Ann. Sc. Math. Québec, 1977, Vol. I, No 1, pp. 87-92

NOTE SUR L'ENTROPIE DE SHANNON

Gilbert Labelle *

§0. INTRODUCTION

Supposons que l'on ait aucune information concernant la probabilité P attachée à un espace probabilisé discret de cardinalité donnée $n \geq 1$. Le principe de la raison insuffisante de Laplace nous autorise à supposer que P est uniformément distribuée dans le simplexe $\mathbb{P}_n = \{(p_1,\ldots,p_n) | p_1 +\ldots+ p_n = 1,\ p_i \geq 0,\ i = 1,\ldots,n\}$. L'entropie $H_n : \mathbb{P}_n \to \mathbb{R}$ de C.E. Shannon [3] donnée par la formule**

$$H_n(p_1,\ldots,p_n) = - \sum_{i=1}^{n} p_i \,\ell n\, p_i \qquad 0.1$$

devient alors une variable aléatoire définie sur l'espace $\mathbf{P}_n$ lequel est muni de la probabilité uniforme.

Le but de cette note est d'expliciter chacun des moments $\mu_k^{(n)} = E(H_n^k)$, $k = 0,1,2,\ldots$ de H_n sous la forme d'une somme finie qui montre, en outre, que leur nature extrêmement complexe est en fait reliée à la fonction zêta (ζ) de Riemann. En particulier, nous obtenons des formules compactes pour la moyenne et la variance de H_n. La constante d'Euler

$$\gamma = \lim_{n\to\infty} (1 + \frac{1}{2} +\ldots+ \frac{1}{n} - \ell n\ n)$$

apparaît alors directement reliée à la grandeur de la différence entre l'entropie maximale $H_n(\frac{1}{n},\ldots,\frac{1}{n}) = \ell n\ n$ et l'entropie moyenne $E(H_n)$ lorsque n tend vers l'infini.

---

* Travail fait dans le cadre de la subvention CNR A 7849 du Conseil National de Recherches du Canada.

** On a convenu ici de prendre les logarithmes naturels pour simplifier les calculs.

§1. LA FONCTION GENERATRICE DES MOMENTS DE H_n

Le théorème suivant fournit une représentation compacte de la fonction génératrice $G_n(t) = E(\exp t H_n)$ des moments de H_n.

THEOREME A.

Soit $n \geq 1$ *alors*

$$G_n(t) = (n-1)!\ t^{-n+1} t^t \cdot \underbrace{(t^{-t} * \ldots * t^{-t})}_{n \text{ facteurs}} \qquad 1.1$$

où $(*)$ *désigne le produit de convolution usuel.*

Démonstration:

Soit $\Delta_{n-1} = \{(x_1,\ldots,x_{n-1}) | x_1 + \ldots + x_{n-1} \leq 1,\ x_i \geq 0,\ i = 1,\ldots,n-1\}$ le simplexe plein standard de dimension $n-1$ et définissons $F_n = F_n(s,t)$ par

$$F_n(s,t) = \int\ldots\int_{t\Delta_{n-1}} x_1^{-sx_1} \ldots x_{n-1}^{-sx_{n-1}} (t-x_1-\ldots-x_{n-1})^{-s(t-x_1-\ldots-x_{n-1})} dx_1 \ldots dx_{n-1}. \qquad 1.2$$

La forme même de $G_n(t)$ montre que $G_n(t) = (n-1)!\ F_n(t,1)$. D'autre part, la substitution $x_i = tu_i$ $i = 1,\ldots,n-1$ entraîne via un court calcul que F_n vérifie l'équation fonctionnelle

$$F_n(s,t) = t^{n-1-st} F_n(st,1). \qquad 1.3$$

En particulier, en posant $s=1$ dans 1.3 on obtient $F_n(1,t) = t^{n-1-t} F_n(t,1)$ d'où l'on tire

$$G_n(t) = (n-1)!\ t^{-n+1} t^t F_n(1,t). \qquad 1.4$$

Comme $F_n(1,t) = t^{-t} * \ldots * t^{-t}$ (n facteurs) le théorème s'ensuit immédiatement.

§2. STRUCTURE DES MOMENTS DE H_n

Nous utiliserons la transformation L de Laplace pour expliciter les moments de H_n. D'abord deux lemmes:

LEMME 2.1.

Soient μ *et* ν *deux entiers* ≥ 0 *alors*

i) $L(t^\mu(-\ell n\ t)^\nu) = s^{-\mu-1} p_{\mu,\nu}(\ell n\ s) = s^{-\mu-1}(\ell n\ s - \frac{d}{dx})^\nu \Gamma(x)|_{x=\mu+1}$ 2.1.1

ii) $L^{-1}(s^{-\mu-1}(\ell n\ s)^{\nu}) = t^{\mu}q_{\mu,\nu}(-\ell n\ t) = t^{\mu}(-\ell n\ t - \frac{d}{dx})^{\nu}\ \frac{1}{\Gamma(x)}|_{x=\mu+1}$ 2.1.2

où

$$p_{\mu,\nu}(S) = \sum_{i=0}^{\nu} (-1)^{i}\binom{\nu}{i}\Gamma^{(i)}(\mu+1)S^{\nu-i}$$

et

$$q_{\mu,\nu}(T) = \sum_{i=0}^{\nu} (-1)^{i}\binom{\nu}{i}\left(\frac{1}{\Gamma}\right)^{(i)}(\mu+1)T^{\nu-i}$$

sont deux polynômes.

<u>Démonstration</u>:

i) Posons

$$\omega_{\mu,\nu}(s) = L(t^{\mu}(-\ell n\ t)^{\nu})$$

et

$$\omega_{\mu}(s,x) = \sum_{\nu=0}^{\infty} \omega_{\mu,\nu}(s)\ \frac{x^{\nu}}{\nu!}\ .$$

Une intégration terme à terme montre que $\omega_{\mu}(s,x) = L(t^{\mu-x}) = s^{-\mu-1+x}\Gamma(\mu+1-x)$. La formule 2.1.1 découle alors en *dérivant* ν fois cette dernière relation par rapport à x via la formule classique de Newton et en posant x=0.

ii) Posons

$$\theta_{\mu,\nu}(t) = L^{-1}(s^{-\mu-1}(\ell n\ s)^{\nu})$$

et

$$\theta_{\mu}(t,x) = \sum_{\nu=0}^{\infty} \theta_{\mu,\nu}(t)\ \frac{x^{\nu}}{\nu!}\ .$$

Une intégration terme à terme montre que $\theta_{\mu}(t,x) = L^{-1}(s^{-\mu-1+x}) = t^{\mu-x}/\Gamma(\mu+1-x)$. La formule 2.1.2 découle encore une fois par ν dérivations successives par rapport à x en x=0.

<u>LEMME 2.2</u>.

$$L(t^{-t}) = \sum_{\mu\geq\nu\geq 0} A_{\mu,\nu}\ \frac{(\ell n\ s)^{\nu}}{s^{\mu+1}}$$

où 2.2

$$A_{\mu,\nu} = (-1)^{\mu-\nu}\ \frac{\Gamma^{(\mu-\nu)}(\mu+1)}{(\mu-\nu)!\ \nu!}\ .$$

<u>Démonstration</u>:

Il suffit de remarquer que

$$L(t^{-t}) = \sum_{k=0}^{\infty} \frac{1}{k!} L((-t\ \ell n\ t)^k),$$

d'appliquer le Lemme 2.1 i) et de réarranger correctement les indices.

Définissons maintenant le symbole $\{{i \atop j}\}$ pour $i,j \in \mathbb{N}$ par

$$\{{i \atop j}\} = \frac{\Gamma^{(j)}(i+j+1)}{i!\ j!}\ . \tag{2.3}$$

On a alors la représentation finie explicite suivante pour les moments de la variable aléatoire H_n.

THEOREME B.

Soient $n \geq 1$ *et* $k \geq 0$ *deux entiers alors*

$$\mu_k^{(n)} = (-1)^k k!(n-1)!\ \sum_{i=0}^{k} \left[\sum_{\substack{i_1+\ldots+i_n=i \\ j_1+\ldots+j_n=k-i}} \begin{Bmatrix} i_1 \\ j_1 \end{Bmatrix} \cdots \begin{Bmatrix} i_n \\ j_n \end{Bmatrix} \right] L^{(i)}(n+k) \tag{2.4}$$

où $\mu_k^{(n)}$ *désigne le moment d'ordre* k *de* H_n *et* $L(x) = 1/\Gamma(x)$.

Démonstration:

On a, à cause du Théorème A,

$$\sum_{k=0}^{\infty} \frac{\mu_k^{(n)} t^k}{k!} = G_n(t) = (n-1)!\ t^{-n+1} t^t \cdot (t^{-t} * \ldots * t^{-t})\ \text{(n facteurs)}. \tag{2.5}$$

Mais $L(t^{-t} * \ldots * t^{-t}) = (L(t^{-t}))^n$ à cause de la propriété classique $L(f_1 * f_2) = L(f_1)L(f_2)$. Elevant l'expression donnée par le Lemme 2.2 à la puissance n, utilisant le Lemme 2.1 et le fait que

$$t^t = \sum_{r=0}^{\infty} (t\ \ell n\ t)^r / r!,$$

on arrive finalement à isoler le coefficient cherché $\mu_k^{(n)}/k!$ après des manipulations d'indices complexes mais élémentaires.

Remarquons que la formule 2.4, en plus de fournir un procédé de calcul ordonné des $\mu_k^{(n)}$ (qui autrement serait presqu'inextricable), nous montre que ces derniers sont des expressions du type

$$f_{n,k}\left(n;\gamma,\zeta(2),\zeta(3),\ldots,\zeta(k);\ \sum_{\nu=1}^{n}\frac{1}{\nu},\ \sum_{\nu=1}^{n}\frac{1}{\nu^2},\ldots,\sum_{\nu=1}^{n}\frac{1}{\nu^k}\right) \qquad 2.6$$

où $f_{n,k}(x;y_1,y_2,\ldots,y_k;z_1,z_2,\ldots,z_k)$ est une fonction polynômiale en $y_1,\ldots,y_k$, $z_1,\ldots,z_k$ dont les coefficients sont des fonctions rationnelles en x et où γ est la constante d'Euler et

$$\zeta(r) = \sum_{\nu=1}^{\infty}\frac{1}{\nu^r}$$

est la fonction zêta de Riemann.

En effet, la fonction digamma Ψ d'Euler [1] définie par $\Psi = \Gamma'/\Gamma$ permet d'exprimer les dérivées des fonctions Γ et $L = 1/\Gamma$ sous les formes

$$\Gamma^{(\nu)} = P_\nu(\Psi,\Psi',\ldots,\Psi^{(\nu-1)})\cdot\Gamma$$

et $\qquad 2.7$

$$L^{(\nu)} = Q_\nu(\Psi,\Psi',\ldots,\Psi^{(\nu-1)})/\Gamma$$

où P_ν et Q_ν sont des polynômes à coefficients entiers définis inductivement. L'expression 2.6 suit alors de l'observation [1, p. 182]

$$\Psi^{(r)}(s+1) = (-1)^r r!\{-\zeta(r+1) + \sum_{\nu=1}^{s}\frac{1}{\nu^{r+1}}\} \text{ si } r > 0 \qquad 2.8$$

et

$$\Psi(s+1) = -\gamma + \sum_{\nu=1}^{s}\frac{1}{\nu} \text{ où } r,s \in \mathbb{N}. \qquad 2.9$$

<u>THEOREME C.</u>

Soit n ≥ 1 *un entier, alors*

i) $$E(H_n) = \sum_{\nu=2}^{n}\frac{1}{\nu} \qquad 2.10$$

ii) $$\mathrm{Var}(H_n) = \sum_{\nu=2}^{n}\frac{1}{\nu^2} - \left(\frac{n-1}{n+1}\right)\left(\frac{\pi^2}{6} - 1\right). \qquad 2.11$$

<u>Démonstration:</u>

Il suffit de substituer les valeurs k=1 et k=2 dans la formule 2.4, d'utiliser 2.8 et 2.9, de remarquer que $\zeta(2) = \frac{\pi^2}{6}$ et de centraliser.

<u>COROLLAIRE</u>:

i) $$\lim_{n\to\infty} \{H_n(\tfrac{1}{n},\ldots,\tfrac{1}{n}) - E(H_n)\} = 1-\gamma \approx 0.4227843351 \qquad 2.12$$

ii) $$\mathrm{Var}(H_n) = \frac{\pi^2-9}{3}\, n^{-1} + O(n^{-2}) \text{ si } n \to \infty. \qquad 2.13$$

<u>Démonstration</u>:

i) Ceci découle de la définition même de la constante d'Euler qui est

$$\gamma = \lim_{n\to\infty} \left(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} - \ell n\ n\right)$$

et de la formule 2.10.

ii) Ceci provient d'une application de la formule sommatoire d'Euler-Maclaurin [2] et de la formule 2.11.

<u>REFERENCES</u>

1. CAMPBELL, R., Les intégrales eulériennes et leurs applications, Collection Universitaire de mathématiques, Vol. 20, Dunod, Paris 1966.

2. GUELFOND, A.O., Calcul des différences finies, Collection Universitaire de mathématiques, Vol. 12, Dunod, Paris 1963.

3. THEODORESCU, R. et GUIASU, S., Incertitude et information, Les Presses de l'Université Laval, 1971.

Université du Québec à Montréal
Montréal, Canada