MORE GENERAL, FURTHER UNIFIED, YET MORE ACCESSIBLE SHARING PROBLEM RESULTS

AARON CHILDS, * McMaster University

Abstract

In this paper we present results of the sharing problem when the population is finite and sampling is without replacement. Our results generalize and further unify the results for sampling with replacement given by Sobel and Frankowski (1994). By using the Dirichlet probability generating function introduced by Sobel and Childs (2002), we unify the probability results with the expected waiting time results. Furthermore, computations for this more general situation using the Dirichlet probability generating function have the advantage that they do not require any special functions to compute, and their construction is quite intuitive.

Keywords: Sharing problem; probability generating function; Dirichlet C function; Dirichlet D function; multinomial; hypergeometric

2010 Mathematics Subject Classification: Primary 60C05
Secondary 60-03; 01-08

1. Introduction

The sharing problem, or the problem of points, can be described as follows. A total of \( k \) players have a contest in which they play a series of independent games where Player \( i \) has probability \( p_i \) of winning on a single game and

\[
\sum_{i=1}^{k} p_i = 1,
\]

so that each game produces a single winner. The games continue until one of the players has reached a total of \( n \) wins, at which time he wins the prize money. Now, an interruption requires the game to stop at a point where Player \( i \) needs \( r_i \) (\( 1 \leq r_i \leq n \)) more wins to be the eventual winner. The question is how should the prize money be divided? This question was posed to Pascal in 1654 who determined that the prize money should be shared among the players in such a way that each player receives an amount proportional to his probability of being the eventual winner if the contest were to continue until completion. However, the origins of the sharing problem date back even further. The problem first appeared in print in 1494 (see Pacioli (1494, p. 197)), and has even been found in Italian manuscripts dating as far back as 1380 (see Ore (1960)). The problem of points was the subject of the correspondence between Fermat and Pascal considered by many to be the birth date of probability theory (see Ross (1998)). For details of this correspondence as well as other historical developments, the reader may refer to Todhunter (1865), Edwards (1982), Hald (1990), or Táňács (1994).

Received 3 January 2006; revision received 15 January 2006.

* Postal address: Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario L8S 4K1, Canada. Email address: childsa@mcmaster.ca
On the 500th anniversary of the sharing problem Sobel and Frankowski (1994) used Dirichlet integrals to give results for the general case of $k$ players with unequal probabilities $p_i$ (as described above), thus unifying all of the partial results that were available up until that time. They also extended the problem in various ways and used Dirichlet integrals to provide solutions to these extensions. Their results are expressed in terms of Dirichlet $C$ and $D$ functions, but to the best of our knowledge there is no computer package (commercial or otherwise) that has a built-in routine able to compute these special functions. Thus, the price of Sobel and Frankowski’s unified solutions was some loss in accessibility.

In this paper we further extend and unify the results of Sobel and Frankowski (1994) by considering finite population analogues of the sharing problem, but in doing so we actually bring the solutions to the sharing problem back into the realm of wide accessibility. The construction of our general results uses the Dirichlet probability generating function (see Sobel and Childs (2002)), which requires only a basic knowledge of probability and counting. In addition, the computation of probabilities and waiting times associated with the sharing problem using our results can in every case be done by simply evaluating the single integral of a polynomial, which can be carried out by the reader with only a basic knowledge of MAPLE® (or any other software able to do such integrals). Furthermore, since numerical results for the usual (multinomial) sharing problem can be obtained to a high degree of accuracy by a numerical limiting process from the finite population without replacement (i.e. hypergeometric) results, this will in practice allow us to calculate answers for the multinomial sharing problem without the use of any special functions.

Specifically, we consider the case where the population of size $N$ has $k$ types of objects of interest (corresponding to $k$ different players), with $M_i$ objects of type $i$ ($i = 1, 2, \ldots, k$). We allow for the possibility of other types of objects in the population that are not of any consequence to the contest. Objects are selected one at a time, without replacement, and we wait until we have $n$ of any one of the $k$ types of objects of interest, at which time the corresponding player is declared the winner. Since $\sum_{i=1}^{k} M_i \leq N$, not every game necessarily produces a winner. In the same situation, if sampling is with replacement and $\sum_{i=1}^{k} M_i = N$ then we obtain the multinomial sharing problem with $p_i = M_i/N, \ i = 1, 2, \ldots, k$. We can also obtain results for the multinomial sharing problem from the hypergeometric results by letting $N$ and $M_i, \ i = 1, 2, \ldots, k$, approach infinity in such a way that $M_i/N = p_i, \ i = 1, 2, \ldots, k$, remains constant.

We begin in Section 2 by introducing the Dirichlet probability generating function (PGF). Our main unifying results are presented in Section 3 where we show how to use the Dirichlet PGF to calculate both probabilities and expected waiting times associated with the sharing problem. Probabilities for finishing orders are discussed in Section 4. We conclude in Section 5 by using our Dirichlet PGFs to obtain waiting time results for yet another generalization of the sharing problem, which is more naturally discussed in the context of the Banach matchbox problem.

2. The Dirichlet probability generating function

Consider associating with each of the $N$ objects in the population a uniform random number in the interval $(0, 1)$, and then ordering these random numbers from smallest to largest. Then sampling one at a time without replacement from the population is equivalent to selecting observations, in order, along this sequence. Suppose that we select objects one at a time, without replacement, until some specified event $E$ occurs. Let $u$ be the value of the random variable corresponding to the last object selected, i.e. the object that results in the event $E$
occurring for the first time. Given the value of $u$, the probability that the event $E$ occurs after $\alpha$ trials is given by

$$P(WT = \alpha) = C_{E,\alpha}u^{\alpha-1}(1 - u)^{N-\alpha},$$

where ‘WT’ denotes ‘waiting time’ and $C_{E,\alpha}$ is the number of partitions of the $N$ objects into distinct groups of size $\alpha - 1$, $1$, and $N - \alpha$, which result in an ordering of the objects in which the event $E$ occurs for the first time on the $\alpha$th trial.

Given the value of the random number $u$ corresponding to the last object selected, the PGF of the waiting time is then given by

$$\phi(t) = \sum_{\alpha=1}^{N} t^\alpha P(WT = \alpha) = \sum_{\alpha=1}^{N} C_{E,\alpha}t(u)^{\alpha-1}(1 - u)^{N-\alpha}.$$  

Therefore, we can obtain the PGF for the event $E$ (without ever calculating $C_{E,\alpha}$) simply by writing the following sum,

$$\phi(t) = \sum_{P} t(ut)^{j-1}(1 - u)^{N-j}, \quad (2.1)$$

where the sum is over all partitions of the uniform random variables that result in the event $E$ occurring (after any number of trials) and each term in the sum corresponds to a partition in which the event $E$ occurs after $j$ trials. The resulting generating function is called the Dirichlet PGF, and was first introduced by Sobel and Childs (2002). To be consistent with the notation used by Sobel and Childs (2002), we replace $u$ by $C$ and $1 - u$ by $D$ in the inert form of the Dirichlet PGF (2.1), and we write

$$\phi(t, C, D) = \sum_{P} t(Ct)^{j-1}D^{N-j}.$$

For any event $E$, the probability that the event $E$ occurs is then given by

$$P(E) = \int_{0}^{1} \phi(1, u, 1 - u) \, du.$$

If the event $E$ must eventually occur then the expected waiting time for $E$ to occur is given by

$$E(WT) = \int_{0}^{1} \phi'(1, u, 1 - u) \, du, \quad (2.2)$$

and the variance of the waiting time is

$$\text{var}(WT) = \int_{0}^{1} [\phi''(1, u, 1 - u) + \phi'(1, u, 1 - u)] \, du - \left(\int_{0}^{1} \phi'(1, u, 1 - u) \, du\right)^2. \quad (2.3)$$

**Problem 2.1.** Cards are drawn one at a time without replacement until all four aces, jacks, kings, or queens are obtained. Find the Dirichlet PGF.

**Solution.** There are 16 possibilities for the last card obtained, and all three of the remaining cards of that type must be already selected. This contributes a factor of $16t(Ct)^3$ to the Dirichlet PGF.
The other three types of cards must each have less than four at stopping time, which contributes a factor of
\[
\left[ \binom{4}{0} (Ct)^0 D^4 + \binom{4}{1} (Ct)^1 D^3 + \binom{4}{2} (Ct)^2 D^2 + \binom{4}{3} (Ct)^3 D \right]^3
\]
or, equivalently, \([(D + Ct)^4 - (Ct)^4]^3\) to the Dirichlet PGF.
There are no restrictions among the remaining 36 cards, so any of them could be chosen before or after the last card selected. To account for all possibilities simultaneously, we simply multiply by \((D + Ct)^{36}\). Therefore, the Dirichlet PGF is given by
\[
\phi(t, C, D) = 16t(Ct)^3[(D + Ct)^4 - (Ct)^4]^3(D + Ct)^{36}.
\]
The expected waiting time and variance are then calculated from (2.2) and (2.3) as follows:
\[
\begin{align*}
E(WT) &= \frac{108\,544}{3\,315} \approx 32.743\,288\,1, \\
\text{var}(WT) &= \frac{446\,719\,808}{76\,924\,575} \approx 58.105\,225\,9.
\end{align*}
\]

3. Main unifying results for the sharing problem

3.1. Winning probabilities

Let \(k\) be the number of players in the contest and let \(M_i\) be the number of elements in the population corresponding to Player \(i\), for \(i = 1, 2, \ldots, k\). The quota for Player \(i\) will be denoted by \(r_i\), and is the number of wins needed by Player \(i\) in order to win the contest. Furthermore, let
\[
M = (M_1, M_2, \ldots, M_k) \quad \text{and} \quad r = (r_1, r_2, \ldots, r_k).
\]
The total number of items in the population is denoted by \(N\), and is not necessarily equal to \(\sum_{i=1}^{k} M_i\).
To obtain the Dirichlet PGF for the event that Player \(i\) wins the contest, we note that the last object selected must be one of the \(M_i\) objects of type \(i\), and \(r_i - 1\) objects of type \(i\) must have already been selected at stopping time. This contributes a factor of
\[
M_i (M_i - 1) \binom{M_i}{r_i - 1} (Ct)^{r_i - 1} D^{M_i - r_i}
\]
to the Dirichlet PGF. The remaining \(k - 1\) players must each have less than their quota at stopping time, which contributes a factor of
\[
\prod_{j=1, j \neq i}^{k} \sum_{l_j=0}^{r_j-1} \binom{M_j}{l_j} (Ct)^{l_j} D^{M_j - l_j}
\]
to the Dirichlet PGF. Finally, we multiply by \((D + Ct)^{N - \sum_{i=1}^{k} M_i}\) to allow for the possibility of quota-free cells, i.e. to allow for the possibility of an additional type of object in the population.
that does not count towards the quota of any of the \( k \) players. This additional factor will not affect the probability calculations, but it will affect the waiting time problems considered below.

Therefore, the Dirichlet PGF for the event that Player \( i \) wins the contest is as follows:

\[
\phi_{W_i}(t, C, D) = \left[ M_i t \left( \frac{M_i - 1}{r_i - 1} \right) (C t)^{r_i - 1} D^{M_i - r_i} \prod_{j=1, j \neq i}^{k} \sum_{l_j=0}^{r_j-1} \left( \frac{M_j}{l_j} \right) (C t)^{l_j} D^{M_j - l_j} \right] \\
\times (D + C t)^{N - \sum_{i=1}^{k} M_i}.
\] (3.1)

More generally, the Dirichlet PGF for the event that Player \( i \) is the \( s \)th contest winner is given by

\[
\phi^{(i,s)}(t, C, D) = \left[ M_i t \left( \frac{M_i - 1}{r_i - 1} \right) (C t)^{r_i - 1} D^{M_i - r_i} \right] \\
\times \sum_{\beta=1}^{(k-1)!} \prod_{l=1}^{s-1} \left[ (D + C t)^{M_{j_l}} - \sum_{m_l=0}^{r_{j_l}-1} \left( \frac{M_{j_l}}{m_l} \right) (C t)^{m_l} D^{M_{j_l} - m_l} \right] \\
\times \prod_{l=s}^{k-1} \sum_{m_l=0}^{r_{j_l}-1} \left( \frac{M_{j_l}}{m_l} \right) (C t)^{m_l} D^{M_{j_l} - m_l} \} \right] \right] \\
\times (D + C t)^{N - \sum_{i=1}^{k} M_i}.
\] (3.2)

where the sum indexed by \( \beta \) is over all \( \binom{k-1}{s-1} \) distinct splittings \( \{j_1, \ldots, j_{s-1}\} \) and \( \{j_s, \ldots, j_{k-1}\} \) of \( \{1, \ldots, i-1, i+1, \ldots, k\} \) into groups of size \( s-1 \) and \( k-s \).

Using the Dirichlet PGF (3.2), the probability that Player \( i \) is the \( s \)th contest winner is given by

\[
W_{i,s} = \int_{0}^{1} \phi^{(i,s)}(1, u, 1-u) \, du, \quad i = 1, 2, \ldots, k.
\] (3.3)

<table>
<thead>
<tr>
<th>Table 1.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( W_1 )</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>( M )</td>
</tr>
<tr>
<td>10( M )</td>
</tr>
<tr>
<td>10^2( M )</td>
</tr>
<tr>
<td>10^3( M )</td>
</tr>
<tr>
<td>10^4( M )</td>
</tr>
<tr>
<td>10^5( M )</td>
</tr>
<tr>
<td>10^6( M )</td>
</tr>
<tr>
<td>( \infty )</td>
</tr>
</tbody>
</table>
Example 3.1. Suppose that \( k = 5 \), \( r = (1, 2, 3, 6, 3) \), and \( M = (10, 15, 12, 51, 21) \). Then the probabilities for each player to be the eventual winner of the contest are (using (3.3)) as follows:

\[
W_{1,1} = \frac{1045033010079}{21066318055692} \approx 0.4943927,
\]
\[
W_{2,1} = \frac{644905575629}{3089474507956} \approx 0.2142918,
\]
\[
W_{3,1} = \frac{354024863837}{10333159027846} \approx 0.0365105,
\]
\[
W_{4,1} = \frac{95641280420}{752368501989} \approx 0.1271203,
\]
\[
W_{5,1} = \frac{98247872692}{752368501989} \approx 0.1305848.
\]

As a check, we note that the above probabilities sum to 1.

If we increase \( M \) then in the limit we should obtain the results for sampling with replacement. Indeed, Table 1 shows that increasing \( M \) by a factor of \( 10^6 \) gives answers that agree to at least seven decimal places with the corresponding results for sampling with replacement (which were calculated by the author using the formulae in Sobel and Frankowski (1994)).

3.2. Waiting for the \( s \)th contest winner

In order to obtain the Dirichlet PGF for the case when the contest is carried out until the \( s \)th contest winner, the work is already done. Recall that to obtain the Dirichlet PGF we simply consider all of the possibilities for the last object selected. Since the last object selected can correspond to any one of the \( k \) players, and any \( s - 1 \) of the players must have less than their quota at stopping time while the remaining \( k - s \) players have at least their quota, we see that the Dirichlet PGF is simply the sum of the Dirichlet PGFs in (3.2), i.e.

\[
\phi(s)(t, C, D) = \sum_{i=1}^{k} \phi(i,s)(t, C, D).
\]  

(3.4)

The expected value and variance of the waiting time is then obtained using (2.2) and (2.3) with \( \phi = \phi(s) \).

Example 3.2. Continuing with Example 3.1, suppose that we wait until the third contest winner. Now that we are considering the calculation of a waiting time rather than a probability, we have to also specify the number of elements in the population. If there are no other elements in the population, i.e. if

\[
N = \sum_{i=1}^{5} M_i = 109,
\]

then the expected value and variance of the waiting time are given by \( \mu[3] = 13.6363774 \) and \( \sigma^2[3] = 7.9525743 \) respectively. If, on the other hand, there are \( N = 120 \) elements in the population, i.e. if there are 11 quota-free cells, then these numbers increase to \( \mu[3] = 15.0000151 \) and \( \sigma^2[3] = 10.9169464 \) respectively.

3.3. The expected total number of wins by a specified player

If the contest is continued until the \( s \)th winner is obtained, then the Dirichlet PGF for the number of wins by a specified player is simply obtained from the expression for \( \phi(s) \) in (3.4)
by changing \( t \) into a vector, i.e.

\[
\begin{align*}
\phi_{E_{i,s}}(t_1, t_2, C, D) &= \sum_{j=1}^{k} \left\{ M_j t_j \left( \frac{M_j - 1}{r_j - 1} \right) (Ct_j)^{r_j-1} D^{M_j-r_j} \right. \\
&\times \sum_{\beta=1}^{s-1} \left\{ \prod_{l=1}^{\beta-1} \left[ (D + Ct_l)^{M_l} - \sum_{m_l=0}^{r_l-1} \left( \frac{M_l}{m_l} \right) (Ct_l)^{m_l} D^{M_l-m_l} \right] \right. \\
&\left. \times \prod_{l=s}^{\beta} \left[ \sum_{m_l=0}^{ \sum_{l=1}^{s-1} M_l } \left( \frac{M_l}{m_l} \right) (Ct_l)^{m_l} D^{M_l-m_l} \right] \right\} \\
&\times (D + Ct_2)^{N-\sum_{i=1}^{s} M_i}.
\end{align*}
\]

where

\[
t_j = \begin{cases} 
  t_1 & \text{if } j = i, \\
  t_2 & \text{if } j \neq i.
\end{cases}
\]

We now have the following theorem.

**Theorem 3.1.** If the contest is continued until the \( s \)th winner is obtained then the expected number of wins by Player \( i \) is given by

\[
E_{i,s}(W) = \int_0^1 \frac{\partial}{\partial t_1} \phi_{E_{i,s}}(t_1, 1, u, 1-u) \bigg|_{t_1=1} du,
\]

where \( \phi_{E_{i,s}}(t_1, t_2, C, D) \) is given in (3.5). The variance of the number of wins \( \text{var}_{i,s}(W) \) can similarly be obtained by taking \( \phi(t, C, D) = \phi_{E_{i,s}}(t, 1, C, D) \) and using (2.3).

The proof of Theorem 3.1 can be obtained by starting with the Dirichlet PGF in (3.5) and carrying out the indicated operations. We omit the details.

**Example 3.3.** Continuing with Example 3.2, suppose that we wait until the third contest winner (with \( N = 109 \)). Then the expected number of wins by each player, along with the corresponding variances, are given in Table 2. As a check we note that the total of the expected number of wins for each individual player is exactly equal to the expected waiting time for the third contest winner with \( N = 109 \) calculated in Example 3.2.

### 4. Probabilities for finishing orders

In this section we consider the problem of finding the probabilities for each of the \( k! \) possible finishing orders, or vector rankings, of the \( k \) players when the contest is carried fully to
completion. We use the Dirichlet PGF to obtain explicit results for the cases \( k = 2 \) and \( 3 \), and then show how results for larger values of \( k \) can be obtained recursively.

Let \( x = (x_1, x_2, \ldots, x_k) \) be a permutation of \((1, 2, \ldots, k)\), and let \( P_{r,M}(x) \) denote the probability that the player with quota \( r_x \), who has \( M_x \) elements in the population comes in \( i \)th place, for \( i = 1, 2, \ldots, k \). So, for example, when \( k = 3 \), \( P_{(3,4,2),(10,8,5)}(2, 1, 3) \) denotes the probability that Player 2, who has a quota of \( r_2 = 4 \) and \( M_2 = 8 \) elements in the population, comes in first place, Player 1, with \((r_1, M_1) = (3, 10)\), comes in second place, and Player 3 comes in third place. Let \( \phi_{r,M,x}(t,C,D) \) be the corresponding Dirichlet PGF.

When \( k = 2 \), \( P_{r,M}(1, 2) \) is simply the probability that Player 1 wins the contest, and can therefore be obtained from the Dirichlet PGF given in (3.1). When \( k = 3 \), the Dirichlet PGF \( \phi_{r,M,(1,2,3)} \) is obtained by noting that when Player 2 reaches his quota \( r_2 \), Player 1 must have at least \( r_1 \) wins, while Player 3 must have less than \( r_3 \) wins. Therefore,

\[
\phi_{r,M,(1,2,3)}(t,C,D) = M_2 t \left( \frac{M_2 - 1}{r_2 - 1} \right) (Ct)^{r_2-1} D^{M_2-r_2} \\
\times \left[ (D+Ct)^{M_1} - \sum_{j=0}^{r_1-1} \binom{M_1}{j} (Ct)^j D^{M_1-j} \right] \\
\times \sum_{j=0}^{r_3-1} \binom{M_3}{j} (Ct)^j D^{M_3-j},
\]

from which we obtain

\[
Pr,M(1, 2, 3) = \int_0^1 \phi_{r,M,(1,2,3)}(1, u, 1-u) \, du. \tag{4.1}
\]

For general \( k \), in order to calculate \( P_{r,M}(1, 2, \ldots, k) \) we first calculate the probability that, when Player 2 reaches his quota, Player 1 has already reached his quota while the remaining \( k-2 \) players have a fixed number of wins, \( j_3, j_4, \ldots, j_k \) respectively, that are each less than their respective quotas. This guarantees that Player 1 comes in first and Player 2 comes in second. The Dirichlet PGF for this event is given by

\[
\phi_{r,M,(1,2,[j_3,\ldots,j_k])}(t,C,D) = M_2 t \left( \frac{M_2 - 1}{r_2 - 1} \right) (Ct)^{r_2-1} D^{M_2-r_2} \\
\times \left[ (D+Ct)^{M_1} - \sum_{j=0}^{r_1-1} \binom{M_1}{j} (Ct)^j D^{M_1-j} \right] \\
\times \prod_{j=3}^{k} \binom{M_j}{j} (Ct)^j D^{M_j-j}. \tag{4.2}
\]

Then, conditional on Players 3 to \( k \) having \( j_3, j_4, \ldots, j_k \) wins respectively, we calculate the probability that the ensuing game between Players 3 to \( k \), the finishing order is 3, 4, \ldots, \( k \). In this ensuing game, Player \( i \) needs \( r_i - j_i \) more wins and has \( M_i - j_i \) elements remaining in the population, for \( i = 3, 4, \ldots, k \). Therefore, we obtain

\[
Pr,M(1, 2, \ldots, k) = \sum_{j_3=0}^{r_3-1} \cdots \sum_{j_k=0}^{r_k-1} \int_0^1 \phi_{r,M,(1,2,[j_3,\ldots,j_k])}(1, u, 1-u) \, du \\
\times Pr,M(1, 2, \ldots, k-2). \tag{4.3}
\]
Table 3.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Probability</th>
<th>Rank</th>
<th>Probability</th>
<th>Rank</th>
<th>Probability</th>
<th>Rank</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>12345</td>
<td>0.0113759</td>
<td>12345</td>
<td>0.0065946</td>
<td>14235</td>
<td>0.0194595</td>
<td>15234</td>
<td>0.0064926</td>
</tr>
<tr>
<td>12354</td>
<td>0.0058576</td>
<td>13254</td>
<td>0.0034278</td>
<td>14253</td>
<td>0.0517652</td>
<td>15243</td>
<td>0.0364327</td>
</tr>
<tr>
<td>12435</td>
<td>0.0246583</td>
<td>13425</td>
<td>0.0079399</td>
<td>14325</td>
<td>0.0116412</td>
<td>15324</td>
<td>0.0042696</td>
</tr>
<tr>
<td>12453</td>
<td>0.0669299</td>
<td>13452</td>
<td>0.0085798</td>
<td>14352</td>
<td>0.0129099</td>
<td>15342</td>
<td>0.0088517</td>
</tr>
<tr>
<td>12534</td>
<td>0.0082665</td>
<td>13524</td>
<td>0.0031414</td>
<td>14523</td>
<td>0.0505001</td>
<td>15423</td>
<td>0.0501453</td>
</tr>
<tr>
<td>12543</td>
<td>0.0463716</td>
<td>13542</td>
<td>0.0063805</td>
<td>14532</td>
<td>0.0216373</td>
<td>15432</td>
<td>0.0207649</td>
</tr>
</tbody>
</table>

where

\[ r' = (r_3 - j_3, \ldots, r_k - j_k), \quad M' = (M_3 - j_3, \ldots, M_k - j_k), \]

and \( \phi_{r', M', (1, 2, \ldots, j_k)}(t, C, D) \) is given in (4.2).

Example 4.1. Continuing with Example 3.1, we use (4.3), which in turn makes use of (4.1), to calculate all of the vector rankings corresponding to Player 1 being the winner of the contest (see Table 3). As a check, we can sum the exact fractions (not given here) corresponding to the 24 probabilities given in Table 3; we obtained

\[ \frac{10415033010079}{27066318055692}, \]

which is equal to \( W_{1,1} \) from Example 3.1.

5. A further generalization

In the Banach matchbox problem, a mathematician has \( k \) different boxes each containing a different type of match, with \( r_i \) matches in each box. He selects box \( i \) with probability \( p_i \) and then takes a match from that box. This setting is equivalent to that of the multinomial sharing problem. If instead the matches are all together in a pile and matches are selected at random (without replacement) from that pile, then the problem becomes equivalent to the hypergeometric sharing problem with \( r_i = M_i, \quad i = 1, 2, \ldots, k. \)

Now, suppose that we consider a further generalization of the Banach matchbox problem in which an environmentally friendly mathematician puts the used matches back in the pile for recycling at a later time. Thus, sampling is with replacement, but a previously used match cannot be used again. Therefore, selection of previously used matches will count toward the waiting time until depletion of one of the match types, but will not bring that match type any closer to reaching depletion. Under this new sampling scheme, if we wait until depletion of any one of the types of matches then the expected waiting time is given by

\[ E(WT) = N \int_0^1 \int_0^y \phi^{(1)}(1, 1 - x, x) \, dx \, dy, \]

where \( \phi^{(1)}(t, C, D) \) is the Dirichlet PGF for the case where the contest is carried out until the first contest winner in the hypergeometric sharing problem given in (3.4). Furthermore,

\[ E[W(T + 1)] = 2N^2 \int_0^1 \int_0^z \int_0^y \phi^{(1)}(1, 1 - x, x) \, dx \, dy \, dz, \]

from which the variance of the waiting time can easily be obtained.
This remarkable connection between the Dirichlet PGF for the hypergeometric waiting time problem and the corresponding expected waiting time when sampling is with replacement, but uses no repetitions, is completely general and is proved by Sobel and Childs (2002). Furthermore, it is not required that \( r_i = M_i \).

Acknowledgements

The author would like to thank the Natural Sciences and Engineering Research Council of Canada for funding this research, and an anonymous referee whose comments led to an improvement in the presentation of this paper.

References