CHAPTER 7

MEAN COVER TIMES FOR COUPON COLLECTORS AND STAR GRAPHS

Erol Pekoz and Sheldon M. Ross

7.1 Introduction and Summary

Suppose that there are m distinct types of coupons and that each coupon collected is type j with probability P_j, $j = 1, \ldots, m$. Let N_k denote the number of coupons one needs to collect in order to have at least one of each of k distinct types. We are interested in using simulation to efficiently estimate the mean and variance of N_k, for each $k = 1, \ldots, m$. Whereas we could simulate the successive types of coupons obtained and then utilize the observed values of N_k over many runs to obtain our estimates, we will attempt to obtain estimators having smaller variances than these raw estimators.

In Section 7.2 we present simulation estimators based on conditional expectations, and in Sections 7.3 and 7.4 we show how these estimators can themselves be improved by use of stratified sampling (in Section 7.3) and control variates (in Section 7.4).

In Section 7.5, we specialize to the case where $k = m$, that is, where one is interested in $N \equiv N_m$, the number of coupons that need be collected until a complete set is obtained. Whereas our previous approach remains available, we present, in Section 7.5, new unbiased estimators of $E[N]$ that have smaller variances than our earlier estimator. In Section 7.6, we present analytical bounds on $E[N]$. In Section 7.7, we show how our results can be used to analyze the mean time until a random walk on a star graph visits k distinct leafs and then returns to the origin.

There is a large literature on the coupon-collecting problem, although we have not come across other papers that are concerned, as we are, with the number needed to have at least one coupon of each of k distinct types. The literature is mostly concerned with N, the number needed to obtain a full set. See Nath [3], Stadje [6], and the references therein for some background on the problem.

7.2 Variance Reduction by Use of Conditional Expectation

Suppose we simulate the above model until a complete collection is obtained. Let I_j, $j = 1, \ldots, m$ denote the jth type of coupon that is obtained. That is, I_1 is the type of the first coupon, I_2 is the type of the first coupon not of type I_1, and so on. Also let A_j denote the additional number of coupons, after the first type I_j coupon is obtained, until a type I_{j+1} is obtained. (Thus, for instance, if the types of the first eight coupons collected are 1, 1, 3, 1, 3, 1, 2, 1, then $I_1 = 1$, $I_2 = 3$, $I_3 = 2$, $A_1 = 2$, $A_2 = 4$.)

Proposition 7.1. Given $\mathbf{I} = (I_1, \ldots, I_m)$, $A_1, \ldots, A_{m-1}$ are independent geometric random variables, with

$$E[A_j] = (1 - Q_j)^{-1},$$

where

$$Q_j = \sum_{S=1}^{j} P_{I_S}, \quad j = 1, \ldots, m-1.$$

Proof.

$$\begin{aligned} P\{A_j = n_j,\quad j = 1, \ldots, m-1 | \mathbf{I} = (i_1, \ldots, i_m)\} \\ = CP\{A_j = n_j,\, I_j = i_j,\, j = 1, \ldots, m-1\} \\ = CP_{i_1} Q_1^{n_1 - 1} P_{i_2} Q_2^{n_2 - 1} P_{i_3} Q_3^{n_3 - 1} \ldots P_{i_{m-1}} Q_{m-1}^{n_{m-1} - 1} P_{i_m}, \end{aligned}$$

which, since C does not depend on $(n_1, \ldots, n_{m-1})$, proves the result.

Now, with N_k equal to the number of coupons needed to obtain at least one from k distinct types, it follows that

$$N_k = 1 + \sum_{j=1}^{k-1} A_j$$

Therefore, from Proposition 7.1,

$$E[N_k | \mathbf{I}] = 1 + \sum_{j=1}^{k-1} (1 - Q_j)^{-1}$$

and

$$\begin{aligned} E[N_k^2 | \mathbf{I}] &= \mathrm{Var}(N_k | \mathbf{I}) + E^2[N_k | \mathbf{I}] \\ &= \sum_{j=1}^{k-1} Q_j (1 - Q_j)^{-2} + \left(1 + \sum_{j=1}^{k-1} (1 - Q_j)^{-1}\right)^2. \end{aligned}$$

We propose using the average, over many simulation runs, of $E[N_k | \mathbf{I}]$ and $E[N_k^2 | \mathbf{I}]$ to estimate, respectively, $E[N_k]$ and $E[N_k^2]$. It is well known (see [5]) that these estimators have smaller variances than the raw simulation estimators N_k and N_k^2. In addition, to use them we need only to simulate the value of $\mathbf{I}$, and this can be easily accomplished as follows: Generate independent exponential random variables X_j with respective rates P_j, $j = 1, \ldots, m$. If we now order these random variables and let I_j denote the index of the jth smallest of them, then $\mathbf{I} = (I_1, \ldots, I_m)$ has the appropriate distribution. That this is so can be easily seen by imagining that the coupons are collected at random times distributed according to a Poisson process with rate 1. Given this, we can assert that the times until the first occurrences of the different types of coupons are independent exponential random variables with respective (for a type j coupon) rates P_j, $j = 1, \ldots, m$. Since our supposition as to when the coupons are collected does not affect the order of occurrence of the different types, the result follows.

Even though the use of the estimators $E[N_k|\mathbf{I}]$ and $E[N_k^2|\mathbf{I}]$ should result in a substantial variance reduction over the raw estimators N_k and N_k^2 (for instance, in the case where the $P_j \equiv 1/m$, the variances of these conditional estimators are both 0), additional variance reduction can still be obtained.

7.3 Additional Variance Reduction by Use of Stratified Sampling

For an ordering vector $\mathbf{I} = (I_1, \ldots, I_m)$ let $\mathbf{I}_j$ be the m-vector

$$\mathbf{I}_j = (j, I_1, \ldots, I_m),$$

where the component of $\mathbf{I}$ that is equal to j is deleted. (For instance, if $\mathbf{I} = (1, 3, 2)$, then $\mathbf{I}_3 = (3, 1, 2)$.)

Suppose now that $\mathbf{I}$ has been simulated. Then, rather than using $E[N_k|\mathbf{I}]$ and $E[N_k^2|\mathbf{I}]$, $k = 1, \ldots, m$, as our estimators from that simulation run, we can use the estimators

$$\sum_{j=1}^{m} P_j E[N_k|\mathbf{I}_j],$$

$$\sum_{j=1}^{m} P_j E[N_k^2|\mathbf{I}_j], \quad k = 1, \ldots, m.$$

These are stratified sampling estimators that take into account the fact that $P\{I_1 = j\} = P_j$. It is known (see [5]) that these estimators not only will remain unbiased but also will have smaller variances than will the estimators $E[N_k|\mathbf{I}]$ and $E[N_k^2|\mathbf{I}]$.

7.4 Additional Variance Reduction by Control Variates

Suppose that we have simulated the value of $\mathbf{I}$. Let M_r, $r = 1, \ldots, m$ denote the position of r in this ordering. That is, $M_r = j$ if the type r coupon was the jth type to be collected (i.e., $I_j = r$). Then

$$M_r = 1 + \sum_{i \neq r} I\{i \text{ before } r\}$$

and so

$$E[M_r] = 1 + \sum_{i \neq r} P_i/(P_r + P_i)$$

Since their expectations are known, we can thus utilize certain of the M_r as control variates (see [5]).

To see which of these M_r might be effective control variates, let us first suppose that we number the types so that $P_1 \geq P_2 \geq \cdots \geq P_m$. Now a control variate will be effective in reducing the variance of the simulation estimator if it is strongly correlated (either positively or negatively) with the estimator. As an illustration, suppose that P_1 is quite a bit larger than the other P_j and that these others are roughly equal in value. For instance, we might have that $P_1 = 0.1, P_j \approx 0.01, j = 2, \ldots, 91$. Then M_1 should be strongly (negatively) correlated with both $E[N_k|\mathbf{I}]$ and $E[N_k^2|\mathbf{I}]$, as well as with $\Sigma_{j=1}^m P_j E[N_k|\mathbf{I}_j]$ and $\Sigma_{j=1}^m P_j E[N_k^2|\mathbf{I}_j]$, and thus should be an effective control variate. That is, in estimating $E[N_k^S]$, $s = 1, 2, k \geq 1$, we can, in this situation, use estimators of the form

$$E[N_k^S|\mathbf{I}] + c_S(M_1 - E[M_1])$$

or, even better, one of the form

$$\sum_{j=1}^{m} P_j E[N_k^S|\mathbf{I}_j] + c_S(M_1 - E[M_1]),$$

where the value of c_s that minimizes the variance is estimated from the simulation (see [5]). If both P_1 and P_2 are quite a bit larger than the others, then we might use both as control variates — that is, we would use estimators of the form

$$\sum_{j=1}^{m} P_j E[N_k^S|\mathbf{I}_j] + c_{S,1}(M_1 - E[M_1]) + c_{S,2}(M_2 - E[M_2]).$$

Similarly, if P_m is quite a bit smaller than the others, then M_m should be strongly (positively) correlated with $E[N_k^S|\mathbf{I}]$ and so can be used as a control. Therefore, depending on the values of the $P_j, j = 1, \ldots, m$, we should be able to further improve upon our simulation estimators by using certain of the M_j as control variates. The use of all three variance reduction schemes should serve to greatly reduce the variance of the simulation estimator.

7.5 Simulation Estimators of $E[N]$

In this section we obtain new estimators of $E[N]$, the expected number of coupons one must collect in order to obtain at least one of each type.

To begin, let $i_1, \ldots, i_m$ be a permutation of $1, \ldots, m$. Let T_1 denotes the number of coupons it takes to obtain a type i_1, and for $j > 1$, let T_j denote the number of additional coupons after having at least one of each type $i_1, \ldots, i_{j-1}$ until one also has at least one of type i_j. (Thus, if a type i_j coupon is obtained before at least one of the types $i_1, \ldots, i_{j-1}$, then $T_j = 0$.) Then $N = \Sigma_{j=1}^m T_j$, and so

$$E[N] = \sum_{j=1}^{m} P\{i_j \text{ is the last of } i_1, \ldots, i_j \text{ to be collected}\}/P_{i_j} \tag{7.1}$$

We will now indicate three different ways of utilizing (7.1) to obtain a simulation estimator of $E[N]$. To begin, note that if we let $I(n_j : n_1, \ldots, n_{j-1})$ be the indicator for the event that a type n_j coupon is the last of the types $n_1, \ldots, n_j$ to be collected, then from (7.1) we have that $\sum_{j=1}^{m} I(i_j : i_1, \ldots, i_{j-1})/P_{i_j}$ is an unbiased estimator of $E[N]$. Since this is true for all permutations of $1, \ldots, m$, it thus follows that

$$\text{Est}(1) \equiv \frac{1}{m!} \sum \sum_{j=1}^{m} I(i_j : i_1, \ldots, i_{j-1})/P_{i_j}$$

is an unbiased estimator of $E[N]$, where the leftmost sum is over all $m!$ permutations.

Now, as before, let I_r be the rth type of coupon to be collected. Since there are exactly $\frac{(r-1)!}{(r-j)!}(m-j)!$ terms of the form $I(I_r : i_1, \ldots, i_{j-1})$ that are equal to 1, it follows that we can express the preceding estimator as

$$\begin{aligned} \text{Est}(1) &= \frac{1}{m!} \sum_{r=1}^{m} \sum_{j=1}^{r} \frac{(r-1)!}{(r-j)!}(m-j)!/P_{I_r} \\ &= \frac{1}{m} \sum_{r=1}^{m} P_{I_r}^{-1} \sum_{j=1}^{r} \binom{r-1}{j-1} \bigg/ \binom{m-1}{j-1}. \end{aligned} \tag{7.2}$$

Like the estimator of Section 7.2, Est (1) also depends on the simulated data only through $\mathbf{I}$. In addition, since we only need to compute (and save) the $m(m+1)/2$ sums $\binom{r-1}{j-1} \Big/ \binom{m-1}{j-1}$, $j \le r$, once, it involves roughly the same amount of computational time as does the estimator of $E[N]$ given in Section 7.2.

Another estimator can be obtained from (7.1) by fixing the permutation $i_1, \ldots, i_m$ and then using simulation to estimate the unknown probabilities in (7.1). So let us assume, without loss of generality, that the coupons are numbered so that P_i is nondecreasing in i, $i = 1, \ldots, m$. With X_i, $i = 1, \ldots, m$, being independent exponential random variables with rates P_i, we have that

$$\begin{aligned} P\{j \text{ is last of } 1, \ldots, j\} &= P\{X_j = \max(X_1, \ldots, X_j)\} \\ &= \int_0^\infty P_j \exp\{-P_j x\} \prod_{i=1}^{j-1} (1 - \exp\{-P_i x\}) dx \\ &= \int_0^1 \prod_{i=1}^{j-1} (1 - y^{P_i/P_j}) dy \\ &= E\left[\prod_{i=1}^{j-1} (1 - U^{P_i/P_j})\right], \end{aligned}$$

where U is a uniform (0, 1) random variable.

Hence, we can estimate $E[N]$ by generating a single random number U and then using the estimator

$$1/P_1 + \sum_{j=2}^{m} \frac{1}{P_j} \prod_{i=1}^{j-1} (1 - U^{P_i/P_j}).$$

However, because of the positive correlations between the successive products, we recommend using separate random numbers to estimate each of the products. In addition to estimate $P\{j \text{ is last of } 1, \ldots, j\}$, we can utilize $(1 - U)^{j-1}$ as a control variate. That is, we recommend estimating $P\{j \text{ is last of } 1, \ldots, j\}$ by generating a sequence of random numbers $U_1, \ldots, U_k$ and then using the estimator

$$\frac{1}{k}\sum_{r=1}^{k}\left\{\prod_{i=1}^{j-1}(1-U_r^{P_i/P_j})+c_j[(1-U_r)^{j-1}-1/j]\right\}$$

where the appropriate value of c_j is to be obtained from the simulated data by standard means. A separate set of random numbers is to be used for each different value of j. Call this estimator Est(2).

Example 7.1. Suppose that $m = 2$. Let $P_1 = p = 1 - P_2$. Then $N = 1 + X$, where

$$X = \begin{cases} \text{Geo}(p) & \text{w.prob. } 1-p, \\ \text{Geo}(1-p) & \text{w.prob. } p. \end{cases}$$

A simple computation gives that the variance of the raw simulation estimator is

$$\text{Var}(N) = (1-p)/p^2 + p/(1-p)^2 - 2.$$

With $\mathbf{I} = (I_1, I_2)$ equal to the order of appearance of the types, the conditional expectation estimator is

$$E[N|\mathbf{I}] = 1 + 1/P_{I_2}$$

and so the variance of this estimator is

$$\begin{aligned}\text{Var}(1+1/P_{I_2}) &= (1-p)/p^2 + p/(1-p)^2 - [(1-p)/p + p/(1-p)]^2 \\ &= (1-p)/p + p/(1-p) - 2.\end{aligned}$$

The first estimator of this section can be expressed as

$$\begin{aligned}\text{Est}(1) &= \frac{1}{2}[1/p + I\{I_2 = 2\}/P_{I_2} + 1/(1-p) + I\{I_2 = 1\}/P_{I_2}] \\ &= \frac{1}{2}[1/p + 1/(1-p)] + \frac{1}{2}1/P_{I_2}.\end{aligned}$$

Hence,

$$\text{Var}[\text{Est}(1)] = \text{Var}(1/P_{I_2})/4.$$

That is, Est(1) has one fourth the variance of the conditional expectation estimator.

To determine Var[(Est(2)], suppose that $p \leq 1/2$. Then

$$\text{Est}(2) = \frac{1}{p} + \frac{1}{1-p}\{1 - U^{p/(1-p)} + c[1 - U - 1/2]\},$$

where $c = -\text{Cov}(U, U^{p/(1-p)})/\text{Var}(U)$. A simple calculation now yields that

$$\begin{aligned}\text{Var}[\text{Est}(2)] &= \text{Var}[U^{p/(1-p)}] - \text{Cov}^2(U, U^{p/(1-p)})/\text{Var}(U) \\ &= p^2/(1-p^2) - 3p^2/(2-p)^2.\end{aligned}$$

Table 7.1 provides the variances of these four estimators for $p = 0.5, 0.4$, and 0.2.

Another approach to estimating $E[N]$ is to again suppose that the P_i are nondecreasing, utilize (7.1), but now estimate $P\{j \text{ is last of } 1, \ldots, j\}$ by first conditioning on the order of appearance of $1, \ldots, j-1$. Letting $I_{i,j}$ be the ith one of types $1, \ldots, j-1$ to appear, $i = 1, \ldots, j-1$, then

$$P\{j \text{ is last of } 1, \ldots, j|\mathbf{I}_j\} = \prod_{r=1}^{j-1}\left\{\sum_{i=r}^{j-1} P_{I_{i,j}} \Big/ \left(P_j + \sum_{i=r}^{j-1} P_{I_{i,j}}\right)\right\}.$$

Hence, we can simulate $\mathbf{I}_j$ (by simulating and then ordering $X_1, \ldots, X_{j-1}$, where X_i is exponential with rate P_i) and then use the preceding to estimate $P\{j \text{ is last of } 1, \ldots, j\}$. Let us call this estimator Est(3). Table 7.2 compares the variance of Est(2) and Est(3) in the case of $m = 3$ and specified values of the P_i.

Table 7.1. Variance of Estimators

p	N	C. Expectation	Est(1)	Est(2)
0.5	2	0	0	0
0.4	103/36	1/6	1/24	1/336
0.2	18.3125	2.25	0.5625	$1/216 \approx 0.00463$

Table 7.2. Comparison of Var[Est(2)] and Var[Est(3)]

P_1, P_2, P_3	Var[Est(2)]	Var[Est(3)]
0.2, 0.3, 0.5	3.081×10^{-3}	4.783×10^{-4}
0.1, 0.2, 0.7	1.466×10^{-3}	1.890×10^{-4}
0.1, 0.4, 0.5	2.094×10^{-3}	3.086×10^{-3}
0.3, 0.3, 0.4	1.095×10^{-3}	0
0.2, 0.4, 0.4	1.331×10^{-3}	2.222×10^{-3}
0.1, 0.3, 0.6	2.070×10^{-3}	1.089×10^{-3}

Remark 7.1. The problem of estimating the moments of the number of coupons one need collect to obtain a *fixed* set of k types, say types $1, 2, \ldots, k$, can be solved by considering the coupon-collecting problem having only those k types with the probabilities $P_j^* \equiv P_j/\Sigma_{i=1}^k P_i$, $j = 1, \ldots, k$. If N is the number that one needs to obtain a full set in this new problem, then, with $N(1, \ldots, k)$ equal to the number needed in the original problem,

$$N(1,\ldots,k) = \sum_{i=1}^{N} X_i,$$

where $X_1, \ldots$ is a sequence of independent and identically distributed geometric random variables with mean $1/\Sigma_{i=1}^k P_j$ that is independent of N. Hence, the moments of $N(1, \ldots, k)$ are easily obtained from those of N.

Remark 7.2. If $P_j \equiv 1/m$, then Est(1) is constant and must thus (since it is unbiased) equal $E[N]$, giving rise to the interesting identity

$$\sum_{j=1}^{m} m/j = \sum_{r=1}^{m}\sum_{j=1}^{r} \binom{r-1}{j-1} \Big/ \binom{m-1}{j-1}.$$

7.6 Bounds on $E[N]$

Suppose $p_1 \le P_2 \le \ldots \le P_m$. For fixed u, $u > 0$, it is easily shown that the function f, defined by

$$f(a_1,\ldots,a_{j-1}) = \prod_{i=1}^{j-1}(1-u^{a_i}),$$

is a Schur concave function. Hence, letting $\overline{P}_k = \Sigma_{i=1}^k P_i/k$, it follows that

$$\prod_{i=1}^{j-1}(1-U^{P_i/P_j}) \le \left(1-U^{\overline{P}_{j-1}/P_j}\right)^{j-1}.$$

Therefore,

$$\begin{aligned} P\{j \text{ is last of } 1,\ldots,j\} &= E\left[\prod_{i=1}^{j-1}(1-U^{P_i/P_j})\right] \\ &\le E\left\{\left(1-U^{\overline{P}_{j-1}/P_j}\right)^{j-1}\right\} \\ &= \prod_{i=1}^{j-1}\{i\overline{P}_{j-1}/(P_j+i\overline{P}_{j-1})\}, \end{aligned}$$

where the final equality follows since $E[(1-U^{\overline{P}_{j-1}/P_j})^{j-1}]$ is the probability that j is the last of $1, \ldots, j$ to appear when a type j occurs with probability P_j and each of the others occur with the same probability $\overline{P}_{j-1}$. Hence, we obtain the upper bound

$$E[N] \le 1/P_1 + \sum_{j=2}^{m}\frac{1}{j}\prod_{i=1}^{j-1}\{i\overline{P}_{j-1}/(P_j+i\overline{P}_{j-1})\}.$$

A lower bound for $E[N]$ can be obtained by letting $R_1, \ldots, R_m$ be a random permutation of $1, \ldots, m$. (That is, $R_1, \ldots, R_m$ is equally likely to be any of the $m!$ permutations.) Then, analogous to (7.1), we have

$$\begin{aligned} E[N] &= \sum_{j=1}^{m} j^{-1} E[1/P_{R_j} | R_j \text{ is the last of } R_1, \ldots, R_j] \\ &\geq \sum_{j=1}^{m} j^{-1} E[1/P_{R_j}] \\ &= \frac{1}{m} \sum_{i=1}^{m} 1/P_i \sum_{j=1}^{m} 1/j, \end{aligned}$$

where the inequality follows, since the conditional distribution of P_{R_j} given that R_j is the last of $R_1, \ldots, R_j$ is stochastically smaller than the unconditional distribution of P_{R_j}.

Another lower bound can be obtained by utilizing (7.1) along with the permutation $m, m-1, \ldots, 1$. Since $P_m \geq P_{m-1} \geq \ldots \geq P_1$, a simple coupling argument shows that

$$P\{j \text{ is the last of } m, m-1, \ldots, j \text{ to be obtained}\} \geq 1/(m-j+1),$$

implying that

$$E[N] \geq \sum_{j=1}^{m} [(m-j+1)P_j]^{-1}.$$

(Of course, a similar argument can be used to show that $E[N] \leq \Sigma_{j=1}^{m} 1/(jP_j)$, but this is not as strong as our previously derived upper bound.)

Remark 7.3. In the special case where $P_j = 1/m$ for all j, both bounds give the exact value.

Remark 7.4. Flajolet et al., in [2], present formulas for $E[N]$. However, an amount of time that is exponential in m is required to evaluate these formulas.

7.7 A Star Graph

Consider a star graph consisting of m rays, with ray i containing n_i vertices, $i = 1, \ldots, m$ (Figure 7.1). Let leaf i denote the leaf on ray i.

Assume that a particle moves along the vertices of the graph in the following manner. Whenever it is at the central vertex 0, it then moves to its neighboring vertex on ray i with probability α_i, $\Sigma_i \alpha_i = 1$. Whenever it is on a nonleaf vertex of ray i, then with probability p_i it moves to its neighbor in the direction of the leaf, and with probability $1 - p_i$ it moves to its neighbor in the direction of 0. When at a leaf, it next moves to its neighbor. Starting at vertex 0, we are interested in the mean number of transitions that it takes to visit k distinct leafs and then to return to 0,

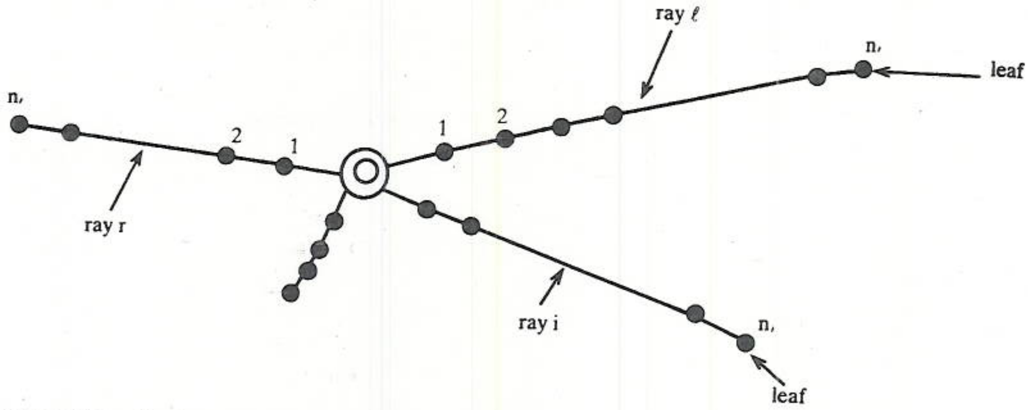

Figure 7.1. A star graph.

for $1 \le k \le m$. When $k = m$, this quantity is the mean cover time, defined to equal the mean time to visit all vertices and return to 0.

To begin, consider first a graph with a finite number of vertices and nonnegative weights defined on its edges, and suppose that a particle moves along the vertices of this graph in the following manner. Whenever it is at vertex i, it next moves to vertex j with probability

$$P_{ij} = w_{ij} \Big/ \sum_k w_{ik},$$

where w_{ij} is the weight on the edge (i, j). (If (i, j) is not an edge of the graph, then take w_{ij} to equal 0.) It is well known that the successive vertices visited constitute a time-reversible Markov chain with stationary probabilities

$$\pi_i = \sum_k w_{ik} \Big/ \sum_i \sum_k w_{ik}.$$

The random walk on the star graph can be regarded as a special case of the preceding model. Just let α_i denote the weight on the edge from 0 to its neighbor on ray i, and let $\alpha_i w_i^j$ be the weight on the edge connecting vertex j and $j + 1$ on ray i, $j = 1, \ldots, n_i - 1$ where $w_i = p_i/(1 - p_i)$. Hence, if we let μ_{00} denote the mean time between visits to 0, then from the result quoted it follows that

$$\begin{aligned} \mu_{00} = \pi_0^{-1} &= 1 + \sum_{i=1}^{m} \alpha_i \left[w_i^{n_i - 1} + \sum_{j=1}^{n_i - 1} (w_i^j - w_i^{j-1}) \right] \\ &= 1 + \sum_{i=1}^{m} [\alpha_i (1 + w_i - 2w_i^{n_i})]/(1 - w_i). \end{aligned} \tag{7.3}$$

If β_i denotes the probability that the particle, when at the 0-neighbor vertex on ray i, will reach leaf i before returning to 0, then by the gambler's ruin problem,

$$\beta_i = (1 - 1/w_i)/[1 - (1/w_i)^{n_i}].$$

Now, say that a cycle is completed every time the particle returns to vertex 0, and let X_j denote the time of the jth cycle, $j \geq 1$. If we let N_k denote the number of cycles needed to visit k distinct leafs, then $T_k = \Sigma_{j=1}^{N_k} X_j$ represents the time that it takes to visit k distinct leafs and then return to 0. By Wald's equation, we have that

$$E[T_k] = E[X]E[N_k] = \mu_{00}E[N_k]. \tag{7.4}$$

To determine $E[N_k]$, note that each cycle will take place on ray i with probability α_i, $i = 1, \ldots, m$. Calling a cycle an i-success if it reaches leaf i, $i = 1, \ldots, m$, and calling it a failed cycle if it returns to 0 before reaching a leaf, it follows that N_k is the number of cycles needed to obtain at least one i-success for k distinct values of i. Letting G_k denote the number of nonfailed cycles needed to obtain at least one i-success for k distinct values of i, then

$$N_k = \sum_{j=1}^{G_k} Y_j,$$

where Y_j is the number of failed cycles between the $(j-1)$st and the jth nonfailed cycle. Since each cycle will be a nonfailed cycle with probability$\Sigma\alpha_i\beta_i$, it follows that

$$E[Y_j] = (\sum \alpha_i\beta_i)^{-1} - 1$$

and, by Wald's equation,

$$E[N_k] = \{((\sum \alpha_i\beta_i)^{-1} - 1)\}E[G_k].$$

Thus, from (7.3) and (7.4), we see that the expected time to visit k distinct leafs and then return to 0 is

$$E[T_k] = E[G_k]\{((\sum \alpha_i\beta_i)^{-1} - 1)\}\left\{1 + \sum_{i=1}^{m}[\alpha_i(1 + w_i - 2w_i^{n_i})]/(1 - w_i)\right\}. \tag{7.5}$$

Now, each nonfaild cycle will be an i-success cycle with probability P_i, where

$$P_i = \alpha_i\beta_i \bigg/ \sum_{j=1}^{m} \alpha_j\beta_j, \quad i = 1, \ldots, m.$$

Hence, $E[G_k]$ is equal to the expected number of coupons one needs collect in order to have at least one of each of k distinct types, when each coupon is a type i with probability P_i, $i = 1, \ldots, m$. It can thus be estimated (and bounded) by the methods of the previous sections.

References

[1] Blom, G., and Sandell, D. Cover times for random walks on graphs. *Math. Scientist* **17**, 111–119, 1992.

[2] Flajolet, Gardy, and Thimonier, "Birthday paradox, coupon collectors, caching algorithms, and self organizing search." *Discrete Appl. Math.* **39**, 207–229, 1992.

[3] Nath, H. B. Waiting time in the coupon collector's problem. *Austral. J. Stat.* **15**, 132–135, 1973.

[4] Palacios, J. L. Cover times for stars. *Math. Scientist* **18**, 103–107, 1994.

[5] Ross, S. M. *Simulation*. Academic Press, Cambridge, 1997.

[6] Stadje, W. The collector's problem with group drawings. *Adv. Appl. Prob.* **22**, 866–874, 1990.