

# Coupon collector’s problem and its extensions in extreme value framework

JELENA JOCKOVIĆ<sup>\*,†</sup> AND PAVLE MLADENOVIĆ<sup>‡</sup>

Let  $\mathbb{N}_n = \{1, 2, \dots, n\}$ . We sample with replacement from the set  $\mathbb{N}_n$ , assuming that each element has the same probability of being drawn. We determine the limiting distributions and some related properties of the waiting time  $M_n$  until all triples  $j_1j_2j_3, j \in \mathbb{N}_n$  are sampled.

KEYWORDS AND PHRASES: Extreme values, Limit theorems, Waiting time, Excesses over the threshold, Recurrence relation.

## 1. INTRODUCTION

Many combinatorial problems in probability have been considered by using different tools and approaches. Usually, the goal of the investigations in this area is to obtain exact and asymptotic results concerning random variables that appear in such problems. References that should be pointed out as pioneer works on asymptotic probability distributions in combinatorial problems are the paper of Goncharov (1944) and the series of papers written by Erdős and his coauthors.

Coupon collector’s problem is a classical combinatorial problem that can be formulated as follows: We sample with replacement from the set  $\mathbb{N}_n = \{1, 2, \dots, n\}$ , assuming that each element has the same probability of being drawn and we are interested in the waiting time until all elements of  $\mathbb{N}_n$  or some other patterns are sampled. Erdős and Rényi (1961) obtained the following asymptotic result concerning the waiting time  $M_n$  until all elements of  $\mathbb{N}_n$  are sampled:

$$\lim_{n \rightarrow \infty} P\{M_n \leq n(x + \log n)\} = \exp(-e^{-x}).$$

Many variations of the coupon collector’s problem have been considered, and different approaches and techniques used, see, for example, Baum and Bilingsley (1965), Holst (1986), Mladenović (2008) and references therein.

In this paper we consider the waiting time  $M_n$  until all triples  $j_1j_2j_3, j \in \mathbb{N}$ , are sampled. The remaining part of the paper is organized as follows. In Section 2 we give some preliminaries, notation and auxiliary results on the exact

distributions of random variables related to the problem considered. Main results are formulated in Section 3. We provide the limiting distribution of the random variable  $M_n$  and also the limiting distribution of the excess of  $M_n$  over high thresholds. The rate of convergence is also determined. Proofs are given in Section 4. The conclusion is formulated in Section 5.

## 2. PRELIMINARIES, NOTATION AND AUXILIARY RESULTS

Let  $Z_1, Z_2, Z_3, \dots$  be a sequence of independent random variables with the uniform distribution over the set  $\mathbb{N}_n = \{1, 2, \dots, n\}$  and

$$(1) \quad X_{nj} = \min\{k : Z_{k-2} = Z_{k-1} = Z_k = j\},$$

where  $j \in \mathbb{N}_n$  is a fixed number.

Let

$$(2) \quad \tilde{Y}_{nj} = \min\{k : Z_{k-2} = Z_{k-1} = Z_k = a \text{ for some } a \in A \subset \mathbb{N}_n, |A| = j\}.$$

We introduce random variables  $Y_{nn}, Y_{n,n-1}, \dots, Y_{n1}$  as follows:  $Y_{nn}$  is the waiting time until the first triple  $j_1j_1j_1$ , where  $j_1 \in \mathbb{N}_n$ , occurs;  $Y_{n,n-1}$  is the waiting time for the second triple  $j_2j_2j_2, j_2 \in \mathbb{N}_n \setminus \{j_1\}$ , after the occurrence of the first triple  $j_1j_1j_1$ ; etc. Therefore,  $Y_{nj} \stackrel{d}{=} \tilde{Y}_{nj}$  for any  $j \in \mathbb{N}_n$ , where  $X \stackrel{d}{=} Y$  means that random variables  $X$  and  $Y$  have the same distribution.

It is obvious that for any  $j \in \mathbb{N}_n$  and any positive integer  $k$  the following equality holds:

$$(3) \quad P\{X_{n1} > k, X_{n2} > k, \dots, X_{nj} > k\} = P\{Y_{nj} > k\}.$$

Let  $M_n$  be the waiting time until all triples  $j_1j_2j_3, j \in \mathbb{N}_n$  are sampled. Therefore,

$$(4) \quad M_n = \max\{X_{n1}, X_{n2}, \dots, X_{nn}\}.$$

Using inclusion-exclusion principle we obtain:

$$(5) \quad P\{M_n \leq k\} = \sum_{m=0}^n (-1)^m \binom{n}{m} P\{Y_{nm} > k\}.$$

Probability distributions and some related properties of random variables  $X_{nj}$  and  $Y_{nj}, j \in \mathbb{N}_n$ , are determined in the following theorems.

\*Corresponding author.

<sup>†</sup>The work is supported by the Ministry of Science and Education of the Republic of Serbia, Grant Nos. 174012 and TR34007.

<sup>‡</sup>The work is supported by the Ministry of Science and Education of the Republic of Serbia, Grant No. 174012.

**Theorem 2.1.** (a) The distribution of the random variable  $X_{nj}$  for  $k \geq 3$  is given by

$$(6) \quad P\{X_{nj} = k\} = \sum_{\substack{a \geq 0, s \geq a \\ a+2s \leq k-3}} \binom{k-a-s-3}{s} \binom{s}{a} \times \\ \times \frac{1}{n^{s+a+3}} \left(1 - \frac{1}{n}\right)^{k-s-a-3}.$$

(b) If  $u_n = n^3(x + \log n)$ , then the following equality holds:

$$(7) \quad \lim_{n \rightarrow \infty} n(1 - F_n(u_n)) = \lim_{n \rightarrow \infty} nP\{X_{nj} > u_n\} = e^{-x}.$$

**Theorem 2.2.** (a) The distribution of the random variable  $Y_{nj}$  is given by

$$(8) \quad P\{Y_{nj} = k\} = \frac{a_{k-1}}{n^k},$$

where the sequence  $a_k$  satisfies the recurrence relation

$$(9) \quad a_{k+4} + (1-n^2)a_{k+2} + (-n^2 + 2nj - 2j + 1)a_k - (n-j)^2 a_{k-2} = 0,$$

for any  $k \geq 2$ , with initial values:

$$(10) \quad \begin{aligned} a_0 = a_1 = 0, \quad a_2 = j, \quad a_3 = (n-1)j, \\ a_4 = n(n-1)j, \quad a_5 = (n^3 - n^2 + n)j. \end{aligned}$$

(b) For any fixed  $j$ , the following relation holds, as  $n \rightarrow \infty$ :

$$(11) \quad P\{Y_{nj} > n^3(x + \log n)\} = \frac{e^{-jx}}{n^j} (1 + o(1)).$$

### 3. MAIN RESULTS

Limiting distribution of the maximum  $M_n$  as  $n \rightarrow \infty$  is determined in the following theorem.

**Theorem 3.1.** Let  $M_n$  be the waiting time until all triples  $jjj$ ,  $j \in \mathbb{N}_n$ , are sampled. For any real number  $x$ , the following equality holds:

$$(12) \quad \lim_{n \rightarrow \infty} P\{M_n \leq n^3(x + \log n)\} = \exp(-e^{-x}).$$

The following theorem gives the limiting distribution of the excess of  $M_n$  over high thresholds, and determines the rate of convergence to the limiting distribution for the corresponding normalizing constants and the threshold.

**Theorem 3.2.** Let  $M_n$  be the waiting time until all triples  $jjj$ ,  $j \in \mathbb{N}_n$ , are sampled. Let  $(c_n)_{n \geq 1}$  be a sequence of real numbers such that  $c_n \rightarrow +\infty$  as  $n \rightarrow \infty$ .

(a) For any  $x > 0$  the following equality holds:

$$(13) \quad \lim_{n \rightarrow \infty} P\{M_n \leq n^3(x + c_n + \log n) | M_n > n^3(c_n + \log n)\} = 1 - e^{-x}.$$

(b) For any  $x > 0$  the following equality holds, as  $n \rightarrow \infty$ :

$$(14) \quad P\{M_n \leq n^3(x + c_n + \log n) | M_n > n^3(c_n + \log n)\} \\ - (1 - e^{-x}) \sim e^{-x}(e^{-x} - 1) \frac{1}{2e^{c_n}}.$$

*Remark 3.3.* If  $c_n = c$  for all  $n \in \mathbb{N}$  and some constant  $c \in \mathbb{R}$ , then the limit on the left hand side of (13) does not belong to the family of generalized Pareto distributions and depends on the threshold. Therefore, the threshold given in the equality (13) is the lowest one such that it is possible to obtain a limit distribution in the sense of Theorem Balkema – de Haan (1974).

### 4. PROOFS

**Proof of the Theorem 2.1:**

(a) Let  $X_{nj}$  be defined as in (1) and let us define the events  $A_k$  and  $A_{k,a,b}$  as follows:

- $A_k$ : no three adjacent of the random variables  $Z_1, Z_2, \dots, Z_k$  take the value  $j$ ;
- $A_{k,a,b}$ : among the random variables  $Z_1, Z_2, \dots, Z_k$ , there are exactly  $a$  pairs  $Z_{s-1} = Z_s = j$ ,  $Z_{s-2} \neq j$ ,  $Z_{s+1} \neq j$ , exactly  $b$  singles  $Z_s = j$ ,  $Z_{s-1} \neq j$ ,  $Z_{s+1} \neq j$ , and no three adjacent of the random variables  $Z_1, Z_2, \dots, Z_k$  take the value  $j$ .

In that case the following equalities hold:

$$(15) \quad P(A_{k,a,b}) = \binom{k-2a-b+1}{a+b} \binom{a+b}{a} \frac{1}{n^{2a+b}} \left(1 - \frac{1}{n}\right)^{k-2a-b},$$

$$(16) \quad P(A_k) = \sum_{\substack{a \geq 0, b \geq 0 \\ a+b \leq k-2a-b+1}} P(A_{k,a,b}) = \sum_{\substack{a \geq 0, b \geq 0 \\ 3a+2b \leq k+1}} P(A_{k,a,b}),$$

and

$$(17) \quad P\{X_{nj} = k\} = P(A_{k-4}) \cdot \frac{n-1}{n} \cdot \frac{1}{n}, \quad k \geq 4.$$

The statement of Theorem 2.1 follows from (15), (16) and (17), after applying the substitution  $s = a + b$ .

(b) From Theorem 2.1 (a) it follows:

$$(18) \quad 1 - F_n(k) = \sum_{k=m+1}^{+\infty} \frac{1}{n^k} (b_{k-3} + a_{k-3}),$$

where the sequences  $a_k$  and  $b_k$  are defined as follows:

$$(19) \quad a_k = \sum_{\substack{a=0, s \geq a \\ a+2s \leq k}} \binom{k-a-s}{s} \binom{s}{a} (n-1)^{k-a-s},$$

$$(20) \quad b_k = \sum_{\substack{a \geq 1, s \geq a \\ a+2s \leq k}} \binom{k-a-s}{s} \binom{s}{a} (n-1)^{k-a-s}.$$

From problem 7(d), page 76 of Riordan (1968) it follows:

$$(21) \quad a_k = (n-1)^k \frac{1}{\alpha} \left( \left( \frac{1+\alpha}{2} \right)^{k+1} - \left( \frac{1-\alpha}{2} \right)^{k+1} \right),$$

where  $\alpha = \sqrt{1 + \frac{4}{n-1}}$ . The following equality:

$$(22) \quad a_{k+1} = (n-1)(a_k + a_{k-1})$$

follows easily. On the other hand, the following relations hold:

$$\begin{aligned} b_k &= \sum_{\substack{a \geq 1, s \geq a \\ a+2s \leq k}} \binom{k-a-s}{s} \binom{s}{a} (n-1)^{k-a-s} \\ &= \sum_{\substack{a \geq 1, s \geq a \\ a+2s \leq k}} \binom{k-a-s-1}{s} \binom{s}{a} (n-1)^{k-a-s} \\ &+ \sum_{\substack{a \geq 1, s \geq a \\ a+2s \leq k}} \binom{k-a-s-1}{s-1} \binom{s-1}{a} (n-1)^{k-a-s} \\ &+ \sum_{\substack{a \geq 1, s \geq a \\ a+2s \leq k}} \binom{k-a-s-1}{s-1} \binom{s-1}{a-1} (n-1)^{k-a-s} \\ &= (n-1)b_{k-1} + (n-1)b_{k-2} \\ (23) \quad &+ (n-1)(b_{k-3} + a_{k-3}). \end{aligned}$$

Using (22) and (23) we obtain the following equality:

$$\begin{aligned} &b_{k+2} - (n-1)b_{k+1} - (n-1)b_k \\ &= (n-1)b_{k+1} + ((n-1) - (n-1)^2)b_k \\ &\quad + ((n-1) - 2(n-1)^2)b_{k-1} \\ (24) \quad &- 2(n-1)^2b_{k-2} - (n-1)^2b_{k-3}. \end{aligned}$$

Finally, from (24),

$$\begin{aligned} &b_{k+2} - 2(n-1)b_{k+1} + ((n-1)^2 - 2(n-1))b_k \\ &\quad + (2(n-1)^2 - (n-1))b_{k-1} \\ (25) \quad &+ 2(n-1)^2b_{k-2} + (n-1)^2b_{k-3} = 0. \end{aligned}$$

The corresponding characteristic equation is

$$\begin{aligned} &t^5 - (2n-2)t^4 + ((n-1)^2 - 2n+2)t^3 \\ &\quad + (2(n-1)^2 - n+1)t^2 \\ (26) \quad &+ 2(n-1)^2t + (n-1)^2 = 0, \end{aligned}$$

which can be factorized as

$$(27) \quad (-t^3 + t^2n - t^2 + tn - t + n - 1)(tn + n - 1 - t - t^2) = 0.$$

Solutions of the equation (27) are:

$$\begin{aligned} t_1 &= \frac{1}{2}(n-1) + \frac{1}{2}\sqrt{(n-1)^2 + 4(n-1)}, \\ t_2 &= \frac{1}{2}(n-1) - \frac{1}{2}\sqrt{(n-1)^2 + 4(n-1)}, \\ t_3 &= \frac{1}{6}C + \frac{2}{3C}(n^2 + n - 2) + \frac{1}{3}(n-1), \\ t_4 &= -\frac{1}{12}C - \frac{1}{3C}(n^2 + n - 2) + \frac{1}{3}(n-1) \\ &\quad + \frac{1}{2}i\sqrt{3} \left( \frac{1}{6}C - \frac{2}{3C}(n^2 + n - 2) \right), \\ t_5 &= -\frac{1}{12}C - \frac{1}{3C}(n^2 + n - 2) + \frac{1}{3}(n-1) \\ (28) \quad &- \frac{1}{2}i\sqrt{3} \left( \frac{1}{6}C - \frac{2}{3C}(n^2 + n - 2) \right), \end{aligned}$$

where

$$\begin{aligned} A &= 12n^2 + 60n - 80 + 8n^3, \\ B &= 6n^3 + 9n^2 + 9n^4 - 72n + 48, \\ (29) \quad C &= \sqrt[3]{A + 12\sqrt{B}}. \end{aligned}$$

As  $n \rightarrow \infty$ , we obtain:

$$(30) \quad C = 2n \left( 1 + \frac{2}{n} - \frac{1}{n^2} - \frac{4}{3n^3} + o\left(\frac{1}{n^2}\right) \right),$$

and, consequently:

$$\begin{aligned} t_1 &= n - \frac{1}{n} + o\left(\frac{1}{n}\right), \\ t_2 &= -1 + \frac{1}{n} + o\left(\frac{1}{n}\right), \\ t_3 &= n - \frac{1}{n^2} + o\left(\frac{1}{n^2}\right), \\ t_4 &= -\frac{1}{2} + \frac{1}{2n^2} + o\left(\frac{1}{n^2}\right) \\ &\quad + i\frac{\sqrt{3}}{2} \left( 1 - \frac{2}{3n} - \frac{32}{9n^2} + o\left(\frac{1}{n^2}\right) \right), \\ t_5 &= -\frac{1}{2} + \frac{1}{2n^2} + o\left(\frac{1}{n^2}\right) \\ (31) \quad &- i\frac{\sqrt{3}}{2} \left( 1 - \frac{2}{3n} - \frac{32}{9n^2} + o\left(\frac{1}{n^2}\right) \right). \end{aligned}$$

Initial values are:

$$b_0 = \sum_{\substack{a \geq 1, s \geq a \\ a+2s \leq 0}} \binom{0-a-s}{s} \binom{s}{a} (n-1)^{0-a-s} = 0,$$

$$\begin{aligned}
b_1 &= \sum_{\substack{a \geq 1, s \geq a \\ a+2s \leq 1}} \binom{1-a-s}{s} \binom{s}{a} (n-1)^{1-a-s} = 0, \\
b_2 &= \sum_{\substack{a \geq 1, s \geq a \\ a+2s \leq 2}} \binom{2-a-s}{s} \binom{s}{a} (n-1)^{2-a-s} = 0, \\
b_3 &= \sum_{\substack{a \geq 1, s \geq a \\ a+2s \leq 3}} \binom{3-a-s}{s} \binom{s}{a} (n-1)^{3-a-s} = n-1, \\
(32) \quad b_4 &= \sum_{\substack{a \geq 1, s \geq a \\ a+2s \leq 4}} \binom{4-a-s}{s} \binom{s}{a} (n-1)^{4-a-s} = 2(n-1)^2.
\end{aligned}$$

Using (18), we obtain:

$$\begin{aligned}
1 - F_n(m) &= \sum_{k=m+1}^{+\infty} \frac{1}{n^k} (b_{k-3} + a_{k-3}) \\
&= \sum_{k=m+1}^{+\infty} \frac{1}{n^k} (C_1 t_1^{k-3} + \dots + C_5 t_5^{k-3}) \\
&\quad + \sum_{k=m+1}^{+\infty} \frac{(n-1)^{k-3}}{n^k} \frac{1}{\alpha} \left( \left( \frac{1+\alpha}{2} \right)^{k-2} - \left( \frac{1-\alpha}{2} \right)^{k-2} \right) \\
&= \frac{1}{n^3} \left( C_1 + \frac{1+\alpha}{2\alpha} \right) \frac{\left( \frac{t_1}{n} \right)^{m-2}}{1 - \frac{t_1}{n}} \\
&\quad + \frac{1}{n^3} \left( C_2 - \frac{1-\alpha}{2\alpha} \right) \frac{\left( \frac{t_2}{n} \right)^{m-2}}{1 - \frac{t_2}{n}} + \frac{1}{n^3} C_3 \frac{\left( \frac{t_3}{n} \right)^{m-2}}{1 - \frac{t_3}{n}} \\
(33) \quad &\quad + \frac{1}{n^3} C_4 \frac{\left( \frac{t_4}{n} \right)^{m-2}}{1 - \frac{t_4}{n}} + \frac{1}{n^3} C_5 \frac{\left( \frac{t_5}{n} \right)^{m-2}}{1 - \frac{t_5}{n}},
\end{aligned}$$

where  $(C_1, C_2, C_3, C_4, C_5)^T$  is the solution of the system of linear equations:

$$(34) \quad \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ t_1 & t_2 & t_3 & t_4 & t_5 \\ t_1^2 & t_2^2 & t_3^2 & t_4^2 & t_5^2 \\ t_1^3 & t_2^3 & t_3^3 & t_4^3 & t_5^3 \\ t_1^4 & t_2^4 & t_3^4 & t_4^4 & t_5^4 \end{pmatrix} \cdot \begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ n-1 \\ 2(n-1)^2 \end{pmatrix},$$

where  $t_1, t_2, \dots, t_5$  are given by (28). Solving the system (34) and using (31), we obtain:

$$(35) \quad C_1 \sim -1, C_2 \sim 0, C_3 \sim 1, C_4 \sim 0, C_5 \sim 0,$$

as  $n \rightarrow \infty$ . Now, let us determine  $m$  from the condition

$$(36) \quad n(1 - F_n(m)) \rightarrow e^{-x}.$$

In the sum (33), all the terms except the term

$$\frac{1}{n^3} C_3 \frac{\left( \frac{t_3}{n} \right)^{m-2}}{1 - \frac{t_3}{n}}$$

can be neglected. Therefore, the condition (36) reduces to:

$$(37) \quad \frac{1}{n^2} C_3 \frac{\left( \frac{t_3}{n} \right)^{m-2}}{1 - \frac{t_3}{n}} \rightarrow e^{-x}.$$

Using (31), (35) and (37), we obtain

$$(38) \quad \frac{1}{n^2} (n^3 + o(n^3)) \left( 1 - \frac{1}{n^3} + o\left( \frac{1}{n^3} \right) \right)^{m-2} \rightarrow e^{-x}.$$

Consequently,

$$(39) \quad -2 \log n + 3 \log n - \frac{m-2}{n^3} \rightarrow -x,$$

therefore,  $m \sim n^3(x + \log n)$ , as  $n \rightarrow \infty$ .  $\square$

### Proof of the Theorem 2.2:

(a) Let  $A$  be a subset of  $\mathbb{N}_n = \{1, 2, \dots, n\}$ ,  $|A| = j$ . For any positive integer  $l$ , let us consider the set  $S$  of all sequences of the form

$$c_1 c_2 \dots c_l, \quad c_1, c_2, \dots, c_l \in \mathbb{N}_n,$$

such that no sequence from  $S$  contains a subsequence of the form  $aaa$ ,  $a \in A$ . Let

$$(40) \quad a_0 = b_0 = p_0 = q_0 = 0, r_0 = 1,$$

and

$$(41) \quad a_1 = p_1 = q_1 = 0, b_1 = j, r_1 = n - j.$$

For  $k \geq 2$ , we define the following sequences:

- $a_l$  is the number of sequences from  $S$  such that  $c_{l-1} = c_l \in A$ ;
- $b_l$  is the number of sequences from  $S$  such that  $c_l \in A, c_{l-1} \in A, c_{l-1} \neq c_l$ ;
- $p_l$  is the number of sequences from  $S$  such that  $c_{l-1} \notin A, c_l \in A$ ;
- $q_l$  is the number of sequences from  $S$  such that  $c_l \notin A, c_{l-1} \in A$ ;
- $r_l$  is the number of sequences from  $S$  such that  $c_{l-1}, c_l \notin A$ .

Then, for any  $k \geq 2$ , the following relations hold:

$$\begin{aligned}
(42) \quad a_k &= j(q_{k-2} + r_{k-2}) + (j-1)(a_{k-2} + b_{k-2} + p_{k-2}), \\
b_k &= (j-1)^2 a_{k-2} + (j-1)j(b_{k-2} + p_{k-2} + q_{k-2} + r_{k-2}), \\
p_k &= j(n-j)(a_{k-2} + b_{k-2} + p_{k-2} + q_{k-2} + r_{k-2}), \\
q_k &= (n-j)j(b_{k-2} + p_{k-2} + q_{k-2} + r_{k-2}) \\
&\quad + (n-j)(j-1)a_{k-2}, \\
r_k &= (n-j)^2(a_{k-2} + b_{k-2} + p_{k-2} + q_{k-2} + r_{k-2}).
\end{aligned}$$

Let  $s_l = a_l + b_l + p_l + q_l + r_l$  for  $l \geq 0$ . Then, from (40), (41) and (42) follows that  $s_0 = 1, s_1 = n$  and, for any  $k \geq 2$ :

$$a_k = j s_{k-2} - (a_{k-2} + b_{k-2} + p_{k-2}),$$

$$\begin{aligned}
b_k &= (j-1)js_{k-2} - (j-1)a_{k-2}, \\
p_k &= j(n-j)s_{k-2}, \\
q_k &= (n-j)js_{k-2} - (n-j)a_{k-2}, \\
(43) \quad r_k &= (n-j)^2s_{k-2}.
\end{aligned}$$

Using relations (43) we obtain

$$(44) \quad b_k + p_k = (-j + jn)s_{k-2} - (j-1)a_{k-2}$$

and

$$(45) \quad b_{k-2} + p_{k-2} = (-j + jn)s_{k-4} - (j-1)a_{k-4}.$$

Substituting (45) into the first equality in (43) leads to:

$$(46) \quad a_k = js_{k-2} - a_{k-2} + j(1-n)s_{k-4} + (j-1)a_{k-4}.$$

Summing all the equalities (43) and using (46), we obtain:

$$(47) \quad s_k = n^2s_{k-2} - na_{k-2} + j(1-n)s_{k-4} + (j-1)a_{k-4}.$$

Subtracting (46) from (47), we obtain:

$$(48) \quad s_k - a_k = (n^2 - j)s_{k-2} - (n-1)a_{k-2}.$$

Therefore,

$$(49) \quad s_k = a_k + (n^2 - j)s_{k-2} - (n-1)a_{k-2}$$

and

$$(50) \quad s_{k-2} = a_{k-2} + (n^2 - j)s_{k-4} - (n-1)a_{k-4}.$$

Substituting (50) into (46), we obtain

$$(51) \quad a_k = (j-1)a_{k-2} + (jn^2 - j^2 + j - jn)s_{k-4} + (-jn + 2j - 1)a_{k-4},$$

and the equalities

$$(52) \quad s_{k-4} = \frac{a_k - (j-1)a_{k-2} + (jn - 2j + 1)a_{k-4}}{jn^2 - j^2 + j - jn},$$

$$(53) \quad s_{k-2} = \frac{a_{k+2} - (j-1)a_k + (jn - 2j + 1)a_{k-2}}{jn^2 - j^2 + j - jn},$$

and

$$(54) \quad s_k = \frac{a_{k+4} - (j-1)a_{k+2} + (jn - 2j + 1)a_k}{jn^2 - j^2 + j - jn}$$

follow easily. After the substitution of (53) and (54) into (48), the equality

$$(55) \quad a_{k+4} + (1-n^2)a_{k+2} + (-n^2 + 2nj - 2j + 1)a_k - (n-j)^2a_{k-2} = 0,$$

for  $k \geq 2$ , follows. Initial values can be easily checked.  $\square$

(b) The corresponding characteristic equation for the recurrence relation (9) is

$$(56) \quad t^6 + (1-n^2)t^4 + (-n^2 + 2nj - 2j + 1)t^2 - (n-j)^2 = 0.$$

Applying the substitution  $s = t^2$  the equation (56) is reduced to the equation of the third degree, which can be solved algebraically. Solutions of the equation (56) are:

$$\begin{aligned}
t_1 &= \frac{1}{6}C + \frac{2}{3}\frac{n^2 + n - 2}{C} + \frac{1}{3}(n-1), \\
t_2 &= -\frac{1}{12}C - \frac{1}{3}\frac{n^2 + n - 2}{C} + \frac{1}{3}(n-1) \\
&\quad + \frac{\sqrt{3}}{2}i \left( \frac{1}{6}C - \frac{2}{3}\frac{n^2 + n - 2}{C} \right), \\
t_3 &= -\frac{1}{12}C - \frac{1}{3}\frac{n^2 + n - 2}{C} + \frac{1}{3}(n-1) \\
&\quad - \frac{\sqrt{3}}{2}i \left( \frac{1}{6}C - \frac{2}{3}\frac{n^2 + n - 2}{C} \right), \\
t_4 &= \frac{1}{6}D + \frac{2}{3}\frac{n^2 + n - 2}{D} - \frac{1}{3}(n-1), \\
t_5 &= -\frac{1}{12}D - \frac{1}{3}\frac{n^2 + n - 2}{D} - \frac{1}{3}(n-1) \\
&\quad + \frac{\sqrt{3}}{2}i \left( \frac{1}{6}D - \frac{2}{3}\frac{n^2 + n - 2}{D} \right), \\
t_6 &= -\frac{1}{12}D - \frac{1}{3}\frac{n^2 + n - 2}{D} - \frac{1}{3}(n-1) \\
&\quad - \frac{\sqrt{3}}{2}i \left( \frac{1}{6}D - \frac{2}{3}\frac{n^2 + n - 2}{D} \right),
\end{aligned} \tag{57}$$

where

$$\begin{aligned}
A &= 12n^2 + 60n + 28 - 108j + 8n^3, \\
B &= 18n^3 + 27n^2 + 9n^4 + 18n + 9 - 90nj \\
&\quad + 81j^2 - 42j - 18n^2j - 12jn^3, \\
C &= \sqrt[3]{A + 12\sqrt{B}}, \\
D &= \sqrt[3]{-A + 12\sqrt{B}}.
\end{aligned}$$

From Theorem 2.2 follows:

$$\begin{aligned}
P\{Y_{nj} > k\} &= \sum_{m=k+1}^{+\infty} P\{Y_{nj} = m\} = \sum_{m=k+1}^{+\infty} \frac{a_{m-1}}{n^m} \\
&= \sum_{m=k+1}^{+\infty} \frac{1}{n^m} (C_1 t_1^{m-1} + C_2 t_2^{m-1} + \dots + C_6 t_6^{m-1}) \\
&= \frac{1}{n} \left( \frac{C_1 \left(\frac{t_1}{n}\right)^k}{1 - \frac{t_1}{n}} + \frac{C_2 \left(\frac{t_2}{n}\right)^k}{1 - \frac{t_2}{n}} + \dots + \frac{C_6 \left(\frac{t_6}{n}\right)^k}{1 - \frac{t_6}{n}} \right),
\end{aligned}$$

where  $(C_1, C_2, C_3, C_4, C_5, C_6)^T$  is the solution of the system

$$(58) \quad \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ t_1 & t_2 & t_3 & t_4 & t_5 & t_6 \\ t_1^2 & t_2^2 & t_3^2 & t_4^2 & t_5^2 & t_6^2 \\ t_1^3 & t_2^3 & t_3^3 & t_4^3 & t_5^3 & t_6^3 \\ t_1^4 & t_2^4 & t_3^4 & t_4^4 & t_5^4 & t_6^4 \\ t_1^5 & t_2^5 & t_3^5 & t_4^5 & t_5^5 & t_6^5 \end{pmatrix} \cdot \begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ j \\ (n-1)j \\ n(n-1)j \\ (n^3 - n^2 + 1)j \end{pmatrix},$$

where  $t_1, t_2, \dots, t_6$  are given by (57). For any fixed  $j$ , as  $n \rightarrow \infty$ , we obtain:

$$C = 2n \left( 1 + \frac{2}{n} - \frac{j}{n^2} - \frac{3j + j^2}{3n^3} + o\left(\frac{1}{n^3}\right) \right),$$

$$D = -2n \left( 1 - \frac{1}{n} + \frac{j}{n^2} + \frac{j^2 - 6j}{3n^3} + o\left(\frac{1}{n^3}\right) \right),$$

and the solutions (57) of the characteristic equation become:

$$\begin{aligned} t_1 &= n \left( 1 - \frac{j}{n^3} + o\left(\frac{1}{n^3}\right) \right), \\ t_2 &= -\frac{1}{2} + \frac{j}{2n^2} + o\left(\frac{1}{n^2}\right) \\ &\quad + \frac{\sqrt{3}}{2}i \left( 1 - \frac{2j}{3n} + \frac{3j - 2j^2}{9n^2} + o\left(\frac{1}{n^2}\right) \right), \\ t_3 &= -\frac{1}{2} + \frac{j}{2n^2} + o\left(\frac{1}{n^2}\right) \\ &\quad - \frac{\sqrt{3}}{2}i \left( 1 - \frac{2j}{3n} + \frac{3j - 2j^2}{9n^2} + o\left(\frac{1}{n^2}\right) \right), \\ t_4 &= -n \left( 1 - \frac{j}{n^3} + o\left(\frac{1}{n^3}\right) \right), \\ t_5 &= \frac{1}{2} - \frac{j}{2n^2} + o\left(\frac{1}{n^2}\right) \\ &\quad + \frac{\sqrt{3}}{2}i \left( 1 - \frac{2j}{3n} + \frac{3j - 2j^2}{9n^2} + o\left(\frac{1}{n^2}\right) \right), \\ t_6 &= \frac{1}{2} - \frac{j}{2n^2} + o\left(\frac{1}{n^2}\right) \\ (59) \quad &- \frac{\sqrt{3}}{2}i \left( 1 - \frac{2j}{3n} + \frac{3j - 2j^2}{9n^2} + o\left(\frac{1}{n^2}\right) \right). \end{aligned}$$

Solving the system (58) and using (59), we obtain:

$$C_1 \sim \frac{j}{n^2}, \quad C_2 = C_3 = O\left(\frac{1}{n}\right),$$

$$(60) \quad C_4 = C_5 = C_6 = o\left(\frac{1}{n^2}\right),$$

as  $n \rightarrow \infty$ . From (59) and (60) follows that in asymptotic behavior of the probability  $P\{Y_{nj} > k\}$  as  $n \rightarrow \infty$ , all terms except the term  $\frac{C_1(t_1/n)^k}{1-t_1/n}$  can be neglected. Denote  $u_n = n^3(x + \log n)$ ,  $r_n = u_n - [u_n]$ , and let us determine the asymptotic behavior of  $P\{Y_{nj} > u_n\}$  for any fixed  $j$  as  $n \rightarrow \infty$ . Using (59) and (60) we obtain that

$$\begin{aligned} P\{Y_{nj} > u_n\} &= \frac{j}{n^3} \left( \frac{t_1(m)}{n} \right)^{[u_n]} \frac{n}{n - t_1(m)} \\ &= \frac{j}{n^3} \frac{e^{-mx} n^3}{n^m j} (1 + o(1)) \\ (61) \quad &= \frac{e^{-mx}}{n^m} (1 + o(1)), \end{aligned}$$

as  $n \rightarrow \infty$ .  $\square$

### Proof of Theorem 3.1:

First, we prove a few lemmas.

**Lemma 4.1.** Let  $X_{nj}^*$ ,  $j \in \{1, 2, \dots, n\}$ , be independent random variables with the same probability distribution,

$$(62) \quad P\{X_{nj}^* = k\} = \sum_{\substack{a \geq 0, s \geq a \\ a+2s \leq k-3}} \binom{k-a-s-3}{s} \binom{s}{a} \\ \times \frac{1}{n^{s+a+3}} \left( 1 - \frac{1}{n} \right)^{k-s-a-3},$$

for  $k \geq 3$ , and let  $M_n^* = \max\{X_{n1}^*, X_{n2}^*, \dots, X_{nn}^*\}$ . For any real number  $x$  the following equality holds:

$$(63) \quad \lim_{n \rightarrow \infty} P\{M_n^* \leq n^3(x + \log n)\} = \exp(-e^{-x}).$$

**Proof:** follows from Theorem 2.1 (b).

**Lemma 4.2.** Let  $p_{nj} = P\{Y_{nj} > n^3(x + \log n)\}$ . For any positive integer  $j \geq 2$ , the following relation holds as  $n \rightarrow \infty$ :

$$(64) \quad p_{nj} - p_{n,j-1}p_{n1} = \frac{e^{-jx}}{n^{j+1}} \cdot o(1).$$

**Proof:** simply follows from Theorem 2.2 (b).

**Lemma 4.3.** (Condition  $D(u_n)$ ) Let  $x$  be a real number and  $k$  and  $l$  positive integers, such that  $k + l \leq n$ . Then, there exists a constant  $C(x)$  such that for  $u_n = n^3(x + \log n)$ , the inequality

$$\begin{aligned} &\left| P\left(\bigcap_{s=1}^{k+l} \{X_{j_s} \leq u_n\}\right) \right. \\ &\quad \left. - P\left(\bigcap_{s=1}^k \{X_{j_s} \leq u_n\}\right) P\left(\bigcap_{s=k+1}^{k+l} \{X_{j_s} \leq u_n\}\right) \right| \\ &\leq C(x) \min\{k, l\} \frac{1}{n^2} \leq \frac{C(x)}{n} \end{aligned}$$

holds.

**Proof** is the same as in the paper Mladenović (2008), where the problem of collecting pairs was considered.

**Lemma 4.4.** (Condition  $D'(u_n)$ ) For  $u_n = n^3(x + \log n)$ , the following relation holds:

$$(65) \quad \lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} n \cdot \sum_{j=2}^{[n/k]} P\{X_{n1} > u_n, X_{nj} > u_n\} = 0.$$

**Proof:** It follows from (3) that  $P\{X_{n1} > u_n, X_{nj} > u_n\} = P\{Y_{n2} > u_n\}$ , for  $j \geq 2$ . The statement of the lemma follows from the relations:

$$n \sum_{j=2}^{[n/k]} P\{X_{n1} > u_n, X_{nj} > u_n\}$$

$$\begin{aligned}
&= n \left( \binom{n}{k} - 1 \right) \frac{e^{-2x}}{n^2} (1 + o(1)) \\
(66) \quad &= \frac{e^{-2x}}{k} (1 + o(1)), \quad \text{as } n \rightarrow \infty.
\end{aligned}$$

**Proof of Theorem:** The statement of the Theorem follows from Lemma 4.1, Lemma 4.2, Lemma 4.3, Lemma 4.4 and Theorem 5.3.1 from Leadbetter, Lindgren and Rootzen (1983).  $\square$

**Proof of Theorem 3.2** (a) Bonferroni inequalities imply:

$$\begin{aligned}
n P\{Y_{n1} > k\} - \binom{n}{2} P\{Y_{n2} > k\} \\
(67) \quad &\leq P\{M_n > k\} \leq n P\{Y_{n1} > k\}.
\end{aligned}$$

Denote  $u_n = n^3(x + c_n + \log n)$ ,  $r_n = u_n - [u_n]$ , and let us determine the asymptotic behavior of  $\binom{n}{m} P\{Y_{nm} > u_n\}$  for any fixed  $m$  as  $n \rightarrow \infty$ .

Using (59) and (60), we obtain:

$$\begin{aligned}
\binom{n}{m} P\{Y_{nm} > u_n\} &= \binom{n}{m} P\{Y_{nm} > [u_n]\} \\
&= \binom{n}{m} \frac{j}{n^3} \left( \frac{t_1(m)}{n} \right)^{[u_n]} \frac{n}{n - t_1(m)} \\
&= \binom{n}{m} \frac{j}{n^3} \frac{e^{-mx}}{n^m e^{mc_n}} \frac{n^3}{j} (1 + o(1)) \\
(68) \quad &= \binom{n}{m} \frac{e^{-mx}}{n^m e^{mc_n}} (1 + o(1)),
\end{aligned}$$

as  $n \rightarrow \infty$ .

For  $m = 1$  and  $m = 2$  it follows from (68) that:

$$\begin{aligned}
n P\{Y_{n1} > u_n\} &= \frac{e^{-x}}{e^{c_n}} (1 + o(1)), \\
(69) \quad \binom{n}{2} P\{Y_{n2} > u_n\} &= \frac{e^{-2x}}{2e^{2c_n}} (1 + o(1)).
\end{aligned}$$

Using (67) and (69) we obtain:

$$\begin{aligned}
P\{M_n > u_n\} &= P\{M_n > n^3(x + c_n + \log n)\} \\
(70) \quad &= \frac{e^{-x}}{e^{c_n}} (1 + o(1)),
\end{aligned}$$

as  $n \rightarrow \infty$ .

Finally, using (70) we obtain the following relations:

$$\begin{aligned}
(71) \quad &P\{M_n \leq n^3(x + c_n + \log n) | M_n > n^3(c_n + \log n)\} \\
&= \frac{P\{M_n > n^3(c_n + \log n)\} - P\{M_n > n^3(x + c_n + \log n)\}}{P\{M_n > n^3(c_n + \log n)\}} \\
&= \frac{\frac{1}{e^{c_n}} (1 + o(1)) - \frac{e^{-x}}{e^{c_n}} (1 + o(1))}{\frac{1}{e^{c_n}} (1 + o(1))} \rightarrow 1 - e^{-x},
\end{aligned}$$

as  $n \rightarrow \infty$ .

(b) For  $m = 3$ , from (68) we obtain:

$$(72) \quad \binom{n}{3} P\{Y_{n3} > [u_n]\} = \frac{e^{-3x}}{6e^{3c_n}} (1 + o(1)),$$

as  $n \rightarrow \infty$ . Bonferroni inequalities imply:

$$\begin{aligned}
n P\{Y_{n1} > k\} - \binom{n}{2} P\{Y_{n2} > k\} + \binom{n}{3} P\{Y_{n3} > k\} \\
(73) \quad &\geq P\{M_n > k\} \geq n P\{Y_{n1} > k\} - \binom{n}{2} P\{Y_{n2} > k\}.
\end{aligned}$$

Using (69), (72) and (73), we obtain:

$$\begin{aligned}
\left( \frac{e^{-x}}{e^{c_n}} - \frac{e^{-2x}}{2e^{2c_n}} + \frac{e^{-3x}}{6e^{3c_n}} \right) (1 + o(1)) &\geq P\{M_n > [u_n]\} \\
(74) \quad &\geq \left( \frac{e^{-x}}{e^{c_n}} - \frac{e^{-2x}}{2e^{2c_n}} \right) (1 + o(1)).
\end{aligned}$$

Consequently,

$$(75) \quad P\{M_n > [u_n]\} = \left( \frac{e^{-x}}{e^{c_n}} - \frac{e^{-2x}}{2e^{2c_n}} + o\left(\frac{1}{e^{2c_n}}\right) \right) (1 + o(1)),$$

as  $n \rightarrow \infty$ . Plugging (75) into the left hand side of the equality (14), we obtain the following relations:

$$\begin{aligned}
&P\{M_n \leq n^3(x + c_n + \log n) | M_n > n^3(c_n + \log n)\} \\
&\quad - (1 - e^{-x}) \\
&= 1 - \frac{P\{M_n > n^3(x + c_n + \log n)\}}{P\{M_n > n^3(c_n + \log n)\}} - (1 - e^{-x}) \\
&= e^{-x} - \frac{P\{M_n > n^3(x + c_n + \log n)\}}{P\{M_n > n^3(c_n + \log n)\}} \\
&= e^{-x} - \left( \frac{e^{-x}}{e^{c_n}} - \frac{e^{-2x}}{2e^{2c_n}} + o\left(\frac{1}{e^{2c_n}}\right) \right) (1 + o(1)) \times \\
&\quad \times \left( \left( \frac{1}{e^{c_n}} - \frac{1}{2e^{2c_n}} + o\left(\frac{1}{e^{2c_n}}\right) \right) (1 + o(1)) \right)^{-1} \\
&= e^{-x} - \frac{e^{-x}}{e^{c_n}} \left( 1 - \frac{e^{-x}}{2e^{c_n}} + o\left(\frac{1}{e^{c_n}}\right) \right) e^{c_n} \times \\
&\quad \times \left( 1 - \frac{1}{2e^{c_n}} + o\left(\frac{1}{e^{c_n}}\right) \right)^{-1} (1 + o(1)) \\
&= e^{-x} \left( 1 - \left( 1 - \frac{e^{-x}}{2e^{c_n}} + \frac{1}{2e^{c_n}} + o\left(\frac{1}{e^{c_n}}\right) \right) \right) (1 + o(1)) \\
&\sim e^{-x} (e^{-x} - 1) \frac{1}{2e^{c_n}},
\end{aligned}$$

as  $n \rightarrow \infty$ .  $\square$

## 5. CONCLUSION

In this paper we obtain the limiting distributions of the random variable  $M_n$  in the coupon collector's problem, where  $M_n$  is defined as the waiting time until all triples



of elements  $jjj, j \in \mathbb{N}_n$ , are sampled. Corresponding results for the waiting time until all pairs of elements  $jj, j \in \mathbb{N}_n$ , are sampled are given in the papers Mladenović (2008) and Jocković, Mladenović (2011).

Although the technique is the same, the case of triples involves a lot more effort in obtaining the exact distributions of random variables  $M_n$  and  $Y_n$ , precisely, solving the difference equations and the corresponding characteristic equations of the fifth and sixth degree. Both equations are reduced to the equations of the second and third degree and solved algebraically.

The work presented here may be extended to the general case, where  $M_n$  is the waiting time until all  $k$ -tuples of elements are sampled,  $k \in \mathbb{N}$ . However, it is not clear if the corresponding characteristic equations can be reduced to the product of polynomials of second and third degree, and, therefore, solved algebraically, or their solutions can be obtained only by using numerical techniques. Is the “solvability” of the corresponding characteristic equations a general property of the problem? This is the subject that requires further research.

*Received 26 September 2013*

## REFERENCES

- BALKEMA, A. A. AND DE HAAN, L. (1974). Residual life time at great age. *Ann. Probab.* **2**, 792–804. [MR0359049](#)
- BAUM, L. E. AND BILINGSLEY, P. (1965). Asymptotic distributions for the coupon collector’s problem. *Ann. Math. Statist.* **36**, 1835–1839. [MR0182039](#)

- GONCHAROV, V. L. (1944). On the field of combinatorial analysis. *Izv. Akad. Nauk SSSR Ser. Mat.* **8**, 3–48 and *Transl. Amer. Math. Soc. Ser. 2* **19** (1962), 1–46. [MR0131369](#)
- ERDÖS, P. AND RÉNYI, A. (1961). On a classical problem of probability theory. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **6**, (1–2), 215–220. [MR0150807](#)
- HOLST, L. (1986). On birthday, collector’s, occupancy and other classical urn problems. *International Statistical Review* **54**, (1), 15–27. [MR0959649](#)
- JOCKOVIĆ, J. AND MLADENOVIĆ, P. (2011). Coupon collector’s problem and generalized Pareto distributions. *Journal of Statistical Planning and Inference* **141**, 2348–2352. [MR2775213](#)
- LEADBETTER, M. R., LINDGREN, G. AND ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York. [MR0691492](#)
- MLADENOVIĆ, P. (2008). Limit distributions for the problem of collecting pairs. *Bernoulli* **14**, (2), 419–439. [MR2544095](#)
- RIORDAN, J. (1968). *Combinatorial Identities*. New York, Wiley. [MR0231725](#)

Jelena Jocković  
University of Belgrade  
Faculty of Pharmacy  
Vojvode Stepe 450  
11000 Belgrade  
Serbia  
E-mail address: [haustor@pharmacy.bg.ac.rs](mailto:haustor@pharmacy.bg.ac.rs)

Pavle Mladenović  
University of Belgrade  
Faculty of Mathematics  
Studentski trg 16  
11000 Belgrade  
Serbia  
E-mail address: [paja@matf.bg.ac.rs](mailto:paja@matf.bg.ac.rs)