Probabilités rationnelles

1 Mesures de probabilité sur les mots

Soit A un alphabet fini, A^* le monoïde libre sur A.

Une *mesure de probabilité* sur A^* est une application

$$\mu : A^* \to R_+$$

vérifiant les conditions

$$\sum_{a \in A} \mu(a) = 1 \quad (1)$$

$$\sum_{a \in A} \mu(ua) = \mu(u),\ u \in A^* \quad (2)$$

L'application μ s'étend par additivité à toutes les parties de A^* en posant

$$\mu(U) = \sum_{u \in U} \mu(u),\ U \subset A^*$$

En particulier, il résulte de (1) et (2) que, pour tout $n \in N$,

$$\mu(A^n) = 1$$

La mesure μ est *invariante* si, outre (1) et (2), elle vérifie les conditions

$$\sum_{a \in A} \mu(au) = \mu(u),\ u \in A^* \quad (3)$$

Transparents de l'exposé de Georges Hansel, "Séries rationnelles positives et probabilités", au Groupe de Travail "Algèbres Tropicales", 17 Dec. 1999.

2 Exemples

1) Une mesure μ est *de Bernoulli* si c'est un morphisme :

$$\mu(a_1 a_2 \dots a_n) = \prod_{i=1}^{n} \mu(a_i) \tag{4}$$

Une telle mesure est (évidemment) invariante.

2) Une mesure *de Markov* est donnée par un couple (λ, P), où

$P = (p(a,b))$ est une $A \times A$-matrice stochastique (ligne)

$\lambda = (\lambda(a))_{a \in A}$ est un A-vecteur ligne stochastique

La mesure μ est alors définie par les formules

$$\mu(b) = \sum_{a \in A} \lambda(a) p(a,b),\ b \in A$$

puis, par récurrence sur la longueur du mot $u \in A^*$

$$\mu(uab) = \mu(ua) p(a,b),\ u \in A^*,\ a, b \in A$$

La mesure μ est invariante si le vecteur λ est « invariant » , c'est-à-dire si

$$\lambda P = \lambda$$

Une mesure de Bernoulli est (évidemment) de Markov (avec $p(a,b) = \lambda(b)$)

3 Mesures de probabilité et séries formelles positives

Una application $\mu : A^* \to R_+$ est une *série formelle (positive)* en les variables non commutatives $a \in A$. On la note classiquement

$$\mu = \sum_{u \in A^*} \mu(u)u$$

Sur l'ensemble $R_+ < A >$ de ces séries on définit des opérations

• La somme

• Le produit (qui généralise le produit des polynômes en variables non commutatives) : si σ et τ sont deux séries formelles, leur produit $\sigma\tau$ est défini par la formule

$$\sigma\tau(w) = \sum_{uv=w} \sigma(u)\tau(v),\ w \in A^*$$

• L'étoile (qui correspond au passage à l'inverse pour les fractions rationnelles) : soit μ une série *propre*, c'est-à-dire telle que $\mu(1) = 0$ (1 le mot vide) ; *l'étoile* de μ est la série

$$\mu^* = 1 + \mu + \mu^2 + \ldots + \mu^n + \ldots$$

μ^* est bien définie grâce au fait que μ est propre.

Une série formelle est *rationnelle* si elle peut être obtenue à partir des polynômes au moyen d'un nombre fini d'opérations.

D'où la définition : Une mesure de probabilité $\mu : A^* \to R_+$ est *rationnelle* si la série formelle correspondante l'est.

Exemple. Une mesure de Bernoulli est rationnelle :

$$\mu = \sum_{u \in A^*} \mu(u)u = (\sum_{a \in A} \mu(a)a)^*$$

On va montrer que les probabilités de Markov sont également rationnelles. Pour cela on a besoin du

Théorème 1 (Schutzenberger)

Soit A un alphabet fini et soit $\mu : A^* \to R_+$ une série formelle. Les conditions suivantes sont équivalentes :

a) μ est rationnelle

b) μ est "reconnaissable" : il existe

- *un ensemble fini Q (les "états")*
- *un morphisme $\phi : A^* \to R_+^{Q \times Q}$ dans le monoïde des $Q \times Q$-matrices à coefficients dans R_+*
- *un vecteur-ligne $\lambda = (\lambda(q))_{q \in Q}$*
- *un vecteur-colonne $\nu = (\nu(q))_{q \in Q}$*

tels que

$$\mu(u) = \lambda \phi(u) \nu, \ u \in A^*$$

Le triplé (λ, ϕ, ν) est une *représentation* de la série μ.

Exemple Si μ est une mesure de Bernoulli, tout est trivial.

On prend pour Q un singleton $Q = \{\omega\}$. On pose $\lambda(\omega) = \nu(\omega) = 1$

Le morphisme ϕ est défini par $\phi(u) = [\mu(u)]$, $u \in A^*$.

Proposition 2 *Une mesure de probabilité de Markov μ est rationnelle.*

Démonstration. Soit μ une probabilité de Markov définie par le couple (λ, P) $(\lambda(a)_{a\in A}$ vecteur stochastique, $P = (p(a,b)$ matrice de transition). On définit alors une représentation de μ :

• Ensemble des états : $Q = A$

• Le morphisme $\phi : A^* \to R_+^{Q\times Q}$ est défini en se donnant les matrices $\phi(a)$, $a \in A$:

$$\phi(a)(a', a'') = \begin{cases} p(a', a'') & \text{si } a = a'' \\ 0 & \text{sinon} \end{cases}$$

• Le vecteur colonne ν est défini par les conditions $\nu(a) = 1$ pour tout $a \in A$.

On vérifie que (λ, ϕ, ν) est une représentation de μ.

Remarque. On a $\sum_{a\in A} \phi(a) = P$.

4 Généralisation

Proposition 3 *Soit* $\mu : A^* \to R_+$ *une série formelle. Les conditions suivantes sont équivalentes*

a) μ *est une mesure de probabilité rationnelle*

b) Il existe une représentation (λ, ϕ, ν) *de* μ *telle que*

(i) $m = \sum_{a \in A} \phi(a)$ *est une matrice stochastique*

(ii) $(\lambda(q))$ *est un vecteur stochastique*

(iii) $\nu(q) = 1$ *pour tout* $q \in Q$.

De plus μ *est invariante si et seulement si* λ *peut être choisi tel que* $\lambda m = \lambda$.

Exemple On prend $A = \{a, b\}$ et soient $p, q \geq 0$ tels que $p + q = 1$.

Je pose :

$$m = \begin{pmatrix} p & q \\ 1 & 0 \end{pmatrix}, \quad \lambda = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad \nu = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

On définit le morphisme ϕ par

$$\phi(a) = \begin{pmatrix} p & 0 \\ 1 & 0 \end{pmatrix}, \quad \phi(b) = \begin{pmatrix} 0 & q \\ 0 & 0 \end{pmatrix}$$

Le triplet (λ, ϕ, ν) est la représentation d'une mesure de probabilité rationnelle invariante sur A^*.

5 Mesure de probabilité image

Définition. Soient A et B deux alphabets et soit $f : B^* \to A^*$ un morphisme "alphabétique" ($f(B) \subset A$).
Soit $\mu' : B^* \to R_+$ une mesure de probabilité.
Alors (presque évidemment) l'application $\mu : A^* \to R_+$ définie par

$$\mu(u) = \mu'(f^{-1}(u)),\ u \in A^*$$

est une mesure de probabilité appelée *image de* μ'.
Si μ' est invariante, il en est de même de μ.

Proposition 4 *L'image d'une mesure de probabilité rationnelle est rationnelle.*

Démonstration. Si μ' est une probabilité rationnelle, on remplace dans l'expression finie donnant μ' chaque lettre $b \in B$ par la lettre $f(b)$ et on obtient une expression finie donnant μ.

Grâce à la proposition 3, on peut démontrer la

Proposition 5 *Soit* $\mu : A^* \to R_+$ *une mesure de probabilité. Les conditions suivantes sont équivalentes :*

a) μ *est rationnelle*

b) μ *est image d'une mesure de probabilité de Markov.*

6 Codes et mesures de probabilité rationnelles

Définition. Une partie X de A^* est un *code* si tout mot $u \in A^*$ admet au plus une factorisation $u = x_1 x_2 \dots x_n$ en éléments de X. On note X^* tous les mots de A^* qui sont des produits d'éléments de X.

Exemples

$X = \{a, ba, bb\}$ est un code (préfixe)

$Y = \{a, ba, ab\}$ n'est pas un code : $a(ba) = (ab)a$

Définition. Soit $X \subset A^*$. Une mesure de probabilité μ sur A^* est dite *X-multiplicative* si on a

$$\mu(xy) = \mu(x)\mu(y) \text{ pour tous } x, y \in X^*$$

Proposition 6 *Soit X un code et μ une mesure de probabilité X-multiplicative. Alors $\mu(X) \leq 1$.*

Démonstration. Pour tout $Y \subset A^*$, on définit la série formelle en la variable t

$$f_Y(t) = \sum_n \mu(Y \cap A^n)t^n$$

Comme $\mu(Y \cap A^n) \leq 1$, elle converge pour tout $t < 1$.
Comme X est un code, les ensemble X^n, $n \in N$, sont disjoints et donc

$$f_{X^*} = \sum_n f_{X^n}$$

De ce que μ est X-multiplicative, on déduit

$$f_{X^n} = (f_X)^n$$

Donc pour tout $t < 1$, on a

$$f_{X^*}(t) = \sum_n (f_X(t))^n = \frac{1}{(1 - f_X(t))}$$

ce qui implique

$$f_X(t) = \sum_n \mu(X \cap A^n)t^n \leq 1$$

En faisant tendre t vers 1, on obtient

$$\mu(X) \leq 1 \qquad \square$$

On est intéressé au cas où l'égalité est atteinte, de sorte que, X étant pris pour nouvel alphabet, μ définit une probabilité de Bernoulli sur X^*.

Définition. Un code X est *maximal* s'il n'existe pas de partie $Y \subset A^*$ contenant strictement X qui soit encore un code. Il est *coupant* s'il existe au moins un mot $u \in A^*$ qui n'est facteur d'aucun mot de X.

Exemple : tout code fini est coupant.

On peut alors démontrer (généralisation d'un résultat de Shutzenberger) :

Proposition 7 *Soit X un code coupant, μ une mesure de probabilité sur A^* invariante et X-multiplicative.*

Alors X est un code maximal si et seulement si $\mu(X) = 1$

Exemple : Le « petit code »

$$X = \{a^2, ba, b^2, ba^2, b^2a\}$$

est un code maximal : pour n'importe quelle mesure de Bernoulli sur A^*, on a $\mu(X) = 1$.

Remarque. Toute probabilité de Bernoulli sur A^* est (évidemment) toujours X-multiplicative, quel que soit le code X.

On peut donc se demander :

1) Existe-t-il une probabilité X-multiplicative invariante qui ne soit pas Bernoulli, X étant un code maximal coupant ?

Réponse : Oui.

$A = \{a, b, c\}$, μ une probabilité de Markov invariante définie par une matrice bi-stochastique de la forme

$$P = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ p & q & r \\ s & t & u \end{pmatrix}$$

et on prend pour code X = ensemble des mots terminant par un a sans en avoir avant : $X = \{b, c\}^* a$.

2) Même question, mais avec X **fini**.

Dans le cas général, ce problème est ouvert.

Cependant, si X est un code préfixe, il est résolu négativement :

Théorème 8 (Langlois) *Soit $X \subset A^*$ un code préfixe maximal fini. Toute mesure de probabilité sur A^* qui est invariante et X-multiplicative est une mesure de Bernoulli.*

(Principe de la) **Démonstration.**

Soit une lettre $a \in A$. Comme X est maximal fini, il existe $n \in N$ tel que $a^n \in X$.

Mieux : comme X est préfixe maximal fini, il existe un multiple $m = kn$ tel que

$$ba^m \subset X^* a^* \text{ pour toute lettre } b \in A$$

Pour toute $b \in A$, on définit une matrice $M(b)$ carrée d'ordre n par

$$M(b)(i,j) = \begin{cases} \mu(ba^{m+i-j}) & \text{si } ba^{m+i-j} \in X^* \\ 0 & \text{sinon} \end{cases} \quad (0 \leq i,j \leq n-1)$$

$$\text{On pose} \quad M = \sum_{b \in A} M(b)$$

On va maintenant appliquer le théorème de Perron-Frobenius.

1) On vérifie que la matrice M est irréductible.

2) Soit $w \in A^*$ un mot **quelconque**. On montre que le vecteur colonne $\mathbf{v} = (\mu(a^i w))_{0 \leq i \leq n-1}$ est un vecteur propre associé à la valeur propre $\mu(a^n)$.

3) D'après Perron-Frobenius, le vecteur $\mathbf{v}$ est unique à une constante multiplicative près, ce qui entraîne que les rapports $\mu(a^i w)/\mu(a^i)$, $0 \leq i \leq n-1$, sont tous égaux, d'où

$$\mu(a^i w) = \mu(a^i)\mu(w)$$

Donc μ une mesure de Bernoulli. $\square$

7 Problèmes d'extension

Le problème :

Soit $X \subset A^*$ un code. On considère X *comme un alphabet* et soit μ' une mesure de probabilité sur X^*.

1) Existe-t-il une mesure de probabilité μ sur A^* qui « prolonge » μ' sur A^*.

2) Si oui, est-elle unique ?

3) Si μ' est rationnelle, μ est-elle rationnelle ?

Pour commencer Dans le cas où X est préfixe, tout se passe bien.

Proposition 9 *Soit X un code préfixe et μ' une mesure de probabilité sur X^**

1) Il existe un prolongement unique μ de μ' sur A^*.

2) Si X est un langage rationnel et si μ' est une mesure rationnelle, il en est de même de μ.

(Idée de la) **Démonstration.**

Disons qu'un mot v est un *contexte droit* d'un mot u si

$$uv = x_1 x_2 \dots x_n \in X^* \text{ avec } |v| < |x_n|$$

On note $C_d(u)$ l'ensemble des contextes droits de u.

On pose

$$\mu(u) = \sum_{v \in C_d(u)} \mu'(uv),\ u \in A^*$$

On vérifie alors tout ce qu'il faut :

μ prolonge μ' et est une mesure de probabilité sur A^*

μ est l'unique prolongement possible de μ

Si X est rationnel ainsi que μ', il en est de même de μ.

8 Cas général

Dans le cas général d'un code non préfixe, ni l'existence d'un prolongement, ni son unicité ne sont valides.

Pour commencer, on fait une hypothèse supplémentaire, à savoir que la « longueur moyenne » de X est finie :

$$l(X) = \sum_{x \in X} |x| \mu'(x) < +\infty$$

On généralise la notion de contexte : un *contexte* du mot $u \in A^*$ est un couple de mot (v, w) tel que

1) $vuw = x_1 x_2 \dots x_n \in X^*$

2) $|v| < x_1$ et $|w| < x_n$

On note $C(u)$ l'ensemble des contextes de u.

On définit la fonction $\mu : A^* \to R^+$ par la formule

$$\mu(u) = \frac{1}{l(X)} \sum_{(v,w) \in C(u)} \mu'(vuw)$$

Proposition 10

1) μ est une mesure de probabilité sur A^*

2) Si μ' est invariante, il en est de même de μ

3) Si X est fini (ou même rationnel) et si μ' est rationnelle, alors μ est rationnelle

4) **Mais rien ne dit encore que μ prolonge μ'.**

9 Où μ prolonge μ'

Définition. Une mesure de probabilité μ sur A^* est *tempérée* s'il existe une constante $M > 0$ telle que

$$\mu(aub) \geq M\mu(u),\ u \in A^*,\ a, b \in A$$

Théorème 11 *Soit X un code maximal coupant, μ une mesure de probabilité sur A^*, invariante, tempérée, X-multiplicative. Soit ν la mesure sur A^* définie par la formule*

$$\nu(u) = \frac{1}{l(X)} \sum_{(v,w) \in C(u)} \mu(vuw)$$

Alors $\mu = \nu$.

(Principe de la) **Démonstration.**

On construit un certain système dynamique (Ω, T). A chaque mesure de probabilité μ sur A^* on associe de manière bijective une mesure de probabilité (au sens classique) $\overline{\mu}$ sur la tribu borélienne de Ω. On montre alors que

- si μ est invariante, $\overline{\mu}$ est T-invariante)
- la mesure $\overline{\nu}$ du théorème est ergodique
- $\overline{\mu}$ est absolument continue par rapport à ν (provient de ce que μ est tempérée).

Il résulte de tout ça que $\overline{\mu} = \overline{\nu}$ et donc $\mu = \nu$. $\square$

Une question : L'hypothèse tempérée est-elle nécessaire ?

Corollaire 12 *Soit X un code maximal fini. Toute mesure de probabilité invariante et X-multiplicative sur A^* est rationnelle.*

« **Exercice** » : Trouver une preuve de ce résultat qui ne fasse pas appel à la théorie ergodique.