

Article

Coupon Collector Problem with Reset Button

Jelena Jocković ^{*,†}  and Bojana Todić [†] 

Faculty of Mathematics, University of Belgrade, 11000 Belgrade, Serbia; bojana.todic@matf.bg.ac.rs

* Correspondence: jelena.jockovic@matf.bg.ac.rs

† These authors contributed equally to this work.

Abstract: We consider the following generalization of the classical coupon collector problem. We assume that, in addition to the initial collection of standard coupons, there is one more coupon that acts as a reset button, removing all coupons from the part of the collection that has already been drawn. For the case where standard coupons have unequal probabilities of being drawn, we obtain the distribution of the waiting time until the end of the collection process. For the case where standard coupons have equal probabilities, we derive a simple formula for the expected waiting time in terms of the beta function, and discuss the asymptotic properties of this expected waiting time, when the number of standard coupons tends toward infinity.

Keywords: coupon collector problem; reset coupon; expected waiting time; Markov chain; beta function

MSC: 60C05; 60J10

1. Introduction

In the classical coupon collector problem (CCP) the goal is to analyze the waiting time (in days) for a collector, who buys a coupon each day, in order to complete a full collection of n distinct coupons from the set $\mathbb{N}_n = \{1, 2, \dots, n\}$, assuming that each coupon is drawn with probability $\frac{1}{n}$. Similarly, the coupon collector problem with unequal probabilities (CCPU) refers to the situation where the coupon $j \in \mathbb{N}_n$ is drawn with probability p_j and $\sum_{j=1}^n p_j = 1$.

The coupon collector problem can be formulated in different ways. In its simplest form, it appears in the well known work [1], but it can also be formulated and treated as a special kind of urn model (see, for example, [2,3]) or in the context of formal languages [4]. On the other hand, several results related to the waiting time problems have been obtained using Markov chain techniques (see, for example, [5,6]). The coupon collector problem and its generalizations also led to various types of asymptotic results (see, for example, [7–9]).

We consider the following generalization of the coupon collector problem, which, to our knowledge, has not been considered before. We assume that there is a special coupon (we call it the reset coupon) that does not belong to \mathbb{N}_n and acts as a reset button, in the sense that the set of coupons drawn up to time (day) t becomes empty when the reset coupon is drawn on day $t + 1$. After that, the collection can start again from the beginning (or not). Therefore, we work with an augmented set of available coupons,

$$\mathbb{N}_n^{(\otimes)} = \{1, 2, \dots, n, \otimes\}, \quad (1)$$

where \otimes denotes the reset coupon.

We call this version of the problem coupon collector with reset button and refer to it as CCPRB in the rest of the text.

We assume, in sampling with replacement, that the probability of obtaining the reset coupon is p_R , $0 \leq p_R < 1$, and that the probabilities of obtaining standard coupons (from



Citation: Jocković, J.; Todić, B. Coupon Collector Problem with Reset Button. *Mathematics* **2024**, *12*, 239. <https://doi.org/10.3390/math12020239>

Academic Editor: Steve Drekić

Received: 11 December 2023

Revised: 4 January 2024

Accepted: 9 January 2024

Published: 11 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

the set \mathbb{N}_n) sum up to $1 - p_R$. Obviously, the problem reduces to CCP or CCPU when $p_R = 0$.

The CCPRB problem we consider belongs to the group of generalizations of CCP that are based on the idea of introducing additional coupons with special purposes into the coupon set. Other generalizations of CCP of this type are analyzed in [5,10–13].

The variant of the coupon collector problem, where each coupon can have several purposes (so called, goals), is considered in [13]. In this case, the experiment ends when the sum of the numbers of goals reaches a certain limit.

Another generalization of the classical coupon collector problem is proposed in [11], where the appearance of an additional coupon (so called, bonus coupon) leads to obtaining one more coupon.

In [12], the author considers the case where the additional coupon (so called, penalty coupon) interferes with collecting standard coupons, in the sense that the collection process ends when the absolute difference between the number of collected standard coupons and the number of collected penalty coupons is equal to the total number of standard coupons.

In this work, we refer to results related to the coupon collector problem with a null coupon, as considered in [5,10]. This is the situation where the probabilities of the standard coupons sum up to $\sum_{j=1}^n p_j = 1 - p_N < 1$, or, equivalently, there is a null coupon, that can be drawn with probability p_N , but does not belong to any collection. This variant of the problem reduces to CCPU when $p_N = 0$.

It is well known that the coupon collector problem has various applications in engineering (see, for example, [14]). In particular, the CCPU has recently been used in biology, to model parasitism (explained in [15]), and in telecommunications, to model the transmission of information in computer networks [16], and to analyze Internet security problems (analyzed in [5,10,17]).

The structure of this paper is as follows. In Section 2, we obtain the general properties of the waiting time for a full collection of standard coupons, in the case of CCPRB with unequal probabilities. In Section 3, we obtain the expected waiting time for a full collection in the case of equal probabilities, and derive its relation to the beta function. In Section 4, we provide some numerical examples. In Section 5, we discuss the asymptotic properties of the expected waiting time for a full collection in the case of equal probabilities, for different values of p_R and when n tends to infinity, and give some specific examples. The conclusions are given in Section 6.

2. Distribution of the Waiting Time for a Full Collection in General Case

Here, we derive the distribution of the waiting time until a full collection of standard coupons is sampled in the case of unequal probabilities (where each coupon $j \in \mathbb{N}_n$ is drawn with probability p_j , such that $\sum_{j=1}^n p_j = 1 - p_R$).

Let $W_n^{(\otimes)}$ denote the waiting time until a full collection of standard coupons in CCPRB are collected, and W_n denote the corresponding waiting time in the coupon collector problem with unequal probabilities and no additional coupons. We will also use the notation $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ and $P_J = \sum_{j \in J} p_j$.

The distribution of the waiting time W_n is a well known result,

$$P\{W_n > k\} = \sum_{i=0}^{n-1} (-1)^{n-1-i} \sum_{\substack{J \subset \mathbb{N}_n, \\ |J|=i}} P_J^k, \quad k \geq 0, \tag{2}$$

(see, for example, [5], Theorem 1, p. 409).

The corresponding result related to the waiting time $W_n^{(\otimes)}$ is obtained in the next theorem. In the rest of the text, we will use the abbreviation $D_{i,k}$ for

$$(k_1, k_2, \dots, k_{i+1}) \in \mathbb{N}_0^{i+1}, k_1 + k_2 + \dots + k_{i+1} = k - i.$$

Theorem 1. For every $k \geq 0$, for the waiting time $W_n^{(\otimes)}$, the following relations hold:

1.

$$P\{W_n^{(\otimes)} > k\} = \sum_{i=0}^k p_R^i (1 - p_R)^{k-i} \sum_{D_{i,k}} \prod_{j=1}^{i+1} P\{W_n > k_j\}, \tag{3}$$

2.

$$P\{W_n^{(\otimes)} = k\} = \sum_{i=0}^k p_R^i (1 - p_R)^{k-i} \sum_{D_{i,k}} \prod_{j=1}^i P\{W_n > k_j\} (P\{W_n > k_{i+1} - 1\} - P\{W_n > k_{i+1}\}). \tag{4}$$

Proof.

1. Each sequence draws up to time k can be presented in the form

$$B_1 R B_2 R \dots B_i R B_{i+1}, \tag{5}$$

where R represents a single reset coupon and $B_j, j \in \{1, 2, \dots, i + 1\}$ represent blocks of standard coupons, such that the length of the block B_j is $k_j \geq 0, \sum_{j=1}^{i+1} k_j = k - i$, and none of the blocks B_j consist of a full collection of standard coupons.

We define the events as follows:

- $S_{m,t}, t \geq 1, 0 \leq m \leq t$: m reset coupons appeared up to time t ;
- $Q_k, k \geq 0$: block of the length k does not consist of a full collection of standard coupons.

Therefore,

$$P\{W_n^{(\otimes)} > k\} = \sum_{i=0}^k \sum_{D_{i,k}} P\left(S_{i,k} \cap \bigcap_{j=1}^{i+1} Q_{k_j}\right). \tag{6}$$

On the other hand, we have

$$P\left(S_{i,k} \cap \bigcap_{j=1}^{i+1} Q_{k_j}\right) = P(S_{i,k}) P\left(\bigcap_{j=1}^{i+1} Q_{k_j} | S_{i,k}\right) = p_R^i (1 - p_R)^{k-i} \prod_{j=1}^{i+1} P\{W_n > k_j\}, \tag{7}$$

since all the blocks B_j consist of incomplete collections of standard coupons, their lengths sum up to $k - i$, and the appearance of the blocks of any combination of lengths are mutually independent events. This completes the proof of the statement.

2. This is a simple modification of the first part of the theorem. Each realization of the experiment has the form

$$B_1 R B_2 R \dots B_i R B_{i+1}, \tag{8}$$

where none of the blocks $B_j, 1 \leq j \leq i$ consists of a full collection of standard coupons, and the block B_{i+1} consist of the full collection of standard coupons. Using the fact that

$$P\{W_n = k_{i+1}\} = P\{W_n > k_{i+1} - 1\} - P\{W_n > k_{i+1}\},$$

we complete the proof of the theorem.

□

Remark 1. The sequence of draws in the CCPRB can be seen as a renewal process, as the coupon collection starts over after each reset. More precisely, the events of the type $B_j R$, defined in (5), can be regarded as recurring events, in the sense of Definition 1, p. 308 in [1].

Example 1. If $n = 2, p_R = \frac{1}{7}$ and $p_1 = p_2 = \frac{3}{7}$, the probability that the full collection of standard coupons has not been drawn by the time (day) $k = 3$ is equal to

$$P\{W_2^{(\otimes)} > 3\} = \sum_{i=0}^3 \left(\frac{1}{7}\right)^i \left(\frac{6}{7}\right)^{3-i} \sum_{D_{i,3}} \prod_{j=1}^{i+1} P\{W_2 > k_j\} = \sum_{i=0}^3 \left(\frac{1}{7}\right)^i \left(\frac{6}{7}\right)^{3-i} F_{3,i}, \tag{9}$$

where

$$\begin{aligned} F_{3,0} &= P\{W_2 > 3\} = 0.25, \\ F_{3,1} &= 2P\{W_2 > 0\}P\{W_2 > 2\} + P\{W_2 > 1\}^2 = 2, \\ F_{3,2} &= 3P\{W_2 > 1\}P\{W_2 > 0\}^2 = 3, \\ F_{3,3} &= P\{W_2 > 0\}^4 = 1. \end{aligned}$$

Therefore,

$$P\{W_2^{(\otimes)} > 3\} = 0.4227.$$

Remark 2. The expected waiting time for a full collection can be obtained as

$$E(W_n^{(\otimes)}) = \sum_{k \geq 0} P\{W_n^{(\otimes)} > k\}, \tag{10}$$

where $P\{W_n^{(\otimes)} > k\}$ is given in (3).

Remark 3. It is well known that, for large n , the computation of probabilities associated with the coupon collector problem with unequal probabilities, such as (3), becomes computationally intensive, and requires some sort of approximation or bounds. This is even more obvious in the case of CCPRB. However, the upper and lower bounds for the probability (3) can be obtained directly by applying the corresponding upper and lower bounds for the probability (2) in (3). For a detailed discussion on this topic, and a comprehensive list of upper and lower bounds on the probability (2), see [17].

Observing that

$$P\{W_n > k - i\} \leq \sum_{D_{i,k}} \prod_{j=1}^{i+1} P\{W_n > k_j\}, \tag{11}$$

we obtain an additional, simple lower bound for the probability $P\{W_n^{(\otimes)} > k\}$:

$$\sum_{i=0}^k p_R^i (1 - p_R)^{k-i} P\{W_n > k - i\} \leq P\{W_n^{(\otimes)} > k\}. \tag{12}$$

Remark 4. Let $W_{n,c}^{(\otimes)}$ denote the waiting time until c , $1 \leq c \leq n$, out of n coupons in CCPRB are collected. For this waiting time, results analogous to Theorem 1 and its consequences can be derived using the same technique.

3. Expected Waiting Time for a Full Collection in the Case of Equal Probabilities

The expected waiting time for a full collection, or a subcollection, can be obtained from (10). However, if we assume that all the standard coupons have an equal probability $p = \frac{1-p_R}{n}$ of being drawn, the expected waiting time for a full collection has a simpler form, which is conveniently derived using the Markov chain technique.

Let X_t be the number of different types of standard coupons sampled after t units of time (days). We can notice that $\{X_t, t \in \mathbb{N}\}$ is a Markov chain on the state space:

$$S = \{0, 1, \dots, n - 1, n\}. \tag{13}$$

Depending on how we define the end of the collection process, we can distinguish between two characteristic cases.

3.1. Case 1: The Collector Gives up Collecting after the First Reset

In the first case, the collector starts with a certain number of coupons, and buys coupons until he completes his collection, or the first reset happens.

The transition probability matrix is

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & \dots & \dots & n-1 & n \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ \vdots \\ n-2 \\ n-1 \\ n \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & \dots & 0 & 0 \\ p_R & p & (n-1)p & 0 & \dots & \dots & 0 & 0 \\ \vdots & & & & & & & \vdots \\ p_R & 0 & 0 & 0 & \dots & (n-2)p & 2p & 0 \\ p_R & 0 & 0 & 0 & \dots & 0 & (n-1)p & p \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix} \end{matrix}, \quad (14)$$

and

$$v_k = E(\min\{t \geq 0 : X_t = 0, \text{ or } X_t = n\} | X_0 = k), \quad k \in S \quad (15)$$

is the waiting time until absorption, starting from the state k .

3.2. Case 2: The Collector Keeps Collecting after the Reset

In this case, the collector buys coupons until he completes his collection, regardless of how many resets occur in the meantime.

The transition probability matrix is

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & \dots & \dots & n-1 & n \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ \vdots \\ n-2 \\ n-1 \\ n \end{matrix} & \begin{pmatrix} p_R & 1-p_R & 0 & 0 & \dots & \dots & 0 & 0 \\ p_R & p & (n-1)p & 0 & \dots & \dots & 0 & 0 \\ \vdots & & & & & & & \vdots \\ p_R & 0 & 0 & 0 & \dots & (n-2)p & 2p & 0 \\ p_R & 0 & 0 & 0 & \dots & 0 & (n-1)p & p \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix} \end{matrix}, \quad (16)$$

and

$$u_k = E(\min\{t \geq 0 : X_t = n\} | X_0 = k), \quad k \in S \quad (17)$$

is the waiting time until absorption, starting from the state k .

The expected waiting times v_k and $u_k, k \in S$ are obtained in the next theorem.

Theorem 2.

1. For the expected waiting time v_k in Case 1, Section 3.1, the following relations hold:

$$\begin{aligned}
 v_k &= \frac{1}{p} \frac{\Gamma(n-k+1)}{\Gamma(\frac{p_R}{p} + n - k + 1)} \sum_{i=0}^{n-k-1} \frac{\Gamma(\frac{p_R}{p} + n - k - i)}{\Gamma(n - k - i + 1)}, \quad k = 1, \dots, n-1, \\
 v_0 &= 0, \\
 v_n &= 0,
 \end{aligned} \quad (18)$$

where $\Gamma(\cdot)$ denotes the gamma function.

2. For the expected waiting time u_k in Case 2, Section 3.2, the following relations hold:

$$\begin{aligned}
 u_k &= \frac{v_k}{(1-p_R)(1-p_R v_1)}, \quad k = 1, 2, \dots, n-1, \\
 u_0 &= \frac{1 + (1-p_R)v_1}{(1-p_R)(1-p_R v_1)}, \\
 u_n &= 0.
 \end{aligned} \quad (19)$$

Proof.

1. Applying the first step analysis to the Markov chain with the transition probability matrix (14), we obtain that

$$v_k = 1 + kpv_k + (n - k)pv_{k+1}, \quad v_0 = v_n = 0. \tag{20}$$

The recurrence (20) is solved in [18] (problem 6.2., p. 165):

$$v_k = \frac{1}{p_R + (n - k)p} + \frac{h_{k,k+1}}{p_R + (n - k - 1)p} + \dots + \frac{h_{k,n-1}}{p_R + p}, \tag{21}$$

where

$$h_{k,j} = \frac{(n - k)p}{p_R + (n - k)p} \cdot \frac{(n - k - 1)p}{p_R + (n - k - 1)p} \dots \frac{(n - j + 1)p}{p_R + (n - j + 1)p}, \quad k < j. \tag{22}$$

Next, we have

$$h_{k,j} = \frac{n - k}{\frac{p_R}{p} + n - k} \cdot \frac{n - k - 1}{\frac{p_R}{p} + n - k - 1} \dots \frac{n - j + 1}{\frac{p_R}{p} + n - j + 1} = \frac{\Gamma(n - k + 1)\Gamma(\frac{p_R}{p} + n - j + 1)}{\Gamma(n - j + 1)\Gamma(\frac{p_R}{p} + n - k + 1)}. \tag{23}$$

Now, we can rewrite (21) as

$$\begin{aligned} v_k &= \frac{1}{p} \sum_{i=0}^{n-k-1} \frac{\Gamma(n - k + 1)\Gamma(\frac{p_R}{p} + n - k - i + 1)}{(\frac{p_R}{p} + n - k - i)\Gamma(n - k - i + 1)\Gamma(\frac{p_R}{p} + n - k + 1)} \\ &= \frac{1}{p} \frac{\Gamma(n - k + 1)}{\Gamma(\frac{p_R}{p} + n - k + 1)} \sum_{i=0}^{n-k-1} \frac{\Gamma(\frac{p_R}{p} + n - k - i)}{\Gamma(n - k - i + 1)}, \end{aligned} \tag{24}$$

which completes the proof of the statement.

2. Applying the first step analysis to the Markov chain with the transition probability matrix (16), we conclude that

$$u_k = 1 + p_R u_0 + kp u_k + (n - k)pu_{k+1}, \quad k = 1, 2, \dots, n - 1, u_n = 0, \tag{25}$$

and

$$u_0 = 1 + p_R u_0 + (1 - p_R)u_1. \tag{26}$$

Applying the substitution

$$v_k = \frac{u_k}{1 + p_R u_0},$$

we obtain the Equation (20), and the solution is given by (21). From (26), it follows that

$$u_0 = \frac{1 + (1 - p_R)v_1}{(1 - p_R)(1 - p_R v_1)}, \tag{27}$$

and

$$u_k = \frac{v_k}{(1 - p_R)(1 - p_R v_1)}, \quad k = 1, 2, \dots, n - 1, \tag{28}$$

which completes the proof of the theorem.

□

Example 2. The case $p_R = 0$ is the CCP, and the waiting time u_0 becomes the expected waiting time in CCP. From (24) with $k = 1$ we have

$$v_1 = n \sum_{j=1}^{n-1} \frac{1}{j} = nH_{n-1}^{(1)}(0), \tag{29}$$

where

$$H_n^{(a)}(b) = \sum_{k=1}^n \frac{1}{(k+b)^a}. \tag{30}$$

Using $p_R = 0$ in (29) and (27) leads to the well known result, related to harmonic numbers:

$$u_0 = n \sum_{j=1}^n \frac{1}{j} = nH_n^{(1)}(0). \tag{31}$$

The expected waiting times v_1 and u_0 are the most general, in the sense that the collector has to wait for the almost full or full collection. Next, we will provide simplified expressions for the waiting times v_1 and u_0 for the case $p_R > 0$. For that purpose, we need the next lemma.

Lemma 1. Let $(a)_k$ denote the falling factorial:

$$(a)_k = a(a-1) \dots (a-k+1). \tag{32}$$

1. For any $m \in \mathbb{N}$, the following equality holds:

$$\sum_{k=1}^m \frac{(a)_k}{(b)_k} = \frac{a}{b-a+1} \left(1 - \frac{(a-1)_m}{(b)_m} \right). \tag{33}$$

2. For any $m \in \mathbb{N}$, $a, b \in \mathbb{R}^+$, $b \neq a+1$, the following equality holds:

$$\sum_{k=0}^m \frac{\Gamma(a+k)}{\Gamma(b+k)} = \frac{\Gamma(a+m)}{\Gamma(b+m)} \frac{(a+m)_{m+1} - (b+m-1)_{m+1}}{(a-b+1)(a+m-1)_m}. \tag{34}$$

Proof.

1. We use a mathematical induction on m . For $m = 1$, we easily confirm that the equality (33) is valid. Next, assuming that (33) holds for $m - 1$, we check that (33) holds for m . We have

$$\begin{aligned} \sum_{k=1}^m \frac{(a)_k}{(b)_k} &= \frac{a}{b-a+1} \left(1 - \frac{(a-1)_{m-1}}{(b)_{m-1}} \right) + \frac{(a)_m}{(b)_m} \\ &= \frac{a}{b-a+1} \left(1 - \frac{(a-1)_{m-1}}{(b)_{m-1}} + (b-a+1) \frac{(a-1)_{m-1}}{(b)_m} \right) \\ &= \frac{a}{b-a+1} \left(1 - (a-m) \frac{(a-1)_{m-1}}{(b)_m} \right) \\ &= \frac{a}{b-a+1} \left(1 - \frac{(a-1)_m}{(b)_m} \right), \end{aligned} \tag{35}$$

which completes the proof of the statement.

2. Using (33), we obtain

$$\begin{aligned} \sum_{k=0}^m \frac{\Gamma(a+k)}{\Gamma(b+k)} &= \sum_{k=0}^m \frac{\Gamma(a+m-k)}{\Gamma(b+m-k)} \\ &= \frac{\Gamma(a+m)}{\Gamma(b+m)} \left(1 + \sum_{k=1}^m \frac{(b+m-1)_k}{(a+m-1)_k} \right) \\ &= \frac{\Gamma(a+m)}{\Gamma(b+m)} \left(1 + \frac{b+m-1}{a-b+1} \left(1 - \frac{(b+m-2)_m}{(a+m-1)_m} \right) \right) \\ &= \frac{\Gamma(a+m)}{\Gamma(b+m)} \left(\frac{a+m}{a-b+1} - \frac{b+m-1}{a-b+1} \frac{(b+m-2)_m}{(a+m-1)_m} \right) \\ &= \frac{\Gamma(a+m)}{\Gamma(b+m)} \frac{(a+m)_{m+1} - (b+m-1)_{m+1}}{(a-b+1)(a+m-1)_m}, \end{aligned} \tag{36}$$

which completes the proof of the lemma.

□

Theorem 3.

1. If $p_R > 0$, for the expected waiting time v_1 , the following equality holds:

$$v_1 = \frac{1}{p_R} \left(1 - \frac{np_R}{1 - p_R} B \left(n, \frac{np_R}{1 - p_R} \right) \right), \tag{37}$$

where $B(\cdot)$ denotes the beta function.

2. If $p_R > 0$, for the expected waiting time u_0 , the following equality holds:

$$u_0 = \frac{1}{p_R} \left(\frac{1}{np_R B \left(n, \frac{np_R}{1 - p_R} \right)} - 1 \right). \tag{38}$$

Proof.

1. From (24) it follows that

$$v_1 = \frac{1}{np} \frac{\Gamma(n+1)}{\Gamma(\frac{p_R}{p} + n)} \sum_{j=1}^{n-1} \frac{\Gamma(\frac{p_R}{p} + n - j)}{\Gamma(n - j + 1)} = \frac{1}{np} \frac{\Gamma(n+1)}{\Gamma(\frac{p_R}{p} + n)} \sum_{j=0}^{n-2} \frac{\Gamma(\frac{p_R}{p} + j + 1)}{\Gamma(j + 2)}. \tag{39}$$

If $p_R > 0$, using Lemma 1, we obtain

$$\begin{aligned} v_1 &= \frac{1}{np} \frac{\Gamma(n+1)}{\Gamma(\frac{p_R}{p} + n)} \frac{\Gamma(\frac{p_R}{p} + n - 1)}{\Gamma(n)} \frac{(\frac{p_R}{p} + n - 1)_{n-1} - (n-1)_{n-1}}{\frac{p_R}{p} (\frac{p_R}{p} + n - 2)_{n-2}} \\ &= \frac{1}{np} \frac{\Gamma(n+1)}{\Gamma(\frac{p_R}{p} + n)} \frac{\Gamma(\frac{p_R}{p} + 1)}{(n-1)!} \frac{(\frac{p_R}{p} + n - 1)_{n-1} - (n-1)!}{\frac{p_R}{p}} \\ &= \frac{1}{np} \frac{\Gamma(n+1)}{\Gamma(\frac{p_R}{p} + n)} \left(\frac{\Gamma(\frac{p_R}{p} + n)}{\frac{p_R}{p} \Gamma(n)} - \frac{\Gamma(\frac{p_R}{p} + 1)}{\frac{p_R}{p}} \right) \\ &= \frac{1}{p_R} \left(1 - \frac{p_R}{p} \frac{\Gamma(n)\Gamma(\frac{p_R}{p})}{\Gamma(\frac{p_R}{p} + n)} \right) \\ &= \frac{1}{p_R} \left(1 - \frac{p_R}{p} B \left(n, \frac{p_R}{p} \right) \right). \end{aligned} \tag{40}$$

Using the equality $p = \frac{1-p_R}{n}$, we obtain the required statement.

2. The equality (27) can be written as follows:

$$u_0 = \frac{1}{p_R} \left(\frac{1}{(1 - p_R)(1 - p_R v_1)} - 1 \right). \tag{41}$$

Using (40) in (41), we obtain

$$u_0 = \frac{1}{p_R} \left(\frac{1}{(1 - p_R) \frac{p_R}{p} B \left(n, \frac{p_R}{p} \right)} - 1 \right) = \frac{1}{p_R} \left(\frac{1}{np_R B \left(n, \frac{np_R}{1 - p_R} \right)} - 1 \right), \tag{42}$$

which completes the proof of the theorem.

□

4. Numerical Examples

In this section, we provide numerical examples for the CCPRB with equal probabilities, as analyzed in Section 3. We assume that the set of available coupons is \mathbb{N}_{10} . We consider different values of the probability p_R and calculate the expected waiting time $u_0 = u_0^{(10)}$ for this case using formula (27). The results are shown in Table 1. Statistical Software R, version 2023.03.0+386 was used for all calculations.

Table 1. Expected waiting time $u_0^{(10)}$.

p_R	$\frac{3}{4}$	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{10}$	$\frac{1}{20}$	$\frac{1}{50}$	$\frac{1}{100}$	0
$u_0^{(10)}$	1.13×10^9	4.51×10^7	3.69×10^5	1.77×10^3	1.92×10^2	1.27×10^2	5.87×10^1	3.83×10^1	3.34×10^1	2.93×10^1

Next, we show how the expected waiting time u_0 depends on the probability p_R for different values of n .

Note that the behavior of the expected waiting time u_0 , depicted in Figure 1, is consistent with the intuition we have about CCPRB: u_0 increases as n increases (as having more standard coupons to collect extends the waiting time), and u_0 increases as p_R increases (as resets remove the coupons already collected, and therefore extend the waiting time).

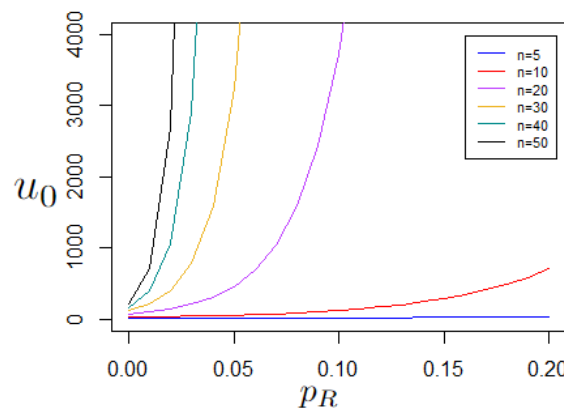


Figure 1. Expected waiting time u_0 in terms of p_R for different values of n .

In some cases considered, we can also notice some kind of exponential growth, which we discuss in more detail in the next section.

5. Asymptotic Properties of the Waiting Times v_1 and u_0

Here, we analyze the properties of the expected waiting times until the end of the collection process, as the number of standard coupons n tends to infinity, for different values of the probability p_R . We can distinguish between the case when p_R is fixed, and the case when p_R depends on n .

For fixed $p_R \in (0, 1)$, we can apply the Stirling approximation of the term $B\left(n, \frac{np_R}{1-p_R}\right)$ in Theorem 3, and obtain the asymptotic estimate for u_0 , as $n \rightarrow \infty$, formulated in the next proposition.

Proposition 1. For any $p_R \in (0, 1)$, the following asymptotic relation holds as $n \rightarrow \infty$:

$$u_0 \sim \frac{1}{p_R} \left(\frac{1}{\sqrt{2p_R\pi n}(1-p_R)^n p_R^{\frac{np_R}{1-p_R}}} - 1 \right). \tag{43}$$

Remark 5. The expression (43) can be alternatively written in the following form:

$$u_0 \sim \frac{1}{p_R} \left(\frac{\alpha^{-n}}{\sqrt{2p_R\pi n}} - 1 \right), \text{ where } \alpha = (1 - p_R)p_R^{\frac{p_R}{1-p_R}}. \tag{44}$$

The relations (43) and (44) also hold when p_R depends on n , in the case when the ratio $\frac{np_R}{1-p_R} = \frac{p_R}{p}$ tends to infinity as $n \rightarrow \infty$.

For the case when the ratio $\frac{np_R}{1-p_R}$ is bounded (which means that the collection process is not interrupted too often by resets), we have the following asymptotic result.

Proposition 2. For $\frac{np_R}{1-p_R} = O(1)$, the following asymptotic relation holds, $n \rightarrow \infty$:

$$u_0 \sim \frac{1}{p_R} \left(\frac{n^{\frac{np_R}{1-p_R}}}{(1 - p_R)\Gamma(\frac{np_R}{1-p_R} + 1)} - 1 \right). \tag{45}$$

Proof. The statement follows from Theorem 3 and the relation

$$\lim_{x \rightarrow \infty} \frac{\Gamma(x + a)}{x^{a-b}\Gamma(x + b)} = 1 \tag{46}$$

(see, for example, [19], page 257). \square

Next, we analyze some particular cases of the problem.

Example 3. The case $p_R = p = \frac{1}{n+1}$ corresponds to the situation where the reset coupon has the same probability of being drawn as any other coupon. Using Theorem 3, we obtain the expressions

$$v_1 = (n + 1)(1 - B(n, 1)) = \frac{n^2 - 1}{n} \sim n, \text{ as } n \rightarrow \infty, \tag{47}$$

and

$$u_0 = (n + 1) \left(\frac{1}{\frac{n}{n+1}B(n, 1)} - 1 \right) = n(n + 1) \sim n^2, \text{ as } n \rightarrow \infty. \tag{48}$$

This case is “exactly solvable”, in the sense that, for a given value u_0 , we can simply obtain $n = n_{u_0}$ such that the expected waiting time is less or equal to u_0 . Precisely, from the inequality

$$n^2 + n \leq u_0$$

and the fact that $n_{u_0} \geq 0$, we obtain that

$$n_{u_0} \in \left[0, \frac{-1 + \sqrt{1 + 4u_0}}{2} \right].$$

Example 4. The case $p_R = \frac{1}{2}$ corresponds to the situation where the ratio $\frac{np_R}{1-p_R}$ is equal to n . Using Theorem 3, we obtain the expressions

$$v_1 = 2(1 - B(n, n)) \sim 2 \left(1 - \frac{\sqrt{\pi n}}{2^{2n-1}} \right), \text{ as } n \rightarrow \infty, \tag{49}$$

and

$$u_0 = 2 \left(\frac{2}{nB(n, n)} - 1 \right) \sim 2 \left(\frac{2^{2n}}{\sqrt{\pi n^3}} - 1 \right), \text{ as } n \rightarrow \infty. \tag{50}$$

6. Conclusions

In this paper, we considered a generalization of the coupon collector problem where there is an additional coupon that resets the collection process and is drawn with probability $0 \leq p_R < 1$.

We distinguished between cases where standard coupons are drawn with arbitrary probabilities that sum up to $1 - p_R$ (which generalizes CCPU), and the case where standard coupons are drawn with the same probability $\frac{1-p_R}{n}$ (which generalizes CCP). The object of interest was the waiting time until the end of the coupon collection procedure.

For the case of equal probabilities, we specified the distribution of this waiting time in a very general way, in terms of the corresponding distribution in CCPU.

For the case of equal probabilities, applying the first step analysis for the appropriately constructed Markov chains, we have obtained the expression for the expected waiting time in both possible cases (when the collection process continues after the reset or not), and derived its simple form in terms of the beta function. We determined the asymptotic behavior (when the size of the collection tends to infinity) of the expected waiting time, considering possible values of the probability p_R (fixed or depending on n).

We have also discussed some characteristic examples of this problem.

The possible applications (or interpretations) of this work relate to the detection of distributed denial-of-service (DDoS) attacks, which are explained, for example, in [5]. The authors conclude that, since the solution to these attacks is to continuously monitor Internet traffic, the occurrence of standard coupons in the coupon collector problem corresponds to tracking c (out of n) recent high traffic flows, where the portion size c is determined by server capacity, and the probability of null coupon, p_N , corresponds to the case of flows with small probabilities that sum up to p_N . In this context, the appearance of the reset coupon at the moment t_0 can be interpreted as a system crash, which leads to losing information about high traffic flows that have been monitored up to that moment. Having information about the frequency of such crashes is important for maintaining proper protection.

Author Contributions: Conceptualization, J.J.; methodology, J.J. and B.T.; software, B.T.; validation, B.T.; formal analysis, J.J. and B.T.; writing—original draft preparation, J.J. and B.T.; supervision, J.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia: Grant No. 451-03-47/2023-01/200104.

Data Availability Statement: The data are contained within the article.

Acknowledgments: The authors would like to thank the reviewers for carefully checking the manuscript and for providing extremely useful comments and suggestions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Feller, W. *An Introduction to Probability Theory and Its Applications*, 3rd ed.; John Wiley and Sons Inc.: New York, NY, USA, 1968; Volume 1.
2. Samuel-Cahn, E. Asymptotic distributions for occupancy and waiting time problems with positive probability of falling through the cells. *Ann. Probab.* **1974**, *2*, 515–521. [[CrossRef](#)]
3. Mahmoud, H. *Polya urn Models*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2009.
4. Flajolet, P.; Gardy, D.; Thimonier, L. Birthday paradox, coupon collectors, caching algorithms and self-organizing search. *Discret. Appl. Math.* **1992**, *39*, 207–229. [[CrossRef](#)]
5. Anceaume, E.; Busnel, Y.; Sericola, B. New results on a generalized coupon collector problem using Markov chains. *J. Appl. Probab.* **2015**, *52*, 405–418. [[CrossRef](#)]
6. Kim, B.; Kim, J. Sooner waiting time problems in a sequence of multi-state trials with random rewards. *Statist. Probab. Lett.* **2019**, *153*, 171–179. [[CrossRef](#)]
7. Baum, L.E.; Billingsley, P. Asymptotic distributions for the coupon collector's problem. *Ann. Math. Statist.* **1965**, *36*, 1835–1839. [[CrossRef](#)]
8. Mladenović, P. Limit distributions for the problem of collecting pairs. *Bernoulli* **2008**, *14*, 419–439. [[CrossRef](#)]

9. Neal, P. The generalized coupon collector problem. *J. Appl. Probab.* **2008**, *45*, 621–629. [[CrossRef](#)]
10. Anceaume, E.; Busnel, Y.; Schulte-Geers, E.; Sericola, B. Optimization results for a generalized coupon collector problem. *J. Appl. Probab.* **2016**, *53*, 622–629. [[CrossRef](#)]
11. Nakata, T.; Kubo, I. A coupon collector's problem with bonuses. In Proceedings of the 4th Colloquium on Mathematics and Computer Science, Nancy, France, 18–22 September 2006; pp. 215–224.
12. Todić, B. Coupon collector problem with penalty coupon. *Mat. Vesn.* **2023**, *in press*. [[CrossRef](#)]
13. Wild, M.; Janson, S.; Wagner, S.; Laurie, D. Coupon collecting and transversals of hypergraphs. *Discrete Math. Theor.* **2013**, *15*, 259–270. [[CrossRef](#)]
14. Boneh, A.; Hofri, M. The coupon-collector problem revisited—A survey of engineering problems and computational methods. *Comm. Statist. Stoch. Models* **1997**, *13*, 39–66. [[CrossRef](#)]
15. Zoroa, N.; Lesigne, E.; Fernández-Sáez, M.J.; Zoroa, P.; Casas, J. The coupon collector urn model with unequal probabilities in ecology and evolution. *J. R. Soc. Interface* **2016**, *14*, 20160643. [[CrossRef](#)] [[PubMed](#)]
16. Comets, F.; Delarue, F.; Schott, R. Information transmission under random emission constraints. *Combin. Probab. Comput.* **2014**, *23*, 973–1009.
17. Shioda, S. Some upper and lower bounds on the coupon collector problem. *J. Comput. Appl. Math.* **2007**, *200*, 154–167. [[CrossRef](#)]
18. Taylor, H.M.; Karlin, S. *An Introduction to Stochastic Modeling*, 3rd ed.; Academic Press: San Diego, CA, USA, 1998.
19. Abramowitz, M.; Stegun, I.A. (Eds.) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*; Dover Publications Inc.: New York, NY, USA, 1965.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.