Séminaire Lotharingien de Combinatoire **47** (2001), Article B47a

Les nombres hyperharmoniques et la fratrie du collectionneur de vignettes (*)

Dominique Foata, Guo-Niu Han et Bodo Lass

Résumé. — Le problème traditionnel du collectionneur de vignettes est prolongé au cas où le collectionneur fait partager sa moisson avec les membres de sa fratrie. Il reste le seul acheteur, mais donne à ses frères les images qu'il obtient en double. Quand son album est fini, les albums de ses frères ont un certain nombre d'emplacements vides. En moyenne, combien ? Nous apportons une réponse à cette question et obtenons, en outre, une expression pour la fonction génératrice multivariée des variables aléatoires en question. Le problème fait apparaître les nombres hyperharmoniques, qu'il faut étudier sous certains aspects, comme solutions d'équations aux différences notamment.

Abstract. — The traditional coupon collector's problem is extended to the case where the collector shares his harvest with his own phratry. He remains the only coupon provider, but gives his brothers the coupons that have already appeared in his picture-book. When his book is completed, the books of the other brothers have certain numbers of empty spots. On the average, how many ? We bring an answer to this question and also derive the multivariable generating function for the random variables involved. The problem gives rise to the hyperharmonic numbers that are to be studied under various aspects, in particular as solutions of finite-difference equations.

1. Introduction. — Le problème du collectionneur de vignettes («coupon collector's problem») est un vieux problème étudié en Calcul des Probabilités (voir Feller (1968)). Chaque tablette de chocolat d'une marque donnée contient une vignette, qui peut être collée dans un album. Celui-ci contient m emplacements qui correspondent aux m vignettes éditées. Il s'agit d'estimer le nombre moyen de tablettes que doit acheter un

(*) English title : *The hyperharmonic numbers and the phratry of the coupon collector.*

AMS 2000 Classification : 05A15, 05E05, 60C05, 60G42.

Keywords : Hyperharmonic numbers, Coupon collector's problem, Surjection calculus, Formal Laplace Transform, Finite-difference equations.

collectionneur pour remplir son album. Naturellement, toutes les vignettes sont différenciées et à chaque achat le collectionneur a la probabilité $1/m$ d'obtenir une vignette donnée. S'il a déjà cette vignette, il la jette, s'il ne l'a pas, il la colle dans son album à l'emplacement réservé pour celle-ci. Soit H_m le *nombre harmonique* défini par $H_m := 1 + \frac{1}{2} + \cdots + \frac{1}{m}$, suivant l'appellation désormais classique (voir Graham et al. (1989), § 6.3). On démontre facilement, par diverses méthodes (simple comptage, en bâtissant une chaîne de Markov adaptée, ...) que le nombre moyen d'achats est : $m\,H_m$.

On doit à Pintacuda (1980) de s'être intéressé au *frère* du collectionneur, celui qui n'achète pas de vignette, mais les reçoit de son grand frère, lorsque celui-ci a des doublons. Il a aussi un album, où il colle les vignettes obtenues. Le problème est alors d'estimer le nombre moyen d'emplacements vides dans son album, lorsque son grand frère a terminé son propre album. En utilisant le théorème des temps d'arrêt pour les martingales, Pintacuda (*op. cit.*) montre que ce nombre moyen vaut H_m.

Nous devons à notre collègue Fuchs de nous avoir expliqué la méthode probabiliste de Pintacuda, telle qu'elle lui avait été rapportée par Letta (1992), pour trouver cette dernière valeur. Une discussion s'est alors engagée entre les trois auteurs, laquelle les a conduit à penser que ce problème, bien que tombé dans l'escarcelle des probabilistes, méritait un traitement combinatoire complet. C'est ce que nous nous proposons de faire, en donnant aussi un prolongement au cas d'une fratrie. Pourquoi, en effet, ne considérer qu'un seul petit frère ?

Dans le problème de la fratrie, le collectionneur est l'aîné de la famille, il a r frères et reste toujours le seul acheteur. Chaque frère a aussi un exemplaire personnel de l'album et colle aussi les vignettes lorsqu'il en obtient une nouvelle. La source d'approvisionnement est la suivante : chaque fois que l'aîné achète une tablette et qu'il obtient une nouvelle vignette, il la colle dans son album. Sinon, il la donne au plus âgé de ses frères. Si celui-ci n'a pas encore cette vignette, il la colle dans son album ; sinon, il la donne au frère suivant, et ainsi de suite. Si donc une vignette apparaît plus de r fois avant que l'aîné ait fini son album, cette vignette ira à la poubelle !

Soit T le nombre de tablettes que doit acheter l'aîné pour compléter son album. A cet instant, les frères, à qui on donne les numéros 1, 2, ... , r, ont encore, respectivement, $M_T^{(1)}$, $M_T^{(2)}$, ... , $M_T^{(r)}$ emplacements vides dans leurs albums. On sait que $\mathbb{E}[T] = m\,H_m$ et $\mathbb{E}[M_T^{(1)}] = H_m$, d'après Pintacuda. Nous nous proposons de retrouver ce dernier résultat, mais aussi d'établir, pour $k \geq 2$, que l'on a

$$\mathbb{E}[M_T^{(k)}] = 1 + K_m^{(1)} + \cdots + K_m^{(k)}, \tag{1.1}$$

où, pour tout $k \geq 0$ et $m \geq 1$, le nombre $K_m^{(k)}$ est le *nombre hyperharmonique*, défini par les relations suivantes :

(1.2) $\quad K_1^{(k)} := 0,$ pour tout $k \geq 1$; $\quad K_m^{(0)} := 1,$ pour tout $m \geq 1$;
et pour $k \geq 1$ et $m \geq 2$

$$K_m^{(k)} := \frac{K_2^{(k-1)}}{2} + \cdots + \frac{K_{m-1}^{(k-1)}}{m-1} + \frac{K_m^{(k-1)}}{m}, \tag{1.3}$$

que l'on peut encore récrire :

$$K_m^{(k)} - \frac{K_m^{(k-1)}}{m} = K_{m-1}^{(k)}. \tag{1.4}$$

En particulier, pour $k \geq 0$, on a : $K_2^{(k)} = 1/2^k$. Le tableau suivant donne une liste des premières valeurs de ces nombres.

m	1	2	3	4	5	10	15	20	25	30	35	40	45	50
$K_m^{(0)}$	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
$K_m^{(1)}$	0	0,50	0,83	**1,08**	**1,28**	**1,93**	**2,31**	**2,60**	**2,82**	**3,00**	**3,15**	**3,28**	**3,39**	**3,50**
$K_m^{(2)}$	0	0,25	0,53	0,80	1,06	**2,14**	**2,98**	**3,67**	**4,27**	**4,79**	**5,26**	**5,68**	**6,07**	**6,43**
$K_m^{(3)}$	0	0,13	0,30	0,50	0,71	1,79	2,82	**3,76**	**4,64**	**5,46**	**6,23**	**6,96**	**7,65**	**8,30**
$K_m^{(4)}$	0	0,06	0,16	0,29	0,43	1,27	2,20	3,14	4,08	4,99	5,89	6,77	7,63	**8,47**
$K_m^{(5)}$	0	0,03	0,09	0,16	0,24	0,81	1,51	2,28	3,08	3,91	4,75	5,59	6,44	7,29
$K_m^{(6)}$	0	0,02	0,04	0,08	0,13	0,48	0,95	1,49	2,09	2,73	3,40	4,09	4,80	5,52

Comme tous les nombres $K_m^{(0)}$ valent 1, on a :

$$K_m^{(1)} = H_m - 1 \quad (m \geq 1).$$

On a ainsi « harmonicisé » la suite dont tous les termes sont égaux à 1, pour obtenir la suite des nombres harmoniques (diminués d'une unité). L'appellation *hyperharmonique* vient du fait que pour obtenir la suite $K_m^{(k)}$ $(m \geq 2)$, on « harmonicise », de la même façon, la suite $K_m^{(k-1)}$ $(m \geq 2)$.

En utilisant les séries génératrices, la récurrence (1.4) s'écrit sous la forme :

$$(1 - t/m) \cdot \sum_{k \geq 0} K_m^{(k)} t^k = \sum_{k \geq 0} K_{m-1}^{(k)} t^k.$$

Puisque $\sum\limits_{k \geq 0} K_1^{(k)} t^k = 1$, on obtient :

$$\sum_{k \geq 0} K_m^{(k)} t^k = \frac{1}{(1 - t/2)(1 - t/3) \cdots (1 - t/m)}. \tag{1.5}$$

Dans le présent article, nous nous proposons de redémontrer le résultat $\mathbb{E}[M_T^{(1)}] = H_m$ de Pintacuda, puis d'établir les formules (1.1). En plus, nous donnons une expression pour la *distribution* (voir Proposition 4.1) du vecteur aléatoire $(T, M_T^{(1)}, M_T^{(2)}, \ldots, M_T^{(r)})$, d'où l'on tire la distribution de chacune des variables et naturellement leurs espérances mathématiques. En fait, nous préférons introduire une suite de variables aléatoires $(X_T^{(k)})$ $(k = 1, 2, \ldots)$ telles que, pour tout $k \geq 1$, on ait :

$$M_T^{(k)} = X_T^{(1)} + X_T^{(2)} + \cdots + X_T^{(k)},$$

calculons la fonction génératrice du vecteur $(T, X_T^{(1)}, X_T^{(2)}, \ldots, X_T^{(r)})$, puis montrons que

$$\text{(1.6)} \qquad \mathbb{E}[X_T^{(1)}] = 1 + K_m^{(1)} = H_m; \quad \mathbb{E}[X_T^{(k)}] = K_m^{(k)} \quad \text{pour } k \geq 2.$$

Le calcul de la fonction génératrice multivariée fait appel aux techniques classiques de la Combinatoire Énumérative, en particulier, des propriétés de la *transformation de Laplace formelle*. La méthode de Pintacuda est reprise ensuite dans le cas de la fratrie, pour permettre d'obtenir deux autres démonstations des formules (1.1). En fait, cette technique du théorème des temps d'arrêt pour les martingales ne fournit qu'une *équation aux différences* pour les temps moyens. La partie la plus délicate n'est plus probabiliste : elle consiste à trouver une solution de cette *équation aux différences* et là, il n'existe pas de méthode universelle.

Nous donnons encore deux autres méthodes de calcul, permettant d'obtenir directement la fonction génératrice des nombres moyens $\mathbb{E}[X_T^{(k)}]$. La première de ces méthodes fournit même une formule dans le cas où la vente des vignettes n'est plus uniforme, où la probabilité d'acheter la vignette i est $p_i > 0$ $(p_1 + \cdots + p_m = 1)$. Comme on le verra, toutes les méthodes combinatoires développées reposent sur des dénombrements de *surjections*.

Nous terminons l'article par une étude asymptotique des nombres hyperharmoniques. Comme il est bien connu, le nombre harmonique H_m est asymptotiquement équivalent à $\ln m$. On montre que le nombre $k!\,K_m^{(k)}$ est asymptotiquement équivalent à $(\ln m)^k$.

Déjà Feller (1968, p. 225) écrivait qu'il y avait une « énorme littérature traitant de variantes du problème du collectionneur de vignettes ». En consultant *MathSciNet*, nous nous sommes rendu compte que cet intérêt n'avait pas décru. Pas moins de dix-sept titres. Ce modèle a trouvé des applications en informatique (Flajolet et al. (1992), Banderier et al. (2000)), mais aussi en recherche opérationnelle (Boneh et al. (1997)). Nous renvoyons à *MathSciNet* pour la liste complète de ces articles.

2. Les nombres hyperharmoniques. — Ils ont été définis en (1.2) et (1.3), de façon itérative. En décomposant la fraction rationnelle, qui apparaît dans (1.5), en éléments simples, on déduit, pour $k \geq 0$ et $m \geq 2$, la formule explicite

$$K_m^{(k)} = m(m-1) \sum_n \frac{(-m+2)_n}{n!} \frac{1}{(n+2)^{k+1}}, \tag{2.1}$$

où $(a)_n$ désigne la *factorielle montante* :

$$(a)_n := \begin{cases} 1, & \text{si } n = 0\,; \\ a(a+1)\cdots(a+n-1), & \text{si } n \geq 1. \end{cases}$$

On reconnaît dans la fraction rationnelle de l'identité (1.5) la fonction génératrice des *fonctions symétriques homogènes complètes* $h_k(x_1, x_2, \dots)$ (*cf.* Macdonald (1995) p. 21), lorsque l'on substitue les fractions $\frac{1}{2}$, $\frac{1}{3}$, … , $\frac{1}{m}$ aux variables x_1, x_2, … , x_{m-1}, ce qui établit l'identité

$$K_m^{(k)} = h_k\Bigl(\frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{m}\Bigr). \tag{2.2}$$

On peut également obtenir une expression pour la fonction génératrice des nombres $K_m^{(1)}$ $(m \geq 2)$, à savoir,

$$\sum_{m \geq 2} K_m^{(1)} s^m = (s-1)^{-1}\bigl(\ln(1-s) + s\bigr). \tag{2.3}$$

Les propriétés des nombres $K_m^{(k)}$ données dans la proposition suivante sont faciles à établir et données sans démonstration. On observe, en particulier, que les nombres $K_m^{(k)}$ $(k \geq 0)$ vont d'abord en croissant (valeurs reproduites en gras dans le tableau de l'introduction).

Proposition 2.1.

(a) *Pour tout $m \geq 2$, on a :* $\sum_{k \geq 0} K_m^{(k)} = m$.

(b) *Pour tout $k \geq 1$, on a :* $K_m^{(k)} \to +\infty$, *lorsque m tend vers l'infini.*

(c) *Pour tout $k \geq 1$, on a* $K_m^{(0)} < K_m^{(1)} < \cdots < K_m^{(k)} < m$ *pour m assez grand.*

Remarque. — La récurrence (1.4), à savoir $K_m^{(k)} - K_m^{(k-1)}/m = K_{m-1}^{(k)}$ $(k \geq 1, m \geq 2)$ apparaît déjà dans Bentley et al. (1978) et Buchta (1989) pour permettre le calcul du nombre moyen de maxima dans un ensemble de vecteurs. Ces auteurs utilisent des conditions initiales *différentes* de (1.2), à savoir : $K_1^{(k)} := 1$ $(k \geq 1)$ et $K_m^{(0)} := 1$ $(m \geq 1)$ et obtiennent donc d'autres suites de nombres. Tous ces nombres, cependant, y compris les nombres hyperharmoniques, ne sont que des variations des *nombres harmoniques généralisés* $H_m^{(k)} := \sum_{1 \leq j \leq m} (1/j^k)$ (*cf.* Graham et al. (1989)). Une autre extension, avec l'adjonction d'un nouveau paramètre, a été introduite et utilisée dans Flajolet et al. (1995) dans l'étude statistique des « quadtrees ».

3. Un calcul de surjections. — Le problème décrit dans l'introduction, bien qu'il ait été offert aux probabilistes, a une très forte connotation combinatoire. Il s'agit, en fait, de dénombrer des classes particulières de *surjections*. Pour tout entier n, posons $[n] := \{1, 2, \ldots, n\}$ et notons $\mathrm{Surj}(l, m)$ l'ensemble des surjections de $[l]$ sur $[m]$. Si f appartient à $\mathrm{Surj}(l, m)$, pour tout $i \geq 1$, on note $\nu_i(f)$ le nombre d'éléments $j \in [m]$ tels que l'image réciproque $f^{-1}(j)$ est de cardinal i. A tout vecteur d'entiers positifs $\mathbf{n} = (n_1, n_2, \ldots)$, on associe le sous-ensemble $\mathrm{Surj}(l, m; \mathbf{n})$ des surjections $f \in \mathrm{Surj}(l, m)$ telles que pour tout $i \geq 1$ on ait $\nu_i(f) = n_i$ et on note $\mathrm{surj}(l, m; \mathbf{n})$ le cardinal de $\mathrm{Surj}(l, m; \mathbf{n})$.

Soit $(t, s_1, s_2, \ldots)$ une suite infinie de variables commutatives. Le *poids* $\pi(f)$ de f est défini par : $\pi(f) := \prod_{i\geq 1} s_i^{\nu_i(f)}$; on a ainsi les relations : $\sum_i \nu_i(f) = m$ et $\sum_i i\,\nu_i(f) = l$.

L'identité

$$(3.1) \qquad \Bigl(\sum_{i\geq 1} s_i\,\frac{t^i}{i!}\Bigr)^m = \sum_{l\geq m} \frac{t^l}{l!} \sum_{f\in \mathrm{Surj}(l,m)} \pi(f),$$

qui peut encore s'écrire

$$(3.2) \qquad \Bigl(\sum_{i\geq 1} s_i\,\frac{t^i}{i!}\Bigr)^m = \sum_{l\geq m} \frac{t^l}{l!} \sum_{\mathbf{n}} \mathrm{surj}(l, m; \mathbf{n}) \prod_{i\geq 1} s_i^{n_i},$$

peut être établie par différentes méthodes, par exemple, par les techniques du composé partitionnel. Elle est démontrée dans Foata (1974), p. 59-60. On peut aussi utiliser la technique des espèces de nos amis canadiens (*cf.* Bergeron et al. (1994)) ou les fonctions d'ensembles (*cf.* Lass (2001)).

Introduisons maintenant le sous-ensemble $A(l, m; \mathbf{n})$ de $\mathrm{Surj}(l, m, \mathbf{n})$ formé par toutes les surjections f ayant la propriété suivante :

la restriction $f\,|_{[l-1]}$ de f à l'intervalle $[l-1]$ est elle-même une surjection de $[l-1]$ sur $[m] \setminus \{f(l)\}$, *mais pas* sur $[m]$, i.e. $f(j) \neq f(l)$ pour $j \leq l-1$.

Si donc f est dans $A(l, m; \mathbf{n})$, on a $\nu_1(f) \geq 1$ et son poids $\pi(f)$ est divisible par s_1. Désignons par $a(l, m; \mathbf{n})$ le nombre des éléments de $A(l, m; \mathbf{n})$.

Si f est dans $A(l, m; \mathbf{n})$, posons $g := f\,|_{[l-1]}$. Alors $\pi(f) = \pi(g)\,s_1$, comme il a été noté plus haut. D'autre part, on a la relation

$$(3.3) \qquad \sum_{\mathbf{n}} \sum_{f\in A(l,m;\mathbf{n})} \pi(f) = m\,s_1 \sum_{g\in \mathrm{Surj}(l-1,m-1)} \pi(g).$$

On en tire :

$$\sum_{l\geq m} \frac{t^{l-1}}{(l-1)!} \sum_{\mathbf{n}} \sum_{f\in A(l,m;\mathbf{n})} \pi(f) = \sum_{l\geq m} \frac{t^{l-1}}{(l-1)!} \times m\,s_1 \sum_{g\in \mathrm{Surj}(l-1,m-1)} \pi(g)$$

$$= m\, s_1 \sum_{k\ge m-1} \frac{t^k}{k!} \sum_{g\in \mathrm{Surj}(k,m-1)} \pi(g) = m\, s_1 \Bigl(\sum_{i\ge 1} s_i \frac{t^i}{i!}\Bigr)^{m-1}.$$

Avec $s_i := 1$ pour tout $i \ge r+1$, l'identité précédente prend la forme

$$m\, s_1 (e^t - 1 + t(s_1-1) + \frac{t^2}{2!}(s_2-1) + \cdots + \frac{t^r}{r!}(s_r-1))^{m-1} = \sum_{l\ge m} \frac{t^{l-1}}{(l-1)!} \sum_{\mathbf{n}} \sum_{f\in A(l,m;\mathbf{n})} \pi(f).$$

En développant le membre de gauche par la formule du multinôme, on en déduit :

$$ms_1 \sum_{a+b+c_1+\cdots+c_r=m-1} \binom{m-1}{a,b,c_1,\ldots,c_r} e^{at}(-1)^b \prod_{i=1}^{r} \Bigl(\frac{t^i(s_i-1)}{i!}\Bigr)^{c_i} = \sum_{l\ge m} \frac{t^{l-1}}{(l-1)!} \sum_{\mathbf{n}} \sum_{f\in A(l,m;\mathbf{n})} \pi(f).$$

La *transformation de Laplace formelle*, $\mathcal{L}$, par rapport à la variable t remplace le terme $d(l)t^l/l!$ d'une série par le terme $d(l)t^l$. En particulier, $\mathcal{L}(e^{at}t^n) = n!\, t^n/(1-at)^{1+n}$. On applique la transformation $\mathcal{L}$ à la dernière identité et on trouve :

$$\text{(3.4)}\quad ms_1 t \sum \binom{m-1}{a,b,c_1,\ldots,c_r} (-1)^b \left(\prod_{k=1}^{r} \Bigl(\frac{s_k-1}{k!}\Bigr)^{c_k}\right) \times \frac{t^{\Sigma k c_k}}{(1-at)^{1+\Sigma k c_k}} \ (\sum_k k c_k)! = \sum_{l\ge m} t^l \sum_{\mathbf{n}} a(l,m;\mathbf{n}) \prod_{k=1}^{r} s_k^{n_k}.$$

4. La fonction génératrice multivariée. — Comme déjà indiqué dans l'introduction, T désigne l'instant où l'aîné termine son album. Pour chaque $n=0,1,\ldots$, notons $X_n^{(0)}$ le nombre d'emplacements vides dans l'album de l'aîné à l'instant n, de sorte que $X_0^{(0)} = m$ et $X_T^{(0)} = 0$. Pour $k=1,2,\ldots$, notons aussi $X_n^{(k)}$ le nombre de vignettes qui sont apparues exactement k fois, jusqu'à la date n incluse. Alors $\sum_{k\ge 1} X_T^{(k)} = m$. De plus, le nombre d'emplacements vides dans l'album du frère d'indice $k\ge 1$, à cet instant T, un nombre qui avait été noté $M_T^{(k)}$ dans l'introduction, est égal à : $M_T^{(k)} = X_T^{(1)} + \cdots + X_T^{(k)}$. La distribution du vecteur $(T, M_T^{(1)}, \ldots, M_T^{(r)})$ se déduit donc immédiatement de celle du vecteur $(T, X_T^{(1)}, \ldots, X_T^{(r)})$. C'est celle-ci que nous déterminons.

Notons $Y_1, Y_2, \ldots$ les numéros (compris entre 1 et m) des vignettes que trouve l'aîné dans les achats successifs de ses tablettes. On suppose ainsi que les Y_n ($n \geq 1$) sont des variables aléatoires indépendantes, chacune uniformément répartie sur $\{1, 2, \ldots, m\}$. Alors l'aîné finit son album après l'achat de la $l^{\text{ième}}$ tablette et pour chaque $k = 1, \ldots, r$, le nombre $X_n^{(k)}$ de vignettes qui sont apparues exactement k fois, jusqu'à la date n incluse est n_k, si et seulement si l'application $f : i \mapsto Y_i$ ($1 \leq i \leq l$) est une surjection appartenant à $A(l, m; \mathbf{n})$. Cet évènement s'exprime comme $\{T = l, X_T^{(1)} = n_1, \ldots, X_T^{(r)} = n_r\}$ et sa probabilité est évidemment égale à $a(l, m; \mathbf{n})/m^l$. La proposition suivante est donc une conséquence de la relation (3.4).

Proposition 4.1. — *La fonction génératrice du vecteur aléatoire $(T, X_T^{(1)}, \ldots, X_T^{(r)})$ est donnée par :*

$$\sum_{l \geq m, \mathbf{n}} \mathrm{P}\{T = l, X_T^{(1)} = n_1, \ldots, X_T^{(r)} = n_r\} t^l s_1^{n_1} \ldots s_r^{n_r}$$

$$= s_1 t \sum \binom{m-1}{a, b, c_1, \ldots, c_r} (-1)^b \left(\prod_{k=1}^{r} \Bigl(\frac{s_k - 1}{k!}\Bigr)^{c_k} \right) \frac{(t/m)^{\Sigma k c_k}}{(1 - at/m)^{1 + \Sigma k c_k}} (\sum_k k c_k)!$$

De cette expression, on retrouve la fonction génératrice usuelle de T et donc immédiatement son espérance $\mathbb{E}[T] = m\, H_m$. Cependant, le fait important est, qu'une fois obtenues les fonctions génératrices des variables $X_T^{(k)}$ ($k \geq 1$), on puisse, grâce à la formule explicite (2.1), *sommer* leurs dérivées en 1, et ainsi déterminer leurs espérances. En effet, on a

$$\text{(4.1)}\quad G_{X_T^{(1)}}(s) := \sum_{n \geq 1} \mathrm{P}\{X_T^{(1)} = n\} s^n = s \sum_{a+b+c=m-1} \binom{m-1}{a, b, c} (-1)^b \frac{m(s-1)^c}{(m-a)^{1+c}} c!\,;$$

et pour $k \geq 2$

$$\text{(4.2)}\quad G_{X_T^{(k)}}(s) := \sum_{n \geq 1} \mathrm{P}\{X_T^{(k)} = n\} s^n = \sum_{a+b+c=m-1} \binom{m-1}{a, b, c} (-1)^b \frac{(s-1)^c}{k!^c} \frac{m}{(m-a)^{1+kc}} (kc)!$$

D'où

$$\mathbb{E}[X_T^{(1)}] = G'_{X_T^{(1)}}(1) = -\sum_{b=0}^{m-1} \frac{(-m)_{b+1}}{(b+1)!} + m(m-1) \sum_b \frac{(-m+2)_b}{b!} \frac{1}{(b+2)^2} = 1 + K_m^{(1)} = H_m,$$

le résultat déjà prouvé par Pintacuda (1980).

De plus, pour $k \geq 2$,

$$\mathbb{E}[X_T^{(k)}] = G'_{X_T^{(k)}}(1) = m(m-1)\sum_b \frac{(-m+2)_b}{b!}\frac{1}{(b+2)^{1+k}} = K_m^{(k)}.$$

5. La méthode des martingales. — Nous réutilisons la méthode probabiliste de Pintacuda (*op. cit.*), telle qu'elle nous avait été rapportée par Letta, pour l'appliquer au calcul des nombres moyens $\mathbb{E}[X_T^{(k)}]$, pour *tout* $k \geq 1$. Conservons les notations du paragraphe précédent et utilisons les notations suivantes : si $(x^{(0)}, \ldots, x^{(r)})$ est un vecteur de r composantes réelles, posons, pour $k \geq 0$,

$$\delta^{(k)}(x^{(0)}, \ldots, x^{(r)}) = \begin{cases} (x^{(0)}, \ldots, x^{(k)} - 1, x^{(k+1)} + 1, \ldots, x^{(r)}), & \\ & \text{si } 0 \leq k \leq r-1; \\ (x^{(0)}, \ldots, x^{(r)} - 1), & \text{si } k = r; \\ (x^{(0)}, \ldots, x^{(r)}), & \text{si } k \geq r+1. \end{cases}$$

Maintenant si f est une fonction réelle définie sur $\mathbb{N}^{r+1}$, formons, pour chaque entier $n \geq 0$, la variable aléatoire

$$V_n := f(X_n^{(0)}, X_n^{(1)}, \ldots, X_n^{(r)}).$$

Traditionnellement, $W := V^{\rceil T}$ désigne le *processus arrêté* en T, c'est-à-dire la suite des variables aléatoires (W_n) $(n \geq 0)$ définies par :

$$W_n := V_{T\wedge n} := V_n\, I_{\{n<T\}} + V_T\, I_{\{n\geq T\}},$$

où le symbole I_A désigne l'*indicatrice* de l'évènement A. Or, $W_{n+1} - W_n = V_{n+1}\, I_{\{n+1<T\}} + V_T\, I_{\{n+1\geq T\}} - V_n\, I_{\{n<T\}} - V_T\, I_{\{n\geq T\}} = (V_{n+1} - V_n)\, I_{\{n<T\}} + V_{n+1}(I_{\{n+1<T\}} - I_{\{n<T\}}) + V_T(I_{\{n+1\geq T\}} - I_{\{n\geq T\}}) = (V_{n+1} - V_n)\, I_{\{n<T\}} = (V_{n+1} - V_n)\, I_{\{X_n^{(0)}>0\}}$. Ainsi, l'accroissement des variables W_n, qu'il est commode de noter $\Delta W_n := W_{n+1} - W_n$, est le même que l'accroissement ΔV_n des V_n, tant que l'album de l'aîné n'est pas terminé, soit

$$\Delta W_n = \Delta V_n\, I_{\{X_n^{(0)}>0\}}.$$

Considérons la variation du vecteur $X_n := (X_n^{(0)}, X_n^{(1)}, \ldots, X_n^{(r)})$ de l'instant n à l'instant $(n+1)$. Si la vignette achetée à l'instant $(n+1)$ a déjà été acquise $k \geq 0$ fois précédemment, alors

$$X_{n+1} = \delta^{(k)} X_n.$$

Les évènements $\{\Delta X_n^{(k)} = -1\}$ $(0 \le k \le r)$ et $\{\Delta X_n = 0\}$ forment ainsi un système complet d'évènements. On peut donc écrire

$$\Delta V_n = \sum_{k=0}^{r} I_{\{\Delta X_n^{(k)}=-1\}} \big(f(\delta^{(k)} X_n) - f(X_n)\big)$$

puisque $\Delta V_n \, I_{\{\Delta X_n = 0\}} = 0$.

Comme l'évènement $\{\Delta X_n^{(k)} = -1\}$ est réalisé si et seulement si la vignette, achetée à l'instant $(n+1)$, a été précédemment acquise k fois $(0 \le k \le r)$, l'espérance conditionnelle de la variable $I_{\{\Delta X_n^{(k)}=-1\}}$, par rapport à la suite $(Y_0, Y_1, \ldots, Y_n)$ est simplement $X_n^{(k)}/m$. (Par convention, $Y_0 := 0$.)

L'espérance conditionnelle de la variation ΔV_n, par rapport à la suite $(Y_0, Y_1, \ldots, Y_n)$, est donc égale à :

$$\mathbb{E}[\Delta V_n \mid Y_0, \ldots, Y_n] = \sum_{k=0}^{r} \frac{X_n^{(k)}}{m} \big(f(\delta^{(k)} X_n) - f(X_n)\big).$$

Supposons que l'on trouve une fonction f telle que cette expression soit nulle. Alors, on aura aussi $\Delta W_n = 0$ et W sera une *martingale* (consulter, par exemple, Bauer (1995), § 17).

PROPOSITION 5.1. — *Supposons que pour chaque $k = 1, \ldots, r$ il existe une fonction réelle $f^{(k)}$, définie sur $\mathbb{N}^{k+1}$, ayant les propriétés suivantes :*

(a) $\displaystyle\sum_{i=0}^{k} x_i \big(f^{(k)}(\delta^{(i)}(x_0, \ldots, x_k)) - f^{(k)}(x_0, \ldots, x_k)\big) = 0$ *pour* $x_0 \ge 1$ *;*

(b) $f^{(k)}(0, x_1, \ldots, x_k) = x_k$.

Alors, $\mathbb{E}[X_T^{(k)}] = f^{(k)}(m, 0, \ldots, 0)$ *pour* $k = 1, \ldots, r$.

Démonstration. — En effet, pour tout $k = 1, \ldots, r$, on a $W_0 = V_0 = f^{(k)}(X_0^{(0)}, X_0^{(1)}, \ldots, X_0^{(k)}) = f^{(k)}(m, 0, \ldots, 0)$ et $W_T = V_T = f^{(k)}(X_T^{(0)}, X_T^{(1)}, \ldots, X_T^{(k)}) = f^{(k)}(0, X_T^{(1)}, \ldots, X_T^{(k)}) = X_T^{(k)}$, d'après (b). Le théorème d'arrêt pour les martingales donne alors $\mathbb{E}[X_T^{(k)}] = \mathbb{E}[W_T] = \mathbb{E}[W_0] = W_0 = f^{(k)}(m, 0, \ldots, 0)$. $\square$

Lorsque $k = 1$, la méthode ci-dessus est due à Pintacuda (*op. cit.*). Il a trouvé, de plus, pour $k = 1$, la fonction $f^{(1)}(x_0, x_1) := H_{x_0} + \dfrac{x_1}{1 + x_0}$. En effet, la propriété (a) s'écrit, en posant $f := f^{(1)}$, $x := x_0$, $y := x_1$:

$$\begin{aligned}
&x(f(x-1, y+1) - f(x,y)) + y(f(x, y-1) - f(x,y)) \\
&\qquad = x\Big(H_{x-1} + \frac{y+1}{x} - H_x - \frac{y}{1+x}\Big) + y\Big(H_x + \frac{y-1}{1+x} - H_x - \frac{y}{1+x}\Big) \\
&\qquad = x\Big(-\frac{1}{x} + \frac{y+1}{x} - \frac{y}{1+x}\Big) + y\,\frac{-1}{1+x} = y\Big(1 - \frac{x}{1+x} - \frac{1}{1+x}\Big) = 0.
\end{aligned}$$

Aussi, $f(m, 0) = H_m$, d'où $\mathbb{E}[X_T^{(1)}] = H_m$.

Pour résoudre le problème de la fratrie du collectionneur par cette méthode, il suffit donc de trouver pour chaque $k = 1, 2, \ldots, r$ une fonction réelle $f^{(k)}$ de $(k+1)$ variables entières ayant les propriétés suivantes :

(a) $\sum_{i=0}^{k} x_i\big(f^{(k)}(\delta^{(i)}(x_0, \ldots, x_k)) - f^{(k)}(x_0, \ldots, x_k)\big) = 0$ pour $x_0 \geq 1$;

(b) $f^{(k)}(0, x_1, \ldots, x_k) = x_k$;

(c) $f^{(1)}(m, 0) = H_m$ (c'est déjà fait !), mais pour $k = 2, \ldots, r$ telle que $f^{(k)}(m, 0, \ldots, 0) = K_m^{(k)}$.

6. La solution de l'équation aux différences finies. — Résoudre une équation aux différences finies, telle que (a), qui plus est, à plusieurs variables, n'est pas aisé. D'abord, qu'entend-on par « résoudre » ? Selon les canons traditionnels, il s'agit de trouver une solution qui s'exprime dans l'algèbre des fonctions spéciales. En l'absence de techniques universelles, il faut beaucoup tâtonner. C'est la dure méthode *symbolique* décrite par Cartier (2000) dans son article de synthèse. Déjà la solution trouvée par Pintacuda relevait du *deus ex machina*. En fait, une longue familiarité avec l'équation elle-même va mener à la solution. Nous donnons deux solutions, une solution « locale » (qui fournit la valeur de $f^{(k)}$), puis « globale » (qui donne l'expression de la fonction génératrice $\sum_{k>0} f^{(k)} t^k$), en attendant que nos amis du Calcul Formel (*cf.* Zeilberger (1990) ou Strehl (2001a), (2001b)) nous fournissent une solution « automatique ».

6.1. *La solution locale.* — Elle consiste à résoudre l'équation (a), que l'on récrit sous la forme

$$(6.1)\quad (x_0 + \cdots + x_k) f^{(k)}(x_0, \ldots, x_k) = \sum_{i=0}^{k} x_i f^{(k)}(\delta^{(i)}(x_0, \ldots, x_k)),$$

pour chaque $k \geq 2$ avec $x_0 \geq 1$. D'après cette équation, on voit que la valeur de $f^{(k)}$ en $x = (x_0, x_1, \ldots, x_k)$ est parfaitement déterminée, si l'on connaît la valeur de $f^{(k)}$ en tous les points $x' = (x'_0, x'_1, \ldots, x'_k)$ plus petits, pour l'*ordre lexicographique*. On voit aussi, par itération, que chaque $f^{(k)}(x_0, \ldots, x_r)$, où $x_0 + \cdots + x_k \geq 1$, s'exprime, de façon unique, comme une combinaison linéaire de valeurs de la forme $f^{(k)}(0, x'_1, \ldots, x'_k)$ avec $x'_1 + \cdots + x'_k \leq x_0 + x_1 + \cdots + x_k$ et $x'_k \geq 1$. Si on se donne les valeurs de $f^{(k)}$ en les points $(0, x'_1, \ldots, x'_k)$, où $(x'_1, \ldots, x'_k) \in \mathbb{N}^k$ et $x'_k \geq 1$, la fonction $f^{(k)}$ est *parfaitement déterminée*.

Proposition 6.1. — *Pour $k \geq 2$, la fonction $f^{(k)}$ définie par*

$$(6.2)\quad f^{(k)}(x_0, x_1, \ldots, x_k) := K_{x_0}^{(k)} + \frac{x_1 K_{x_0+1}^{(k-1)} + \cdots + x_{k-1} K_{x_0+1}^{(1)} + x_k}{x_0 + 1}$$

$$:= K_{x_0}^{(k)} + (K_{x_0+1}^{(k)} - K_{x_0}^{(k)})x_1 + (K_{x_0+1}^{(k-1)} - K_{x_0}^{(k-1)})x_2 + \cdots + (K_{x_0+1}^{(1)} - K_{x_0}^{(1)})x_k$$

est l'unique fonction satisfaisant l'équation (6.1), *de conditions initiales*

$$f^{(k)}(0, x_1, \ldots, x_k) = x_k,$$

pour tout $(x_1, \ldots, x_k) \in \mathbb{N}^k$. *Elle satisfait, de plus,*

$$f^{(k)}(x_0, 0, \ldots, 0) = K_{x_0}^{(k)}.$$

Démonstration. — L'unicité a déjà été établie par l'argument précédant la proposition. Il suffit donc de vérifier que la fonction définie en (6.2) satisfait l'équation (6.1). Il s'agit d'une simple vérification, qui utilise pleinement la récurrence (1.4) des nombres hyperharmoniques. Nous nous dispensons d'écrire les détails de cette vérification facile. $\square$

Les Propositions 5.2 et 6.1 fournissent ainsi une seconde démonstration des identités $\mathbb{E}[X_T^{(k)}] = f^{(k)}(m, 0, \ldots, 0) = K_m^{(k)}$ pour $k = 2, \ldots, r$.

6.2. *La solution globale.* — On récrit l'identité (a) du paragraphe 6 sous la forme

$$\sum_{i=0}^{k} x_i\, N_i\, f^{(k)}(x_0, \ldots, x_k) = 0,$$

de sorte que, pour tout $i \geq 0$, on a défini l'opérateur N_i par :

$$N_i\, f^{(k)}(x_0, \ldots, x_k) := f^{(k)}(\delta^{(i)}(x_0, \ldots, x_k)) - f^{(k)}(x_0, \ldots, x_k).$$

On forme, ensuite, l'algèbre des séries formelles $f := \sum_{k \geq 0} f^{(k)}\, t^k$, où chaque coefficient $f^{(k)}$ est une fonction réelle définie sur $\mathbb{N}^{k+1}$. On prolonge la définition de N_i à cette algèbre, en posant

$$N_i\, f := \sum_{k \geq 0} N_i\, f^{(k)} t^k = \sum_{k \geq i} N_i\, f^{(k)} t^k,$$

puis, on définit l'opérateur :

$$N := \sum_{i \geq 0} x_i\, N_i.$$

Il s'agit alors de résoudre

$$N\, f = 0. \tag{6.3}$$

D'abord, on étudie l'action de N sur des séries formelles particulières. Si $g(x_0)$ est une série formelle dont les coefficients $g^{(k)}(x_0)$ ne dépendent que de l'argument x_0 et si $1 \leq i \leq k$, alors $N_i\, g^{(k)}(x_0) = 0$; d'où

$N_i\, g(x_0) = \sum\limits_{k\ge i} N_i\, g^{(k)}(x_0)\, t^k = 0$ pour $i \ge 1$. De plus, pour tout $k \ge 0$, on a : $N_0\, g^{(k)}(x_0) = g^{(k)}(x_0-1) - g^{(k)}(x_0)$; d'où $N_0\, g(x_0) = g(x_0-1) - g(x_0)$ et

$$N\, g(x_0) = x_0(g(x_0-1) - g(x_0)). \tag{6.4}$$

D'autre part, si s est une série formelle et si $i \ge 1$, on a :

$$N_i\, g(x_0)s = g(x_0)N_i\, s. \tag{6.5}$$

L'argument x_0 jouant un rôle privilégié, on cherche une solution de (6.3) de la forme $f = g(x_0)(h(x_0)+s)$, où $g(x_0)$ et $h(x_0)$ ne dépendent que de x_0 et la série s que de $x_1, x_2, \ldots$ La série s la plus simple à considérer est $s := \sum\limits_{k\ge 1} x_k\, t^k$. En évaluant $N\, g(x_0)s$ pour une telle série, on trouve :

$$N\, g(x_0)s = x_0 g(x_0-1)s - x_0 g(x_0)s + x_0 t\, g(x_0-1) + g(x_0)(t-1)s. \tag{6.6}$$

Si donc il existe des solutions f de (6.3) de la forme $f = g(x_0)(h(x_0)+s)$, on doit avoir : $N\, g(x_0)h(x_0) + N\, g(x_0)s = 0$; soit

$$-x_0 g(x_0)(h(x_0)+s) + (t-1)sg(x_0) + x_0 g(x_0-1)(h(x_0-1)+s+t) = 0.$$

On voit alors qu'en prenant $h(x_0) := x_0 + 1 - t$, on réduit l'équation précédente à

$$\frac{g(x_0)}{g(x_0-1)} = \frac{x_0}{x_0+1-t}.$$

Prenant alors $g(0) := 1$, on obtient, par itération, si x_0 est un entier positif,

$$g(x_0) = \frac{1}{(1-t/2)(1-t/3)\cdots(1-t/x_0)(x_0+1-t)}$$

et la solution

$$f = g(x_0)\cdot\big((x_0+1) + (x_1-1)t + x_2t^2 + x_3t^3 + \cdots\big). \tag{6.7}$$

Proposition 6.2. — *La série formelle donnée en* (6.7) *est solution de l'équation $N\, f = 0$ et satisfait les conditions initiales $f^{(1)}(0, x_1) = x_1 - 1$ et $f^{(k)}(0, x_1, \ldots, x_k) = x_k$ pour $k \ge 2$. De plus, avec $x_0 = m$ et les autres x_k égaux à 0, on obtient*

$$f = \frac{1}{(1-t/2)\cdots(1-t/m)} = \sum_{k\ge 0} K_m^{(k)} t^k. \tag{6.8}$$

Démonstration. — La vérification de ces propriétés a été faite précédemment et l'identité résulte de (2.2). □

La précédente proposition donne ainsi une autre démonstration de $\mathbb{E}[X_T^{(k)}] = K_m^{(k)}$ pour $k \geq 2$. Pour $k = 1$, on reprend la démonstration de la Proposition 5.1. On a $W_0 = f^{(1)}(X_0^{(0)}, X_1^{(0)}) = f^{(1)}(m, 0)$ et $W_T = f^{(1)}(X_t^{(0)}, X_T^{(1)}) = f^{(1)}(0, X_T^{(1)}) = X_T^{(1)} - 1$. D'où $\mathbb{E}[X_T^{(1)}] - 1 = \mathbb{E}[W_T] = \mathbb{E}[W_0] = W_0 = f^{(1)}(m, 0) = K_m^{(1)}$. D'où $\mathbb{E}[X_T^{(1)}] = K_m^{(1)} + 1 = H_m$.

7. Distribution non-uniforme des vignettes. — On suppose maintenant qu'à tout instant n, la probabilité que la variable Y_n soit égale à i est un nombre $p_i > 0$ ($i = 1, 2, \ldots, m$) et non plus nécessairement $1/m$. Par commodité, pour tout sous-ensemble $A \subset [m] := \{1, 2, \ldots, m\}$, on pose : $\mathbf{p}_A := \sum_{i \in A} p_i$. Pour $k \geq 1$, $n \geq 1$ et $i = 1, 2, \ldots, m$, on pose également $X_{n,i}^{(k)} := 1$ si la vignette numérotée i est apparue exactement k fois jusqu'à et y compris la date n et $X_{n,i}^{(k)} := 0$, sinon. De là, $X_n^{(k)} = \sum\limits_{1 \leq i \leq m} X_{n,i}^{(k)}$ est le nombre de vignettes qui sont apparues exactement k fois jusqu'à et y compris la date n.

PROPOSITION 7.1. — *Soit T l'instant où l'aîné termine son album. Alors, la fonction génératrice $G(t)$ des nombres $\mathbb{E}[X_T^{(k)}]$ est donnée par :*

$$G(t) := 1 + \sum_{k \geq 1} \mathbb{E}[X_T^{(k)}]\, t^k = t + \sum_{i=1}^{m} \sum_{A \subset [m] \setminus \{i\}} (-1)^{m-|A|} \frac{1 - p_i - \mathbf{p}_A}{1 - tp_i - \mathbf{p}_A}. \tag{7.1}$$

Démonstration. — D'abord,

$$\begin{aligned} G(t) &= 1 + \sum_{k \geq 1} \sum_{1 \leq i \leq m} \mathbb{E}[X_{T,i}^{(k)}]\, t^k = 1 + \sum_{k \geq 1} \sum_{1 \leq i \leq m} \mathrm{P}\{X_{T,i}^{(k)} = 1\}\, t^k \\ &= 1 + \sum_{k \geq 1} \sum_{1 \leq i \leq m} \sum_{n \geq 1} \sum_{1 \leq j \leq m} \mathrm{P}\{X_{n,i}^{(k)} = 1, T = n, Y_n = j\}\, t^k, \end{aligned}$$

puisque T est presque sûrement fini. Pour tout entier $n \geq 1$ et tout sous-ensemble $A \subset [m]$, notons $\mathrm{Surj}(n, A)$ l'évènement « $f : l \mapsto Y_l$ est, pour $l = 1, 2, \ldots, n$, une *surjection* de $[n]$ sur A ». Alors, pour $k \geq 1$ et $i \neq j$, l'évènement $\{X_{n,i}^{(k)} = 1, T = n, Y_n = j\}$ s'écrit encore $\{\mathrm{Surj}(n-1, [m] \setminus \{j\}), X_{n-1,i}^{(k)} = 1, Y_n = j\}$. Pour $k \geq 1$ et $i = j$, on a, en revanche,

$$\{X_{n,i}^{(k)} = 1, T = n, Y_n = j\} = \begin{cases} \{\mathrm{Surj}(n-1, [m] \setminus \{i\}), Y_n = i\}, & \text{si } k = 1\,; \\ \emptyset, & \text{si } k \geq 2. \end{cases}$$

Comme T est presque sûrement fini, on a aussi

$$\sum_{n\ge 1}\sum_{i} \mathrm{P}\{\mathrm{Surj}(n-1,[m]\setminus\{i\}), Y_n=i\}=1.$$

D'où

(7.2) $\quad G(t)-1-t$

$$\begin{aligned}
&=\sum_{n\ge1}\sum_{k\ge1}\sum_{j}\sum_{i\ne j}\mathrm{P}\{\mathrm{Surj}(n-1,[m]\setminus\{j\}), X^{(k)}_{n-1,i}=1, Y_n=j\}\,t^k\\
&=\sum_{n\ge1}\sum_{j} p_j\sum_{i\ne j}\sum_{k\ge1}\mathrm{P}\{\mathrm{Surj}(n-1,[m]\setminus\{j\}), X^{(k)}_{n-1,i}=1\}\,t^k\\
&=\sum_{j} p_j\sum_{i\ne j}\sum_{n\ge1}\sum_{f\in\mathrm{Surj}(n-1,[m]\setminus\{j\})} w_i(f),
\end{aligned}$$

avec

$$w_i(f):=t^{|f^{-1}(i)|}\prod_{h\in f([n-1])} p_h^{|f^{-1}(h)|},$$

pour chaque fonction $f:[n-1]\to[m]$. Le principe d'inclusion-exclusion permet d'écrire :

$$\sum_{f\in\mathrm{Surj}(n-1,[m]\setminus\{j\})} w_i(f)=\sum_{A\subset[m]\setminus\{j\}}(-1)^{m-1-|A|}\sum_{f:[n-1]\to A} w_i(f),$$

où

$$\sum_{f:[n-1]\to A} w_i(f)=\begin{cases}\Big(tp_i+\displaystyle\sum_{h\in A\setminus\{i\}} p_h\Big)^{n-1}, & \text{si } i\in A\,;\\[2ex] \Big(\displaystyle\sum_{h\in A} p_h\Big)^{n-1}, & \text{si } i\notin A.\end{cases}$$

On en tire :

$$\begin{aligned}
&G(t)-1-t\\
&=\sum_{j} p_j\sum_{i\ne j}\sum_{n\ge1}\sum_{A\subset[m]\setminus\{i,j\}}(-1)^{m-1-|A|}\Big(\mathbf{p}_A^{n-1}-(tp_i+\mathbf{p}_A)^{n-1}\Big)\\
&=\sum_{j} p_j\sum_{i\ne j}\sum_{A\subset[m]\setminus\{i,j\}}(-1)^{m-1-|A|}\Big(\frac{1}{1-\mathbf{p}_A}-\frac{1}{1-tp_i-\mathbf{p}_A}\Big),
\end{aligned}$$

d'où la formule (7.1) en sommant sur j et sur le terme $1/(1-\mathbf{p}_A)$. $\square$

La formule (7.1) se spécialise évidemment en la formule obtenue pour l'équirépartition. Nous ne reproduisons pas le calcul.

8. Autre méthode de calcul de la fonction génératrice. — Dans le cas de l'équirépartition, la formule (7.2) devient :

$$G(t) = 1 + t + (m-1)\sum_n \sum_{k\ge 1} \mathrm{P}\{\mathrm{Surj}(n-1,[m-1]), X^{(k)}_{n-1,1}=1\} t^k.$$

Une surjection de l'évènement $\mathrm{Surj}(n-1,[m-1])$ est caractérisée par une suite $(\sigma,\mathbf{c},\mathbf{g}) := (\sigma,(c_1,\ldots,c_{m-1}),(g_1,\ldots,g_{m-1}))$, où σ est la suite des vignettes telles qu'elles apparaissent pour la première fois, disons aux dates $d_1, d_2, \ldots, d_{m-1}$ $(1 = d_1 < d_2 < \cdots < d_{m-1} \le n-1)$, les c_j sont les nombres $d_2 - d_1 - 1$, $d_3 - d_2 - 1$, … , $n - 1 - d_{m-1}$ et g_j est l'application qui envoie l'intervalle $[d_j + 1, d_{j+1} - 1]$ (par convention, $d_m := n$) sur l'ensemble des numéros des vignettes déjà sorties $(j = 1, \ldots, m-1)$. On peut alors écrire

$$\begin{aligned} G(t) = 1 + t &+ (m-1)(m-2)! \\ &\times \sum_n \sum_{k\ge 1} \mathrm{P}\{\mathrm{Surj}'(n-1,[m-1]), X^{(k)}_{n-1,1}=1\} t^k, \end{aligned}$$

où l'évènement $\mathrm{Surj}'(n-1,[m-1])$ est restreint aux seules surjections $(\sigma,\mathbf{c},\mathbf{g})$ telles que les vignettes 2, 3, …, $(m-1)$ arrivent dans cet ordre (il y en a bien $(m-2)!$). Pour $n \ge m$ fixé, on a alors

$$\begin{aligned} &\sum_{k\ge 1} \mathrm{P}\{\mathrm{Surj}'(n-1,[m-1]), X^{(k)}_{n-1,1}=1\} t^k \\ &\quad = \sum_{c_1+\cdots+c_{m-1}=n-m} \Bigl(\frac{t}{m}\Bigr)^{c_1} \frac{1}{m}\Bigl(\frac{t+1}{m}\Bigr)^{c_2} \cdots \frac{1}{m}\Bigl(\frac{t+m-2}{m}\Bigr)^{c_{m-1}} \\ &\qquad - \sum_{c_2+\cdots+c_{m-1}=n-m} \frac{1}{m}\Bigl(\frac{1}{m}\Bigr)^{c_2} \cdots \frac{1}{m}\Bigl(\frac{m-2}{m}\Bigr)^{c_{m-1}}, \end{aligned}$$

puisque les seules vignettes qui contribuent un facteur t sont les vignettes notées 1 et que la variable t doit apparaître au moins une fois. En sommant cette somme pour $n \ge m$, on obtient :

$$\Bigl(\frac{1}{m}\Bigr)^{m-2} \frac{1}{1-t/m}\,\frac{1}{1-(t+1)/m} \cdots \frac{1}{1-(t+m-2)/m} - \frac{1}{(m-1)!}.$$

D'où

$$G(t) = t + \frac{1}{(1-t/2)(1-t/3)\cdots(1-t/m)},$$

qui est bien la formule cherchée.

9. Les entiers hyperharmoniques. — Multiplions le nombre harmonique $K_m^{(k)}$ par $m!^k$. Par construction, nous obtenons un entier positif, noté

$$J_m^{(k)} := K_m^{(k)}\, m!^k, \tag{9.1}$$

que nous désignons par *entier hyperharmonique.* Les premières valeurs de ces entiers sont reproduites dans le tableau ci-après.

m	2	3	4	5	6	7
$J_m^{(0)}$	1	1	1	1	1	1
$J_m^{(1)}$	1	5	26	154	1044	8028
$J_m^{(2)}$	1	19	460	15196	672336	38724624
$J_m^{(3)}$	1	65	6920	1229704	346296384	146661388992
$J_m^{(4)}$	1	211	95536	89222896	157188439296	483005642823936
$J_m^{(5)}$	1	665	1254176	6060649504	65990223258624	$\cdots$
$J_m^{(6)}$	1	2059	15958720	394810588096	26339109589241856	$\cdots$

Une de nos premières tâches a été de regarder si certaines lignes, colonnes ou diagonales de ce tableau figurent déjà dans l'*On-Line Encyclopedia of Integer Sequences* de Sloane. Dans cette encyclopédie en ligne, on trouve, en effet, la *ligne* se référant aux entiers $J_m^{(1)}$ $(m \ge 2)$ et les *colonnes* se rapportant aux entiers $J_3^{(k)}$ $(k \ge 0)$ et $J_4^{(k)}$ $(k \ge 0)$. Sloane indique que le développement

$$(1-x)^{-2}\ln(1/(1-x)) = x + \frac{x^2}{2!}5 + \frac{x^3}{3!}26 + \frac{x^4}{4!}154 + \frac{x^5}{5!}1044 + \cdots \tag{9.2}$$

apparaît dans un article des Mitrinović (1962), qui en ont donné une interprétation combinatoire en termes de nombres de Stirling généralisés. On voit que le précédent développement s'obtient par dérivation, par rapport à s, de l'identité (2.3). Le membre de droite de (9.2) est donc égal à $\sum_{m\ge 2} J_m^{(1)}\, x^{m-1}/(m-1)!$

Par ailleurs, la série

$$\frac{1}{(1-2x)(1-3x)} = 1 + 5x + 19x^2 + 65x^3 + 211x^4 + 665x^5 + \cdots \tag{9.3}$$

apparaît dans un article de Kreweras (1969), qui en a donné une interprétation en termes de relations binaires connexes. C'est naturellement, d'après l'identité (1.5), un développement qui est égal à : $\sum_{k\ge 0} J_3^{(k)} x^k$ (*cf.*

colonne $m = 3$ dans le tableau précédent). Enfin, la colonne $m = 4$, dont la fonction génératrice, d'après l'identité (1.5), est égale à

$$\frac{1}{(1-4!\,x/2)(1-4!\,x/3)(1-4!\,x/4)},$$

est simplement dans la liste de l'encyclopédie sans référence particulière.

10. Une asymptotique pour les nombres hyperharmoniques. Cette asymptotique découle, évidemment, de l'asymptotique des nombres harmoniques généralisés. Nous tenons simplement à indiquer comment on peut l'obtenir à l'aide des formules de changement de base des fonctions symétriques homogènes complètes aux fonctions symétriques de puissances $p_k = \sum x_i^k$ (*cf.* Macdonald (1995) p. 28). Pour m fixé, le nombre $K_m^{(k)}$ est la fonction symétrique homogène complète h_k de degré k en les variables $\frac{1}{2}, \ldots, \frac{1}{m}$.

D'abord, $K_m^{(1)} = h_1 = p_1$ et $p_1 - \ln m$ tend vers $\gamma - 1 \approx -0,4227$, où γ est la constante d'Euler. D'où $K_m^{(1)} / \ln m$ tend vers 1. D'autre part, pour $k \geq 2$, on a $p_k \to \zeta(k) - 1$. Comme

$$K_m^{(k)} = h_k = \frac{1}{k!} p_1^k + \sum a(i_1, i_2, \ldots, i_k)\, p_1^{i_1} p_2^{i_2} \cdots p_k^{i_k},$$

où la somme est sur toutes les suites $(i_1, i_2, \ldots, i_k)$ d'entiers telles que $1 \cdot i_1 + 2 \cdot i_2 + \cdots + k \cdot i_k = k$ et $i_2 + \cdots + i_k \geq 1$, on a, pour chaque terme de cette somme,

$$\frac{p_1^{i_1} p_2^{i_2} \cdots p_k^{i_k}}{(\ln m)^{i_1} (\ln m)^{i_2 + \cdots + i_k}} \to 0.$$

D'où

$$\frac{k!\, K_m^{(k)}}{(\ln m)^k} \to 1 \qquad (m \to +\infty). \tag{10.1}$$

Connaissant les valeurs de la fonction ζ en les entiers et aussi les coefficients de la matrice de passage de la base des h_λ à la base des p_λ, on peut aussi obtenir les formules asymptotiques suivantes, où les constantes sont des valeurs *approchées* :

$$\begin{aligned}
K_m^{(1)} &\approx \ln m - 0,4227;\\
K_m^{(2)} &\approx 0,5 \ln^2 m - 0,4227 \ln m + 0,4118;\\
K_m^{(3)} &\approx 0,1666 \ln^3 m - 0,2113 \ln^2 m + 0,4118 \ln m - 0,0815;\\
K_m^{(4)} &\approx 0,0417 \ln^4 m - 0,0704 \ln^3 m + 0,2059 \ln^2 m\\
&\qquad\qquad - 0,0815 \ln m + 0,0742.
\end{aligned}$$

Remerciements. — Les auteurs remercient Aimé Fuchs de leur avoir communiqué la méthode des martingales pour le calcul de $\mathbb{E}[X_T^{(1)}]$ d'après la version qu'en avait donnée Giorgio Letta. Ils remercient aussi Philippe Flajolet, qui a attiré leur attention sur les variations des nombres harmoniques généralisés et les travaux de Bentley et al. (1978) et Buchta (1989), ainsi que Christian Krattenthaler pour sa relecture attentive.

BIBLIOGRAPHIE

Banderier, Cyril ; Dobrow, Robert P. (2000). — A Generalized Cover Time for Random Walks on Graphs, Proc. FPSAC'00, Springer-Verlag.

Bauer, Heinz (1995). — *Wahrscheinlichkeitstheorie.* — Walter de Gruyter, Berlin.

Bentley, J.L. ; Kung, H.T. ; M., Schkolnick, M. ; Thompson, C.D. (1978). On the average number of maxima in a set of vectors and applications, *J. ACM*, **25**, p. 536–543.

Bergeron, François ; Labelle, Gilbert ; Leroux, Pierre (1994). — *Théorie des espèces et combinatoire des structures arborescentes.* — Publ. LACIM, vol. 19, Montréal.

Boneh, Arnon ; Hofri, Micha (1997). — The coupon-collector problem revisited—a survey of engineering problems and computational methods, *Comm. Statist. Stochastic Models*, **13**, no. 1, p. 39–66.

Buchta, Christian (1989). — On the average number of maxima in a set of vectors, *Information Proc. Letters*, **33**, p. 63–65.

Cartier, Pierre (2000). — Mathemagics, *Sém. Lothar. Combin.*, B42d, 71 pp. (`http://www.mat.univie.ac.at/~slc/`).

Feller, William (1968). — *An Introduction to Probability Theory and its Applications*, Third edition, vol. 1. — John Wiley & Sons, New York.

Flajolet, Philippe ; Gardy, Danièle ; Thimonier, Loÿs (1992). — Birthday paradox, coupon collectors, caching algorithms and self-organizing search, *Discrete Appl. Math.*, **39**, p. 207–229.

Flajolet, Philippe ; Labelle, Gilbert ; Laforest, Louise ; Salvy, Bruno (1995). — Hypergeometrics and the Cost Structure of Quadtrees, *Random Structures and Algorithms*, **7**, p. 117–144.

Foata, Dominique (1974). — *La série génératrice exponentielle dans les problèmes d'énumération.* — Les Presses de l'Université de Montréal, Montréal.

Graham, Ronald L. ; Knuth, Donald E. ; Patashnik, Oren (1989). *Concrete Mathematics.* — Addison-Wesley, Reading.

Kreweras, Germain (1969). — Inversion des polynômes de Bell bidimensionnels et application au dénombrement des relations binaires connexes, *C. R. Acad. Sci. Paris Ser. A-B*, **268**, A577-A579.

Lass, Bodo (2001). — Calcul combinatoire ensembliste. Thèse doctorat Aachen-Strasbourg, en préparation.

Letta, Giorgio (1992). — Communication personnelle à Aimé Fuchs.

Macdonald, I. G. (1995). — *Symmetric Functions and Hall Polynomials*, Second Edition. — Clarendon Press, Oxford.

Mitrinović, D. S., Mitrinović, R. S. (1962). — Tableaux d'une classe de nombres reliés aux nombres de Stirling, *Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fiz.*, **77**, 77 pp.

Pintacuda N. (1980). — Coupon Collectors via the Martingales, *Boll. Un. Mat. Ital. A*, **17**, p. 174–177.

Sloane, Neil, J. A. — *On-Line Encyclopedia of Integer Sequences*, `http://www.research.att.com/~njas`

Strehl, Volker (2001a). — Calcul automatique des identités hypergéométriques, Colloquium, Univ. Louis Pasteur, Strasbourg.

Strehl, Volker (2001b). — Ten years of special function computer algebra, *Sém. Lothar. Combin.*, en préparation.

Zeilberger, Doron (1990). — Holonomic system approach to special function identities, *J. of Computational and Appl. Math.*, **32**, p. 321-368.

Dominique Foata, Bodo Lass
et Guo-Niu Han
Département de mathématique
et I.R.M.A.-C.N.R.S.
Université Louis Pasteur
7, rue René-Descartes
F-67084 Strasbourg

`foata@math.u-strasbg.fr`
`lass@math.u-strasbg.fr`
`guoniu@math.u-strasbg.fr`