_{k-1} = a_{k-1} r_k + 0 \text{ ou } 1$$

suivant l'algorithme usuel de la division euclidienne. Le nombre d'extrêmes est alors supérieur ou égal à

$$p! q! / (a_1!)^q (a_2!)^{r_1} \dots (a_{k+1}!)^{r_k} r_{k-1}!$$

Il n'est pas impossible que si  $p + q = N$ , le nombre d'extrêmes soit maximum lorsque le rapport  $p / q$  est le plus proche de  $(1 + \sqrt{5}) / 2$ .

### Remarque 4

$F_r \cap F_s = \{0, 1\}$  est la condition nécessaire et suffisante pour que les supports de toutes les extrêmes soient des ensembles d'unicité maximaux.

### Remarque 5

On peut démontrer que si  $(\alpha, \beta)$  est un couple de permutations de

$X$  et  $Y$  tel que  $(x, y) \rightarrow (\alpha(x), \alpha(y))$  conserve un ensemble d'unicité fini  $U$ , alors  $\alpha$  ou  $\beta$  ont un point fixe.

Voici maintenant quelques algorithmes destinés à vérifier qu'une partie  $U$  de  $X \times Y$  est un ensemble d'unicité.

Soit  $U$  une partie de  $X \times Y$ , on considère l'application  $a_U$  de l'ensemble des cubes de  $X \times Y$  dans lui-même qui au cube  $C = X_1 \times Y_1$  associe le cube  $X_1 \times Y_2$ , où  $Y_2$  est l'ensemble des  $y$  de  $Y_1$  tels qu'il existe  $x$  et  $x'$  dans  $X_1$  tels que  $(x, y)$  et  $(x', y)$  soient dans  $U$ . On définit  $b_U$  en inversant les rôles de  $X$  et  $Y$  et on pose

$$c_U = a_U \circ b_U$$

### Proposition 6

$U$  est d'unicité si et seulement si il existe  $n$  tel que  $C_U^n(X \times Y) = \emptyset$ .

Transcrite sur le graphe associé à  $U$ , l'opération  $C_U$  revient à enlever les bouts pendants du graphe, et il est clair que toutes les arêtes disparaissent si et seulement si il n'y a pas de cycle. Cet algorithme est très rapide pour des ensembles compliqués. Naturellement, c'est pour les extrémales de Fréchet qu'il est le plus long.

	2					3	
2		5	4				
					1		
			3		2		
	1						
				2			2
		5				3	
1							

Exemple n° 1

					2		
	1		x				x
		1				1	
x			x	1			x
					2		
x			x				
4					3		
2							

Exemple n° 2

Les chiffres des exemples indiquent l'instant où le point correspondant a été balayé, l'ensemble des croix la limite de  $c_U^n(X \times Y)$ .

Pour énoncer la dernière proposition, quelques rappels d'algèbre de Boole sont nécessaires. Soit  $E$  un ensemble à  $n$  éléments, identifié à  $\{1, \dots, n\}$ ;  $F_2 = \{0, 1\}$  est le corps à deux éléments,  $\mathcal{P}$  est l'ensemble des polynomes (dits booléens) sur le corps  $F_2$  à  $n$  variables  $x_1, \dots, x_n$  et de degré  $\leq 1$  par rapport à chaque variable. Les éléments  $P$  de  $\mathcal{P}$  sont identifiés aux fonctions-polynomes correspondantes qui appliquent  $F_2^E$  dans  $F_2$ . Il est donc évident que toute famille  $\mathcal{F}$  de parties de  $E$  peut-être identifiée par un unique polynôme booléen  $P$ , en ce sens que :

$$P(x_1, \dots, x_n) = 1 \text{ si et seulement si } \{i ; x_i = 1\} \in \mathcal{F}$$

Un polynôme booléen s'écrit donc :

$$P(x) = \sum_{T \subset E} c(T) \prod_{i \in T} x_i$$

où  $T$  décrit les sous-ensembles de  $E$  et  $c(T) \in F_2$ . Le degré de comple-

xité d'une famille  $\mathcal{F}$  s'évalue à l'aide de constantes variées :

$$d(\mathcal{F}) = \sup_{T \subset E} |T| \overline{c(T)}$$

$$s(\mathcal{F}) = \sum_{T \subset E} |T| \overline{c(T)}$$

$$m(\mathcal{F}) = \sum_{T \subset E} \overline{c(T)}$$

(Ici  $|T|$  est le nombre d'éléments de  $T$ ,  $\overline{c(T)} = 0$  ou  $1$  dans  $\mathbb{R}$  suivant que  $c(T) = 0$  ou  $1$  dans  $F_2$ ). On remarque que ces constantes sont insensibles au changement  $\mathcal{F} \rightarrow p(E) \setminus \mathcal{F}$ , qui change  $P$  en  $1 + P$ . Toutes ces notions sont extraites de l'article [20] et on peut généraliser le théorème 3 de [20] pour obtenir le lemme suivant, que nous ne démontrons pas.

#### Lemme

Si  $\mathcal{F}$  est une famille de parties telle que  $A \in \mathcal{F}$  et  $B \supset A$  entraîne  $B \in \mathcal{F}$ , et si  $M$  est un élément maximal de  $p(E) \setminus \mathcal{F}$ , alors

$$d(\mathcal{F}) > n - |M|, \quad s(\mathcal{F}) > (n - |M|) 2^{n-|M|} \quad \text{et} \quad m(\mathcal{F}) > 2^{n-M}.$$

Nous allons appliquer ces considérations au cas où

$$E = X \times Y \quad \text{avec} \quad |X| = p \quad \text{et} \quad |Y| = q$$

#### Proposition 7

Si  $\mathcal{F}$  est la famille des ensembles d'unicité, alors  $d(\mathcal{F}) > (p - 1)(q - 1)$ . Si  $\mathcal{F}$  est la famille des ensembles d'existence, alors  $d(\mathcal{F}) > (p - 1)(q - 1) - 1$ . Les inégalités pour  $s$  et  $m$  sont analogues.

### Démonstration

Si  $\mathcal{F}$  est la famille des ensembles d'unicité, notons  $\mathcal{F}' = \mathcal{P}(X \times Y) \setminus \mathcal{F}$ . On a vu que  $d(\mathcal{F}) = d(\mathcal{F}')$ . Soit  $U$  un ensemble d'unicité maximal. Il a nécessairement  $p + q - 1$  éléments car un arbre à  $n$  sommets à  $n-1$  arêtes. En appliquant le lemme à  $\mathcal{F}'$  et  $U$  on a le résultat.

Si  $\mathcal{F}$  est la famille des ensembles d'existence, prenons  $(x_0, y_0) \in X \times Y$  et définissons

$$M = [X \setminus \{x_0\}] \times \{y_0\} \cup \{x_0\} \times [Y \setminus \{y_0\}]$$

Il est clair que quel que soit  $(x, y) \notin M$  le graphe associé à  $M \cup \{(x, y)\}$  est connexe ; en appliquant le lemme à  $\mathcal{F}$  et  $M$  on a le résultat.

## 12) Remarques concernant l'existence

Soient  $(\Omega_n, \mathcal{A}_n, P_n)_{n \in I}$  des espaces de probabilité, où  $I$  est un ensemble fini ou dénombrable,  $(\Omega, \mathcal{A}, P')$  leur produit  $\mathcal{M}$  l'ensemble convexe des probabilités sur  $(\Omega, \mathcal{A})$  de marges  $P_n$ . Nous avons le théorème suivant :

### Théorème 15

Soit  $C = \prod_{n \in I} A_n$ ,  $A_n \in \mathcal{A}_n$ . Alors  $P(C) > 0$  pour tout  $P$  de  $\mathcal{M}$  si et seulement si

$$\sum_{n \in I} [1 - P_n(A_n)] < 1.$$

### Corollaire

Si le complémentaire de  $E \in \mathcal{A}$  contient un  $\prod_{n \in I} A_n$ , avec  $\sum_{n \in I} [1 - P_n(A_n)] < 1$ , il n'existe aucun  $P \in \mathcal{M}$  porté par  $E$ .

### Démonstration

Si  $\sum_{n \in I} [1 - P_n(A_n)] < 1$ , supposons qu'il existe  $P$  dans  $\mathcal{M}$  tel que  $P(C) = 0$ .

Alors :

$$1 = P(\Omega \setminus C) \leq \sum_{n \in I} [1 - P_n(A_n)] < 1,$$

ce qui est impossible.

La démonstration de la réciproque est plus intéressante. Supposons que  $P(C) > 0$  pour tout  $P$  de  $\mathcal{M}$  et que  $\sum_{n \in I} [1 - P_n(A_n)] \geq 1$ .

Nous devons distinguer deux cas :

1) ou bien il existe un ensemble fini  $I' \subset I$  tel que

$$\sum_{n \in I'} [1 - P_n(A_n)] \geq 1$$

2) ou non

Nous étudions d'abord le 2ème cas ;  $I$  est alors infini =  $\{1, 2, \dots\}$ .

On a alors :

$$\sum_{n=1}^{\infty} P_n(\Omega_n \setminus A_n) = 1. \quad (1)$$

Définissons  $A'_n = A_1 \times A_2 \times \dots \times A_{n-1} \times (\Omega_n \setminus A_n) \times A_{n+1} \times \dots$ . Rappelons que  $P'$  désigne la probabilité produit. On définit  $P$  sur  $(\Omega, \mathcal{A})$  par :

$$P(B) = \sum_{n=1}^{\infty} P'(B | A'_n) P_n(\Omega_n \setminus A_n).$$

Si  $B$  est de la forme  $B = \Omega_1 \times \dots \times \Omega_{m-1} \times B_m \times \Omega_{m+1} \times \dots$ , alors :

$$P'(B | A'_n) = P_m(B_m | A_m) \quad \text{si } n \neq m$$

$$P'(B | A'_m) = P_m(B_m | \Omega_m \setminus A_m)$$

$$P(B) = P_m(B_m | A_m) \left[ \sum_{n \neq m} P_n(\Omega_n \setminus A_n) \right] + P_m(B_m \cap \Omega_m \setminus A_m)$$

et donc compte tenu de (1),  $P(B) = P_m(B_m)$ , ce qui entraîne que

$P \in \mathcal{M}$ . Il est évident que  $P(C) = 0$ , ce qui contredit l'hypothèse.

Nous étudions le premier cas ; il existe un entier  $N$  tel que

$$\sum_{n=1}^N P_n(\Omega_n \setminus A_n) \geq 1 \quad (2)$$

Le reste de la démonstration repose sur un lemme qui demande quelques notations.

Soit  $E = \{0, 1\}^N$ . Un élément  $e$  de  $E$  est noté  $e = (X_1(e), \dots, X_N(e))$  où  $X_N(e)$  est 0 ou 1. Si  $E$  est probabilisé,  $X_n$  devient une variable aléatoire. Voici le lemme :

Lemme

Si  $\epsilon_1, \dots, \epsilon_N$  sont des nombres de  $[0, 1]$  tels que  $\epsilon_1 + \dots + \epsilon_N \geq 1$ , il existe une probabilité  $p$  sur  $E$  telle que :

$$p(X_n = 0) = \epsilon_n \text{ pour tout } n \text{ et } p((1, 1, \dots, 1)) = 0$$

Démonstration du lemme

Nous procédons par récurrence sur  $N$ . L'énoncé étant trivialement vrai pour  $N = 1$ , supposons le vrai pour  $N - 1$ . Deux cas peuvent se produire

- 1°) ou bien  $s = \epsilon_1 + \dots + \epsilon_N \leq 1 + \epsilon_n$  pour tout  $n$  (3)
- 2°) ou non

Plaçons nous d'abord dans le premier cas.

Posons  $a_n = \frac{N - s}{N - 1} - 1 + \epsilon_n$ . Il découle de (3) que  $a_n \geq \frac{N-2}{N-1}(s-1) \geq$

d'autre part  $\sum_{n=1}^N a_n = \frac{N-s}{N-1} \leq 1$ . Posons enfin

$$p[(0, 0, \dots, 0)] = \frac{s}{N-1} \text{ et}$$

$$p[(1, \dots, 1, 0, 1, \dots, 1)] = a_n \text{ quand le zéro occupe le rang } n ;$$

on vérifie facilement que  $p$  répond à la question.

Plaçons nous maintenant dans le second cas. Il existe donc un  $n$  - que sans perte de généralité nous prenons égal à  $N$  - pour que l'inégalité soit fautive. On a donc

$$\varepsilon_1 + \dots + \varepsilon_N > 1 + \varepsilon_N \quad (4)$$

D'après l'hypothèse de récurrence, il existe donc une probabilité  $p'$  sur  $E' = \{0, 1\}^{N-1}$  telle que

$$p'(1, \dots, 1) = 0 \text{ et } p'(X_n = 0) = \varepsilon_n \text{ si } n < N .$$

On pose alors :

$$p(e', 0) = p'(e') \varepsilon_N ,$$

$$p(e', 1) = p'(e') (1 - \varepsilon_N) , \text{ avec } e' \in E'$$

et on vérifie facilement que  $p$  répond à la question.

Nous pouvons achever la preuve du théorème. A tout  $e$  de  $E$  on associe

$$A(e) = A_1(e) \times \dots \times A_N(e) \times \Omega_{N+1} \times \dots ,$$

$$\text{où } A_n(e) = A_n \text{ si } X_n(e) = 1$$

$$A_n(e) = \Omega_n \setminus A_n \text{ si } X_n(e) = 0 .$$

Si  $p$  est la probabilité construite sur  $E$  par le lemme, on définit  $P$  sur  $(\Omega, \mathcal{C})$  par

$$P(B) = \sum_{e \in E} P'(B | A(e)) p(e) .$$

On vérifie facilement que  $P \in \mathcal{M}$ , et le fait que  $p((1, \dots, 1)) = 0$  entraîne  $P(C) = 0$ .

Nous allons maintenant tirer d'un théorème de V. Strassen et d'une de ses conséquences due à P.A. Meyer, une condition nécessaire et suffisante pour que si  $E \subset X \times Y$  ( $X$  et  $Y$  finis ou dénombrables munis de probabilités  $r$  et  $s$ ) il existe  $p$  dans  $\mathcal{M}$  dont le support soit dans  $E$  ( $\mathcal{M}$  est l'ensemble des probabilités de marges  $r$  et  $s$  comme d'habitude). Si  $X_1$  est une partie de  $X$ , on définit  $\alpha_E(X_1)$  comme l'ensemble des points  $y$  de  $Y$  tels qu'il existe  $x$  dans  $X_1$  et  $(x, y) \in E$ .

### Théorème 16

Il existe  $p$  dans  $\mathcal{M}$  à support dans  $E$  si et seulement si :

$$r(X_1) \leq s(\alpha_E(X_1)) \text{ pour tout } X_1 \subset X.$$

### Démonstration

Le "seulement si" est trivial, puisque

$$(X_1 \times Y) \cap E \subset X \times \alpha_E(X_1)$$

La réciproque se démontre par le théorème 52 de [2] qui s'énonce ici :

"Il existe  $p$  à support dans  $\mathcal{M}$  si et seulement si pour toute fonction bornée  $f$  sur  $Y$ , alors

$$r(f) = \sum_x r(x) f(x) \leq \sum_y s(y) \hat{f}(y) = s(\hat{f})$$

où  $\hat{f}(y) = \sup \{f(x) ; (x, y) \in E\}$  "

On remarque que  $\hat{1}_{X_1} = 1_{\alpha_E(X_1)}$  ; pour en déduire que  $r(f) \leq s(\hat{f})$  pour tout  $f$ , il suffit de remarquer que

$$\widehat{f_1 + f_2} = \hat{f}_1 + \hat{f}_2 \text{ si } f_1(y) f_2(y) = 0$$

On conclut par un argument de densité.

### Corollaire

$E \subset X \times Y$  est tel que  $p(E) > 0$  pour tout  $p$  de  $\mathcal{M}$  si et seulement si  $E$  contient un ensemble  $X_1 \times Y_1$  avec  $r(X_1) + s(Y_1) > 1$

### Remarque

Il semble qu'il y ait peu de différence entre ce corollaire et le théorème 15 : il y a pourtant nécessairement entre les deux un théorème délicat comme le théorème de Strassen ou bien le théorème de Ford - Fulkerson de la coupe minimale.

### Démonstration du corollaire

La partie "si" est une conséquence immédiate de la partie facile du théorème 15. Inversement supposons  $p(E) > 0$  pour tout  $p$  de  $\mathcal{M}$  et que, si  $X_1 \times Y_1 \subset E$

$$\text{alors } r(X_1) + s(Y_1) \leq 1.$$

Puisque  $p(E) > 0$  pour tout  $p$  de  $\mathcal{M}$ , nécessairement  $E' = (X \times Y) \setminus E$  n'est jamais le support d'un élément de  $\mathcal{M}$  et donc il existe  $X_1 \subset X$  tel que :

$$s(\alpha_{E'}(X_1)) < r(X_1).$$

En posant  $Y_1 = Y \setminus \alpha_{E'}(X_1)$ , on a le couple  $(X_1, Y_1)$  cherché.

### Bibliographie du chapitre III

- [1] AUSTIN T.L.  
"The enumeration of point labelled chromatic graphs and trees"  
Canadian Journal of Mathematics, vol. 12 (1960), pp. 535-545
  
- [2] BARRA Jean-René  
Notions fondamentales de Statistique Mathématique. Dunod (1971)
  
- [3] BASS Jean  
"Sur la compatibilité des fonctions de répartition"  
Comptes rendus de l'Académie des Sciences, vol. 240 (1965)  
pp. 839-843
  
- [4] BERGE Claude  
"Théorie des graphes et applications". Dunod (1958)
  
- [5] BILLINGSLEY Patrick  
"Ergodic Theory and Information". Wiley (New-York) (1965)
  
- [6] BUCK, R. Creighton  
"Linear Spaces and approximation theory" in "On numerical Approximation", R.E. LANGER Editeur, publié par les University of Wisconsin Press (1958) pp. 11-23
  
- [7] CARTIER, Pierre
  
  
  
  
  
  
  
  
  
  
- [8] CHOQUET, Gustave  
"Le problème des Moments". Séminaire d'Initiation à l'Analyse,  
lère année. Secrétariat Mathématique de l'I.H.P. (Paris) 1962

- [9] DILIBERTO, S.P. et STRAUS, E.G.  
"On the approximation of a function of several variables by the sum of functions of fewer variables".  
Pacific Journal of Mathematics, Vol. 1 (1951) pp. 195-210
- [10] DJOKOVIĆ D.Z.  
"Extreme points of certain convex polytopes". Monatshefte für Mathematik, Band 69, Heft 5 (1965), pp. 385-388
- [11] DOUGLAS, R.G.  
"On extremal measures and subspace density". Michigan Mathematical Journal, 11 (1964), pp. 243-46
- [12] FELLER William  
"An introduction to probability theory and its applications"  
Vol. 1, 3<sup>rd</sup> edition, Wiley (New-York) (1968)
- [13] FISHBURN Peter J;  
"Additive representations of Real functions on product sets"  
Journal of combinatorial theory. Vol. 4, n° 4 (1968)  
pp. 397-402
- [14] FOATA, D. et SCHUTZENBERGER, M.  
"Théorie géométrique des polynômes eulériens"  
Lecture notes, vol. 138. Springer Verlag (Berlin), 1970
- [15] FRECHET Maurice  
"Sur les tableaux de corrélation dont les marges et les bornes sont données". Annales de l'Université de Lyon, 3<sup>ème</sup> série, Sciences, section A, 1958, pp. 19-32
- [16] HENNEQUIN P.L. et TORTRAT A.  
"Théorie des probabilités et quelques applications"  
Masson (Paris), 1965
- [17] LETAC Gérard  
"Problèmes de Probabilité", problème n° 70  
Presses Universitaires de France, Paris (1970)

- [18] LETAC Gérard  
"Représentation des mesures de probabilité de deux espaces dénombrables, de marges données"  
Illinois Journal of Mathematics, vol. 10, n° 3, Sept. 1966,  
pp. 497-507
- [19] LINDENSTRAUSS Joram  
"A remark on extreme doubly stochastic measures"  
American Mathematical Monthly 72, 4 (1965) pp. 379-382
- [20] Mc KENZIE R., MYCIELSKI J. and THOMPSON D.  
"On Boolean functions and connected sets". Mathematical systems theory, vol. 5, n° 3 (1971), pp. 259-270
- [21] MEYER Paul-André  
"Probabilité et Potentiel". HERMANN (Paris) 1966
- [22] PHAM DINH Tuan  
"Contributions à l'Analyse de la variance et aux plans d'expérience". Thèse de spécialité. Grenoble 1970.
- [23] PHELPS, Robert B.  
"Lectures on Choquet's theorem". VAN NOSTRAND (New-York) 1966
- [24] RADO R.  
"On a combinatorial theorem on vector spaces"  
Journal of London Mathematical society (1962)
- [25] RYSER Herbert-John  
"Combinatorial Mathematics". Carus mathematical monographs,  
Wiley (New-York) 1965
- [26] SCOINS  
"The number of trees with modes of alternate parity"  
Proceedings of Cambridge Philosophical Society 58 (1962)  
pp. 12-18

- [27] STRASSEN Victor  
"The existence of probability measures with given marginals"  
Annals of Mathematical Statistics. Vol. 36 (1965), pp. 423-439
- [28] TITS Jacques  
"Sur le groupe des automorphismes d'un arbre" In "Mémoires dédiés  
à Georges de Rham" pp. 188-211, Springer Verlag, Berlin 1970
- [29] VAROPOULOS Nicolas  
"Tensor algebras and Harmonic analysis". Acta Mathematica 119  
(1967), pp. 51-112
- [30] VOROBEV, N.N.  
"Consistent families of measures and their extensions"  
Theory of Probability and its applications. Vol. 7, n° 2 (1962)  
pp. 147-163.