SOME NEW ASPECTS OF THE
COUPON COLLECTOR'S PROBLEM[*]

AMY N. MYERS[†] AND HERBERT S. WILF[‡]

Abstract. We extend the classical coupon collector's problem to one in which two collectors are simultaneously and independently seeking collections of d coupons. We find, in finite terms, the probability that the two collectors finish at the same trial, and we find, using the methods of Gessel–Viennot, the probability that the game has the following "ballot-like" character: the two collectors are tied with each other for some initial number of steps, and after that the player who first gains the lead remains ahead throughout the game. As a by-product we obtain the evaluation in finite terms of certain infinite series whose coefficients are powers and products of Stirling numbers of the second kind.

We study the variant of the original coupon collector's problem in which a single collector wants to obtain at least h copies of each coupon. Here we give a simpler derivation of results of Newman and Shepp and extend those results. Finally we obtain the distribution of the number of coupons that have been obtained exactly once ("singletons") at the conclusion of a successful coupon collecting sequence.

Key words. coupon, Stirling number, lattice path

AMS subject classification. 05-02

1. Introduction and results. The classical coupon collector's problem is the following. Suppose that a breakfast cereal manufacturer offers a souvenir ("coupon") hidden in each package of cereal, and there are d different kinds of souvenirs altogether. The collector wants to have a complete collection of all d souvenirs. What is the probability $p(n, d)$ that exactly n boxes of cereal will have to be purchased in order to obtain, for the first time, a complete collection of at least one of each of the d kinds of souvenir coupons?

The theory of this problem has practical applications, such as to the testing of methods for producing pseudorandom numbers for use in extended simulation or other Monte Carlo computations. To do that, suppose we are testing some method for producing pseudorandom numbers. Using that method we would generate a number of coupons in the range $[1, d]$, and when a complete set of them is achieved, we would record the number of trials that were needed. This experiment is then repeated a large number of times, and the observed mean and standard deviation of the number required to get a complete set is compared with their theoretical values. It turns out that this is a sensitive test for the randomness of relatively long sequences of random numbers that are produced by the method that is being tested.

The answer to the question in the first paragraph above is well known (e.g., [5, p. 132]) to be

$$(1.1) \qquad p(n, d) = \frac{d!}{d^n} \begin{Bmatrix} n - 1 \\ d - 1 \end{Bmatrix},$$

where the $\{{n \atop k}\}$'s are the Stirling numbers of the second kind, i.e., the number of partitions of the set $\{1, 2, \ldots, n\}$ into k classes.

We study, in this paper, a number of other aspects of this problem as well as a generalization of it to a two-player game.

First, suppose we have two coupon collectors, drawing coupons simultaneously, and each seeking to obtain a complete collection of d coupons. We ask for the probability that the two games are completed at the same time. The answer is given by (2.6) below. That answer is expressed in finite terms, owing to the closed form evaluation of the ordinary power series generating function for the squares of the Stirling numbers of the second kind, contained in (2.5).

Next, we consider the following two-person game. Again two coupon collectors are simultaneously drawing coupons at random. This time we are interested in a ballot-like problem: What is the probability that the player who first completed a collection (the *winner*) was never behind (i.e., never had fewer distinct coupons) at any intermediate stage of the play? Here we give a complete answer to a slightly easier question, namely the following: What is the probability that, after an initial segment of play in which the players are tied, one of them takes the lead and keeps the lead strictly until the end. The answer is in (2.26) below and is obtained by the Gessel–Viennot theory of nonintersecting lattice paths.

In each of these cases the answer can first be written as an infinite series whose coefficients involve various products of Stirling numbers. What is interesting, though, is that in all such cases we are able to express the answers in finite terms. Indeed, one of our main results here is the observation that infinite series whose coefficients involve various powers and products of Stirling numbers of the second kind can readily be evaluated in finite terms.

In section 4 we return to the original collecting problem of obtaining at least one copy of each coupon, but now we study the variant of the problem in which a single collector wants to obtain at least $h \geq 1$ copies of each coupon. We obtain the generating function (4.5) for the probability that exactly n trials are needed, the exact value of the average number of trials (4.10), and the asymptotic behavior (4.15) of these quantities as $n \to \infty$.

Finally, in section 5 we study the number of coupons that have been collected only once, at the end of a collection sequence. We find the distribution function (5.4) for this number and show that the average number of these singletons is just the harmonic number $H_d = 1 + 1/2 + \cdots + 1/d$.

2. The two-person collecting competition.

2.1. Simultaneous completion. We find now the probability of simultaneous completion of two independent coupon collecting sequences. Evidently this is

$$(2.1) \qquad \sum_{n \geq 0} p(n, d)^2 = \sum_{n \geq 0} \frac{d!^2}{d^{2n}} \left\{ {n-1 \atop d-1} \right\}^2,$$

which expresses the answer as an infinite sum. We can rewrite this as a finite sum by finding a finite expression for the generating function for the squares of the Stirling numbers of the second kind,

$$F_k(x) \stackrel{\text{def}}{=} \sum_{n \geq k} \left\{ {n \atop k} \right\}^2 x^n,$$

analogously to the well-known generating function for these numbers themselves,

$$(2.2) \qquad \sum_{n \geq k} \begin{Bmatrix} n \\ k \end{Bmatrix} x^n = \frac{x^k}{(1-x)(1-2x)\dots(1-kx)}.$$

The easiest way to do this is via the standard explicit formula for these Stirling numbers, viz.

$$(2.3) \qquad \begin{Bmatrix} n \\ k \end{Bmatrix} = \frac{1}{k!} \sum_{r=1}^{k} (-1)^{k-r} \binom{k}{r} r^n \qquad (1 \leq k \leq n)$$

$$\stackrel{\text{def}}{=} \sum_{r=1}^{k} A_{k,r} r^{n-k},$$

where we have written

$$(2.4) \qquad A_{k,r} = \frac{(-1)^{k-r} r^k}{k!} \binom{k}{r}.$$

It follows that

$$F_k(x) \stackrel{\text{def}}{=} \sum_{n \geq k} \begin{Bmatrix} n \\ k \end{Bmatrix}^2 x^n = \sum_{n \geq k} x^n \sum_{r,s=1}^{k} A_{k,r} A_{k,s} r^{n-k} s^{n-k}$$

$$= x^k \sum_{r,s=1}^{k} A_{k,r} A_{k,s} \sum_{n \geq k} (rsx)^{n-k}$$

$$(2.5) \qquad = x^k \sum_{r,s=1}^{k} \frac{A_{k,r} A_{k,s}}{1 - rsx} \qquad \left(|x| < \frac{1}{k^2} \right).$$

Thus for the simultaneous completion probability we obtain, from (2.1),

$$(2.6) \qquad \sum_{n \geq 0} p(n,d)^2 = \frac{d!^2}{d^{2d}} \sum_{r,s=1}^{d-1} \frac{A_{d-1,r} A_{d-1,s}}{1 - \frac{rs}{d^2}}$$

by (2.5), where the A's are given by (2.4). This sequence of probabilities, for $d = 1, 2, \dots$, begins as

$$1, \frac{1}{3}, \frac{11}{70}, \frac{9}{91}, \frac{688877}{9561123}, \frac{358555}{6330324}, \frac{2730269557627901}{58560931675094420}, \frac{146271649897951}{3695016639410525}, \dots,$$

i.e., as

$$1, 0.33333\dots, 0.15714\dots, 0.098901\dots, 0.072049\dots,$$
$$0.056640\dots, 0.046622\dots, 0.039586\dots, \dots.$$

2.2. Neck-and-neck then always ahead. We encode a sequence of n draws as a path ω with n vertices in the lattice $\mathcal{L}$ consisting of vertices (i,j) and edges $\{(i,j),(i+1,j)\}, \{(i,j),(i+1,j+1)\}$ for all $i,j \geq 0$. The first coordinate of a vertex in the path gives the number of draws, or *steps*, and the second coordinate gives the number of distinct coupons the collector has at that step. Thus ω starts at $(0,0)$

indicating the collector has 0 coupons at draw 0, proceeds to $(1,1)$ (the collector has 1 coupon after 1 draw), and ends at (n,d), $n \geq d$ (the collector has a complete collection at step n). We write $\omega = (0,0)\overline{\omega}(n,d)$, where $\overline{\omega}$ is a path from $(1,1)$ to $(n-1,d-1)$, to indicate that ω starts at the vertex $(0,0)$, continues with the first vertex $(1,1)$ in $\overline{\omega}$, then follows $\overline{\omega}$ through to $(n-1,d-1)$, and finally ends with the vertex (n,d).

We assign a weight of i/d to each horizontal edge $\{(i,j),(i+1,j)\}$ in the lattice $\mathcal{L}$. This is the probability that at the $(j+1)$st step, the collector draws one of the i distinct coupons already collected at step j. We assign a weight of $1 - i/d$ to each *northeast* edge $\{(i,j),(i+1,j+1)\}$. The probability that the collector draws the particular sequence of coupons encoded by the path ω is given by the product of the weights on the edges of ω. We let $P(\omega)$ denote this probability.

Suppose one collector, the *winner*, collects all d distinct coupons for the first time at step n. (At step $n-1$ the winner had $d-1$ distinct coupons.) Let ω_1 be the lattice path which encodes the winner's sequence of draws. Let ω_2 encode the other collector's draws. We compute the probability $p(d)$ that ω_1 and ω_2 are identical until some point at which the winner takes the lead and the other collector never catches up.

To do this, we begin by supposing ω_1 is identical to ω_2 until step k, at which point both collectors have d_1 distinct coupons. The argument splits into two cases, namely $k \leq n-2$ and $k = n-1$. In both cases, at step $k+1$ the winner collects one additional distinct coupon while the other collector does not. After step k, the two paths never intersect again. The winner collects all d distinct coupons for the first time at step n. Suppose the other collector has d_2 distinct coupons at this point. The probability we seek is

$$(2.7) \qquad p(d) = \sum_{n=d}^{\infty} \sum_{k=1}^{n-1} \sum_{d_1=1}^{d-1} \sum_{d_2=d_1}^{d-1} \sum_{(\omega_1,\omega_2)} P(\omega_1)P(\omega_2),$$

where the innermost sum ranges over all pairs (ω_1, ω_2) described above.

2.3. The case $k \leq n-2$. Write $\omega_1 = \alpha\overline{\omega_1}(n,d)$, where α denotes a lattice path from $(0,0)$ to (k,d_1), and $\overline{\omega_1}$ denotes a path from $(k+1,d_1+1)$ to $(n-1,d-1)$. Similarly, set $\omega_2 = \alpha\overline{\omega_2}$, where α is as above and $\overline{\omega_2}$ is a path from $(k+1,d_1)$ to (n,d_2). Note that $\overline{\omega_1}$ and $\overline{\omega_2}$ are nonintersecting paths in the lattice $\mathcal{L}$. In terms of these we have $P(\omega_1) = P(\alpha)(1-d_1/d)P(\overline{\omega_1})(1/d)$ and $P(\omega_2) = P(\alpha)(d_1/d)P(\overline{\omega_2})$. Hence from (2.7) we find for the combined probability of all pairs if $k \leq n-2$,

$$p(d)_{k \leq n-2} = \sum_{n=d}^{\infty} \sum_{k=1}^{n-2} \sum_{d_1=1}^{d-1} \sum_{d_2=d_1}^{d-1} \sum_{(\omega_1,\omega_2)} \left[P(\alpha)\left(1-\frac{d_1}{d}\right)P(\overline{\omega_1})\left(\frac{1}{d}\right) \right] \left[P(\alpha)\left(\frac{d_1}{d}\right)P(\overline{\omega_2}) \right]$$

$$(2.8) \qquad = \sum_{n=d}^{\infty} \sum_{k=1}^{n-2} \sum_{d_1=1}^{d-1} \sum_{d_2=d_1}^{d-1} \left(1-\frac{d_1}{d}\right)\left(\frac{1}{d}\right)\left(\frac{d_1}{d}\right) \sum_{\alpha} P(\alpha)^2 \sum_{(\overline{\omega_1},\overline{\omega_2})} P(\overline{\omega_1})P(\overline{\omega_2}).$$

At this point we have translated a question about coupon collecting into a problem involving nonintersecting paths in a lattice. We have set the stage for application of the Gessel–Viennot theorem [3]. This result concerns pairs of nonintersecting lattice paths with no constraints on vertices or edges in the paths. For this reason we have written ω_1 and ω_2 in terms of $\overline{\omega_1}$ and $\overline{\omega_2}$.

The theorem refers to an arbitrary set $\mathcal{L}$, which we will take to be the lattice defined earlier, and a weight (or valuation) v, which we take to be P. The theorem equates a sum of weights of paths with the determinant of a matrix $(a_{ij})_{1 \leq i,j \leq l}$. The entries of this matrix are defined by $a_{ij} = \sum_\omega v(\omega)$, where ω ranges over all paths from A_i to B_j.

The theorem requires that two given sequences, $(A_1, A_2, \ldots, A_l)$ and $(B_1, B_2, \ldots, B_l)$, of vertices in $\mathcal{L}$, the sets Ω_{ij}, $1 \leq i, j \leq l$, of all paths in $\mathcal{L}$ between A_i and B_j, and the weight v satisfy both the *finiteness* and *crossing conditions*. The *finiteness condition* requires the set of paths in Ω_{ij} with nonzero weight be finite. The *crossing condition* requires that paths in $\Omega_{ij'}$ and $\Omega_{i'j}$, $i < i'$ and $j < j'$, with nonzero weight share a common vertex. Both conditions hold for the paths we consider.

THEOREM 2.1 (Gessel–Viennot). *Suppose $\mathcal{L}$, v, $(A_1, A_2, \ldots, A_l)$, and $(B_1, B_2, \ldots, B_l)$ satisfy both the finiteness and crossing conditions. Then the determinant of the matrix $(a_{ij})_{1 \leq i,j \leq l}$ is the sum of the weights of all configurations of paths $(\omega_1, \omega_2, \ldots, \omega_l)$ satisfying the following two conditions:*
(i) *The paths ω_k are pairwise nonintersecting, and*
(ii) *ω_k is a path from A_k to B_k.*
In other words,

$$\det\left(\{a_{ij}\}_{i,j=1}^l\right) = \sum_{(\omega_1, \omega_2, \ldots, \omega_l)} v(\omega_1) v(\omega_2) \ldots v(\omega_l).$$

Application of this theorem to our problem requires the computation of only a 2×2 determinant! Let $A_1 = (k+1, d_1+1)$, $A_2 = (k+1, d_1)$, $B_1 = (n-1, d-1)$, and $B_2 = (n, d_2)$. Then

$$(2.9) \qquad \sum_{(\overline{\omega_1}, \overline{\omega_2})} P(\overline{\omega_1}) P(\overline{\omega_2}) = \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

where a_{ij} is the sum $\sum_\omega P(\omega)$ over all paths ω from A_i to B_j.

2.4. Paths from A to B. In this section we compute the probability $P(\omega)$ of an arbitrary path ω from a vertex $A = (a_1, b_1)$ to a vertex $B = (a_2, b_2)$ as well as the sum over all such paths. Such a path contains $b_2 - b_1$ *northeast* edges $\{(i,j),(i+1,j+1)\}$ and $(a_2 - a_1) - (b_2 - b_1)$ horizontal edges. The weights assigned to northeast edges in order from left to right are $1 - \frac{b_1}{d}, 1 - \frac{b_1+1}{d}, \ldots, 1 - \frac{b_2-1}{d}$. The weight assigned to a horizontal edge depends on its coordinates. Consider the edge $\{(i,j),(i+1,j)\}$. This edge indicates the collector has j distinct coupons at step i and draws one of the same j coupons at step $i+1$. The probability of this (weight of the edge) is $\frac{j}{d}$. Thus the probability of a path ω from A to B is

$$P(\omega) = \left(\frac{b_1}{d}\right)^{e_1} \left(1 - \frac{b_1}{d}\right) \left(\frac{b_1+1}{d}\right)^{e_2} \left(1 - \frac{b_1+1}{d}\right) \ldots \left(1 - \frac{b_2-1}{d}\right) \left(\frac{b_2}{d}\right)^{e_{b_2-b_1+1}}$$

$$= \frac{1}{d^{a_2-a_1}} \frac{(d-b_1)!}{(d-b_2)!} (b_1)^{e_1} (b_1+1)^{e_2} \ldots (b_2)^{e_{b_2-b_1+1}},$$

(2.10)

where $e = (e_1, e_2, \ldots, e_{b_2-b_1+1})$ is an ordered partition, a *composition*, of $(a_2 - a_1) - (b_2 - b_1)$ into $b_2 - b_1 + 1$ nonnegative integer parts. With this we compute the sum of

the probabilities of all paths from A to B.

$$\sum_{\omega=A\cdots B} P(\omega) = \sum_e \left(\frac{b_1}{d}\right)^{e_1} \left(1 - \frac{b_1}{d}\right) \left(\frac{b_1+1}{d}\right)^{e_2} \left(1 - \frac{b_1+1}{d}\right)$$

$$\cdots \left(1 - \frac{b_2-1}{d}\right) \left(\frac{b_2}{d}\right)^{e_{b_2-b_1+1}}$$

$$(2.11) \qquad = \frac{1}{d^{a_2-a_1}} \frac{(d-b_1)!}{(d-b_2)!} \sum_e (b_1)^{e_1}(b_1+1)^{e_2} \ldots (b_2)^{e_{b_2-b_1+1}},$$

where the sum is over all compositions $e = (e_1, e_2, \ldots, d_{b_2-b_1+1})$ of $(a_2-a_1)-(b_2-b_1)$ into $b_2 - b_1 + 1$ nonnegative integer parts. This is the coefficient of $x^{(a_2-a_1)-(b_2-b_1)}$ in the series expansion of

$$\frac{1}{(1-b_1 x)(1-(b_1+1)x)\ldots(1-b_2 x)},$$

so we can find a simpler formula for it by looking at the partial fraction expansion

$$(2.12) \qquad \frac{1}{\prod_{m=a}^{b}(1-mx)} = \sum_{m=a}^{b} \frac{B_m}{1-mx},$$

where

$$B_m = \frac{(-1)^{b-m} m^{b-a}}{(b-a)!} \binom{b-a}{m-a}.$$

From this and (2.11) we obtain

$$\sum_{\omega=A\cdots B} P(\omega) = \frac{1}{d^{a_2-a_1}} \frac{(d-b_1)!}{(d-b_2)!} [x^{(a_2-a_1)-(b_2-b_1)}] \left\{ \sum_{m=b_1}^{b_2} \frac{B_m}{1-mx} \right\}$$

$$= \frac{1}{d^{a_2-a_1}} \frac{(d-b_1)!}{(d-b_2)!} \sum_{m=b_1}^{b_2} B_m m^{(a_2-a_1)-(b_2-b_1)}$$

$$(2.13) \qquad = \frac{1}{d^{a_2-a_1}} \frac{(d-b_1)!}{(d-b_2)!} \sum_{m=b_1}^{b_2} \frac{(-1)^{b_2-m} m^{a_2-a_1}}{(b_2-b_1)!} \binom{b_2-b_1}{m-b_1}.$$

2.5. Evaluating the determinant. We use the results of the previous section to evaluate the determinant in (2.9). To compute a_{11}, we substitute $A = A_1 = (k+1, d_1+1)$ and $B = B_1 = (n-1, d-1)$ into (2.13). This yields

$$(2.14) \quad a_{11} = \frac{1}{d^{n-k-2}} (d-d_1-1)! \sum_{m=d_1+1}^{d-1} \frac{(-1)^{d-m-1} m^{n-k-2}}{(d-d_1-2)!} \binom{d-d_1-2}{m-d_1-1}.$$

In a similar manner we obtain

$$(2.15) \quad a_{12} = \frac{1}{d^{n-k-1}} \frac{(d-d_1-1)!}{(d-d_2)!} \sum_{m=d_1+1}^{d_2} \frac{(-1)^{d_2-m} m^{n-k-1}}{(d_2-d_1-1)!} \binom{d_2-d_1-1}{m-d_1-1},$$

$$(2.16) \quad a_{21} = \frac{1}{d^{n-k-2}} (d-d_1)! \sum_{m=d_1}^{d-1} \frac{(-1)^{d-m-1} m^{n-k-2}}{(d-d_1-1)!} \binom{d-d_1-1}{m-d_1},$$

$$(2.17) \quad a_{22} = \frac{1}{d^{n-k-1}} \frac{(d-d_1)!}{(d-d_2)!} \sum_{m=d_1}^{d_2} \frac{(-1)^{d_2-m} m^{n-k-1}}{(d_2-d_1)!} \binom{d_2-d_1}{m-d_1}.$$

Using (2.14)–(2.17) we compute the determinant of our 2×2 matrix.

$$(2.18) \quad \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \frac{(d-d_1)!(d-d_1-1)!}{d^{2n-2k-3}(d-d_2)!}$$

$$\times \sum_{l=d_1}^{d} \sum_{m=d_1}^{d_2} \frac{m(lm)^{n-k-2}(-1)^{d_1+d_2}(l-d)(m^2-l^2)}{(d-d_1-1)!(d_2-d_1)!} \binom{d-d_1-1}{l-d_1} \binom{d_2-d_1}{m-d_1}$$

$$(2.19) \quad \overset{\text{def}}{=} \det(d, d_1, d_2, k, n).$$

Substituting (2.19) in (2.9), we obtain

$$(2.20) \qquad \sum_{(\overline{\omega_1}, \overline{\omega_2})} P(\overline{\omega_1}) P(\overline{\omega_2}) = \det(d, d_1, d_2, k, n).$$

2.6. The initial common segment. In the previous section we evaluated the determinant in (2.9). In this section we compute the sum $\sum_\alpha P(\alpha)^2$ in (2.8). Recall α is a path from $(0,0)$ to (k, d_1).

Equation (2.10) gives the probability of an arbitrary path from A to B. Substituting $A = (0,0)$ and $B = (k, d_1)$ gives the probability

$$P(\alpha) = \frac{d!}{d^k (d-d_1)!} 1^{e_1} 2^{e_2} \cdots d_1^{e_{d_1}}$$

of an arbitrary path α from $(0,0)$ to (k, d_1). It follows that

$$\sum_{\alpha=(0,0)\cdots(k,d_1)} P(\alpha)^2 = \frac{d!^2}{d^{2k}(d-d_1)!^2} \sum_{e_1+\cdots+e_{d_1}=k-d_1} 1^{2e_1} 2^{2e_2} \cdots d_1^{2e_{d_1}}$$

$$= \frac{d!^2}{d^{2k}(d-d_1)!^2} [x^{k-d_1}] \left\{ \frac{1}{(1-1^2 x)(1-2^2 x)\ldots(1-d_1^2 x)} \right\}$$

$$= \frac{d!^2}{d^{2k}(d-d_1)!^2} \sum_{m=1}^{d_1} C_m m^{2k-2d_1} \qquad \text{(as in (2.12))}$$

$$(2.21) \qquad = \frac{2d!^2}{(d-d_1)!^2 d^{2k}(2d_1)!} \sum_{m\geq 1} (-1)^{d_1-m} \binom{2d_1}{d_1+m} m^{2k}$$

$$(2.22) \qquad \overset{\text{def}}{=} \operatorname{init}(d, d_1, k).$$

2.7. The case $k = n - 1$. Suppose now that the two walks are identical up to the point $(n-1, d-1)$. Since step n is the finish, the next step for the winning player will be to (n, d) and for the losing player to $(n, d-1)$. These last steps have respective probabilities $1/d$ and $1 - 1/d$. Hence the probability of the complete pair of walks in this case is the probability of two identical walks from $(0,0)$ to $(n-1, d-1)$ (which is given by (2.21) with $(k, d_1) := (n-1, d-1)$) multiplied by $(d-1)/d^2$.

2.8. Putting it together. We now substitute (2.19) and (2.22) into (2.8) to obtain the probability of all pairs of paths that we are considering,

$$p(d) = \sum_{n=d}^{\infty} \sum_{k=1}^{n-2} \sum_{d_1=1}^{d-1} \sum_{d_2=d_1}^{d-1} \left(1 - \frac{d_1}{d}\right) \left(\frac{d_1}{d}\right) \left(\frac{1}{d}\right) \operatorname{init}(d, d_1, k) \det(d, d_1, d_2, k, n)$$

$$(2.23) \qquad + \frac{d-1}{d^2} \sum_{n=d}^{\infty} \operatorname{init}(d, d-1, n-1)$$

$$(2.24) \quad \overset{\text{def}}{=} \Sigma_1 + \Sigma_2.$$

It turns out that the sums over the indices d_2, n, k can all be carried out in explicit closed form. Hence we can obtain an expression which is in finite terms for the total probability.

First, the sum on d_2 in Σ_1 above can be done in closed form since

$$\sum_{d_2=d_1}^{d-1} (-1)^{d_2} \binom{d-d_1}{d-d_2}\binom{d_2-d_1}{t-d_1} = (-1)^{d+1}\binom{d-d_1}{d-t}.$$

Next, the remaining sum over the indices n and k in the first summation Σ_1 is
(2.25)

$$\psi(d,r,s,t) \stackrel{\text{def}}{=} \sum_{n=d}^{\infty}\sum_{k=1}^{n-2} \frac{r^{2k}t^{n-k-1}s^{n-k-2}}{d^{2n}} = \begin{cases} \frac{r^{2d}(d^3-2d^2-r^2d+3r^2)}{d^{2d-2}(d^2-r^2)^2s^2t} & \text{if } r^2 = st, \\ \frac{r^4(st)^{d-1}(r^2-d^2)+str^{2d}(d^2-st)}{d^{2d-2}(d^2-st)(d^2-r^2)(r^2-st)r^2s} & \text{otherwise.} \end{cases}$$

The sum over n in Σ_2 is trivial, and so there remain no infinite sums in our final expression for the probability $p(d)$, which is

$$\sum_{d_1=1}^{d-1} \frac{2d!^2 d_1(d-d_1)}{(d-d_1)!^2(2d_1)!} \sum_{r,s,t\geq 1} (-1)^{d_1-r-s-t}(s-t)\binom{2d_1}{d_1+r}$$

$$\binom{d-d_1-1}{s-d_1}\binom{d-d_1}{d-t}\psi(d,r,s,t)$$

(2.26) $$+ \frac{4(d-1)d!^2}{d^{2d-2}(2d-2)!}\sum_{r=1}^{d-1}(-1)^{d-1-r}\binom{2d-2}{d-1+r}\frac{r^{2d-2}}{d^2-r^2} + \delta_{d,1},$$

where ψ is given by (2.25).

This is the probability that the game is of the type we described, namely where the players are tied for some initial segment of trials and then the player who pulls ahead remains ahead always, expressed as a finite sum (albeit a complicated one!). More precisely, the values of $p(d)$ can be calculated, as rational numbers, with $O(d^4)$ evaluations of the above summand. The exact values of $p(d)$ for $d = 1, 2, 3, 4, 5, \ldots$ are

$$\left\{1, \frac{2}{3}, \frac{43}{70}, \frac{986}{2275}, \frac{5672893}{1912246}, \ldots\right\}.$$

As decimals, the values of $\{p(d)\}_{d=1}^{10}$ are

$$\{1.0, 0.66667, 0.61429, 0.43341, 0.29667, 0.21177, 0.16016, 0.12748, 0.10551, 0.08988\}.$$

2.9. One collector never behind. In contrast to the problem of staying ahead as soon as the tie is broken, which we have solved in the preceding sections, the problem in which the ultimate winner has never been behind is unsolved.

Suppose the winner collects all d distinct coupons for the first time at step n, at which point the other collector has $d' < d$ distinct coupons. We discuss the probability $b(d)$ the winner has never been behind. We use b for "ballot" since this version of the problem has a distinct *ballot-problem* flavor (see, e.g., [1]).

Let w_1 be the lattice path which encodes the winner's sequence of draws. Let ω_2 encode the other collector's sequence of draws. Then $b(d)$ is the probability that ω_2 *does not cross* ω_1. To say ω_2 *does not cross* ω_1 means, for each horizontal coordinate

(n, d)

(n, d')
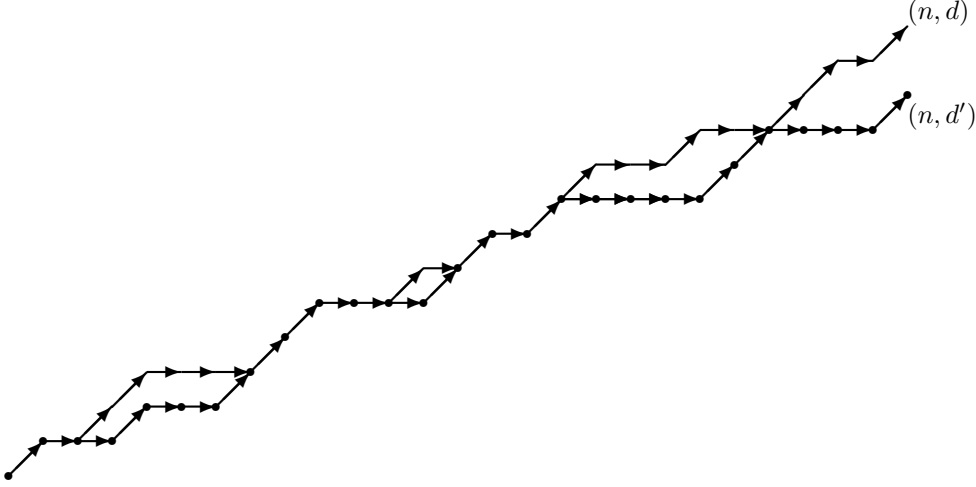
FIG. 2.1. *The winner is never behind.*

i shared by vertices (i, j_1) in ω_1 and (i, j_2) in ω_2, we have $j_2 \leq j_1$. In the case $j_2 = j_1$, we say ω_1 and ω_2 *intersect* at $(i, j_1) = (i, j_2)$. Thus we seek all pairs (ω_1, ω_2) such that ω_1 is a path from $(0, 0)$ to (n, d) including the vertex $(n-1, d-1)$, ω_2 is a path from $(0, 0)$ to (n, d') for $1 \leq d' \leq d$, and ω_2 does not cross ω_1. Such a pair (ω_1, ω_2) is illustrated by Figure 2.1. Note that ω_1 and ω_2 may intersect several times. The probability we seek is

$$b(d) = \sum_{n=d}^{\infty} \sum_{d'=1}^{d-1} \sum_{(\omega_1, \omega_2)} P(\omega_1) P(\omega_2),$$

where the innermost sum ranges over all pairs described above.

Look again at Figure 2.1. A pair (ω_1, ω_2) appears to form a chain of flying kites anchored to the ground at $(0, 0)$. The highest kite has two ribbons attached to its tip. Their loose ends are at (n, d) and (n, d').

Each kite consists of a *frame* together with a *tail*. See Figure 2.2. A *frame* from (i_1, j_1) to (i_2, j_2) consists of a pair of paths from (i_1, j_1), the *lower tip* of the frame, to (i_2, j_2), the *upper tip*, which intersect only at the endpoints. A *tail* from (i_1, j_1) to (i_2, j_2) consists of two identical paths between these endpoints. The *length* of a tail is the number of vertices in the tail minus one, i.e., the number of edges.

A pair (ω_1, ω_2) such that ω_2 does not cross ω_1 forms an alternating sequence of tails and frames, beginning with a tail. Note that tails may have length zero.

The upper tip of the final frame in this sequence is the common endpoint for two paths which intersect only at this common endpoint (these are the "ribbons" described above). One path ends at (n, d), this is the *top ribbon*, and the other ends at (n, d'), the *bottom ribbon*.

Below we compute the probability of a frame from (i_1, j_1) to (i_2, j_2), a tail from (i_1, j_1) to (i_2, j_2), and a pair of ribbons with common initial point (k, d'') and terminal points at (n, d) and (n, d'), respectively.

Let $f_{(i_1, j_1)}^{(i_2, j_2)}(d)$ denote the probability of a frame from (i_1, j_1) to (i_2, j_2). Note for $f_{(i_1, j_1)}^{(i_2, j_2)}(d) \neq 0$, we must have $i_2 \geq i_1 + 2$, $j_2 > j_1$, and $j_2 - j_1 \leq i_2 - i_1 - 1$. Assuming
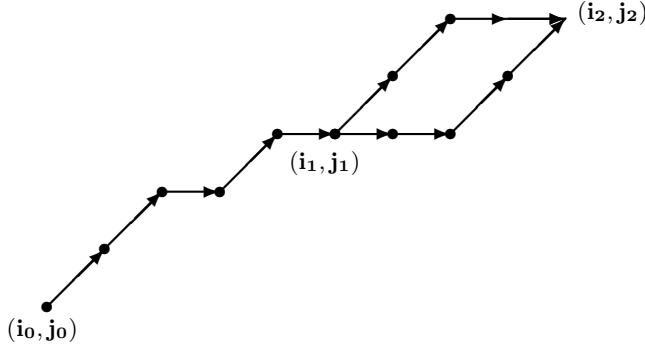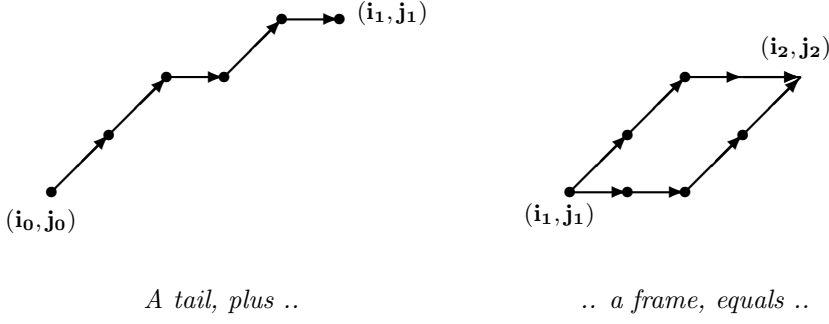
A tail, plus ..   .. a frame, equals ..

FIG. 2.2. *A kite.*

these conditions, we write

$$(2.27) \qquad f_{(i_1,j_1)}^{(i_2,j_2)}(d) = \sum_{(\alpha,\beta)} P(\alpha)P(\beta),$$

where (α, β) is a pair of paths from (i_1, j_1) to (i_2, j_2) intersecting only at the endpoints such that β does not cross α. (That is, α forms the upper edge of the frame, and β forms the lower edge.)

We convert the sum above into a determinant using the Gessel–Viennot theorem. Evaluation of the determinant gives

$$f_{(i_1,j_1)}^{(i_2,j_2)}(d) = \frac{j_1 j_2 (d - j_1)!^2}{d^{2(i_2 - i_1)}(j_2 - j_1)!^2 (d - j_2)!^2}$$

$$\sum_{l,m=j_1}^{j_2} (-1)^{l+m}(lm)^{i_2-i_1-2}(l-j_1)(m-j_2)\binom{j_2-j_1}{m-j_1}\binom{j_2-j_1}{l-j_1}.$$

We compute the probability $t_{(i_1,j_1)}^{(i_2,j_2)}(d)$ of a tail from (i_1, j_1) to (i_2, j_2) in a manner analogous to the computation of $\sum_\alpha P(\alpha)^2$ in section 2.6. We obtain

$$t_{(i_1,j_1)}^{(i_2,j_2)}(d) = \frac{(d - j_1)!^2}{d^{2(i_2 - i_1)}(2j_2)!(d - j_2)!^2}$$

$$\sum_{m=1}^{j_2}(-1)^{j_2-j_1}m^{2(i_2-i_1)}(2m)!\binom{2j_2}{j_2+m}\binom{m+j_1-1}{j_1-m}.$$

Finally we compute the probability $r(d, d', d'', k, n)$ of a pair of ribbons with common initial point (k, d'') and terminal points (n, d') and (n, d). The probability is given by a determinant similar to the one in (2.9). In the present case, we have d'' in place of d_1 and d' in place of d_2. Thus

$$r(d, d', d'', k, n) = \det(d, d', d'', k, n).$$

3. Winning margin. Now we look for the probability distribution of the number of distinct coupons that the second player has collected at the moment the first player completes the collection. Let $g(d, d')$ denote the probability that the second player has collected exactly d' distinct coupons at that moment. The probability that the first player finishes after exactly n trials is $p(n, d)$; see (1.1). The probability that the second player has exactly d' distinct coupons after n trials, given that the first player has just completed the collection at that time, is

$$\binom{d}{d'} \left\{ {n \atop d'} \right\} \frac{d'!}{d^n}$$

if $1 \le d' < d$. Thus for $d' < d$ our distribution g is given by

$$
\begin{aligned}
g(d, d') &= \sum_{n \ge d'} \frac{d!}{d^n} \left\{ {n-1 \atop d-1} \right\} \binom{d}{d'} \left\{ {n \atop d'} \right\} \frac{d'!}{d^n} \\
&= \frac{d!^2}{(d-d')!} \sum_{n \ge d'} \left\{ {n-1 \atop d-1} \right\} \left\{ {n \atop d'} \right\} \frac{1}{d^{2n}} \\
&= \frac{d!^2}{(d-d')! d^{2d'}} \sum_{r=1}^{d-1} r^{d'-d} \sum_{s=1}^{d'} \frac{A_{d-1,r} A_{d',s}}{1 - \frac{rs}{d^2}} - \delta_{d',1},
\end{aligned}
$$

by (2.3). A table of the probabilities $\{g(6, j)\}_{j=1}^5$ is as follows:

$$.000003, .000793, .018444, .118454, .333986.$$

These do not sum to 1 because the second player might have completed a collection at some time before the first player did.

4. The "double dixie cup problem," of Newman and Shepp, revisited. Here we consider a different generalization of the coupon collector's problem. Let integers $h, d \ge 1$ be fixed. Again we are sampling with replacement from d kinds of coupons, but now T is the epoch at which we have collected at least h copies of each of the d coupons for the first time. (For example, my $h - 1$ siblings and I might each want to have our own copy of every one of the available baseball cards.) We study the expectation, the probability generating function, and the asymptotic behavior of the expectation of this generalized problem.

These questions were investigated by Newman and Shepp [4] and the asymptotics were refined by Erdős and Rényi [2]. It is interesting to note that this problem is equivalent to one about the evolution of a random graph. Suppose we fix n vertices, and then we begin to collect from among n kinds of coupons. If we collect a particular sequence, say, $\{c_1, c_2, c_3, \dots\}$, then we add the edges $(c_1, c_2), (c_3, c_4) \dots$. That is, we add an edge each time we choose a new pair of coupons. Our problem about collecting at least h copies of each kind of coupon is thereby equivalent to the question of

obtaining a minimum degree of at least h in an evolving random graph.[1] In this section we will not add anything new to the asymptotics of this problem. Instead we claim only a derivation simpler than the original and an explicit generating function, which gives a nice road to the asymptotics. We deal only with generating functions in one variable, whereas in [4] multivariate generating functions were used. We obtain not only the expectation of the time to reach a collection that has at least h copies of each kind of coupon, but also the complete probability distribution of that time.

For n fixed, consider a sequence of n drawings of coupons that constitutes, for the first time at the nth drawing, a complete collection of at least h copies of each of the d kinds of coupons.

There are d possibilities for the coupon that completes the collection on the nth drawing. There are $\binom{n-1}{h-1}$ ways to choose the set of earlier drawings on which that last coupon type occurred. On the remaining $n-h$ drawings we can define, as usual, an equivalence relation: two drawings i, j are equivalent if the same kind of coupon was drawn at the ith and the jth drawings. The number of such equivalence relations is equal to the number of ordered partitions of a set of $n-h$ elements into $d-1$ classes, each class containing at least h elements. We will denote this latter number by $(d-1)!\left\{{n-h \atop d-1}\right\}_h$, where the $\left\{{n \atop k}\right\}_h$'s count the unordered partitions of an n-set into k classes of at least h elements each.

The number of sequences of n drawings for which we achieve a complete collection for the first time at the nth drawing is therefore

$$d\binom{n-1}{h-1}(d-1)!\left\{{n-h \atop d-1}\right\}_h.$$

Since there are d^n possible drawing sequences of length n, the probability that $T = n$ is

$$(4.1) \qquad p_n = \frac{d!}{d^n}\binom{n-1}{h-1}\left\{{n-h \atop d-1}\right\}_h,$$

and the probability generating function is

$$P_h(x) \overset{\text{def}}{=} \sum_{n \geq 0} p_n x^n = \sum_{n \geq 0} \frac{d!}{d^n}\binom{n-1}{h-1}\left\{{n-h \atop d-1}\right\}_h x^n$$

$$(4.2) \qquad = d!\binom{xD-1}{h-1}\sum_{n \geq 0}\left\{{n-h \atop d-1}\right\}_h \left(\frac{x}{d}\right)^n,$$

where $D = \partial/\partial x$.

It remains to find the ordinary power series generating function of the $\left\{{n \atop k}\right\}_h$'s. The exponential formula immediately gives us their exponential generating function as

$$(4.3) \qquad \sum_{n \geq 0}\left\{{n \atop k}\right\}_h \frac{x^n}{n!} = \frac{1}{k!}\left(e^x - 1 - x - \cdots - \frac{x^{h-1}}{(h-1)!}\right)^k.$$

We can convert this into an ordinary power series generating function by applying the Laplace transform operator

$$\int_0^\infty e^{-sx} \cdots dx$$

---

[1] Our thanks to Ed Bender and to a helpful referee for pointing this out.

to both sides, which yields

$$\sum_{n\geq 0}\left\{{n \atop k}\right\}_h \frac{1}{s^{n+1}} = \frac{1}{k!}\int_0^\infty e^{-sx}\left(e^x - 1 - x - \cdots - \frac{x^{h-1}}{(h-1)!}\right)^k dx,$$

or finally

$$(4.4) \qquad \sum_{n\geq 0}\left\{{n \atop k}\right\}_h t^n = \frac{1}{k!t}\int_0^\infty e^{-x/t}\left(e^x - 1 - x - \cdots - \frac{x^{h-1}}{(h-1)!}\right)^k dx.$$

Now if we substitute (4.4) into (4.2) we obtain the probability generating function of the generalized coupon collector's problem in the form

$$(4.5)$$
$$P_h(x) = \frac{1}{d^{h-2}}\int_0^\infty \left\{\binom{xD-1}{h-1}x^{h-1}e^{-td/x}\right\}\left(e^t - 1 - t - \cdots - \frac{t^{h-1}}{(h-1)!}\right)^{d-1} dt.$$

In the above, $\binom{xD-1}{h-1}$ is the differential operator that is defined by

$$\binom{xD-1}{h-1}f(x) = \frac{1}{(h-1)!}\left(x\frac{d}{dx} - 1\right)\left(x\frac{d}{dx} - 2\right)\cdots\left(x\frac{d}{dx} - h\right)f(x).$$

However, it is easy to establish, by induction on h, the interesting fact that

$$(4.6) \qquad \binom{xD-1}{h-1}x^{h-1}e^{-td/x} = \frac{(td)^{h-1}}{(h-1)!}e^{-td/x}.$$

Hence we have proved the following evaluation.

THEOREM 4.1. *The probability generating function for the coupon collecting problem in which at least h copies of each coupon are needed is given by*

$$(4.7) \qquad P_h(x) = \frac{d}{(h-1)!}\int_0^\infty t^{h-1}e^{-td/x}\left(e^t - 1 - t - \cdots - \frac{t^{h-1}}{(h-1)!}\right)^{d-1} dt.$$

4.1. Two examples. Let's look at the cases $h = 1$, the classical case, and $h = 2$, where we want to collect at least two specimens of each of the d kinds of coupons.

If $h = 1$, then (4.7) takes the form

$$P_1(x) = d\int_0^\infty e^{-td/x}(e^t - 1)^{d-1}dt.$$

If we expand the power of $(e^t - 1)$ by the binomial theorem and integrate termwise, we obtain

$$P_1(x) = xd\sum_{j=0}^{d-1}\binom{d-1}{j}\frac{(-1)^{d-1-j}}{d-jx},$$

which is precisely the partial fraction expansion of the classical generating function (2.2).

To see something new, let $h = 2$. Then

$$(4.8) \qquad P_2(x) = d\int_0^\infty te^{-td/x}\left(e^t - 1 - t\right)^{d-1} dt.$$

Again, by termwise integration this can be made fairly explicit, but since the most interest attaches to the expectation, let's look at the average number of trials that are needed to collect at least two samples of each of d coupons. This is $P_2'(1)$, which after some simplification takes the form

$$(4.9) \qquad P_2'(1) = d^2 \int_0^\infty \left(\frac{t^2}{e^t - 1 - t} \right) (1 - (1+t)e^{-t})^d dt.$$

From this we can go in either of two directions: an exact evaluation or an asymptotic approximation. By termwise integration it is easy to obtain the following exact formula, which is a finite sum, for $\langle T \rangle_2$, the average number of trials needed to collect at least two of each of the d kinds of coupons:

$$(4.10) \qquad \langle T \rangle_2 = d^2 \sum_{m,j} (-1)^m \binom{d-1}{m} \binom{m}{j} \frac{(j+2)!}{(m+1)^{j+3}}.$$

For $d = 2, 3, 4, 5$ these are $2, 11/2, 347/36, 12259/864$. To facilitate comparison with the classical ($h = 1$) case, we show below, for $1 \le d \le 10$, a table of the expected numbers of trials needed when $h = 1, 2$.

$d:$	1	2	3	4	5	6	7	8	9	10
$\langle T \rangle_1:$	1.0000	3.0000	5.5000	8.3333	11.417	14.700	18.150	21.743	25.460	29.290
$\langle T \rangle_2:$	2.0000	5.5000	9.6389	14.189	19.041	24.134	29.425	34.885	40.492	46.230

4.2. Asymptotics. Now we investigate the asymptotic behavior of (4.9), for large d, to compare it with the $d \log d$ behavior of the classical case where $h = 1$.

THEOREM 4.2. *If there are d different kinds of coupons, and if at each step we sample one of the d kinds with uniform probability, let $\langle T \rangle_h$ denote the average number of samples that we must take until, for the first time, we have collected at least h specimens of each of the d kinds of coupons. Then for every $h \ge 1$, we have $\langle T \rangle_h \sim d \log d$ $(d \to \infty)$.*

Consider first the case $h = 2$. In (4.9) we make the substitution

$$(4.11) \qquad e^{-u} = 1 - (1+t)e^{-t},$$

where u is a new variable of integration. We then find that

$$(4.12) \qquad P_2'(1) = d^2 \int_0^\infty t(u) e^{-ud} du,$$

where $t(u)$ is the inverse function of the substitution (4.11), which is well defined since the right side of (4.11) increases steadily from 0 to 1 as t increases from 0 to ∞.

The main contribution to $P_2'(1)$ comes from values of u near $u = 0$, and when u is near 0 we have

$$t(u) = -\log u + O(\log \log u).$$

Following the arguments in [6, sect. 2.2], we see that $P_2'(1)$ of (4.12) has the same asymptotic behavior as

$$d^2 \int_0^c (-\log u) e^{-ud} du \qquad (0 < c < 1),$$

and in [6] this is shown to be

$$\sim d^2 \cdot \frac{\log d}{d} = d \log d.$$

Now we consider the asymptotic behavior of the expected number of trials for general values of h. From (4.7) we see that this expected number of trials can be written in the form

(4.13)
$$\frac{d2}{(h-1)!} \int_0^\infty \left\{ \frac{t^h}{e^t - 1 - t - \cdots - \frac{t^{h-1}}{(h-1)!}} \right\} \left\{ 1 - \left(1 + t + \frac{t^2}{2} + \cdots + \frac{t^{h-1}}{(h-1)!} \right) e^{-t} \right\}^d dt.$$

Again we make the change of variable

(4.14)
$$e^{-u} = 1 - \left(1 + t + \frac{t^2}{2} + \cdots + \frac{t^{h-1}}{(h-1)!} \right) e^{-t}$$

in the integral, and it takes the remarkably simple form (compare (4.12))

$$P_h'(1) = d^2 \int_0^\infty t(u) e^{-ud} du,$$

where $t(u)$ is the inverse function of the substitution (4.14). Again the main contribution to the integral comes from small values of u, and when u is small and positive we have

$$t(u) = -\log u + (h-1)\log(-\log u) + \cdots.$$

Using the method of section II.2 of [6] once more, we find that

(4.15)
$$\langle T \rangle_h = d \log d + (h-1) d \log\log d (1 + o(1)) \qquad (d \to \infty).$$

We remark that in the case of $d = 200$ coupons, the correct expected number of trials to obtain two of each coupon is 1614 trials, the approximation $d \log d$ is 1175, and the approximation $d \log d + (h-1)d \log\log d$ is 1393, each rounded to the nearest integer.

5. The number of singletons. In view of the asymptotics in the preceding section we realize that at the moment when a coupon collector sequence terminates with a complete collection, "most" coupons will have been collected more than once, and only "a few" will have been collected just once. We call a coupon that has been seen just once a *singleton*. We will now look at the distribution of singletons.

In more detail, let j be the number of singletons in a collecting sequence that terminates successfully at the nth step. We first want the joint distribution $f(n, j)$ of n and j, i.e., the probability that a collecting sequence halts successfully at the nth step and has exactly j singletons at that moment. We claim that

(5.1)
$$f(n, j) = \frac{d!}{d^n} \binom{n-1}{j-1} \left\{ \begin{matrix} n-j \\ d-j \end{matrix} \right\}_2.$$

Indeed, the last coupon to be collected can be chosen in d ways; the other $j-1$ singleton coupons can be chosen in $\binom{d-1}{j-1}$ ways and can be presented in an ordered sequence in $(j-1)!\binom{d-1}{j-1}$ ways. This ordered sequence can appear among the first $n-1$ trials in $\binom{n-1}{j-1}$ ways, and the remaining $n-j$ trials constitute an ordered partition of $n-j$ elements into $d-j$ classes, no class having fewer than two elements, which can be chosen in $(d-j)!\left\{ \begin{matrix} n-j \\ d-j \end{matrix} \right\}_2$ ways. If we multiply these together and divide by d^n, the number of n-sequences, we obtain the result (5.1) claimed above.

Next we compute the probability that a completed collecting sequence contains exactly j singletons, whatever the length of the sequence may be. That is, we find $F(j) = \sum_n f(n, j)$, where f is given by (5.1). We have, after using the generating function (4.4),

$$(5.2) \quad F(j) = \sum_n \frac{d!}{d^n} \binom{n-1}{j-1} \left\{ \begin{matrix} n-j \\ d-j \end{matrix} \right\}_2$$

$$(5.3) \quad = \frac{d!}{(d-j)!d^j} \left\{ \int_0^\infty \left\{ \binom{t\frac{\partial}{\partial t} + j - 1}{j-1} \left(\frac{e^{-xt}}{t} \right) \right\} (e^x - 1 - x)^{d-j} dx \right\}_{t \to 1/d}.$$

But using the fact that, analogously to (4.6), we have

$$\binom{t\frac{\partial}{\partial t} + j - 1}{j-1} \left(\frac{e^{-xt}}{t} \right) = \frac{x^{j-1}}{(j-1)!t^j} e^{-x/t},$$

we can simplify the expression for $F(j)$ to

$$(5.4) \quad F(j) = j\binom{d}{j} \int_0^\infty x^{j-1}(e^x - 1 - x)^{d-j} e^{-xd} dx \qquad (j = 1, 2, 3, \dots),$$

which is the desired distribution of the number of singletons in a successfully terminated coupon collecting sequence.

Now if we multiply by j and sum over j, we'll get the average number of singletons that appear in a completed collection of d coupons. This is, after some termwise integration,

$$\bar{j}(d) = d \sum_m (-1)^m \binom{d-2}{m} \frac{d(m+1)+1}{(m+2)^2(m+1)}.$$

If we expand the summand in partial fractions, viz.

$$\bar{j}(d) = d \sum_m (-1)^m \binom{d-2}{m} \left(\frac{1}{m+1} - \frac{1}{m+2} + \frac{d-1}{(m+2)^2} \right),$$

then each of the three sums indicated can be expressed in closed form, in two cases by using the identity

$$(5.5) \quad \sum_k (-1)^k \binom{n}{k} \frac{1}{x+k} = \frac{1}{x\binom{x+n}{n}},$$

directly, with $x = 1$ and $x = 2$, and in the third case by differentiating (5.5) w.r.t. x and using the result with $x = 2$. The identity (5.5) is itself certified, after multiplying by the denominator on the right, by the WZ proof certificate $R(n, k) = k(x+k)/((n+1)(k-n-1))$.

What results is that $\bar{j}(d) = H_d$, the dth harmonic number. That is, *the average number of singleton coupons in a completed collection sequence of d coupons is the harmonic number H_d.*

REFERENCES

[1] L. COMTET, *Advanced Combinatorics*, D. Reidel, Dordrecht, The Netherlands, 1974.
[2] P. ERDŐS AND A. RÉNYI, *On a classical problem of probability theory*, Magyar Tud. Akad. Mat. Kutató Int. Közl., 6 (1961), pp. 215–220 (English. Russian summary).

[3] I. Gessel and G. Viennot, *Binomial determinants, paths, and hook length formulae*, Adv. Math., 58 (1985), pp. 300–321.

[4] D. J. Newman and L. Shepp, *The double dixie cup problem*, Amer. Math. Monthly, 67 (1960), pp. 58–61.

[5] H. S. Wilf, *generatingfunctionology*, 2nd ed., Academic Press, Boston, 1994.

[6] R. Wong, *Asymptotic Approximations of Integrals*, Academic Press, Boston, 1989.