On the coupon-collector's problem with several parallel collections

Marco Ferrante* 1 and Alessia Tagliavini $^{\dagger 2}$

¹Department of Mathematics, University of Padua, Italy ²Department of Information Engineering, University of Padua, Italy

September 15, 2016

Abstract

In this note we evaluate the expectation and variance of the waiting time to complete m parallel collections of coupons, in the case of coupons which arrives independently, one by one and with equal probabilities.

AMS Classification: 60C05

1 Introduction

The coupon-collector's problem is a classical problem in combinatorial probability. The description of the basic problem is easy: consider one person that collects coupons and assume that there is a finite number, say N, of different types of coupons. These items arrive one by one in sequence, with the type of the successive items being independent random variables that are each equal to k with probability p_k .

In the coupons-collector's problem, one is usually interested in answering the following questions: which is the probability to complete the collection (or a given subset of the collection) after the arrival of exactly n coupons $(n \geq N)$? which is the expectation and the variance of the number of coupons that we need to complete the collection (or to complete a given subset of the collection)? which is the expectation and the variance of the number of coupons that we need to complete a set of collections? how these

^{*}marco.ferrante@unipd.it

[†]alessia.tagliavini@dei.unipd.it

probabilities and expectations change if we assume that the coupons arrive with unequal probabilities or in groups of constant size?

The first results, due to De Moivre, Laplace and Euler (see [7] for a comprehensive introduction on this topic), deal with the case of constant probabilities $p_k \equiv \frac{1}{N}$, while the first results on the unequal case have to be ascribed to Von Schelling (see [8]). Many other studies have been carry out on this classical problem ever since (see e.g. [6], [3], [1] and [2]).

In this note we will consider the waiting time to complete a set of m collections, all independent to each other. We will assume that any collection is made of a finite number of different coupons, at any unit of time arrive m new coupons, one for each collection, and that the probability to purchase any type at any time is uniform. We will be able to derive the expectation and variance of the waiting time to complete all the m collections.

2 Single collection with equal probabilities

Let us start by considering a single collection. Assume that this collection consists of N different coupons, which are equally likely, with the probability to purchase any type at any time equal to $\frac{1}{N}$. In this section we will evaluate the expectation and the variance of the random number of coupons that one needs to purchase in order to complete the collection. We will follow two approaches, the first one present in most of the textbooks in Probability, while the second one, based on a Markov Chain approach, will allow us to extend the computation to the case of parallel collections.

2.1 The Geometric Distribution approach

Let X denote the (random) number of coupons that we need to purchase in order to complete our collection. We can write $X = X_1 + X_2 + \ldots + X_N$, where for any $i = 1, 2, \ldots, N$, X_i denotes the additional number of coupons that we need to purchase to pass from i-1 to i different types of coupons in our collection. Trivially $X_1 = 1$ and, since we are considering the case of a uniform distribution, it follows that when i distinct types of coupons have been collected, a new coupon purchased will be of a distinct type with probability equal to $\frac{N-i}{N}$. By the independence assumption, we get that the random variable X_i , for $i \in \{2, \ldots, N\}$, is independent from the other variables and has a geometric law with parameter $\frac{N-i+1}{N}$. The expected number of coupons that we have to buy to complete the collection will be therefore

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \dots + \mathbb{E}[X_N] = N \sum_{i=1}^{N} \frac{1}{i}$$

while its variance

$$Var[X] = \sum_{i=0}^{N-1} \frac{Ni}{(N-i)^2} = N^2 \sum_{i=1}^{N} \frac{1}{i^2} - N \sum_{i=1}^{N} \frac{1}{i}$$
 (1)

2.2 The Markov Chain approach

Even if the previous result is very simple and the formula completely clear, it will not help us in order to deal with the problem of parallel collections. We will introduce then a different approach, that will provide an alternative to the previous computation and that we will be able to extend in the following section, where the situation becomes more complicate.

Assuming as before that one coupon arrives at any unit of time, it is possible to solve the previous problem by using a Markov Chains approach. Indeed, let Y_n be the number of different types of coupons collected after n units of time and assume again that the probability of finding a coupon of any type at any time is $p = \frac{1}{N}$. $\{Y_n, n \in \mathbb{N}\}$ is then a Markov Chain on the state space $S = \{0, 1, \ldots, N\}$ and it is immediate to see that its transition matrix is given by

$$P = \begin{pmatrix} 0 & 1 & 0 & \dots & \dots & 0 \\ 0 & \frac{1}{N} & \frac{N-1}{N} & 0 & \dots & \dots & 0 \\ 0 & 0 & \frac{2}{N} & \frac{N-2}{N} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 0 & \frac{N-1}{N} & \frac{1}{N} \\ 0 & \dots & \dots & \dots & 0 & 0 & 1 \end{pmatrix}$$

Note that $\{N\}$ is the unique closed class and the state N is therefore absorbing, while the other states are transient. If we denote by Q the sub matrix of P relative to the transient states (in this case the first N lines and rows) we can define the fundamental matrix of $\{Y_n, n \in \mathbb{N}\}$ as

$$F = (Id - Q)^{-1}$$

Note that QF = FQ = F - Id (see [4] for more details). If we define the random variable

$$D^N = \inf\{n \ge 0 : Y_n = N\}$$

its conditional expectation $\mathbb{E}_0[D^N] = \mathbb{E}[D^N|Y_0 = 0]$ will be equal to the expected number of coupons needed to complete the collection. If we define the vector $k^N = (k_0^N, \dots, k_N^N) = (\mathbb{E}_0[D^N], \dots, \mathbb{E}_N[D^N])$, a classical result (see [5]) states that k^N is the minimal non negative solution of the linear system:

$$\begin{cases} k_N^N = 0 \\ k_i^N = 1 + \sum_{j \neq N} p_{ij} k_j^N, \quad i \neq N \end{cases}$$

Therefore

$$(k_0^N, \dots, k_{N-1}^N) = F \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

It is immediate to see that in this case

$$F = \begin{pmatrix} 1 & \frac{N}{N-1} & \frac{N}{N-2} & \frac{N}{N-3} & \dots & \dots & N \\ 0 & \frac{N}{N-1} & \frac{N}{N-2} & \frac{N}{N-3} & \dots & \dots & N \\ 0 & 0 & \frac{N}{N-2} & \frac{N}{N-3} & \dots & \dots & N \\ 0 & 0 & 0 & \frac{N}{N-3} & \dots & \dots & N \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 0 & \frac{N}{N-1} & N \\ 0 & \dots & \dots & \dots & 0 & 0 & N \end{pmatrix}$$

Then, for j = 0, ..., N - 1,

$$k_j^N = N \sum_{i=1}^{N-j} \frac{1}{i}$$

and the expected waiting time to collect all the different coupons is given by

$$k_0^N = N \sum_{i=1}^N \frac{1}{i}$$

The computation of the conditional variance is similar (see again [4]). Denoting by

$$v_i^N = Var_i[D^N] = \mathbb{E}[(D^N)^2|Y_0 = i] - (\mathbb{E}[(D^N)^2|Y_0 = i])^2$$

we have

$$v^{N} = (2F - Id) \begin{pmatrix} k_{0}^{N} \\ \vdots \\ k_{N-1}^{N} \end{pmatrix} - ((k_{0}^{N})^{2}, \dots, (k_{N-1}^{N})^{2})$$

A simple computation gives that the variance of the waiting time to collect all the different coupons is given by

$$v_0^N = N \sum_{i=1}^N \frac{1}{i} - \left(N \sum_{i=1}^N \frac{1}{i} \right)^2 + \sum_{k=1}^{N-1} \left[\frac{2N^2}{N-k} \sum_{i=1}^{N-k} \frac{1}{i} \right]$$
 (2)

It is possible to see that (1) and (2) are indeed the same.

3 Parallel collections with equal probabilities

Let us now consider the case that m different collections are available. Assume that the i-th collection is made of N_i different coupons and that the probability to purchase any type at any time is equal to $\frac{1}{N_i}$. Moreover, let us assume that we purchase simultaneously one coupon of every collection. In this section we will derive the expectation and the variance of the waiting time to complete all the m different collections.

With the notation of the previous section, let X^i denote be the random number of coupons needed to complete the *i*-th collection. The random number of coupons needed to complete all the m collections will be therefore $\max(X^1,\ldots,X^m)$. It looks not simple to determine the law of this random variable, or even just to provide a direct computation of its expectation. On the converse, adapt to this situation the Markov Chain's approach is quite simple.

Let $Y_n = (Y_n^1, \dots, Y_n^m)$ be the number of different types of coupons of any collection, collected after n units of time. $\{Y_n, n \in \mathbb{N}\}$ is again a Markov Chain on the state space

$$S = \{(0, \dots, 0)\} \cup \{(i_1, \dots, i_m) : i_j \in \{1, \dots, N_i\}, j \in \{1, \dots, m\}\}$$

We have $|S| = 1 + \prod_{j=1}^{m} N_j$. Defined $(\alpha_1, \ldots, \alpha_m) \in \{0, 1\}^m$ and assuming that $(i_1, \ldots, i_m) \in S$ and $(i_1 + \alpha_1, \ldots, i_m + \alpha_m) \in S$, we have that the transition matrix is defined as

$$\mathbb{P}[Y_{t+1} = (i_1 + \alpha_1, \dots, i_m + \alpha_m) | Y_{t+1} = (i_1, \dots, i_m)] =$$

$$= \prod_{j=1}^m \left(1 - \frac{i_j}{N_j}\right)^{\alpha_j} \left(\frac{i_j}{N_j}\right)^{1 - \alpha_j}$$

(Note that $\mathbb{P}[Y_{t+1} = (1, \dots, 1)|Y_t = (0, \dots, 0)] = 1$). The transition matrix has again a unique absorbing state $\{(N_1, \dots, N_m)\}$. If we denote by Q the sub matrix of P relative to the transient states we can define the fundamental matrix F of $\{Y_n, n \in \mathbb{N}\}$ as before and compute the conditional expectations

$$k^{\{(N_1,\dots,N_m)\}} = \left(\mathbb{E}_{(0,\dots,0)}[D^{\{(N_1,\dots,N_m)\}}],\dots,\mathbb{E}_{(N_1,\dots,N_m-1)}[D^{\{(N_1,\dots,N_m)\}}]\right)$$

as

$$k^{\{(N_1,\dots,N_m)\}} = F \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

Example 1 Consider three collections, each with six different coupon to be collected. With the previous notation we get m=3, $N_1=N_2=N_3=6$ and the matrix $F \in [0,1]^{215\times 215}$, even if at most eight entries of every row

are not zeros. Using Matlab[®] we are able to compute F in this case and the expectation of the waiting time to complete all the three collections is equal to 20.01.

In order to compute the variance of the waiting time to complete all the parallel collections, we shall operate as in the previous section. Denoting, for $i \in S \setminus \{(N_1, \ldots, N_m)\}$,

$$\begin{aligned} v_i^N &= Var_i[D^{\{(N_1,\dots,N_m)\}}] \\ &= \mathbb{E}[(D^{\{(N_1,\dots,N_m)\}})^2|Y_0 = i] - (\mathbb{E}[(D^{\{(N_1,\dots,N_m)\}})^2|Y_0 = i])^2 \end{aligned}$$

the conditional variances, we have

$$v^{N} = (2F - Id) \begin{pmatrix} k_{(0,\dots,0)}^{N} \\ \vdots \\ k_{(N_{1},\dots,N_{m}-1)}^{N} \end{pmatrix} - ((k_{(0,\dots,0)}^{N})^{2}, \dots, (k_{(N_{1},\dots,N_{m}-1)}^{N})^{2})$$

Example 2 (Continue) Considering again three collections, each with six different coupon to be collected, the variance of the waiting time to complete all the three collections is equal to 44.8975.

References

- [1] Adler, Ilan and Oren, Shmuel and Ross, Sheldon M.: *The coupon-collector's problem revisited* J. Appl. Probab. **40** (2013), 513-518.
- [2] FERRANTE, MARCO AND SALTALAMACCHIA, MONICA: The coupon collector's problem MATerials MATemàtics **2014** (2014), 35.
- [3] Holst, Lars: On birthday, collectors', occupancy and other classical urn problems Internat. Statist. Rev. **54** (1986), 15-27.
- [4] Kemeny, John G. and Snell, J. Laurie: Finite Markov chains Springer-Verlag, New York-Heidelberg (1976).
- [5] NORRIS, J. R.: Markov chains Cambridge University Press (1998).
- [6] ROSÉN, BENGT: On the coupon collector's waiting time Ann. Math. Statist. 41 (1970),1952-1969.
- [7] STADJE, WOLFGANG: The collector's problem with group drawings Adv. in Appl. Probab. 22 (1990), 866-882.
- [8] VON SCHELLING, HERMANN: Coupon collecting for unequal probabilities Amer. Math. Monthly **61** (1954), 306-311.