Adv. Appl. Prob. **22**, 758–760 (1990)
Printed in N. Ireland

A GENERALIZATION OF A WAITING TIME PROBLEM

B. S. EL-DESOUKY* AND
S. A. HUSSEN,* *Faculty of Science, Aswan*

Abstract

An urn contains m types of balls of unequal numbers. Let n_i be the number of balls of type $i, i = 1, 2, \cdots, m$. Balls are drawn with replacement until first duplication. In the case of finite memory of order k, the distribution of $Y_{m,k}$, the number of drawings required, is discussed. Special cases are obtained.

1. Introduction

An urn contains m distinguishable balls which are sampled one at a time with replacement. The sampling is continued until the first duplication. Let X_m be the number of drawings required.

This problem, which was solved by McCabe [4], is a special case of the problem of waiting time until first duplication with finite memory of order k (Arnold [1]), in which sampling is continued until a ball is drawn to duplicate one of the k preceding balls drawn. Let X_{mk} be the number of draws required when there are m balls in the urn and there is finite memory of order k. The distribution of X_{mk} is found by Arnold [1].

In this paper we consider two cases.

Case 1. We generalize McCabe [4] as follows. Suppose that we have an urn containing m types of balls with n_i the number of balls of type $i, i = 1, 2, \cdots, m$. Assume that balls are sampled one at a time with replacement and the sampling is continued until the first duplication (i.e., until a ball of the same type has been drawn twice), and Y_m is the number of drawings performed.

Case 2. Case 1 can be considered as a special case of the problem of waiting time until first duplication with finite memory of order k.

In this case sampling is continued until a ball is drawn to duplicate one of the k immediately preceding balls (one of each type) drawn. For example, when $k = 1$, sampling stops only when two successive drawings yield a ball of the same type.

Let $Y_{m,k}$ be the number of draws required when there are m types of balls of unequal numbers and there is a finite memory of order k, which can be considered as a generalization of Arnold's problem [1]. It is clear that Y_m is identical with $Y_{m,m}$ and if $k > m$, then $Y_{m,k}$ also has the same distribution as $Y_{m,m}$.

2. The distribution of $Y_{m,k}$

If $k = m$ it is clear that the random variable $Y_{m,m}$ cannot be smaller than 2 or larger than $m + 1$. When $k < m$, the random variable $Y_{m,k}$ may assume any value greater than or equal to 2. Hence we discuss the two cases. First we consider the distribution of $Y_{m,k}$ (Case 2) which gives us the distribution of $Y_{m,m}$ (Case 1) as a special case.

If $k < m$, for any integer j satisfying $1 \leqq \leqq k + 1$,

$$P(Y_{m,k} > j) = P \text{ (the first } j \text{ balls are all distinct, one of each type)}$$

$$P(Y_{m,k} > j) = \sum_{\pi} p_{\pi(1)} p_{\pi(2)} \cdots p_{\pi(j)}, \tag{2.1}$$

Received 18 September 1989; revision received 25 May 1990.
* Postal address for both authors: Department of Mathematics, Faculty of Science, Aswan, Egypt.

where the summation is over all permutations $(\pi_{(1)}, \pi_{(2)}, \cdots, \pi_{(j)})$ of the integers $1, 2, \cdots, m$ taken j at a time, and $p_i = n_i/N$, $N = \sum_{i=1}^m n_i$, is the probability that a ball of type i is drawn.

In the special case $k = m$, this yields the following result.

Theorem 2.1. The generating function of the probabilities $P(Y_{m,m} > j)$, $j = 1, 2, \cdots, m$ is given by

$$\sum_{j=0}^{m} P(Y_{m,m} > j)\frac{t^j}{N^j} = \prod_{i=1}^{m}(1 + p_i t). \tag{2.2}$$

Proof. Note that

$$\sum_{\pi} p_{\pi_{(1)}} p_{\pi_{(2)}} \cdots p_{\pi_{(j)}} = j! \sum_{c(j)} \left[\prod_{k=1}^{j} p_{i_k}\right],$$

where the second summation is over all unordered subsets $(i_1, i_2, \cdots, i_j)$ of size j of the integers $1, 2, \cdots, m$. But the coefficient of t^j on the right-hand side of (2.2) is clearly

$$\sum_{c(j)} \left[\prod_{i=1}^{j} p_{i_k}\right],$$

hence the theorem is proved.

From (2.2)

$$\sum_{j=0}^{m} P(Y_{m,m} > j)\frac{t^j}{j!} = \left(1 + n_1\frac{t}{N}\right)\left(1 + n_2\frac{t}{N}\right)\cdots\left(1 + n_m\frac{t}{N}\right), \qquad N = \sum_{i=1}^{m} n_i$$

$$= \sum_{i=0}^{m} \mathbb{S}_n(m, m-i)\left(\frac{t}{N}\right)^i,$$

hence

$$P(Y_{m,m} > j) = \frac{j!}{N^j}\mathbb{S}_n(m, m-j) = \frac{(-1)^j j!}{N^j} s_n(m, m-j), \tag{2.3}$$

where $\mathbb{S}_n(m, k)$ is the unsigned generalized Stirling number of the first kind and $\mathbb{S}_n(m, k) = (-1)^{m-k} s_n(m, k)$ and $s_n(m, k)$ is the generalized Stirling number of the first kind associated with the real numbers $n_1, n_2, \cdots, n_m$ (see [2] and [3]).

If $n_i = i$, $i = 1, 2, \cdots, m$, then

$$P(Y_{m,m} > j) = \frac{(-1)^j j!}{N^j} s(m+1, m+1-j), \qquad N = m(m+1)/2, \tag{2.4}$$

where $s(m, k)$ is the Stirling number of the first kind.

If $n_i = n$, $i = 1, 2, \cdots, m$, i.e. there is an equal number of balls from each type, then

$$\sum_{j=0}^{m} P(Y_{m,m} > j)\frac{t^j}{j!} = \left(1 + n\frac{t}{N}\right)^m,$$

hence

$$P(Y_{m,m} > j) = \frac{(m)_j}{m^j},$$

in agreement with Arnold's result.

Acknowledgement

The first author thanks Professor B. C. Arnold for his advice and help.

References

[1] Arnold, B. C. (1972) The waiting time until first duplication. *J. Appl. Prob.* **9,** 841–846.

[2] Comtet, L. (1972) Nombres de Stirling généraux et fonctions symetriques. *C. R. Acad. Sci. Paris* A **275,** 747–750.

[3] Comtet, L. (1974) *Advanced Combinatorics*. D. Reidel, Dordrecht.

[4] McCabe, B. (1979) Problem E. 2263. *Amer. Math. Monthly* **77,** 1008.