Algorithms Seminar 2000–2001,
F. Chyzak (ed.), INRIA, (2002), pp. 19–22.

Available online at the URL
http://algo.inria.fr/seminars/.

Hyperharmonic Numbers and the Phratry of the Coupon Collector

Dominique Foata
Département de mathématique, Université Louis Pasteur (France)

May 21, 2001

Summary by Cyril Banderier

Abstract

The classical coupon-collector problem is here extended to the case where the collector shares his harvest with other members of his phratry. She (!) remains the single buyer, but she gives to his brothers all the pictures that she got in double. When her album is filled, her brothers' albums have some empty places. How many in average? Dominique Foata (in a joint work with Guoniu Han and Bodo Lass) answers this question via an expression for the multivariate generating function. The problem is related to hyperharmonic numbers, that are studied here as solutions of finite differences equations.

1. Coupons Collector

A cleaver firm sells chocolate, with a picture (or "coupon") of a famous cricket player in each bar. In total, there are m different pictures to collect and each picture appears with probability $1/m$. Mr. and Mrs. Brown have r sons and one daughter, chocolate and cricket addicts. The girl (she's the oldest) is the only one to buy chocolate. She tries to complete her collection. When she gets a new picture, she puts it in her album, and when she gets a double, she gives it to her oldest brother, and when this one gets a double, he gives it to the remaining oldest brother, and so on. After having bought T bars of chocolate, the girl has completed her album, and it remains $M_T^{(i)}$ empty places in the album of the ith brother ($i = 1, \dots, r$). Let $X_n^{(k)}$ be the number of coupons which appeared exactly k times until time n. As $M_T^{(k)} = X_T^{(1)} + \cdots + X_T^{(k)}$, the distribution of $(T, M_T^{(1)}, \dots, M_T^{(r)})$ is then totally determined by the distribution of $(T, X_T^{(1)}, \dots, X_T^{(r)})$.

The question is to find formulae and asymptotics for $M_T^{(i)}$, or equivalently for $X_T^{(k)}$.

There is a lot of ways to solve coupon-collector-like problems. One can distinguish three main approaches:

- formal approach: combinatoricians indeed used a language-theory approach (shuffle products and Laplace transforms [2] or manipulation of regular expressions [6]);
- probabilistic approach: a lot of folklore results are established via basic probabilistic considerations, and more sophisticated tools such as martingales theory were also useful [5];
- matricial approach: this is in fact a mixture of the two precedent approaches, which is exploited to solve a Markovian generalization of the coupon-collector problem in [1] (with Perron–Frobenius theory and approximation of integrals).

This is with a combinatorial approach (enumeration of surjections and formal Laplace transform) that Foata *et al.* obtain in [4] the multivariate generating function of the coupon-collector problem, from which they derive the formulae for the expectations of $\mathbf{E}[T]$ and $\mathbf{E}[X_T^{(k)}]$.

2. The Multivariate Generating Function

Theorem 1. *The generating function of the random variables T, $X_T^{(1)}$, ..., $X_T^{(r)}$ is*

$$\sum_{l\geq m,\mathbf{n}} \mathbf{P}\{T=l, X_T^{(1)}=n_1,\dots,X_T^{(r)}=n_r\}t^l u_1^{n_1}\dots u_r^{n_r} =$$

$$u_1 t\sum\binom{m-1}{a,b,c_1,\dots,c_r}(-1)^b\Big(\prod_{k=1}^{r}\Big(\frac{u_k-1}{k!}\Big)^{c_k}\Big)\frac{(t/m)^{\sum_k kc_k}}{(1-at/m)^{1+\sum_k kc_k}}\Big(\sum_k kc_k\Big)!.$$

The proof follows from setting $u_i=1$ (for $i\geq r+1$), expanding (with Newton multinomial formula), and applying the Laplace transform to

$$\sum_{l\geq m}\frac{t^{l-1}}{(l-1)!}\sum_{\mathbf{n}}\sum_{s\in S(l,m;\mathbf{n})}\pi(s)=mu_1\Big(\sum_{i\geq 1}u_i\frac{t^i}{i!}\Big)^{m-1}.$$

The formal Laplace transform is defined as a linear map such that $\mathcal{L}\big(g_n\frac{t^n}{n!}\big)=g_nt^n$. This implies $\mathcal{L}\big(\exp(at)t^n\big)=n!\frac{t^n}{(1-at)^{n+1}}$. The set $S(l,m;\mathbf{n})$ is defined as a subset of the surjections from $[1,\dots,l]$ to $[1,\dots,m]$ for which $s\in S(l,m;\mathbf{n})$ implies i is reached n_i times and the restriction of s to $[1,\dots,l-1]$ is still a surjection from $[1,\dots,l-1]$ to $[1,\dots,m]\setminus\{s(l)\}$. The weight π of a surjection $S\in S(l,m;\mathbf{n})$ is defined by $\pi(s)=\prod {u_i}^{n_i}$.

3. Hyperharmonic Numbers

In order to complete one collection, it is well known that the average number of needed bars is

$$\mathbf{E}[T]=mH_m \qquad \text{where} \qquad H_m=\sum_{k=1}^{m}\frac{1}{k}.$$

For example, when there are $m=50$ different pictures, $\mathbf{E}[T]=50H_{50}\approx 50\times 4.5\approx 225$ and thus the daughter has a lot of doubles and we can expect that the oldest brother has almost completed his album with the 175 remaining pictures.

Pintacuda [5] proved with martingale theory that $\mathbf{E}[M_T^{(1)}]=H_m$. Foata *et al.* prove

Theorem 2. *For $k\geq 2$, the average number of empty places in the kth brother's album is*

$$\mathbf{E}[M_T^{(k)}]=1+\sum_{i=1}^{k}K_m^{(i)} \quad \textit{where} \quad K_m^{(k)}=\sum_{i=2}^{m}\frac{K_i^{(k-1)}}{i}, \quad (k\geq 1, m\geq 3)$$

with the following initial conditions $K_2^{(k)}=\frac{1}{2^k}$ (for $k\geq 0$) and $K_m^{(0)}=1$ (for $m\geq 2$).

A first derivation of this result follows of Theorem 1. Another proof is in two steps: first get the generating function for the $K_m^{(k)}$ (end of this section) and then prove that this generating function is also the one of the coupon-collector problem (next section).

Consider the rising factorial defined by $(a)_n=a(a+1)\dots(a+n-1)$ if $n\geq 1$ and $(a)_0=1$. An hypergeometric function with respect to two lists $(a_1,\dots,a_r)$ and $(b_1,\dots,b_s)$ is defined as the function given by the series

$${}_rF_s\left({a_1,\dots,a_r \atop b_1,\dots,b_s};x\right):=\sum_{n\geq 0}\frac{(a_1)_n\dots(a_r)_n}{(b_1)_n\dots(b_s)_n}\frac{x^n}{n!}.$$

The authors prove that the numbers $K_m^{(k)}$ (that they call "hyperharmonic numbers") satisfy

$$K_m^{(k)} = \frac{m(m-1)}{2^{k+1}} \, {}_{k+2}F_{k+1}\left(\begin{matrix} -m+2, 2, \ldots, 2 \\ 3, \ldots, 3 \end{matrix} \,\middle|\, 1 \right).$$

Comparing the recurrences satisfied by both sides and then summing gives the generating function

$$\sum_{k\geq 0} K_m^{(k)} t^k = \sum_{n=2}^{m} \frac{(-m)_n}{(n-2)!} \frac{1}{n} \frac{1}{1-t/n} = \frac{1}{(1-t/2)(1-t/3)\ldots(1-t/m)} \tag{1}$$

(the last equality following from a partial fraction decomposition).

Thus $K_m^{(k)} = h_k(\frac{1}{2}, \ldots, \frac{1}{k})$, the symmetric homogeneous polynomial of (total) degree k in $m-1$ variables. Reexpressing h_k in the basis of the power symmetric functions $p_k := \sum x_i^k$ gives

$$K_m^{(k)} \sim \frac{p_1^k}{k!} \sim \frac{(\ln m)^k}{k!}$$

One also has explicit asymptotics (for fixed k), e.g.,

$$K_m^{(3)} \approx 1.1666 \ln^3 m - 0.2113 \ln^2 m + 0.4118 \ln m - 0.0815.$$

4. Martingales Rescue the Phratry

Let $X_n^{(0)}$ be the number of empty places in the daughter's album. Now, define the process X as $X_n = \big(X_n^{(0)}, X_n^{(1)}, \ldots, X_n^{(r)}\big)$. For any function f, the average increase of $f(X)$ (knowing all the previously drawn coupons) is easy to get:

$$\mathbf{E}\big[f(X_{n+1}) - f(X_n) \;\big|\; Y_0, \ldots, Y_n\big] = \sum_{k=0}^{r} \frac{X_n^{(k)}}{m} \big(f(X_n^{(0)}, \ldots, X_n^{(k)} - 1, X_n^{(k+1)} + 1, \ldots, X_n^{(r)}) - f(X_n)\big);$$

this simply reflects the different possible updates ($X_n^{(k)}/m$ is the probability to get a new coupon which was already in k-tuple).

If f is such that the sum is 0, one has also $W_{n+1} - W_n = 0$ and thus W is a martingale, where W is the process $f(X)$ stopped at T, that is $W_n := f(X_n)$ (for $n < T$) and $W_n := f(X_T)$ (for $n \geq T$).

More generally, suppose that for r functions $f^{(1)}, \ldots, f^{(r)}$ from $\mathbb{N}^{k+1}$ to $\mathbb{R}$ one has:

1. $\displaystyle\sum_{i=0}^{k} x_i\big(f(x_0, \ldots, x_i - 1, x_{i+1} + 1, \ldots, x_k) - f^{(k)}(x_0, \ldots, x_k)\big) = 0$ for $x_0 \geq 1$;
2. $f^{(k)}(0, x_1, \ldots, x_k) = x_k$.

Then $\mathbf{E}\big[X_T^{(k)}\big] = f^{(k)}(m, 0, \ldots, 0)$.

Proof. 2. implies that $X_T^{(k)} = f(X_T) = W_T$; 1. gives a martingale property for W, Doob's theorem for stopping time of martingales gives $\mathbf{E}[W_T] = \mathbf{E}[W_0] = W_0$, and 2. implies that $W_0 = f^{(k)}(m, 0, \ldots, 0)$. □

Pintacuda [5] used this result with $k = 1$ and found $f^{(1)}(x_0, x_1) = H_{x_0} + \frac{x_1}{1+x_0}$. Foata *et al.* guessed the general formula:

Proposition 1. *For $k \geq 2$, the function $f^{(k)}$ defined as*

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

is the only solution of 1. and 2. One also has $f^{(k)}(x_0, 0, \ldots, 0) = K_{x_0}^{(k)}$.

They give two proofs in [4], but I prefer to explain what I heard in the meeting Random Structure and Algorithms (Poznan, August 2001), where Doron Zeilberger explained how to use a language theory argument to get a shorter proof. A complete collection of coupons can be written $11^*2\{1,2\}^*3\{1,2,3\}^*4\ldots\{1,2,\ldots,m-1\}^*m$. Let $\mathcal{W}$ be this set of words. This leads to the generating function

$$f(x_1, \ldots, x_m) := \frac{x_1}{m}\frac{1}{1-\frac{x_1}{m}} \cdots \frac{x_2}{m}\frac{1}{1-\frac{x_1+x_2}{m}} \frac{x_{m-1}}{m}\frac{1}{1-\frac{x_1+\cdots+x_{m-1}}{m}} x_m.$$

Recall that $\mathbf{E}[X_T^{(k)}]$ is the expected number of kinds of coupons in k-tuple (at time T, that is when the daughter has completed her album). Thus,

$$\begin{aligned}\sum_{k=1}^{\infty} \mathbf{E}\left[X_T^{(k)}\right] t^k &= \sum_{w\in\mathcal{W}} \sum_{k=1}^{m} \mathbf{P}(w) t^{|w|_k} = \sum_{k=1}^{m} \sum_{w\in\mathcal{W}} \left(\frac{1}{m}\right)^{|w|} t^{|w|_k} \\ &= m!\big(f(t,0,\ldots,0) + f(0,t,0,\ldots,0) + \cdots + f(0,\ldots,0,t)\big) \\ = t + t\sum_{k=1}^{m-1} &\frac{k!}{(2-t)(3-4)\ldots(k+1-t)} = t - 1 + \frac{m!}{\prod_{j=2}^{m}(j-t)}.\end{aligned}$$

As words of $\mathcal{W}$ are ordered (whereas it is in fact irrelevant for the coupon collector), there is a factor $m!$ at the second line takes into account all the permutations. The generating function obtained at the last line shows that the hyperharmonic numbers generated by Equation (1) indeed gives the average value of Theorem 2.

5. **Conclusion**

The coupon-collector problem (like the ménage problem, the birthday paradox) belongs to the large class of problems that can be modeled by simple urns models. It is very likely that, during the next years, the symbolic method will be applied with success to all these urns problems, and analytic combinatorics will then provide enumeration, complete asymptotics expansions and limit laws. The "classical" coupon-collector problem waits for his next revisitor!

This summary is related to Foata's article [4] (the more recent preprint [3] is also relevant). These articles are accessible at `http://www-irma.u-strasbg.fr/~foata/paper`.

Bibliography

[1] Banderier (Cyril) and Dobrow (Robert P.). – A generalized cover time for random walks on graphs. In *Formal power series and algebraic combinatorics (Moscow, 2000)*, pp. 113–124. – Springer, Berlin, 2000.

[2] Flajolet (Philippe), Gardy (Danièle), and Thimonier (Loÿs). – Birthday paradox, coupon collectors, caching algorithms and self-organizing search. *Discrete Applied Mathematics*, vol. 39, n° 3, 1992, pp. 207–229.

[3] Foata (D.) and Zeilberger (D.). – The collector's brotherhood problem using the Newman-Shepp symbolic method. – To appear in *Algebra Universalis*. Special issue dedicated to the memory of Gian-Carlo Rota.

[4] Foata (Dominique), Han (Guo-Niu), and Lass (Bodo). – Les nombres hyperharmoniques et la fratrie du collectionneur de vignettes. *Séminaire Lotharingien de Combinatoire*, vol. 47, n° B47a, 2001. – 20 pages. Available from `http://www.mat.univie.ac.at/~slc/`.

[5] Pintacuda (N.). – Coupons collectors via the martingales. *Bollettino dell'Unione Matematica Italiana. A. Serie V*, vol. 17, n° 1, 1980, pp. 174–177.

[6] Zeilberger (Doron). – How many singles, doubles, triples, etc., should the coupon collector expect? *Personal Journal of Ekhad and Zeilberger*, 2001. – 1 page. Available from `http://www.math.rutgers.edu/~zeilberg/`.