

THE COUPON COLLECTOR’S PROBLEM REVISITED: GENERALIZING THE DOUBLE DIXIE CUP PROBLEM OF NEWMAN AND SHEPP

ARISTIDES V. DOUMAS¹ AND VASSILIS G. PAPANICOLAOU¹

Abstract. The “double Dixie cup problem” of [D.J. Newman and L. Shepp, *Amer. Math. Monthly* **67** (1960) 58–61] is a well-known variant of the coupon collector’s problem, where the object of study is the number $T_m(N)$ of coupons that a collector has to buy in order to complete m sets of all N existing different coupons. More precisely, the problem is to determine the asymptotics of the expectation (and the variance) of $T_m(N)$, as well as its limit distribution, as the number N of different coupons becomes arbitrarily large. The classical case of the problem, namely the case of equal coupon probabilities, is here extended to the general case, where the probabilities of the selected coupons are unequal. In the beginning of the article we give a brief review of the formulas for the moments and the moment generating function of the random variable $T_m(N)$. Then, we develop techniques of computing the asymptotics of the first and the second moment of $T_m(N)$ (our techniques apply to the higher moments of $T_m(N)$ as well). From these asymptotic formulas we obtain the leading behavior of the variance $V[T_m(N)]$ as $N \rightarrow \infty$. Finally, based on the asymptotics of $E[T_m(N)]$ and $V[T_m(N)]$ we obtain the limit distribution of the random variable $T_m(N)$ for large classes of coupon probabilities. As it turns out, in many cases, albeit not always, $T_m(N)$ (appropriately normalized) converges in distribution to a Gumbel random variable. Our results on the limit distribution of $T_m(N)$ generalize a well-known result of [P. Erdős and A. Rényi, *Magyar. Tud. Akad. Mat. Kutató Int. Közl.* **6** (1961) 215–220] regarding the limit distribution of $T_m(N)$ for the case of equal coupon probabilities.

Mathematics Subject Classification. 60F05, 60F99.

Received September 1st, 2015. Revised June 1st, 2016. Accepted June 8, 2016.

1. INTRODUCTION

The “*coupon collector’s problem*” (CCP) pertains to a population whose members are of N different *types* (e.g., baseball cards, viruses, fish, words, etc.). For $1 \leq j \leq N$ we denote by p_j the probability that a member of the population is of type j , where $p_j > 0$ and $\sum_{j=1}^N p_j = 1$. We refer to the p_j ’s as the *coupon probabilities*. The members of the population are sampled independently *with replacement* and their types are recorded. Naturally, a quantity of interest here is the number $T(N)$ of trials needed until all N types are detected (at least once).

CCP belongs to the family of *Urn problems* among with other classical problems, such as the birthday and the matching problem. In its simplest form (*i.e.* when all p_j ’s are equal and the collector aims for one complete

Keywords and phrases. Urn problems, coupon collector’s problem, double Dixie cup problem, rising moments, limit distribution, Gumbel distribution, generalized Zipf law.

¹ Department of Mathematics, National Technical University of Athens, Zografou Campus, 15780 Athens, Greece.
aris.doumas@hotmail.com; papanico@math.ntua.gr

set of coupons) the problem has appeared in many standard probability textbooks (*e.g.*, in W. Feller's classical work [13], as well as in [10, 19, 22, 23], to name a few). Its origin can be traced back to De Moivre's treatise *De Mensura Sortis* of 1712 (see, *e.g.*, [15]) and Laplace's pioneering work *Theorie Analytique de Probabilites* of 1812 (see [7]). The problem was related to the Dixie Cup Company, since in the 1930's the company introduced a highly successful procedure by which children collected Dixie lids to receive "Premiums", beginning with illustrations of their favored Dixie Circus characters, and then Hollywood stars and Major League baseball players (see [19, 25]).

CCP has attracted the attention of various researchers due to its applications to several areas of science (computer science—search algorithms, mathematical programming, optimization, learning processes, engineering, ecology, as well as linguistics—see, *e.g.*, [4]).

For the asymptotics of the moments, as well as for the limit distribution of the random variable $T(N)$, there is a plethora of articles obtaining a variety of results (for the case of equal probabilities see for instance [11, 14, 17]; as for the case of unequal probabilities, see, *e.g.*, [6]—which presents the results of Brayton's Ph.D. thesis under the supervision of Norman Levinson—[8, 9, 20], and the references therein).

Several variants of the original problem have been studied. Among them there is the so-called "*double Dixie cup problem*", which reads: how long does it take to obtain m complete sets of N coupons?

Let $T_m(N)$ be the number of trials a collector needs in order to accomplish this goal (obviously, $T_1(N) = T(N)$, thus the case $m = 1$ reduces to the more "classical" CCP).

Naturally, the simplest case occurs when one takes $p_1 = \dots = p_N = 1/N$. For this case Newman and Shepp [21] obtained that, for fixed m ,

$$E[T_m(N)] = N \ln N + (m - 1) N \ln \ln N + NC_m + o(N) \quad (1.1)$$

as $N \rightarrow \infty$, where C_m is a constant depending on m . Roughly speaking, formula (1.1) tells us that, on the average, the first set "costs" $N \ln N + O(N)$, while each additional set has an additional cost of $N \ln \ln N + O(N)$.

Soonafter, Erdős and Rényi [11] went a step further and determined the limit distribution of $T_m(N)$, as well as the exact value of the constant C_m . They proved that

$$C_m = \gamma - \ln(m - 1)!, \quad (1.2)$$

where $\gamma = 0.5772\dots$ is the Euler–Mascheroni constant, and that for every real y the following limiting result holds:

$$\lim_{N \rightarrow \infty} P \left\{ \frac{T_m(N) - N \ln N - (m - 1)N \ln \ln N}{N} \leq y \right\} = \exp \left(-\frac{e^{-y}}{(m - 1)!} \right) \quad (1.3)$$

or, equivalently,

$$\lim_{N \rightarrow \infty} P \left\{ \frac{T_m(N) - N \ln N - (m - 1)N \ln \ln N + N \ln(m - 1)!}{N} \leq y \right\} = e^{-e^{-y}} \quad (1.4)$$

(in the right-hand side of (1.4) we have the standard Gumbel distribution function; recall that its expectation is γ and its variance is $\pi^2/6$).

Later, and as long as the coupon probabilities remained equal, this result was generalized in [14, 17].

In the present paper we extend the results of Newman–Shepp [21] and Erdős–Rényi [11] to large families of unequal coupon probabilities. Notice that in practically all applications the coupon probabilities are not equal (for example, in several applications the coupon probabilities follow a generalized Zipf law, a case which is covered by our results). As we will see, for many families of coupon probabilities the quantity $T_m(N)$, after an appropriate normalization, converges in distribution to the standard Gumbel random variable as $N \rightarrow \infty$. The correct normalization of $T_m(N)$, which depends on the coupon probabilities, is determined with the help of the asymptotics of $E[T_m(N)]$ and $V[T_m(N)]$. We also present families of coupon probabilities for which the limit distribution of $T_m(N)$ (again after a suitable normalization) is not Gumbel.

1.1. Moments and the moment generating function of $T_m(N)$

Suppose that, for $j = 1, \dots, N$, we denote by W_j the number of trials needed in order to obtain m times the coupon of type j . Then, it is clear that W_j is a negative binomial random variable (with parameters m and p_j) and

$$T_m(N) = \max_{1 \leq j \leq N} W_j.$$

However, the above formula for $T_m(N)$ is not very useful, since the W_j 's are not independent. Instead, one can apply a clever ‘‘Poissonization technique’’ found in [23] in order to get explicit formulas for the moments of $T_m(N)$.

Let $Z(t)$, $t \geq 0$, be a Poisson process with rate $\lambda = 1$. We imagine that each Poisson event associated to Z is a collected coupon, so that $Z(t)$ is the number of collected coupons at time t . Next, for $j = 1, \dots, N$, let $Z_j(t)$ be the number of type- j coupons collected at time t . Then, the processes $\{Z_j(t)\}_{t \geq 0}$, $j = 1, \dots, N$, are independent Poisson processes with rates p_j respectively [23] and, of course, $Z(t) = Z_1(t) + \dots + Z_N(t)$. If X_j , $j = 1, \dots, N$, denotes the time of the m th event of the process Z_j , then X_1, \dots, X_N are obviously independent (being associated to independent processes) and

$$X := \max_{1 \leq j \leq N} X_j \tag{1.5}$$

is the time when all different coupons have arrived at least m times.

Now, for each $j = 1, \dots, N$, X_j is Erlang with parameters m and p_j , meaning that

$$P\{X_j > t\} = S_m(p_j t)e^{-p_j t}, \tag{1.6}$$

where $S_m(y)$ denotes the m th partial sum of e^y , namely

$$S_m(y) := 1 + y + \frac{y^2}{2!} + \dots + \frac{y^{m-1}}{(m-1)!} = \sum_{l=0}^{m-1} \frac{y^l}{l!}. \tag{1.7}$$

Incidentally, let us observe that

$$0 < e^{-y}S_m(y) < 1 \quad \text{for all } y > 0. \tag{1.8}$$

It follows from (1.5) and the independence of the X_j 's that

$$P\{X \leq t\} = \prod_{j=1}^N [1 - S_m(p_j t)e^{-p_j t}]. \tag{1.9}$$

It remains to relate X and $T_m(N)$. Clearly,

$$X = \sum_{k=1}^{T_m(N)} U_k, \tag{1.10}$$

where U_1, U_2, \dots are the interarrival times of the process Z . It is common knowledge that the U_j 's are independent and exponentially distributed random variables with parameter 1. In order to compute the moments of $T_m(N)$ via formula (1.10) we need the formula

$$E \left[\left(\sum_{k=1}^M U_k \right)^r \right] = M(M+1) \dots (M+r-1) =: M^{(r)}, \quad r = 1, 2, \dots \tag{1.11}$$

(the justification of (1.11) is immediate, if we just notice that $U_1 + \dots + U_M$ is Erlang with parameters M and 1). Since $T_m(N)$ is independent of the U_j 's, formulas (1.10) and (1.11) imply

$$E[X^r | T_m(N)] = T_m(N)^{(r)} E[U_j] = T_m(N)^{(r)}, \quad r = 1, 2, \dots, \tag{1.12}$$

hence, by (1.12), with the help of (1.6), we obtain

$$E [T_m(N)^{(r)}] = E [X^r] = r \int_0^\infty \left\{ 1 - \prod_{j=1}^N [1 - S_m(p_j t)e^{-p_j t}] \right\} t^{r-1} dt \tag{1.13}$$

for $r = 1, 2, \dots$. The quantity $E [T_m(N)^{(r)}]$ is, actually, the r th rising moment of $T_m(N)$. In particular,

$$E[T_m(N)] = \int_0^\infty \left\{ 1 - \prod_{j=1}^N [1 - S_m(p_j t)e^{-p_j t}] \right\} dt, \tag{1.14}$$

$$E [T_m(N) (T_m(N) + 1)] = 2 \int_0^\infty \left\{ 1 - \prod_{j=1}^N [1 - S_m(p_j t)e^{-p_j t}] \right\} t dt, \tag{1.15}$$

and, of course,

$$V [T_m(N)] = E [T_m(N) (T_m(N) + 1)] - E [T_m(N)] - E [T_m(N)]^2. \tag{1.16}$$

Formulas (1.14) and (1.15) were first derived in [6] by a more complicated argument. As far as we know, the more general formula (1.13) is new.

Using (1.13) one can easily obtain the moment generating function of $T_m(N)$:

$$G(z) := E [z^{-T_m(N)}] = 1 - (z - 1) \int_0^\infty \left\{ 1 - \prod_{j=1}^N [1 - S_m(p_j t)e^{-p_j t}] \right\} e^{-(z-1)t} dt, \tag{1.17}$$

where $\Re(z) > 1$.

1.2. Discussion

Under the quite restrictive assumption of “nearly equal coupon probabilities”, namely

$$\lambda(N) := \frac{\max_{1 \leq j \leq N} \{p_j\}}{\min_{1 \leq j \leq N} \{p_j\}} \leq M < \infty, \quad \text{independently of } N, \tag{1.18}$$

Brayton [6] obtained detailed asymptotics of the expectation $E[T_m(N)]$, while for the asymptotics of the variance, he only did the case $m = 1$, where he found the formula

$$V [T_1(N)] = N^2 \left[\frac{\pi^2}{6} + O \left(\frac{\ln \ln \ln N}{\ln \ln N} \right) \right] \quad \text{as } N \rightarrow \infty. \tag{1.19}$$

The present paper builds on [8], where the case $m = 1$ was considered. Our results here are valid for all positive integers m , including $m = 1$.

The rest of our work is organized as follows. In Section 2 we calculate the asymptotics of $E[T_m(N)]$ and $V[T_m(N)]$ for the the general case of unequal probabilities. We first show how to create a sequence $\pi_N = (p_1, \dots, p_N)$, $N = 2, 3, \dots$, of probability measures (*i.e.* of coupon probabilities) by successive normalizations of the terms of a given, albeit arbitrary, sequence $\alpha = \{a_j\}_{j=1}^\infty$ of positive real numbers. Thus, we need to focus on the sequence α . First (Case I) we consider a large class of sequences α , such that $a_j \rightarrow \infty$. The main result for this case is presented in Theorem 2.2.

Then (Case II) we consider classes of decaying sequences α such that $a_j \rightarrow 0$. This case is much more challenging. It turns out that in order to obtain the leading term of the variance $V[T_m(N)]$ (see Thm. 2.5) one has to go deep in the asymptotics of $E [T_m(N)]$ (up to the fifth term) and $E [T_m(N) (T_m(N) + 1)]$ (up to the sixth term). These asymptotic formulas are presented in Theorems 2.9 and 2.8 respectively.

Our way of defining the triangular sequence of coupon probabilities may look restrictive. The main advantage of our approach is that we can obtain very explicit asymptotic formulas. If one considers substantially more general triangular sequence of coupon probabilities, then it seems hopeless to get explicit results. Furthermore, among the classes of coupon probabilities we consider there are some popular classes, such as (see Sect. 3) the polynomial law (in particular, the so-called *linear* case falls in this category) and exponential families of coupon probabilities (Case I), as well as, the so-called *generalized Zipf law* (Case II). Those classes arise in various areas of science (see, *e.g.*, [12, 16], and the recent work of Locey and Lennon in the Proceedings of the National Academy of Sciences [18]).

The approach presented in Section 2 can be used to calculate the asymptotics, as $N \rightarrow \infty$, of the r th rising moment of $T_m(N)$, for any positive integer r .

Section 3 is a short section where we present some illustrative examples. Then, in Section 4 we take advantage of our formulas in order to find the limit distribution of $T_m(N)$ (appropriately normalized) for a very large class of coupon probabilities. More precisely, for sequences of Case I the limit distribution is obtained (in Thm. 4.1) by using formula (1.17). As for sequences of Case II, we combine our asymptotic formulas with a limit theorem of Neal [20] (in the spirit of [2]) in order to obtain the appropriate normalization of the random variable $T_m(N)$ and arrive into specific limiting distributions. Our main results are Theorem 4.2 (see also Sect. 4.3, whose content complements Thm. 4.2). It is notable that for the considered class of coupon probabilities the random variable $T_m(N)$, appropriate normalized, converges in distribution to a Gumbel random variable. This is a generalization of the classical result of Erdős and Rényi (see (1.3)) for the case of equal coupon probabilities. For the special case $m = 1$ the statement of Theorem 4.2 had been established in our earlier work [8]. Section 4.3 is a brief discussion of the case of slowly decaying sequences. As an illustration we consider the sequence $a_j = 1/j^p$, where $p > 0$. Such sequences were never studied before, not even in the case $m = 1$. An interesting phenomenon arises regarding decaying sequences: If the decay of α is subexponential, then, at least for the great variety of cases we have analyzed, the limit distribution of $T_m(N)$ is always Gumbel. However, if α decays to zero exponentially, then the limit distribution of $T_m(N)$ is not Gumbel and we expect that the same is true for sequences decaying to zero superexponentially. In the latter case the behavior of $T_m(N)$ seems similar to the corresponding behavior for the case where α tends to infinity.

The proofs of Theorem 2.2 and of some technical lemmas of Section 2 are given in the Appendix (Sect. 5). Finally, Section 6 is a short epilogue containing a comparison with earlier works and some remarks on the construction of the coupon probabilities.

1.3. Some conjectures

We finish this introductory section with two conjectures. Formulas (1.3)–(1.4) suggest the following conjecture:

Conjecture 1. In the case of equal coupon probabilities we have

$$V[T_m(N)] \sim \frac{\pi^2}{6} N^2, \quad N \rightarrow \infty. \quad (1.20)$$

Actually, for $m = 1$, we have already seen formula (1.19), proved in [6], which is a stronger form of (1.20).

Conjecture 2. For fixed positive integers m and N , the case of equal probabilities, has the property that it is the one with the smallest variance $V[T_m(N)]$.

The results of the present work confirm that, for a large class of probabilities, $V[T_m(N)]$ is actually minimized in the case of equal probabilities, as N becomes sufficiently large.

2. ASYMPTOTICS OF $E[T_m(N)]$ AND $V[T_m(N)]$

2.1. Preliminary material

For relatively small values of N it is easy to evaluate $E[T_m(N)]$ and $V[T_m(N)]$ by using formulas (1.14) and (1.15) respectively. However, for large N (this case is typical in applications) it is not clear at all what

information one can obtain for $E[T_m(N)]$ and $V[T_m(N)]$ from the aforementioned formulas, *e.g.*, the bounds for $E[T_m(N)]$ and $V[T_m(N)]$, which can be derived from those formulas are crude to the point of being essentially meaningless. For this reason there is a need to develop efficient ways for deriving asymptotics as $N \rightarrow \infty$. As in [5, 8, 9], let $\alpha = \{a_j\}_{j=1}^\infty$ be a sequence of strictly positive numbers. Then, for each integer $N > 0$, one can create a probability measure $\pi_N = \{p_1, \dots, p_N\}$ on the set of types $\{1, \dots, N\}$ by taking

$$p_j = \frac{a_j}{A_N}, \quad \text{where } A_N = \sum_{j=1}^N a_j. \tag{2.1}$$

Notice that p_j depends on α and N , thus, given α , it makes sense to consider the asymptotic behavior of $E[T_m(N)]$, $E[T_m(N)(T_m(N) + 1)]$, and $V[T_m(N)]$ as $N \rightarrow \infty$.

Remark 2.1. If $\alpha = \{a_j\}_{j=1}^\infty$ is such that $\lim_j a_j = \infty$, by rearranging its terms, it can be assumed, essentially without loss of generality, that α is a nondecreasing sequence (see Sect. 6.2). Likewise, if $\alpha = \{a_j\}_{j=1}^\infty$ is such that $\lim_j a_j = 0$, then by rearranging its terms, it can be assumed essentially without loss of generality, that α is a nonincreasing sequence.

Motivated by (1.14) we introduce the notation

$$E_m(N; \alpha) := \int_0^\infty \left[1 - \prod_{j=1}^N \left(1 - e^{-a_j t} S_m(a_j t) \right) \right] dt \tag{2.2}$$

$$= \int_0^1 \left[1 - \prod_{j=1}^N \left(1 - x^{a_j} S_m(-a_j \ln x) \right) \right] \frac{dx}{x}. \tag{2.3}$$

For a sequence $\alpha = \{a_j\}_{j=1}^\infty$ and a number $s > 0$ we set $s\alpha = \{sa_j\}_{j=1}^\infty$ (notice that α and $s\alpha$ create the same sequence of probability measures π_N , $N = 2, 3, \dots$). Then, (2.2) implies that $E_m(N; s\alpha) = s^{-1}E_m(N; \alpha)$ and hence, in view of (1.14) and (2.1),

$$E[T_m(N)] = A_N E_m(N; \alpha). \tag{2.4}$$

Likewise, motivated by (1.15), let us introduce

$$Q_m(N; \alpha) := 2 \int_0^\infty t \left[1 - \prod_{j=1}^N \left(1 - e^{-a_j t} S_m(a_j t) \right) \right] dt \tag{2.5}$$

$$= -2 \int_0^1 \left[1 - \prod_{j=1}^N \left(1 - x^{a_j} S_m(-a_j \ln x) \right) \right] \frac{\ln x}{x} dx. \tag{2.6}$$

From the above it follows that $Q_m(N; s\alpha) = s^{-2}Q_m(N; \alpha)$, hence

$$E[T_m(N)(T_m(N) + 1)] = A_N^2 Q_m(N; \alpha). \tag{2.7}$$

In view of (2.4) and (2.7), (1.16) yields

$$V[T_m(N)] = A_N^2 Q_m(N; \alpha) - A_N E_m(N; \alpha) - A_N^2 E_m(N; \alpha)^2. \tag{2.8}$$

Under (2.1) the problem of estimating $E[T_m(N)]$ can be treated as two separate problems, namely estimating A_N and estimating $E_m(N; \alpha)$, (see (2.4)). The estimation of A_N can be considered an external matter which can be handled by existing powerful methods, such as the Euler–Maclaurin sum formula, the Laplace method for sums (see, *e.g.*, [3]), or even summation by parts. Hence, our analysis focuses on estimating $E_m(N; \alpha)$. Of course, the same observation applies to the expression of (2.7).

2.2. The Dichotomy

For convenience, we set

$$f_N^\alpha(x) := \prod_{j=1}^N [1 - x^{a_j} S_m(-a_j \ln x)], \quad 0 \leq x \leq 1,$$

in particular, $f_N^\alpha(0) := f_N^\alpha(0+) = 1$ and $f_N^\alpha(1) = 0$.

Since

$$\frac{d}{dy} [e^{-y} S_m(y)] = -\frac{y^{m-1} e^{-y}}{(m-1)!}, \tag{2.9}$$

we get that $f_N^\alpha(x)$ is monotone decreasing in x . Furthermore, (1.8) implies immediately that $f_{N+1}^\alpha(x) \leq f_N^\alpha(x)$. In particular

$$\lim_N f_N^\alpha(x) = \prod_{j=1}^\infty [1 - x^{a_j} S_m(-a_j \ln x)] \quad \text{exists for all } x \in [0, 1].$$

Thus, by applying the Monotone Convergence Theorem in (2.3) and (2.6), we get respectively

$$L_1(\alpha; m) := \lim_N E_m(N; \alpha) = \int_0^1 \left\{ 1 - \prod_{j=1}^\infty [1 - x^{a_j} S_m(-a_j \ln x)] \right\} \frac{dx}{x} \tag{2.10}$$

and

$$L_2(\alpha; m) := \lim_N Q_m(N; \alpha) = -2 \int_0^1 \left\{ 1 - \prod_{j=1}^\infty [1 - x^{a_j} S_m(-a_j \ln x)] \right\} \frac{\ln x}{x} dx. \tag{2.11}$$

Notice that $L_1(\alpha; m), L_2(\alpha; m) > 0$, for any α (since, for every $x \in (0, 1)$, $f_N^\alpha(x) < 1$ and decreases with N). However, we may have $L_1(\alpha; m) = \infty$ and/or $L_2(\alpha; m) = \infty$. In fact, Theorem 2.2 below tells us that $L_1(\alpha; m) = \infty$ if and only if $L_2(\alpha; m) = \infty$.

Theorem 2.2. *Let $L_1(\alpha; m)$ and $L_2(\alpha; m)$ as defined in (2.10) and (2.11) respectively. The following are equivalent (for all positive integers m):*

- (i) $L_1(\alpha; m) < \infty$;
- (ii) $L_2(\alpha; m) < \infty$;
- (iii) *There exist a $\xi \in (0, 1)$ such that*

$$\sum_{j=1}^\infty \xi^{a_j} < \infty. \tag{2.12}$$

The proof of the theorem is given in the Appendix.

The theorem implies that we have the following dichotomy simultaneously for all positive integers m :

$$(i) \ 0 < L_1(\alpha; m), L_2(\alpha; m) < \infty \quad \text{or} \quad (ii) \ L_1(\alpha; m) = L_2(\alpha; m) = \infty. \tag{2.13}$$

Remark 2.3. The word ‘‘dichotomy’’ may be misleading: If $p > 0$, then, the sequence $\alpha = \{e^{pj}\}_{j=1}^\infty$ satisfies $L_i(\alpha; m) < \infty$, while for the sequence $\beta = \{e^{-pj}\}_{j=1}^\infty$ we have $L_i(\beta; m) = \infty$ ($i = 1, 2$). However, it is clear that α and β produce the same coupon probabilities(!), *i.e.* the same sequence of probability measures $\{\pi_N\}_{N=2}^\infty$. Actually, this is an exceptional case as explained in Section 6.2

2.3. Case I: $L_1(\alpha; m) < \infty$

Let A_N and $L_1(\alpha; m)$ be as in (2.1) and (2.10). We note that, by Theorem 2.2 (see (2.12)), $L_1(\alpha; m) < \infty$ implies that $\lim_j a_j = \infty$ (hence $\lim_N A_N = \infty$).

Theorem 2.4. *If $L_1(\alpha; m) < \infty$, then for all positive integers m we have*

$$E [T_m(N)] = A_N L_1(\alpha; m) (1 + \delta_N), \tag{2.14}$$

$$E [T_m(N)(T_m(N) + 1)] = A_N^2 L_2(\alpha; m) (1 + \Delta_N), \tag{2.15}$$

$$V [T_m(N)] = A_N^2 [L_2(\alpha; m) - L_1(\alpha; m)^2] [1 + o(1)] \tag{2.16}$$

(as $N \rightarrow \infty$), where for the error terms

$$\delta_N := L_1(\alpha; m) - E_m(N; \alpha) \quad \text{and} \quad \Delta_N := L_2(\alpha; m) - Q_m(N; \alpha) \tag{2.17}$$

we have $\delta_N = o(1)$ and $\Delta_N = o(1)$ as $N \rightarrow \infty$. Furthermore, in the formula (2.16) it is always true that

$$L_2(\alpha; m) - L_1(\alpha; m)^2 > 0. \tag{2.18}$$

Proof. Formula (2.14) follows immediately from (2.4) and (2.10), while formula (2.15) follows from (2.7) and (2.11).

To prove (2.18) we first notice that

$$G(x) := 1 - \prod_{j=1}^{\infty} [1 - x^{a_j} S_m(-a_j \ln x)]$$

is a nondegenerate distribution function on $[0, 1]$. If X is a random variable with distribution function $G(x)$, then simple integration by parts in (2.11) and (2.10) gives

$$L_2(\alpha; m) = E_G [\ln(x)^2] > E_G [\ln(x)]^2 = L_1(\alpha; m)^2,$$

where $E_G[\cdot]$ denotes the expectation associated to the distribution function $G(x)$. Having established (2.18), formula (2.16) follows by using (2.14) and (2.15) in (1.16). □

Remark 2.5. If a_j grows to infinity sufficiently fast, we can get a better estimate for the errors δ_N and Δ_N of (2.17). By (2.3), (2.10), (A.1) (see Appendix), and Tonelli’s theorem

$$\begin{aligned} \delta_N &= \int_0^1 \prod_{j=1}^N (1 - x^{a_j} S_m(-a_j \ln x)) \left[1 - \prod_{j=N+1}^{\infty} (1 - x^{a_j} S_m(-a_j \ln x)) \right] \frac{dx}{x} \\ &\leq \sum_{j=N+1}^{\infty} \left[\int_0^1 x^{a_j} S_m(-a_j \ln x) \frac{dx}{x} \right] = \sum_{j=N+1}^{\infty} \sum_{k=0}^{m-1} \int_0^1 x^{a_j-1} (\ln x)^k. \end{aligned}$$

Integration by parts yields

$$\delta_N \leq m \sum_{j=N+1}^{\infty} \frac{1}{a_j}. \tag{2.19}$$

In a similar manner one gets

$$\Delta_N \leq m(m+1) \sum_{j=N+1}^{\infty} \frac{1}{a_j^2}. \tag{2.20}$$

Remark 2.6. For $r = 1, 2, \dots$ let us set

$$L_r(\alpha; m) := (-1)^{r-1} r \int_0^1 \left\{ 1 - \prod_{j=1}^{\infty} [1 - x^{a_j} S_m(-a_j \ln x)] \right\} \frac{\ln^{r-1} x}{x} dx,$$

i.e.

$$L_r(\alpha; m) = r \int_0^{\infty} \left\{ 1 - \prod_{j=1}^{\infty} [1 - S_m(a_j t) e^{-a_j t}] \right\} t^{r-1} dt.$$

Then, Theorem 2.2 is valid for $L_r(\alpha; m)$, for any r (the proof is similar). Furthermore, it is not hard to see that

$$E \left[T_m(N)^{(r)} \right] = A_N^r L_r(\alpha; m) [1 + o(1)], \quad N \rightarrow \infty,$$

which is an extension of Theorem 2.2 for all r .

2.4. Case II: $L_1(\alpha; m) = \infty$

2.4.1. Asymptotic behavior of $E[T_m(N)]$

By Theorem 2.2, $L_1(\alpha; m) = \infty$ is equivalent to

$$\sum_{j=1}^{\infty} x^{a_j} = \infty, \quad \text{for all } x \in (0, 1).$$

For our further analysis we follow [5, 8, 9], and write a_j in the form

$$a_j = \frac{1}{f(j)}, \tag{2.21}$$

where

$$f(x) > 0 \quad \text{and} \quad f'(x) > 0. \tag{2.22}$$

In order to proceed we assume that $f(x)$ possesses three derivatives satisfying the following conditions as $x \rightarrow \infty$:

$$\begin{aligned} \text{(i)} \quad & f(x) \rightarrow \infty, & \text{(ii)} \quad & \frac{f'(x)}{f(x)} \rightarrow 0, \\ \text{(iii)} \quad & \frac{f''(x)/f'(x)}{f'(x)/f(x)} = O(1), & \text{(iv)} \quad & \frac{f'''(x) f(x)^2}{f'(x)^3} = O(1). \end{aligned} \tag{2.23}$$

Roughly speaking, $f(\cdot)$ belongs to the class of positive and strictly increasing $C^3(0, \infty)$ functions, which grow to ∞ (as $x \rightarrow \infty$) slower than exponentials, but faster than powers of logarithms. These conditions are satisfied by a variety of commonly used functions. For example,

$$f(x) = x^p (\ln x)^q, \quad p > 0, \quad q \in \mathbb{R}, \quad f(x) = \exp(x^r), \quad 0 < r < 1,$$

or various convex combinations of products of such functions. Notice that the smoothness assumption on f does not impose any restriction on the sequence α , since we only need $f(x)$ to interpolate $1/a_j$ for $x = j, j = 1, 2, \dots$. The restrictions of α come from the growth assumptions (2.23). In Section 4.3 we discuss the case where $f(x)$ grows slower than any (positive) power of x and hence does not satisfy all conditions of (2.23).

For typographical convenience we set

$$F(x) := f(x) \ln \left(\frac{f(x)}{f'(x)} \right) \tag{2.24}$$

(notice that (2.22) and (ii) of (2.23) imply that $F(x) > 0$, for x sufficiently large). The following lemma plays an important role in our analysis:

Lemma 2.7. *Set*

$$J_{\kappa}(N) := \int_1^N f(x)^{\kappa} e^{-\frac{F(N)}{f(x)} s} dx, \quad \kappa \in \mathbb{R}. \tag{2.25}$$

Then, under (2.23) and (2.24), we have

$$J_\kappa(N) = \frac{f(N)^{\kappa+2}}{sF(N)f'(N)} e^{-\frac{F(N)}{f(N)}s} + \omega(N) \frac{f(N)^{\kappa+3}}{s^2F(N)^2f'(N)} e^{-\frac{F(N)}{f(N)}s} \left[1 + O\left(\frac{f(N)}{F(N)}\right) \right], \tag{2.26}$$

where

$$\omega(N) := -2 + \frac{f''(N)/f'(N)}{f'(N)/f(N)}, \tag{2.27}$$

uniformly in $s \in [s_0, \infty)$, for any fixed $s_0 > 0$.

The proof is given in [8] in the case where $\kappa \geq 0$, while it is straightforward to check that the lemma above is still valid when κ is negative. Notice that the condition (iii) of (2.23) says that $\omega(N) = O(1)$ as $N \rightarrow \infty$.

Using Lemma 2.7, as well as (2.22)–(2.24), we have as $N \rightarrow \infty$

$$\int_1^N e^{-\frac{F(N)}{f(x)}s} S_m \left(\frac{F(N)}{f(x)}s \right) dx \sim \frac{1}{(m-1)!} \left[\ln \left(\frac{f(N)}{f'(N)} \right) \right]^{m-2} s^{m-2} \left[\frac{f(N)}{f'(N)} \right]^{1-s}, \tag{2.28}$$

where, as usual, $E_1(N) \sim E_2(N)$ means that $E_1(N)/E_2(N) \rightarrow \infty$ as $N \rightarrow \infty$. From (2.28) we obtain

$$\lim_N \int_1^N e^{-\frac{F(N)}{f(x)}s} S_1 \left(\frac{F(N)}{f(x)}s \right) dx = \begin{cases} \infty, & \text{if } s < 1, \\ 0, & \text{if } s \geq 1, \end{cases} \tag{2.29}$$

$$\lim_N \int_1^N e^{-\frac{F(N)}{f(x)}s} S_2 \left(\frac{F(N)}{f(x)}s \right) dx = \begin{cases} \infty, & \text{if } s < 1, \\ 1, & \text{if } s = 1, \\ 0, & \text{if } s > 1, \end{cases} \tag{2.30}$$

while, for $m \geq 3$

$$\lim_N \int_1^N e^{-\frac{F(N)}{f(x)}s} S_m \left(\frac{F(N)}{f(x)}s \right) dx = \begin{cases} \infty, & \text{if } s \leq 1, \\ 0, & \text{if } s > 1. \end{cases} \tag{2.31}$$

It is easy for one to check that the function $h(x) := e^{-\frac{F(N)}{f(x)}s} S_m \left(\frac{F(N)}{f(x)}s \right)$ is increasing. Hence,

$$\begin{aligned} \int_1^N e^{-\frac{F(N)}{f(x)}s} S_m \left(\frac{F(N)}{f(x)}s \right) dx &\leq \sum_{j=1}^N e^{-\frac{F(N)}{f(j)}s} S_m \left(\frac{F(N)}{f(j)}s \right) \\ &\leq \int_1^N e^{-\frac{F(N)}{f(x)}s} S_m \left(\frac{F(N)}{f(x)}s \right) dx + e^{-\frac{F(N)}{f(N)}s} S_m \left(\frac{F(N)}{f(N)}s \right). \end{aligned}$$

It follows (see (2.24) and (ii) of (2.23)) that the limits in (2.30) and (2.31) are valid, if the integral is replaced by the sum, namely $\sum_{j=1}^N e^{-\frac{F(N)}{f(j)}s} S_m \left(\frac{F(N)}{f(j)}s \right)$. Finally, by the definition of $F(\cdot)$ and the Taylor expansion for the logarithm, namely $\ln(1-x) \sim -x$ as $x \rightarrow 0$, we get

$$\lim_N \sum_{j=1}^N \ln \left(1 - e^{-\frac{F(N)}{f(j)}s} S_1 \left(\frac{F(N)}{f(j)}s \right) \right) = \begin{cases} -\infty, & \text{if } s < 1 \\ 0, & \text{if } s \geq 1, \end{cases} \tag{2.32}$$

$$\lim_N \sum_{j=1}^N \ln \left(1 - e^{-\frac{F(N)}{f(j)}s} S_2 \left(\frac{F(N)}{f(j)}s \right) \right) = \begin{cases} -\infty, & \text{if } s < 1 \\ 0, & \text{if } s > 1 \\ -1, & \text{if } s = 1, \end{cases} \tag{2.33}$$

$$\lim_N \sum_{j=1}^N \ln \left(1 - e^{-\frac{F(N)}{f(j)}s} S_m \left(\frac{F(N)}{f(j)}s \right) \right) = \begin{cases} -\infty, & \text{if } s \leq 1 \\ 0, & \text{if } s > 1, \end{cases} \tag{2.34}$$

for $m = 1, m = 2$, and $m = 3, 4, \dots$ respectively. Next, we take advantage of the above limits. Starting from (2.2), and for any given $\varepsilon \in (0, 1)$ we rewrite $E_m(N; \alpha)$ as

$$E_m(N; \alpha) = F(N) [1 + \varepsilon - I_1(N) - I_2(N) + I_3(N)], \tag{2.35}$$

where

$$I_1(N) := \int_0^{1-\varepsilon} \left[\exp \left\{ \sum_{j=1}^N \ln \left(1 - e^{-\frac{F(N)}{f(j)}s} S_m \left(\frac{F(N)}{f(j)}s \right) \right) \right\} \right] ds, \tag{2.36}$$

$$I_2(N) := \int_{1-\varepsilon}^{1+\varepsilon} \left[\exp \left\{ \sum_{j=1}^N \ln \left(1 - e^{-\frac{F(N)}{f(j)}s} S_m \left(\frac{F(N)}{f(j)}s \right) \right) \right\} \right] ds, \tag{2.37}$$

$$I_3(N) := \int_{1+\varepsilon}^{\infty} \left[1 - \exp \left\{ \sum_{j=1}^N \ln \left(1 - e^{-\frac{F(N)}{f(j)}s} S_m \left(\frac{F(N)}{f(j)}s \right) \right) \right\} \right] ds. \tag{2.38}$$

For typographical convenience we set

$$\delta := \frac{1}{\ln \left(\frac{f(N)}{f'(N)} \right)} = \frac{f(N)}{F(N)} \tag{2.39}$$

(notice that as $N \rightarrow \infty, \delta \rightarrow 0^+$).

Lemma 2.8. *Let $I_1(N), I_2(N), I_3(N)$, and δ be as defined in (2.36)–(2.39) respectively. Then, for any given $\varepsilon \in (0, 1)$ and for all positive integers m we have, as $\delta \rightarrow 0^+$,*

$$I_1(N) = o\left(\delta^{4-m} e^{-\varepsilon/\delta}\right). \tag{2.40}$$

Furthermore,

$$\begin{aligned} I_2(N) &= \varepsilon + (m - 2) \delta \ln \delta + [\ln(m - 1)! - \gamma] \delta + (m - 2)^2 \delta^2 \ln \delta \\ &\quad + [(m - 2) \ln(m - 1)! - (m - 2) \gamma - (m - 1) - \omega(N)(m - 1)!] \delta^2 \\ &\quad + O\left(\delta^3 (\ln \delta)^2\right). \end{aligned} \tag{2.41}$$

and

$$I_3(N) = \frac{(1 + \varepsilon)^{m-2}}{(m - 1)!} \frac{1}{\delta^{m-3}} e^{-\varepsilon/\delta} (1 + O(\delta)) \quad \text{as } \delta \rightarrow 0^+. \tag{2.42}$$

The proof of the lemma is given in the Appendix.

Observation 1. It follows by Lemma 2.8 that both integrals $I_1(N)$ and $I_3(N)$, are negligible compared to the sixth term in the asymptotic expansion of the integral $I_2(N)$. Hence, all the information for the $E[T_m(N)]$, at least for the five first terms, comes from $I_2(N)$.

We are, therefore, ready to present the following theorem.

Theorem 2.9. *Let δ be as defined in (2.39) (hence $\delta \rightarrow 0^+$ as $N \rightarrow \infty$) and $\omega(N)$ as given in (2.27). Then (γ is, as usual, the Euler–Mascheroni constant)*

$$\begin{aligned} E[T_m(N)] &= A_N f(N) \left\{ \frac{1}{\delta} - (m - 2) \ln \delta + [\gamma - \ln(m - 1)!] - (m - 2)^2 \delta \ln \delta \right. \\ &\quad \left. + [(m - 1) + \omega(N)(m - 1)! - (m - 2) \ln(m - 1)! + (m - 2) \gamma] \delta + O(\delta^2 \ln^2 \delta) \right\}. \end{aligned} \tag{2.43}$$

Proof. The result follows immediately by combining (2.4), (2.35), and Lemma 2.8. □

To follow Newman and Shepp [21], although the first set “costs” $A_N f(N)/\delta$, all further sets cost $A_N f(N) \ln \delta$.

2.4.2. Asymptotics of the second rising moment of $T_m(N)$

We will follow a similar approach as in Section 2.4.1, in order to find the sixth(!) term in the asymptotic expansion of the second rising moment of the random variable $T_m(N)$, so that the leading behavior of $V[T_m(N)]$ will be obtained. Let us expand $Q_m(N; \alpha)$ as

$$Q_m(N; \alpha) = 2F(N)^2 \left[\frac{1}{2} + \varepsilon + \frac{\varepsilon^2}{2} - I_4(N) - I_5(N) + I_6(N) \right], \tag{2.44}$$

where

$$I_4(N) := \int_0^{1-\varepsilon} \left[\exp \left\{ \sum_{j=1}^N \ln \left(1 - e^{-\frac{F(N)}{f(j)}s} S_m \left(\frac{F(N)}{f(j)}s \right) \right) \right\} \right] s \, ds, \tag{2.45}$$

$$I_5(N) = \int_{1-\varepsilon}^{1+\varepsilon} \left[\exp \left\{ \sum_{j=1}^N \ln \left(1 - e^{-\frac{F(N)}{f(j)}s} S_m \left(\frac{F(N)}{f(j)}s \right) \right) \right\} \right] s \, ds, \tag{2.46}$$

and

$$I_6(N) = \int_{1+\varepsilon}^{\infty} \left[1 - \exp \left\{ \sum_{j=1}^N \ln \left(1 - e^{-\frac{F(N)}{f(j)}s} S_m \left(\frac{F(N)}{f(j)}s \right) \right) \right\} \right] s \, ds. \tag{2.47}$$

Lemma 2.10. *Let $I_4(N)$, $I_5(N)$, $I_6(N)$, and δ be as defined in (2.45)–(2.47), and (2.39) respectively. Then, for any given $\varepsilon \in (0, 1)$ we have, as $\delta \rightarrow 0^+$,*

$$I_4(N) = o\left(\delta^{4-m} e^{-\varepsilon/\delta}\right). \tag{2.48}$$

Furthermore,

$$\begin{aligned} I_5(N) &= \varepsilon + \frac{\varepsilon^2}{2} + (m-2)\delta \ln \delta + [\ln(m-1)! - \gamma] \delta - \frac{(m-2)^2}{2} \delta^2 \ln^2 \delta \\ &\quad + \left[(m-2)^2 - (m-2)(\ln(m-1)! - \gamma) \right] \delta^2 \ln \delta \\ &\quad + [(m-2)\ln(m-1)! - (m-2)\gamma - \omega(N)(m-1)! - (m-1) \\ &\quad - \frac{1}{2}(\ln(m-1)!)^2 - \frac{1}{2}\left(\gamma^2 + \frac{\pi^2}{6}\right) + \gamma \ln(m-1)!] \delta^2 + O\left(\delta^3 (\ln \delta)^2\right) \end{aligned} \tag{2.49}$$

and

$$I_6(N) = \frac{(1+\varepsilon)^{m-1}}{(m-1)!} \frac{1}{\delta^{m-3}} e^{-\varepsilon/\delta} (1 + O(\delta)). \tag{2.50}$$

The proof of lemma above is given in the Appendix.

Observation 2. By Lemma 2.10 we have that both integrals $I_4(N)$ and $I_6(N)$, are negligible compared to the seventh term in the asymptotic expansion of the integral $I_5(N)$. Hence, all the information regarding $E[T_m(N)(T_m(N)+1)]$, at least for the six first terms, comes from $I_5(N)$.

By combining (2.7), (2.44), and Lemma 2.10 we obtain the following theorem.

Theorem 2.11. *Let δ be as defined in (2.39) (hence $\delta \rightarrow 0^+$ as $N \rightarrow \infty$) and $\omega(N)$ as given in (2.27). Then for all positive integers m*

$$\begin{aligned}
 E [T_m(N) (T_m(N) + 1)] &= A_N^2 f(N)^2 \left\{ \frac{1}{\delta^2} - 2(m-2) \frac{\ln \delta}{\delta} - 2 [\ln(m-1)! - \gamma] \frac{1}{\delta} \right. \\
 &\quad + (m-2)^2 \ln^2 \delta + 2(m-2) [\ln(m-1)! - \gamma - (m-2)] \ln \delta \\
 &\quad + [2(m-2)\gamma - 2(m-2) \ln(m-1)! + 2\omega(N)(m-1)! \\
 &\quad + 2(m-1) + (\ln(m-1)!)^2 + \gamma^2 + (\pi^2/6) - 2\gamma \ln(m-1)!] \\
 &\quad \left. + O(\delta \ln^2 \delta) \right\}. \tag{2.51}
 \end{aligned}$$

We are now ready for our main result regarding the variance (in Case II).

2.4.3. Asymptotics of $V [T_m(N)]$

Theorem 2.12. *Let $\alpha = \{a_j\}_{j=1}^\infty = \{1/f(j)\}_{j=1}^\infty$, where f satisfies (2.22) and (2.23) (hence, $L_1(\alpha; m) = \infty$). Then for all positive integers m we have as $N \rightarrow \infty$*

$$V [T_m(N)] \sim \frac{\pi^2}{6} A_N^2 f(N)^2 = \frac{\pi^2}{6} \cdot \frac{1}{p_N^2} = \frac{\pi^2}{6} \cdot \frac{1}{\min_{1 \leq j \leq N} \{p_j\}^2}, \tag{2.52}$$

where $A_N = \sum_{j=1}^N a_j$ ($p_j = a_j/A_N$ are the coupon probabilities).

Proof. From Theorems 2.9 and 2.11 one gets

$$E [T_m(N)(T_m(N) + 1)] - E [T_m(N)]^2 \sim \frac{\pi^2}{6} A_N^2 f(N)^2 \text{ as } N \rightarrow \infty.$$

In view of (2.8), in order to finish the proof it only remains to show that

$$\frac{E [T_m(N)]}{A_N^2 f(N)^2} \rightarrow 0, \quad N \rightarrow \infty. \tag{2.53}$$

The proof of (2.53) is given in the Appendix. □

Remark 2.13.

(a) It is notable that for the sequences of Case II the leading behavior of the variance of the random variable $T_m(N)$ is *independent* of the value of the positive integer m (which is in agreement with (1.19)). The reader may compare this with (2.16) where the leading behavior of the variance *depends* on m .

(b) Regarding the asymptotics of A_N let us mention that if

$$C_f := \sum_{n=1}^\infty 1/f(n) < \infty,$$

then

$$A_N = C_f [1 + o(1)].$$

On the other hand, if $C_f = \infty$, then, as $N \rightarrow \infty$, we have

$$A_N \sim \int_1^N \frac{dx}{f(x)}.$$

Remark 2.14. Using the techniques presented in this section it can be shown that

$$E [T_m(N)^{(r)}] \sim A_N^r f(N)^r \ln \left(\frac{f(N)}{f'(N)} \right)^r, \quad r \in \mathbb{N} := \{1, 2, \dots\},$$

where, as usual, the coupon probabilities are given by (2.1) with $\alpha = \{a_j\}_{j=1}^\infty = \{1/f(j)\}_{j=1}^\infty$, where f satisfies (2.22) and (2.23).

3. SOME EXAMPLES

Example 1 (The positive power law). Consider the sequence $\alpha = \{j^p\}_{j=1}^\infty$, where $p > 0$. Here we have (see (2.10) and (2.11) and Thm. 2.2),

$$L_1(\alpha; m) = \int_0^\infty \left\{ 1 - \prod_{j=1}^\infty [1 - e^{-j^p t} S_m(j^p t)] \right\} dt < \infty \tag{3.1}$$

and

$$L_2(\alpha; m) = 2 \int_0^\infty \left\{ 1 - \prod_{j=1}^\infty [1 - e^{-j^p t} S_m(j^p t)] \right\} t dt < \infty, \tag{3.2}$$

where $S_m(\cdot)$ is given by (1.7). By Theorem 2.4 it follows that

$$E [T_m(N)] = A_N L_1(\alpha; m) (1 + \delta_N), \tag{3.3}$$

$$E [T_m(N)(T_m(N) + 1)] = A_N^2 L_2(\alpha; m) (1 + \Delta_N), \tag{3.4}$$

where

$$\delta_N = o(1), \quad \Delta_N = o(1) \tag{3.5}$$

and

$$A_N = \sum_{j=1}^N j^p = \frac{N^{p+1}}{p+1} \left[1 + O\left(\frac{1}{N}\right) \right], \quad N \rightarrow \infty. \tag{3.6}$$

Thus, (1.16) yields

$$V [T_m(N)] = \frac{N^{2p+2}}{(p+1)^2} [L_2(\alpha; m) - L_1(\alpha; m)^2] [1 + o(1)] \tag{3.7}$$

as $N \rightarrow \infty$. Actually, by (2.19) and (2.20) of Remark 2.5 we have

$$\delta_N \leq m \sum_{j=N+1}^\infty \frac{1}{j^p} = O\left(\frac{1}{N^{p-1}}\right), \quad \text{if } p > 1 \tag{3.8}$$

and

$$\Delta_N \leq m(m+1) \sum_{j=N+1}^\infty \frac{1}{j^{2p}} = O\left(\frac{1}{N^{2p-1}}\right), \quad \text{if } p > \frac{1}{2} \tag{3.9}$$

(for the equalities in (3.23) and (3.9) see, e.g., [1]). Thus, if $p > 1$, formula (1.16) gives

$$V [T_m(N)] = \frac{N^{2p+2}}{(p+1)^2} [L_2(\alpha; m) - L_1(\alpha; m)^2] \left[1 + O\left(\frac{1}{N^{(p-1) \wedge 1}}\right) \right], \tag{3.10}$$

where $(p-1) \wedge 1 = \min\{p-1, 1\}$.

In the case $p = m = 1$ we can get explicit values for $L_1(\alpha; 1)$ and $L_2(\alpha; 1)$ as well as more accurate estimates for δ_N and Δ_N [5, 8].

Example 2 (The generalized Zipf law). . Here, let us consider the sequence $\alpha = \{1/j^p\}_{j=1}^\infty$, where $p > 0$. Clearly $L_1(\alpha; m) = \infty$. Furthermore, the function

$$f(x) = x^p \tag{3.11}$$

satisfies the conditions of (2.23), thus Theorems 2.9–2.11 can be applied. Formula (2.39) becomes

$$\delta = \frac{1}{\ln\left(\frac{f(N)}{f'(N)}\right)} = \frac{1}{\ln N - \ln p}. \tag{3.12}$$

Hence, Theorems 2.9–2.11 yield

$$E[T_m(N)] = A_N N^p [\ln N + (m - 2) \ln \ln N - \ln p + \gamma - \ln(m - 1)! + o(1)] \tag{3.13}$$

$$E[T_m(N)(T_m(N) + 1)] = A_N^2 N^{2p} (\ln N)^2 \left[1 + O\left(\frac{\ln \ln N}{\ln N}\right)\right], \tag{3.14}$$

and

$$V[T_m(N)] \sim \frac{\pi^2}{6} A_N^2 N^{2p} \tag{3.15}$$

where

$$A_N = \sum_{j=1}^N \frac{1}{j^p}. \tag{3.16}$$

Thus (see, e.g., [1]),

$$A_N = \frac{N^{1-p}}{1-p} + \zeta(p) + O\left(\frac{1}{N^p}\right), \quad \text{if } 0 < p < 1, \tag{3.17}$$

$$A_N = \ln N + \gamma + O\left(\frac{1}{N}\right), \quad \text{if } p = 1, \tag{3.18}$$

and

$$A_N = \zeta(p) - \frac{1}{(p-1)N^{p-1}} + O\left(\frac{1}{N^p}\right), \quad \text{if } p > 1, \tag{3.19}$$

where $\zeta(\cdot)$ is Riemann’s Zeta function (recall that $\zeta(p) < 0$ if $0 < p < 1$). For instance, if $0 < p < 1$, then, in view of (3.17), formulas (3.13) and (3.15) yield

$$E[T_m(N)] = \frac{N}{1-p} [\ln N + (m - 2) \ln \ln N - \ln p + \gamma - \ln(m - 1)!] + O(N^p \ln N) \tag{3.20}$$

and

$$V[T_m(N)] \sim \frac{\pi^2}{6} \frac{N^2}{(1-p)^2} \tag{3.21}$$

respectively. Formula (3.20) should be compared with (1.1)–(1.2); likewise formula (3.21) should be related to formula (1.20) of Conjecture 1.

Example 3 (The exponential law). As in Remark 2.3, let $p > 0$ and consider the sequences $\alpha = \{e^{pj}\}_{j=1}^\infty$ and $\beta = \{e^{-pj}\}_{j=1}^\infty$. We have already observed that α and β produce the same coupon probabilities. We have $L_1(\beta; m) = \infty$. Furthermore $f(x) = e^{px}$ does not satisfy condition (ii) of (2.23), thus Theorems 2.9–2.11 cannot be applied. Let us consider, instead, the sequence α , where we have $L_1(\alpha; m) < \infty$. Here

$$A_N = \sum_{j=1}^N e^{pj} = \frac{e^{p(N+1)} - e^p}{e^p - 1} = \frac{e^{p(N+1)}}{e^p - 1} + O(1), \quad N \rightarrow \infty. \tag{3.22}$$

Also, formulas (2.19) and (2.20) of Remark 2.5 give

$$\delta_N = O(e^{-pN}), \quad \text{and} \quad \Delta_N = O(e^{-2pN}). \tag{3.23}$$

Therefore, Theorem 2.2 yields (as $N \rightarrow \infty$)

$$E [T_m(N)] = \frac{e^{p(N+1)}}{e^p - 1} L_1(\alpha; m) + O(1), \tag{3.24}$$

$$E [T_m(N)(T_m(N) + 1)] = \frac{e^{2p(N+1)}}{(e^p - 1)^2} L_2(\alpha; m) + O(e^{pN}), \tag{3.25}$$

and

$$V [T_m(N)] = \frac{e^{2p(N+1)}}{(e^p - 1)^2} (L_2(\alpha; m) - L_1(\alpha; m)^2) + O(e^{pN}). \tag{3.26}$$

It follows that, regarding the sequence β , the associated asymptotics are also given by (3.24)–(3.26). In this way we get cheaply a counterexample for Theorems 2.9–2.11, in case where $f(\cdot)$ does not satisfy all conditions of (2.23).

4. LIMIT DISTRIBUTIONS

4.1. Case I: $L_1(\alpha; m) < \infty$

Theorem 4.1. *Let $\alpha = \{a_j\}_{j=1}^\infty$ be a sequence such that $L_1(\alpha; m) < \infty$ (recall (2.10) and Thm. 2.2) and, as in Section 2,*

$$p_j = \frac{a_j}{A_N}, \quad \text{where } A_N = \sum_{j=1}^N a_j.$$

Then, for all $s \in [0, \infty)$ we have

$$P \left\{ \frac{T_m(N)}{A_N} \leq s \right\} \rightarrow F(s) := \prod_{j=1}^\infty [1 - S_m(a_j s) e^{-a_j s}], \quad N \rightarrow \infty, \tag{4.1}$$

where $S_m(\cdot)$ is given by (1.7).

Proof. Setting $z = e^\lambda$ with $\Re(\lambda) > 0$, formula (1.17) can be written as

$$E \left[e^{-\lambda T_m(N)} \right] = 1 - (e^\lambda - 1) \int_0^\infty \left\{ 1 - \prod_{j=1}^N [1 - S_m(p_j t) e^{-p_j t}] \right\} \exp(-(e^\lambda - 1)t) dt, \tag{4.2}$$

where $\Re(\lambda) > 0$.

Substituting $t = A_N s$ in the integral of (4.2) we obtain

$$E \left[e^{-\lambda T_m(N)} \right] = 1 - (e^\lambda - 1) A_N \int_0^\infty \left\{ 1 - \prod_{j=1}^N [1 - S_m(a_j s) e^{-a_j s}] \right\} \exp(-(e^\lambda - 1) A_N s) ds,$$

or

$$E \left[e^{-\lambda T_m(N)/A_N} \right] = 1 - (e^{\lambda/A_N} - 1) A_N \int_0^\infty \left\{ 1 - \prod_{j=1}^N [1 - S_m(a_j s) e^{-a_j s}] \right\} \exp(-(e^{\lambda/A_N} - 1) A_N s) ds. \tag{4.3}$$

Finally, in view of Section 2.2 and (4.1) (for the definition of F) dominated convergence gives

$$\lim_N E \left[e^{-\lambda T_m(N)/A_N} \right] = 1 - \lambda \int_0^\infty [1 - F(s)] e^{-\lambda s} ds = \int_0^\infty e^{-\lambda s} dF(s),$$

for all complex λ such that $\Re(\lambda) > 0$. □

Notice that the limit distribution depends on the sequence α .

Remark 4.2. Notice that the function $F(s)$ appearing in (4.1) is the distribution function of the random variable $X = \sup_{j \in \mathbb{N}} X_j$, where X_1, X_2, \dots are independent random variables, such that each X_j is Erlang with parameters m and a_j . It is remarkable that for the *linear case* $a_j = j$, which is of particular interest in applications, we can get other more convenient explicit expressions for $F(s)$ by invoking the celebrated Euler pentagonal-number formula. For more details see [8].

4.2. Case II: $L_1(\alpha; \mathbf{m}) = \infty$

Neal [20] has established a general theorem regarding the limit distribution of $T_m(N)$ (appropriately normalized) as $N \rightarrow \infty$, where $\pi_N = \{p_{N1}, p_{N2}, \dots, p_{NN}\}$, $N = 1, 2, \dots$, is a sequence of (sub)probability measures, not necessarily of the form (2.1).

Theorem N. Suppose that there exist sequences $\{b_N\}$ and $\{k_N\}$ such that $k_N/b_N \rightarrow 0$ as $N \rightarrow \infty$ and that, for $y \in \mathbb{R}$,

$$A_N(y; m) := \frac{b_N^{m-1}}{(m-1)!} \sum_{j=1}^N p_{Nj}^{m-1} \exp \left(-p_{Nj} (b_N + yk_N) \right) \rightarrow g(y), \quad N \rightarrow \infty, \tag{4.4}$$

for a nonincreasing function $g(\cdot)$ with $g(y) \rightarrow \infty$ as $y \rightarrow -\infty$ and $g(y) \rightarrow 0$ as $y \rightarrow \infty$. Then

$$\frac{T_m(N) - b_N}{k_N} \xrightarrow{D} Y, \quad N \rightarrow \infty, \tag{4.5}$$

where Y has distribution function

$$F(y) = P\{Y \leq y\} = e^{-g(y)}, \quad y \in \mathbb{R}. \tag{4.6}$$

Theorem *N* does not indicate at all how to choose the sequences $\{b_N\}$ and $\{k_N\}$. Here our asymptotic formulas can help.

The conclusion (4.5) of Theorem *N* suggests that as $N \rightarrow \infty$

$$b_N \sim E[T_m(N)] \quad \text{and} \quad k_N \sim c\sqrt{V[T_m(N)]}, \quad \text{for some } c \neq 0.$$

Recall that for the Case II the coupon probabilities p_{Nj} , $1 \leq j \leq N$, $N = 1, 2, \dots$, are taken as

$$p_{Nj} = \frac{a_j}{A_N} \quad \text{with} \quad A_N = \sum_{j=1}^N a_j, \quad a_j = \frac{1}{f(j)}, \tag{4.7}$$

where $f(x)$ satisfies (2.23). Then Theorems 2.9 and 2.11 propose the choices

$$b_N = A_N f(N) [\rho(N) + (m-2) \ln \rho(N)] \quad \text{and} \quad k_N = A_N f(N), \tag{4.8}$$

where

$$\rho(N) := 1/\delta = \ln(f(N)/f'(N)) \tag{4.9}$$

(notice that, as $N \rightarrow \infty$ we have that $\rho(N) \rightarrow \infty$, and hence $k_N/b_N \rightarrow 0$ as required). In this case, $A_N(y; m)$ of (4.4) becomes

$$A_N(y; m) = \frac{f(N)^{m-1}}{(m-1)!} [\rho(N) + (m-2) \ln \rho(N)]^{m-1} \tilde{A}_N(y; m), \tag{4.10}$$

where

$$\tilde{A}_N(y; m) := \sum_{j=1}^N \frac{1}{f(j)^{m-1}} \exp\left(-\frac{f(N)}{f(j)} [\rho(N) + (m-2) \ln \rho(N) + y]\right).$$

Since f is increasing and satisfies (2.23) we have for sufficiently large N

$$\begin{aligned} \tilde{A}_N(y; m) &= \int_1^N \frac{1}{f(x)^{m-1}} \exp\left(-\frac{f(N)}{f(x)} [\rho(N) + (m-2) \ln \rho(N) + y]\right) \\ &\quad + O\left(\frac{1}{f(N)^{m-1}} \exp(-[\rho(N) + (m-2) \ln \rho(N) + y])\right). \end{aligned} \tag{4.11}$$

Let us consider the integral

$$\tilde{I}_N(y; m) := \int_1^N \frac{1}{f(x)^{m-1}} \exp\left(-\frac{f(N)}{f(x)} [\rho(N) + (m-2) \ln \rho(N) + y]\right) dx.$$

Integration by parts gives

$$\begin{aligned} \tilde{I}_N(y; m) &= \left[\frac{1}{M} \cdot \frac{f(x)^{3-m}}{f'(x)} \exp\left(-\frac{M}{f(x)}\right) \right]_{x=1}^N \\ &\quad + \int_1^N \frac{f(x)^{2-m}}{M} \left[m-3 + \frac{f''(x)/f'(x)}{f'(x)/f(x)} \right] \exp\left(-\frac{M}{f(x)}\right) dx, \end{aligned} \tag{4.12}$$

where for typographical convenience we have set

$$M := f(N) [\rho(N) + (m-2) \ln \rho(N) + y].$$

By (2.23) it follows that the integral in the right-hand side of (4.12) is $o(\tilde{I}_N(y; m))$ as $N \rightarrow \infty$. The quantity $\frac{1}{M} \cdot \frac{f(1)^{3-m}}{f'(1)} \exp\left(-\frac{M}{f(1)}\right)$ is, also, $o(\tilde{I}_N(y; m))$. Hence, as $N \rightarrow \infty$

$$\tilde{A}_N(y; m) \sim \tilde{I}_N(y; m) \sim \frac{f(N)^{2-m}}{f'(N)} \cdot \frac{\exp(-\rho(N) - (m-2) \ln \rho(N) - y)}{\rho(N) + (m-2) \ln \rho(N) + y}.$$

In view of (4.9) and the fact that $\rho(N) \rightarrow \infty$ as $N \rightarrow \infty$, the above formula becomes

$$\tilde{A}_N(y; m) \sim f(N)^{1-m} \rho(N)^{1-m} e^{-y}.$$

Using the above asymptotics in (4.10) yields

$$A_N(y; m) \rightarrow \frac{e^{-y}}{(m-1)!}, \quad N \rightarrow \infty.$$

Therefore, by invoking Theorem N we obtain the following limit theorem.

Theorem 4.3. *Suppose the coupon probabilities come from a sequence α as in (4.7), where $f(x)$ satisfies (2.23). Then, for all $y \in \mathbb{R}$ we have*

$$P \left\{ \frac{T_m(N) - b_N}{k_N} \leq y \right\} \rightarrow \exp \left(-\frac{e^{-y}}{(m-1)!} \right), \quad N \rightarrow \infty, \tag{4.13}$$

where b_N and k_N are given by (4.8)–(4.9).

Notice that the limiting distribution in (4.13) is independent of the choice of $f(x)$.

Remark 4.4. The fact that for the sequences b_N and k_N of (4.8)–(4.9) the limit $g(y)$ in (4.4) exists and has the right behavior is an independent confirmation that the statements of the Theorems 2.9–2.11 are correct.

Example 4 (continued). For $p > 0$ let us take $f(x) = x^p$, so that $a_j = 1/j^p$. Then, (4.9) becomes

$$\rho(N) = \ln N - \ln p \tag{4.14}$$

and hence b_N and k_N of (4.8) can be taken as

$$b_N = A_N N^p [\ln N + (m-2) \ln \ln N - \ln p] \quad \text{and} \quad k_N = A_N N^p, \tag{4.15}$$

where A_N is given by (3.17)–(3.19) (it is enough to use the leading asymptotic term of A_N). If, in particular, $0 < p < 1$, then formula (4.13) holds with

$$b_N = \frac{N}{1-p} [\ln N + (m-2) \ln \ln N - \ln p] \quad \text{and} \quad k_N = \frac{N}{1-p}. \tag{4.16}$$

This example should be compared with the limiting behavior (1.3) of the case of equal coupon probabilities.

4.3. Slowly decaying sequences

Suppose that our sequence $\alpha = \{a_j\}$ decays to 0 slower than $1/j^p$ for every $p > 0$ (of course, $L_1(\alpha; m) = \infty$). Then, the corresponding function $f(x)$ (for which $a_j = 1/f(j)$) may not satisfy all conditions of (2.23) and, hence, the method presented in Section 2.4 for determining the asymptotics of the expectation and the variance of $T_m(N)$ may not work. Nevertheless, Example 2 together with formulas (1.3)–(1.4) for the case of equal coupon probabilities (*i.e.* when $a_j = \text{constant}$) suggest that the limit distribution of $T_m(N)$, appropriately normalized, should be Gumbel and, furthermore, that the sequences b_N and k_N of Theorem 4.3 should be taken as

$$b_N = N \ln N + (m-1)N \ln \ln N \quad \text{and} \quad k_N = N \tag{4.17}$$

(also, that, as $N \rightarrow \infty$, $V[T_m(N)] \sim (\pi^2/6)N^2$, while $E[T_m(N)] = N \ln N + (m-1)N \ln \ln N + c_m(\alpha)N + o(N)$, where $c_m(\alpha)$ is a constant depending on m and α such that $c_m(\alpha) \geq \gamma - \ln(m-1)!$).

Let us illustrate the above comment with the function $f(x) = (\ln x)^p$, $p > 0$, which does not satisfy conditions (iii) and (iv) of (2.23):

Suppose the coupon probabilities come from the sequence $\alpha = \{a_j = (\ln j)^{-p}\}_{j=2}^\infty$ for some $p > 0$. Then, for all $y \in \mathbb{R}$ we have

$$P \left\{ \frac{T_m(N) - N \ln N - (m-1)N \ln \ln N}{N} \leq y \right\} \rightarrow \exp \left(-\frac{e^{-(y-p)}}{(p+1)(m-1)!} \right) \tag{4.18}$$

as $N \rightarrow \infty$ (the proof of formula (4.18) can be found in the version of the article which appears in [arXiv.org](https://arxiv.org)).

Needless to say that formula (4.18) is equivalent to

$$P \left\{ \frac{T_m(N) - N \ln N - (m-1)N \ln \ln N - [\gamma + p - \ln(p+1) - \ln(m-1)!] N}{N} \leq y \right\} \rightarrow e^{-e^{-y}} \tag{4.19}$$

as $N \rightarrow \infty$.

Finally, let us observe that for $a_j = (\ln j)^{-p}$, $p > 0$, the above example suggests the asymptotic formulas (as $N \rightarrow \infty$)

$$E [T_m(N)] = N \ln N + (m-1)N \ln \ln N + [\gamma + p - \ln(p+1) - \ln(m-1)!] N + o(N), \tag{4.20}$$

and

$$V [T_m(N)] \sim \frac{\pi^2}{6} N^2.$$

Notice that the expected value in (4.20) is slightly bigger than the corresponding expected value for the case of equal coupon probabilities (recall (1.1)–(1.2)), due to the term $p - \ln(p+1)$ which is strictly positive for all $p > 0$.

APPENDIX A.

Here we give the proofs of Theorem 2.2 and some technical lemmas which appeared in Section 2.

Proof of Theorem 2.2. Before proving the theorem we recall the following inequality which can be proved easily by induction and limit:

Let $\{b_j\}_{j=1}^\infty$ be a sequence of real numbers such that $0 \leq b_j \leq 1$, for all j . If $\sum_{j=1}^\infty b_j < \infty$, then

$$1 - \prod_{j=1}^\infty (1 - b_j) \leq \sum_{j=1}^\infty b_j. \tag{A.1}$$

Let us prove the equivalence of (i) and (iii). The equivalence between (ii) and (iii) is similar.

Assume that there is a $\xi \in (0, 1)$ such that (2.12) is true. Then, by (2.10) and (A.1) we have

$$L_1(\alpha; m) \leq \int_0^\xi \sum_{j=1}^\infty (x^{a_j} S_m(-a_j \ln x)) \frac{dx}{x} + \int_\xi^1 \frac{dx}{x}.$$

Using Tonelli’s Theorem we have, in view of (1.7)

$$L_1(\alpha; m) \leq \sum_{j=1}^\infty \sum_{k=0}^{m-1} \left(\frac{(-1)^k a_j^k}{k!} \int_0^\xi x^{a_j-1} (\ln x)^k dx \right) + \ln \xi.$$

For the integral above we have by repeated integration by parts

$$\int_0^\xi x^{a_j-1} (\ln x)^k dx = \frac{1}{a_j} \xi^{a_j} \sum_{i=0}^k (-1)^i (k)_i \frac{1}{a_j^i} (\ln \xi)^{k-i},$$

where $(k)_i = k! / (k - i)!$ is the falling Pochhammer symbol. Hence,

$$L_1(\alpha; m) \leq \sum_{j=1}^\infty \left[\sum_{k=0}^{m-1} \frac{(-1)^k a_j^k}{k!} \left(\frac{1}{a_j} \xi^{a_j} \sum_{i=0}^k (-1)^i (k)_i \frac{1}{a_j^i} (\ln \xi)^{k-i} \right) \right] + \ln \xi.$$

Now, (2.12) implies that $\xi^{a_j} \rightarrow 0$, hence $a_j \rightarrow \infty$. Therefore, $\min_j \{a_j\} = a_{j_0} > 0$. Thus,

$$L_1(\alpha; m) \leq \left(\sum_{j=1}^{\infty} \xi^{a_j} a_j^{m-1} \right) \left[\sum_{k=0}^{m-1} \frac{(-1)^k}{k!} (\ln \xi)^k \left(\sum_{i=0}^k (-1)^i \binom{k}{i} \frac{(\ln \xi)^{-i}}{a_{j_0}^i} \right) \right] + \ln \xi.$$

Since $\xi \in (0, 1)$, (2.12) implies

$$\sum_{j=1}^{\infty} \xi^{a_j} a_j^{m-1} < \infty.$$

It follows that $L_1(\alpha; m) < \infty$. Conversely, if

$$\sum_{j=1}^{\infty} \xi^{a_j} = \infty \quad \text{for all } \xi \in (0, 1)$$

then for any fixed positive integer m we have

$$\sum_{j=1}^{\infty} \xi^{a_j} a_j^{m-1} = \infty, \quad \text{for all } \xi \in (0, 1)$$

and by a standard property of infinite products (see, e.g., [24]) it follows that

$$\prod_{j=1}^{\infty} \left(1 - x^{a_j} S_m(-a_j \ln x) \right) = 0, \quad \text{for all } x \in (0, 1).$$

Hence (2.10) yields $L_1(\alpha; m) = \int_0^1 (dx/x) = \infty$. □

Proof of Lemma 2.8 – PART I (the integral I_1). Regarding the integral of (2.36), given any $\varepsilon \in (0, 1)$ we have

$$\begin{aligned} I_1(N) &:= \int_0^{1-\varepsilon} \exp \left[\sum_{j=1}^N \ln \left[1 - e^{-\frac{F(N)}{f(j)}s} S_m \left(\frac{F(N)}{f(j)}s \right) \right] \right] ds \\ &< \exp \left[- \sum_{j=1}^N e^{-\frac{F(N)}{f(j)}(1-\varepsilon)} S_m \left(\frac{F(N)}{f(j)}(1-\varepsilon) \right) \right] \\ &< \exp \left[- \sum_{k=0}^{m-1} \left[\frac{(1-\varepsilon)^k F(N)^k}{(m-1)!} \left(\sum_{j=1}^N f(j)^{-k} e^{-(1-\varepsilon)\frac{F(N)}{f(j)}} \right) \right] \right], \end{aligned}$$

since $\ln(1-x) < -x$, for $0 < x < 1$. Let us now consider the function

$$g(x) := f(x)^{-k} \exp(-\lambda F(N)/f(x)), \quad x \in [1, N], \quad k = 0, 1, \dots, m-1, \quad \lambda \in (0, 1).$$

It is easy to check that conditions (2.23) imply that for sufficiently large N $g(\cdot)$ is strictly increasing in $[1, N]$. Hence,

$$\int_1^N g(x) dx \leq \sum_{j=1}^N g(j) \leq \int_1^N g(x) dx + g(N).$$

Moreover, by Lemma 2.7 it is easy to see that $g(N) = o\left(\int_1^N g(x) dx\right)$ as $N \rightarrow \infty$. Thus,

$$\sum_{j=1}^N g(j) \sim \int_1^N g(x) dx \quad \text{as } N \rightarrow \infty.$$

Applying Lemma 2.7 for $\kappa = -k$ one arrives at

$$I_1(N) < \exp \left[- \sum_{k=0}^{m-1} \frac{(1-\varepsilon)^k F(N)^k}{(m-1)!} \left[\frac{f(N)^{2-k}}{(1-\varepsilon)F(N)f'(N)} e^{-\frac{F(N)}{f(N)}(1-\varepsilon)} \left(1 + M_1 \frac{f(N)}{F(N)} \right) \right] \right],$$

where M_1 is a positive constant. Using (2.24) and (2.39) *i.e.* the definitions of $F(\cdot)$ and δ , we have

$$\begin{aligned} I_1(N) &< \exp \left[- \sum_{k=0}^{m-1} \frac{(1-\varepsilon)^{k-1}}{(m-1)!} \frac{e^{\varepsilon/\delta}}{\delta^{k-1}} (1 + M_1 \delta) \right] \\ &= \exp \left[- (1-\varepsilon) \frac{\delta^m - (1-\varepsilon)^m}{\delta^m (\delta - (1-\varepsilon))} e^{\varepsilon/\delta} (1 + M_1 \delta) \right]. \end{aligned}$$

Since $\delta \rightarrow 0^+$ and $\varepsilon \in (0, 1)$ we have

$$I_1(N) \ll \delta^{4-m} e^{-\varepsilon/\delta},$$

for sufficiently large N , $m = 1, 2, 3, \dots$ □

Proof of Lemma 2.8 – PART II (the integral I_2). Our first task is to compute a few terms of the asymptotic expansion of the integral of (2.37). For convenience we set

$$B_m(N; s) := \sum_{j=1}^N \ln \left[1 - e^{-\frac{F(N)}{f(j)}s} S_m \left(\frac{F(N)}{f(j)}s \right) \right]. \tag{A.2}$$

Since

$$\frac{F(N)}{f(j)} \rightarrow \infty \quad \text{as } N \rightarrow \infty,$$

and $\ln(1-x) = -x + O(x^2)$ as $x \rightarrow 0$, we have (as long as $s \geq s_0 > 0$)

$$B_m(N; s) = \sum_{j=1}^N \left[-e^{-\frac{F(N)}{f(j)}s} S_m \left(\frac{F(N)}{f(j)}s \right) + O \left(e^{-\frac{2F(N)}{f(j)}s} \left[S_m \left(\frac{F(N)}{f(j)}s \right) \right]^2 \right) \right]. \tag{A.3}$$

Using (1.7), (A.3) yields

$$\begin{aligned} B_m(N; s) &= - \sum_{k=0}^{m-1} \frac{F(N)^k s^k}{k!} \left(\sum_{j=1}^N f(j)^{-k} e^{-\frac{F(N)}{f(j)}s} \right) \\ &\quad + \sum_{j=1}^N O \left(e^{-\frac{2F(N)}{f(j)}s} \left[S_m \left(\frac{F(N)}{f(j)}s \right) \right]^2 \right). \end{aligned} \tag{A.4}$$

Since $f(\cdot)$ is increasing and under conditions (2.23), it follows from the comparison of sums and integrals that for sufficiently large N

$$\sum_{j=1}^N f(j)^{-k} e^{-\frac{F(N)}{f(j)}s} = \int_1^N f(x)^{-k} e^{-\frac{F(N)}{f(x)}s} dx + O \left(f(N)^{-k} e^{-\frac{F(N)}{f(N)}s} \right). \tag{A.5}$$

In view of (A.5) and Lemma 2.7 (for $\kappa = -k$), (A.5) yields (as long as $s \geq s_0 > 0$),

$$B_m(N; s) = - \sum_{k=0}^{m-1} \frac{F(N)^k s^k}{k!} \left[\frac{f(N)^{2-k}}{sF(N)f'(N)} e^{-\frac{F(N)}{f(N)}s} + \omega(N) \frac{f(N)^{3-k}}{s^2F(N)^2f'(N)} e^{-\frac{F(N)}{f(N)}s} \left[1 + O\left(\frac{f(N)}{F(N)}\right) \right] \right]. \tag{A.6}$$

For typographical convenience we set

$$A := \frac{f(N)}{f'(N)}. \tag{A.7}$$

(notice that $A \rightarrow \infty$ as $N \rightarrow \infty$). Using (2.24) and (A.7), (A.6) yields

$$B_m(N; s) = - \frac{1}{(m-1)!} A^{1-s} s^{m-2} (\ln A)^{m-2} - \left(\omega(N) + \frac{1}{(m-2)!} \right) A^{1-s} s^{m-3} (\ln A)^{m-3} \left[1 + O\left(\frac{1}{s \ln A}\right) \right]. \tag{A.8}$$

Hence the quantity (see, (2.37))

$$I_2(N) := \int_{1-\varepsilon}^{1+\varepsilon} e^{B_m(N; s)} ds$$

via the substitutions $s = 1 - t$ and $u = A^t (\ln A)^{m-2}$ (and in view of (1.7)), yields

$$I_2(N) = \delta \int_{\delta^{2-m} \exp(-\varepsilon/\delta)}^{\delta^{2-m} \exp(\varepsilon/\delta)} \exp \left\{ - \frac{1}{(m-1)!} u [1 - \delta \ln u - (m-2) \delta \ln \delta]^{m-2} - \left(\omega(N) + \frac{1}{(m-2)!} \right) u \delta [1 - \delta \ln u - (m-2) \delta \ln \delta]^{m-3} (1 + O(\delta)) \right\} \frac{du}{u},$$

where (see (2.39))

$$\delta := \frac{1}{\ln A} = \frac{1}{\ln \left(\frac{f(N)}{f'(N)} \right)} = \frac{f(N)}{F(N)}$$

(hence, $A \rightarrow \infty$ implies $\delta \rightarrow 0^+$). We have

$$I_2 = \delta \left(\int_{\delta^{2-m} \exp(-\varepsilon/\delta)}^{1/\sqrt{\delta}} + \int_{1/\sqrt{\delta}}^{\delta^{2-m} \exp(\varepsilon/\delta)} \right). \tag{A.9}$$

First we get an upper bound for the second integral of (A.9) as follows:

$$\int_{1/\sqrt{\delta}}^{\delta^{2-m} \exp(\varepsilon/\delta)} \exp \left\{ - \frac{1}{(m-1)!} u (1 - \delta \ln (u \delta^{m-2}))^{m-2} \left[1 + \left(\omega(N) + \frac{1}{(m-2)!} \right) \frac{\delta (1 + O(\delta))}{1 - \delta \ln (u \delta^{m-2})} \right] \right\} \frac{du}{u}, = O\left(\sqrt{\delta} e^{-1/(m-1)! \sqrt{\delta}}\right). \tag{A.10}$$

Let us denote $K_1(\delta)$ the first integral of (A.9). We use the binomial theorem to expand the quantities $[1 - \delta \ln u - (m-2) \delta \ln \delta]^{m-2}$ and $[1 - \delta \ln u - (m-2) \delta \ln \delta]^{m-3}$. Next, we expand the exponentials and get

$$K_1(\delta) = \int_{\delta^{2-m} \exp(-\varepsilon/\delta)}^{1/\sqrt{\delta}} \frac{e^{-u/(m-1)!}}{u} \left\{ 1 + \frac{m-2}{(m-1)!} u \delta \ln (u \delta^{m-2}) + u^2 O\left[\delta \ln (u \delta^{m-2})\right]^2 \right\} \times \left\{ 1 - \left(\omega(N) + \frac{1}{(m-2)!} \right) u \delta (1 + O(\delta)) + u^2 O\left[\delta \ln (u \delta^{m-2})\right]^2 \right\} du$$

(since $e^x = 1 + x + O(x^2)$ as $x \rightarrow 0$). Hence,

$$K_1(\delta) = \int_{\delta^{2-m} \exp(-\varepsilon/\delta)}^{1/\sqrt{\delta}} \frac{e^{-u/(m-1)!}}{u} \left[1 + \frac{m-2}{(m-1)!} u \delta \ln(u \delta^{m-2}) - \left(\omega(N) + \frac{1}{(m-2)!} \right) u \delta (1 + O(\delta)) + u^2 O[\delta \ln(u \delta^{m-2})]^2 \right] du.$$

We split the integral above as

$$K_1(\delta) = \int_{\delta^{2-m} \exp(-\varepsilon/\delta)}^{\infty} - \int_{1/\sqrt{\delta}}^{\infty}. \quad (\text{A.11})$$

However (and this is an easy exercise)

$$\begin{aligned} \int_{1/\sqrt{\delta}}^{\infty} \frac{e^{-u/(m-1)!}}{u} \left[1 + \frac{m-2}{(m-1)!} u \delta \ln(u \delta^{m-2}) - \left(\omega(N) + \frac{1}{(m-2)!} \right) u \delta (1 + O(\delta)) + u^2 O[\delta \ln(u \delta^{m-2})]^2 \right] du \\ = O\left(\sqrt{\delta} e^{-1/(m-1)!\sqrt{\delta}}\right) \quad \text{as } \delta \rightarrow 0^+. \end{aligned} \quad (\text{A.12})$$

It follows that in the expression for $K_1(\delta)$ we can replace the upper limit of the integral by ∞ and therefore as $\delta \rightarrow 0^+$

$$\begin{aligned} I_2(N) = \delta \int_{\delta^{2-m} \exp(-\varepsilon/\delta)}^{\infty} \frac{e^{-u/(m-1)!}}{u} \left[1 + \frac{m-2}{(m-1)!} u \delta \ln(u \delta^{m-2}) - \left(\omega(N) + \frac{1}{(m-2)!} \right) u \delta (1 + O(\delta)) \right. \\ \left. + u^2 O[\delta \ln(u \delta^{m-2})]^2 \right] du. \end{aligned} \quad (\text{A.13})$$

The following asymptotic expansions easy exercises:

$$\int_x^{\infty} \frac{e^{-t}}{t} dt = -\ln x - \gamma + x + O(x^2) \quad \text{as } x \rightarrow 0^+, \quad (\text{A.14})$$

$$\int_x^{\infty} \ln t e^{-t} dt = -\gamma - x \ln x + x + O(x^2 \ln x) \quad \text{as } x \rightarrow 0^+, \quad (\text{A.15})$$

where $\gamma = 0.5772\dots$ is the Euler–Mascheroni constant. Applying (A.14) and (A.15) in (A.13) we get

$$\begin{aligned} I_2(N) = \varepsilon + (m-2) \delta \ln \delta + [\ln(m-1)! - \gamma] \delta + (m-2)^2 \delta^2 \ln \delta \\ + [(m-2) \ln(m-1)! - (m-2) \gamma - (m-1) - \omega(N)(m-1)!] \delta^2 \\ + O(\delta^3 (\ln \delta)^2). \end{aligned}$$

Notice that the error term in the above dominates the terms of (A.10) and (A.12). \square

Proof of Lemma 2.8 – PART III (the integral I_3). Our goal is to compute the leading term of $I_3(N)$. Here we will follow a different approach.

Given $\vartheta \in (0, 1)$, there is a $\eta = \eta(\vartheta)$ such that, for $0 < x < \eta$, we have

$$-(1 + \vartheta)x < \ln(1 - x) < -(1 - \vartheta)x \quad (\text{A.16})$$

and

$$(1 - \vartheta)x < 1 - e^{-x} < (1 + \vartheta)x. \quad (\text{A.17})$$

For $j = 1, \dots, N$, $s \geq 1$, we use the definition of F , conditions (2.23), and (1.7) to get

$$0 < x = e^{-\frac{F(N)}{f(j)}s} S_m \left(\frac{F(N)}{f(j)}s \right) = e^{-\frac{F(N)}{f(j)}s} \sum_{k=0}^{m-1} \frac{1}{k!} \left(\frac{F(N)}{f(j)}s \right)^k \rightarrow 0 \text{ as } N \rightarrow \infty.$$

Hence, for a given $\vartheta \in (0, 1)$, there is $N_0 = N_0(\vartheta)$ such that, for $N \geq N_0$, (A.16) yields

$$\begin{aligned} -(1 + \vartheta) e^{-\frac{F(N)}{f(j)}s} S_m \left(\frac{F(N)}{f(j)}s \right) &< \ln \left[1 - e^{-\frac{F(N)}{f(j)}s} S_m \left(\frac{F(N)}{f(j)}s \right) \right] \\ &< -(1 - \vartheta) e^{-\frac{F(N)}{f(j)}s} S_m \left(\frac{F(N)}{f(j)}s \right), \quad j = 1, \dots, N. \end{aligned}$$

By summing over j and using (A.2) we get

$$-(1 + \vartheta) \sum_{j=1}^N e^{-\frac{F(N)}{f(j)}s} S_m \left(\frac{F(N)}{f(j)}s \right) < B_m(N; s) < -(1 - \vartheta) \sum_{j=1}^N e^{-\frac{F(N)}{f(j)}s} S_m \left(\frac{F(N)}{f(j)}s \right).$$

Using (1.7) we have

$$\sum_{j=1}^N e^{-\frac{F(N)}{f(j)}s} S_m \left(\frac{F(N)}{f(j)}s \right) = \sum_{k=0}^{m-1} \frac{F(N)^k s^k}{k!} \left[\sum_{j=1}^N f(j)^{-k} e^{-\frac{F(N)}{f(j)}s} \right]$$

and from the comparison of sums and integrals (see also (A.5)), we arrive at

$$\begin{aligned} -(1 + \vartheta) \left[f(N)^{-k} e^{-\frac{F(N)}{f(N)}s} + \int_1^N f(x)^{-k} e^{-\frac{F(N)}{f(x)}s} dx \right] &< \sum_{j=1}^N f(j)^{-k} e^{-\frac{F(N)}{f(j)}s} \\ &< -(1 - \vartheta) \int_1^N f(x)^{-k} e^{-\frac{F(N)}{f(x)}s} dx. \end{aligned} \tag{A.18}$$

Hence,

$$\begin{aligned} -(1 + \vartheta) \sum_{k=0}^{m-1} \frac{F(N)^k s^k}{k!} \left[f(N)^{-k} e^{-\frac{F(N)}{f(N)}s} + \int_1^N f(x)^{-k} e^{-\frac{F(N)}{f(x)}s} dx \right] \\ < B_m(N; s) \\ < -(1 - \vartheta) \sum_{k=0}^{m-1} \frac{F(N)^k s^k}{k!} \left[\int_1^N f(x)^{-k} e^{-\frac{F(N)}{f(x)}s} dx \right]. \end{aligned} \tag{A.19}$$

Now, by Lemma 2.7 and from (2.33) and (2.34) we have $B_m(N; s) \rightarrow 0$ as $N \rightarrow \infty$ uniformly in $s \in [1 + \varepsilon, \infty)$, for all positive integers m . Thus, for given $\vartheta > 0$, there exist $N_1 = N_1(\vartheta)$ such that, for $N \geq N_1$, (A.17) gives

$$-(1 - \vartheta) B_m(N; s) < 1 - e^{B_m(N; s)} < -(1 + \vartheta) B_m(N; s).$$

Therefore (see (2.37) and (A.2)),

$$-(1 - \vartheta) \int_{1+\varepsilon}^{\infty} B_m(N; s) ds < I_3(N) < -(1 + \vartheta) \int_{1+\varepsilon}^{\infty} B_m(N; s) ds.$$

Using the bounds of $B(s; N)$ of (A.19) in the above formula we get that for all $N \geq N_2 = \max\{N_0, N_1\}$

$$\begin{aligned}
 & (1 - \vartheta)^2 \int_{1+\varepsilon}^\infty \int_1^N \sum_{k=0}^{m-1} \frac{F(N)^k s^k}{k!} \left[f(x)^{-k} e^{-\frac{F(N)}{f(x)}s} dx \right] ds \\
 & \quad - \vartheta (1 - \vartheta) \sum_{k=0}^{m-1} \frac{F(N)^k}{f(N)^k k!} \left[\int_{1+\varepsilon}^\infty s^k e^{-\frac{F(N)}{f(N)}s} ds \right] \\
 & \qquad \qquad \qquad < I_3(N) \\
 & < (1 + \vartheta)^2 \int_{1+\varepsilon}^\infty \int_1^N \sum_{k=0}^{m-1} \frac{F(N)^k s^k}{k!} \left[f(x)^{-k} e^{-\frac{F(N)}{f(x)}s} dx \right] ds \\
 & \quad + (1 + \vartheta)^2 \sum_{k=0}^{m-1} \frac{F(N)^k}{f(N)^k k!} \left[\int_{1+\varepsilon}^\infty s^k e^{-\frac{F(N)}{f(N)}s} ds \right]. \tag{A.20}
 \end{aligned}$$

Using Lemma 2.7, for $\kappa = -k$ we have

$$\begin{aligned}
 \int_{1+\varepsilon}^\infty s^k \left[\int_1^N f(x)^{-k} e^{-\frac{F(N)}{f(x)}s} dx \right] ds &= \frac{f(N)^{2-k}}{F(N)f'(N)} \int_{1+\varepsilon}^\infty s^{k-1} e^{-\frac{F(N)}{f(N)}s} ds \\
 & \quad + \omega(N) \frac{f(N)^{3-k}}{F(N)^2 f'(N)} \int_{1+\varepsilon}^\infty s^{k-2} e^{-\frac{F(N)}{f(N)}s} \left[1 + O\left(\frac{f(N)}{F(N)}\right) \right] ds.
 \end{aligned}$$

Via the scaling $F(N)s = f(N)u$ and integration by parts we have

$$\begin{aligned}
 \int_{1+\varepsilon}^\infty s^{k-1} e^{-\frac{F(N)}{f(N)}s} ds &= \left(\frac{f(N)}{F(N)}\right)^k \int_{(1+\varepsilon)\frac{F(N)}{f(N)}}^\infty u^{k-1} e^{-u} du \\
 &= (1 + \varepsilon)^{k-1} \frac{f(N)}{F(N)} e^{-(1+\varepsilon)\frac{F(N)}{f(N)}} \left[1 + O\left(\frac{f(N)}{F(N)} e^{-\frac{F(N)}{f(N)}}\right) \right].
 \end{aligned}$$

Hence, (using again, the definition of $F(\cdot)$ and (2.39) we get)

$$\begin{aligned}
 & \sum_{k=0}^{m-1} \frac{F(N)^k}{k!} \left\{ \int_{1+\varepsilon}^\infty s^k \left[\int_1^N f(x)^{-k} e^{-\frac{F(N)}{f(x)}s} dx \right] ds \right\} \\
 &= \frac{f(N)^2}{F(N)^2} \frac{f(N)}{f'(N)} e^{-(1+\varepsilon)\frac{F(N)}{f(N)}} \sum_{k=0}^{m-1} \frac{1}{k!} \left(\frac{F(N)}{f(N)}\right)^k (1 + \varepsilon)^{k-1} \left[1 + O\left(\frac{f(N)}{F(N)} e^{-\frac{F(N)}{f(N)}}\right) \right] \\
 &= \frac{(1 + \varepsilon)^{m-2}}{(m - 1)!} \frac{1}{\delta^{m-3}} e^{-\varepsilon/\delta} (1 + O(\delta)). \tag{A.21}
 \end{aligned}$$

Likewise as $\delta \rightarrow 0^+$

$$\sum_{k=0}^{m-1} \frac{F(N)^k}{k! f(N)^k} \left[\int_{1+\varepsilon}^\infty s^k e^{-\frac{F(N)}{f(N)}s} ds \right] = o\left(\frac{1}{\delta^{m-4}} e^{-\varepsilon/\delta}\right). \tag{A.22}$$

In view of (A.21), (A.22), and since $\vartheta \in (0, 1)$ is arbitrary, (A.20) implies

$$I_3(N) = \frac{(1 + \varepsilon)^{m-2}}{(m - 1)!} \frac{1}{\delta^{m-3}} e^{-\varepsilon/\delta} (1 + O(\delta)) \quad \text{as } \delta \rightarrow 0^+, \quad m = 2, 3, \dots \quad \square$$

Proof of Lemma 2.10. We will discuss briefly, the proof for $I_5(N)$. The proofs for $I_4(N)$ and $I_6(N)$ are similar to the proofs of the results for $I_1(N)$ and $I_3(N)$ respectively of Lemma 2.8. For $I_5(N)$ of (2.46) and in view of (A.2) we have

$$I_5(N) := \int_{1-\varepsilon}^{1+\varepsilon} s e^{B_m(N;s)} ds.$$

We can treat $I_5(N)$ as we treated $I_2(N)$ of Lemma 2.8. One gets (as $N \rightarrow \infty$),

$$\begin{aligned} I_5(N) &= I_2(N) - \delta^2 \int_{\delta^{2-m} \exp(-\varepsilon/\delta)}^{\delta^{2-m} \exp(\varepsilon/\delta)} \ln(u \delta^{m-2}) \\ &\quad \times \exp \left\{ -\frac{1}{(m-1)!} u [1 - \delta \ln(u \delta^{m-2})]^{m-2} \right. \\ &\quad \left. - \left(\omega(N) + \frac{1}{(m-2)!} \right) u \delta [1 - \delta \ln(u \delta^{m-2})]^{m-3} (1 + O(\delta)) \right\} \frac{du}{u}, \end{aligned}$$

We have

$$I_{51}(N) = I_{21}(N) - \delta^2 \left(\int_{\delta^{2-m} \exp(-\varepsilon/\delta)}^{1/\sqrt{\delta}} + \int_{1/\sqrt{\delta}}^{\delta^{2-m} \exp(\varepsilon/\delta)} \right). \tag{A.23}$$

For the second integral of (A.23) one gets an upper bound (see (A.10)), namely $O(\ln \delta \sqrt{\delta} e^{-1/(m-1)! \sqrt{\delta}})$. The first integral of (A.23) is

$$\begin{aligned} K_2(\delta) &:= \int_{\delta^{2-m} \exp(-\varepsilon/\delta)}^{1/\sqrt{\delta}} \ln(u \delta^{m-2}) \exp \left\{ -\frac{1}{(m-1)!} u [1 - \delta \ln(u \delta^{m-2})]^{m-2} \right. \\ &\quad \left. - \left(\omega(N) + \frac{1}{(m-2)!} \right) u \delta [1 - \delta \ln(u \delta^{m-2})]^{m-3} \right. \\ &\quad \left. \times (1 + O(\delta)) \frac{du}{u} \right\}. \end{aligned}$$

If we treat $K_2(\delta)$ as we treated $K_1(\delta)$ of Lemma 2.8, we can replace the upper limit of the integral $K_2(\delta)$ by ∞ . Thus, as $\delta \rightarrow 0^+$,

$$\begin{aligned} I_5(N) &= I_2(N) - \delta^2 \int_{\delta^{2-m} \exp(-\varepsilon/\delta)}^{\infty} \ln(u \delta^{m-2}) \frac{e^{-u/(m-1)!}}{u} \left[1 + \frac{m-2}{(m-1)!} u \delta \ln(u \delta^{m-2}) \right. \\ &\quad \left. - \left(\omega(N) + \frac{1}{(m-2)!} \right) u \delta (1 + O(\delta)) + u^2 O[\delta \ln(u \delta^{m-2})]^2 \right] du. \end{aligned} \tag{A.24}$$

The following asymptotic expansions easy exercises as $x \rightarrow 0^+$:

$$\int_x^\infty \frac{e^{-t}}{t} \ln t \, dt = -\frac{1}{2} \ln^2 x + \frac{1}{2} \left(\gamma^2 + \frac{\pi^2}{6} \right) + O(x \ln x), \tag{A.25}$$

$$\int_x^\infty e^{-t} \ln^2 t \, dt = \left(\gamma^2 + \frac{\pi^2}{6} \right) - x \ln^2 x + O(x \ln x). \tag{A.26}$$

Applying (A.25), (A.26) in (A.24), and using (2.41), one arrives at

$$\begin{aligned}
 I_5(N) &= \varepsilon + \frac{\varepsilon^2}{2} + (m-2)\delta \ln \delta + [\ln(m-1)! - \gamma]\delta - \frac{(m-2)^2}{2}\delta^2 \ln^2 \delta \\
 &\quad + \left[(m-2)^2 - (m-2)(\ln(m-1)! - \gamma) \right] \delta^2 \ln \delta \\
 &\quad + [(m-2)\ln(m-1)! - (m-2)\gamma - \omega(N)(m-1)! - (m-1) \\
 &\quad - \frac{1}{2}(\ln(m-1)!)^2 - \frac{1}{2}\left(\gamma^2 + \frac{\pi^2}{6}\right) + \gamma \ln(m-1)!] \delta^2 + O\left(\delta^3 (\ln \delta)^2\right). \quad \square
 \end{aligned}$$

Proof of formula (2.53). From (2.43) and (2.39) we have

$$E[T_m(N)] \sim A_N f(N) \ln\left(\frac{f(N)}{f'(N)}\right).$$

Due to the above, (2.53) is equivalent to

$$\frac{\ln f(N) - \ln f'(N)}{A_N f(N)} \rightarrow 0, \quad N \rightarrow \infty. \tag{A.27}$$

Using (i) and (ii) of (2.23) it remains to prove that for sufficiently large x

$$\frac{\ln f'(x)}{\ln f(x)} = O(1). \tag{A.28}$$

One arrives at (A.28) starting from (iii) (of (2.23)). There is a positive constant M , such that for sufficiently large x

$$\left|(\ln f'(x))'\right| \leq M |(\ln f(x))'|.$$

Since $f(x), f'(x) > 0$ the above becomes

$$\left|(\ln f'(x))'\right| \leq M (\ln f(x))'.$$

For any fixed $x_0 > 0$ and x sufficiently large, we have

$$\int_{x_0}^x \left|(\ln f'(x))'\right| dx \leq M \int_{x_0}^x (\ln f(x))' dx.$$

Hence,

$$\left| \int_{x_0}^x (\ln f'(x))' dx \right| \leq M \int_{x_0}^x (\ln f(x))' dx,$$

which implies

$$|\ln f'(x) - \ln f'(x_0)| \leq M (\ln f(x) - \ln f(x_0)).$$

If we divide the above inequality with the positive function $\ln f(x)$ and use (i) (of (2.23)) we have the desired result. This completes the proof. \square

APPENDIX B. FINAL COMMENTS

B.1. Comparison with earlier works

The main task of this paper was to obtain the limiting distribution of the random variable $T_m(N)$ (the number of trials a collector needs in order to obtain m complete sets of all N different types of coupons). A key feature in our approach is that for each integer $N > 0$, one can create a probability measure $\pi_N = \{p_1, \dots, p_N\}$ on the set of types $\{1, \dots, N\}$ by taking

$$p_j = \frac{a_j}{A_N}, \quad \text{where } A_N = \sum_{j=1}^N a_j,$$

where $\alpha = \{a_j\}_{j=1}^\infty$ is a given sequence of positive numbers. Under this setup p_j depends on α and N . Thus, given α , it makes sense to consider the asymptotic behavior of the moments and the variance of the random variable $T_m(N)$. Moreover, since the leading term of A_N is, generally, easy to be found, one focuses in the asymptotics of $E_m(N; \alpha)$ and $Q_m(N; \alpha)$ (see formulae (2.3), (2.5)). Theorem 2.2 separates the problem in classes of growing (Case I) and decaying sequences (Case II) α . For Case I the asymptotics of the expected value and the variance of $T_m(N)$ follow easily and *depend* on m (Thm. 2.4). Having those asymptotics, we were able to establish Theorem 4.1, which gives the limiting distribution of $T_m(N)$ (appropriately normalized) as $N \rightarrow \infty$. Notice that the proof of Theorem 4.1 is new, since only for the case $m = 1$ [8] the result follows from known theorems together with the asymptotics of the expected value and the variance.

In Case II we examine sequences $\alpha = \{a_j\}_{j=1}^\infty$ of the form $a_j = f(j)^{-1}$, where $f(\cdot)$ satisfies some rather weak conditions (see (2.22), (2.23)). In order to apply the general Theorem N (see Sect. 4.2), we need to come up with appropriate sequences b_N and k_N . Here, our asymptotics for $E[T_m(N)]$ and $V[T_m(N)]$ indicate specifically how to choose b_N and k_N (see Thm. 4.3). Furthermore, in Section 4.3 we discuss the case where $f(x)$ does not satisfy the conditions of (2.23). Formula (4.18) presented there is completely new. Recall that Erdős and Rényi (1961) [11] proved a limit theorem for the case of equal probabilities. Thus, our paper generalizes that result for a large class of coupon probabilities.

The computation of the asymptotics of $E[T_m(N)]$ and $V[T_m(N)]$ in Case II is quite involved. The heart of our analysis rests in Lemma 2.7, which determines the behavior of the quantity $\sum_{j=1}^N e^{-\frac{F(N)}{f(j)}} S_m(\frac{F(N)}{f(j)}s)$. Hence, it is necessary to rewrite $E_m(N; \alpha)$ and $Q_m(N; \alpha)$ as in (2.35) and (2.44) respectively. It turns out that we have to compute the fifth asymptotic term of $E_m(N; \alpha)$ and the sixth term of $Q_m(N; \alpha)$, so that the leading term of the variance emerges (notice that this term is independent of the number m). In an earlier work [8], the authors established these formulas for $m = 1$. The main difference here comes from the limit $\lim_N \int_1^N e^{-\frac{F(N)}{f(x)}s} S_m(\frac{F(N)}{f(x)}s) dx$, (see (2.29) *versus* (2.30) and (2.31)). This limit causes some difficulties. Its different values (for $m = 1$, $m = 2$, and $m \geq 3$) explain the reason for considering the formulas (2.35) and (2.44).

B.2. Certain issues arising in the construction of coupon probabilities

In this final note we try to re-express in a rigorous way certain statements we have made regarding the triangular sequence of coupon probabilities obtained from a sequence of positive numbers.

Let $\mathcal{T} := \{\tau_1, \tau_2, \dots, \tau_N\}$ be the set of the different types of the members of a population. A probability measure π on \mathcal{T} is an N -tuple

$$\pi = (p_1, p_2, \dots, p_N),$$

where, of course,

$$p_j > 0, \quad j = 1, 2, \dots, N, \quad \text{and} \quad \sum_{j=1}^N p_j = 1$$

(p_j is the probability of being of type τ_j).

We say that two probability measures $\pi = (p_1, p_2, \dots, p_N)$ and $\pi' = (p'_1, p'_2, \dots, p'_N)$ induce equivalent distributions on \mathcal{T} , symbolically

$$\pi \stackrel{d}{=} \pi',$$

if there is a permutation σ on $\{1, 2, \dots, N\}$ such that

$$p'_j = p_{\sigma(j)}, \quad j = 1, 2, \dots, N.$$

Given a measure $\pi = (p_1, p_2, \dots, p_N)$ we denote by $\pi^+ = (p_1^+, p_2^+, \dots, p_N^+)$ the (unique) measure such that $\pi^+ \stackrel{d}{=} \pi$ and $p_1^+ \leq p_2^+ \leq \dots \leq p_N^+$. Moreover, we denote by $\pi^- = (p_1^-, p_2^-, \dots, p_N^-)$ the (unique) measure such that $\pi^- \stackrel{d}{=} \pi$ and $p_1^- \geq p_2^- \geq \dots \geq p_N^-$.

Observe that if $\pi \stackrel{d}{=} \varpi$, then $\pi^+ = \varpi^+$ and $\pi^- = \varpi^-$.

As we have seen, a sequence $\alpha = \{a_j\}_{j=1}^\infty$, where $a_j > 0$ for all j , yields a sequence of probability measures $\{\pi_N(\alpha)\}_{N=1}^\infty$, where

$$\pi_N(\alpha) = (p_{N1}, p_{N2}, \dots, p_{NN}) \quad \text{with} \quad p_{Nj} = \frac{a_j}{A_N}, \quad A_N = \sum_{j=1}^N a_j.$$

If α is increasing, then $\pi_N(\alpha) = \pi_N(\alpha)^+$, while if α is decreasing, then $\pi_N(\alpha) = \pi_N(\alpha)^-$

We now define some equivalence relations between sequences of positive elements, where by “positive” we always mean “strictly positive.”

Definition B.1. Two sequences $\alpha = \{a_j\}_{j=1}^\infty$ and $\beta = \{b_j\}_{j=1}^\infty$ of positive numbers are called equivalent, symbolically

$$\alpha \equiv \beta,$$

if

$$\pi_N(\alpha) \stackrel{d}{=} \pi_N(\beta) \quad \text{for all } N = 1, 2, \dots,$$

while α and β are called essentially equivalent, symbolically

$$\alpha \cong \beta,$$

if there is a positive integer N_0 such that

$$\pi_N(\alpha) \stackrel{d}{=} \pi_N(\beta) \quad \text{for all } N \geq N_0.$$

More generally, if \mathbb{S} is an infinite subset of $\mathbb{N} := \{1, 2, \dots\}$, the sequences α and β are called \mathbb{S} -equivalent, symbolically

$$\alpha \equiv \beta(\mathbb{S}),$$

if

$$\pi_{N_k}(\alpha) \stackrel{d}{=} \pi_{N_k}(\beta) \quad \text{for all } N \in \mathbb{S}.$$

Clearly, if α and β are \mathbb{S} -equivalent and \mathbb{S}' is an infinite subset of \mathbb{S} , then α and β are \mathbb{S}' -equivalent. Two \mathbb{S} -equivalent sequences are essentially equivalent if and only if \mathbb{S}^c (the complement of \mathbb{S}^c) is a finite subset of \mathbb{N} . A typical example of equivalent sequences are the sequences $\alpha = \{a_j\}_{j=1}^\infty$ and $\lambda\alpha := \{\lambda a_j\}_{j=1}^\infty$, where $\lambda > 0$. In particular, $\alpha \equiv a_1^{-1}\alpha$, hence every sequence is equivalent to a sequence whose first element is 1.

Proposition B.2. *Let $\alpha = \{a_j\}_{j=1}^\infty$ and $\beta = \{b_j\}_{j=1}^\infty$ be two increasing sequences of positive numbers. Suppose there is a set $\mathbb{S} \subset \mathbb{N}$ for which α and β are \mathbb{S} -equivalent. Then, there is a $\lambda > 0$ such that $\beta = \lambda\alpha$; in particular, α and β are equivalent.*

Proof. By replacing α by $a_1^{-1}\alpha$ and β by $b_1^{-1}\beta$ we can assume without loss of generality that $a_1 = b_1 = 1$. Let $\mathbb{S} = \{N_1, N_2, \dots\}$, where $\{N_k\}_{k=1}^\infty$ is a strictly increasing sequence of positive integers (obviously every infinite subset \mathbb{S} of \mathbb{N} is of this form). Since $\alpha \equiv \beta(\mathbb{S})$ and both sequences are increasing we must have $\pi_{N_k}(\alpha) = \pi_{N_k}(\alpha)^+ = \pi_{N_k}(\beta)^+ = \pi_{N_k}(\beta)$ for all $k \geq 1$. Therefore,

$$\frac{a_j}{A_{N_k}} = \frac{b_j}{B_{N_k}} \quad 1 \leq j \leq N_k, \quad k = 1, 2, \dots \tag{B.1}$$

For $j = 1$ the above formula implies that $A_{N_k} = B_{N_k}$ for all $k \geq 1$ (since $a_1 = b_1 = 1$). Thus, (B.1) becomes

$$a_j = b_j \quad \text{for all } j \leq N_k,$$

for any fixed N_k . Since $N_k \rightarrow \infty$ we get that $a_j = b_j$ for all $j \in \mathbb{N}$. □

By using essentially the same proof we can show that Proposition B.2 remains true in the case where α and β are decreasing sequences.

Proposition B.3. *Let $\alpha = \{a_j\}_{j=1}^\infty$ and $\beta = \{b_j\}_{j=1}^\infty$ be two sequences of positive numbers such that α is increasing and β is decreasing. If α and β are essentially equivalent, then there is a $\lambda \geq 1$ such that $a_j = a_1\lambda^{j-1}$ and $b_j = b_1\lambda^{1-j}$ for all $j \geq 1$.*

Proof. As in the proof of the previous proposition, we can assume without loss of generality that $a_1 = b_1 = 1$.

Our monotonicity assumptions for α and β imply that there is a positive integer N_0 such that $\pi_N(\alpha) = \pi_N(\alpha)^+ \stackrel{d}{=} \pi_N(\beta)^- = \pi_N(\beta)$ for all $N \geq N_0$. It follows (from the equivalence $\pi_N(\alpha)^+ \stackrel{d}{=} \pi_N(\beta)^-$) that we must have

$$a_{N-j+1} = a_N b_j \quad \text{and} \quad b_{N-j+1} = b_N a_j \quad 1 \leq j \leq N, \tag{B.2}$$

for all $N \geq N_0$. Taking $j = 2$ in the first equation above yields

$$\frac{a_{N-1}}{a_N} = b_2 \quad \text{for all } N \geq N_0,$$

hence, there is a constant $c > 0$ such that

$$a_j = c\lambda^j \quad \text{for all } j \geq N_0, \quad \text{where } \lambda = 1/b_2. \tag{B.3}$$

Next, by putting $j = N - k$ and $j = N - k + 1$ in the second equation of (B.2) we obtain

$$b_{k+1} = b_N a_{N-k} \quad \text{and} \quad b_k = b_N a_{N-k+1} \quad 1 \leq k \leq N - 1,$$

as long as $N \geq N_0$. Therefore, for all N sufficiently large we have

$$\frac{b_{k+1}}{b_k} = \frac{a_{N-k}}{a_{N-k+1}} \quad 1 \leq k \leq N - 1. \tag{B.4}$$

Since (B.3) is valid for all sufficiently large j , formula (B.5) yields

$$\frac{b_{k+1}}{b_k} = \lambda^{-1} \quad \text{for all } k \leq 1 \tag{B.5}$$

and, similarly, $a_{k+1}/a_k = \lambda$ for all $k \geq 1$. □

In general, if α and β are only \mathbb{S} -equivalent, for some $\mathbb{S} \subset \mathbb{N}$ with \mathbb{S}^c infinite, the conclusion of Proposition B.3 may not be true. For example, if we take

$$a_{2k-1} = \frac{1}{b_{2k-1}} = \lambda^{k-1} \quad \text{and} \quad a_{2k} = \frac{1}{b_{2k}} = \alpha \lambda^{k-1}, \quad k = 1, 2, \dots,$$

where $1 < \alpha < \lambda$, then $\alpha \equiv \beta(\mathbb{S})$ with $\mathbb{S} = \{2k : k \in \mathbb{N}\}$.

It will be interesting to characterize the sets \mathbb{S} for which Proposition B.3 is valid under the weaker assumption that $\alpha \equiv \beta(\mathbb{S})$ and, conversely, for a given set \mathbb{S} , to characterize all the possible pairs of monotone sequences α (increasing) and β (decreasing) such that $\alpha \equiv \beta(\mathbb{S})$.

Finally, suppose $\alpha = \{a_j\}_{j=1}^\infty$ is a sequence such that $a_j \rightarrow \infty$. Let

$$N_1 = 1 \quad \text{and} \quad N_{k+1} = \min\{j : a_j > a_{N_k}\}, \quad k \geq 1.$$

Since $a_j \rightarrow \infty$, we have that N_k (exists and) is strictly increasing in k , and hence $N_k \rightarrow \infty$. We define a new sequence $\hat{\alpha} = \{\hat{a}_j\}_{j=1}^\infty$ as follows:

$$\hat{a}_{N_k} = a_{N_k} \quad \text{for all } k \geq 1$$

and if $N_k < j < N_{k+1}$, the elements \hat{a}_j of $\hat{\alpha}$ are obtained by rearranging the elements a_j of α so that $\hat{\alpha}$ is an increasing sequence.

Likewise, if $\beta = \{b_j\}_{j=1}^\infty$ is a sequence (of positive elements) such that $b_j \rightarrow 0$ we can define a new sequence $\check{\beta} = \{\check{b}_j\}_{j=1}^\infty$ as

$$\check{\beta} := \left[(\beta^{-1})^\wedge \right]^{-1},$$

where $\beta^{-1} := \{b_j^{-1}\}_{j=1}^\infty$. Of course, $\check{\beta}$ is a decreasing sequence.

Observe that α and $\hat{\alpha}$ are \mathbb{S}_α -equivalent, with $\mathbb{S}_\alpha = \{N_1, N_2, \dots\}$ and a similar statement holds for the sequences β and $\check{\beta}$. It follows that, if for some sequences $\alpha = \{a_j\}_{j=1}^\infty$, such that $a_j \rightarrow \infty$, and $\beta = \{b_j\}_{j=1}^\infty$, such that $b_j \rightarrow 0$, we have that $\hat{\alpha}$ and $\check{\beta}$ are essentially equivalent (recall the Definition), then by Proposition 2 there is a $\lambda > 1$ such that $\hat{a}_j = a_1 \lambda^{j-1}$ and $\check{b}_j = b_1 \lambda^{1-j}$ for all $j \geq 1$.

Acknowledgements. The authors wish to thank the anonymous reviewer for hers/his constructive comments and suggestions. Furthermore, we would like to thank professor Amir Dembo for reading an earlier version of the manuscript and for making valuable remarks.

REFERENCES

- [1] T.M. Apostol, *Introduction to Analytic Number Theory*. Springer-Verlag, New York (1976).
- [2] A.D. Barbour, L. Holst and S. Janson, Poisson Approximation, *Oxford Studies in Probability - 2*. Clarendon Press, Oxford (1992).
- [3] C.M. Bender and S.A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory*. Springer-Verlag, New York (1999).
- [4] A. Boneh and M. Hofri, The Coupon Collector Problem Revisited—a Survey of Engineering Problems and Computational Methods. *Commun. Stat. Stochastic Models* **13** (1997) 39–66.
- [5] S. Boneh and V.G. Papanicolaou, General Asymptotic Estimates for the Coupon Collector Problem. *J. Comput. Appl. Math.* **67** (1996) 277–289.
- [6] R.K. Brayton, On the asymptotic behavior of the number of trials necessary to complete a set with random selection. *J. Math. Anal. Appl.* **7** (1963) 31–61.
- [7] P. Diaconis and S. Holmes, A Bayesian peek into Feller volume I, Special issue in memory of D. Basu. *Sankhyā Ser. A* **64** (2002) 820–841.
- [8] A.V. Doumas and V.G. Papanicolaou, The Coupon Collector's Problem Revisited: Asymptotics of the Variance. *Adv. Appl. Probab.* **44** (2012) 166–195.
- [9] A.V. Doumas and V.G. Papanicolaou, Asymptotics of the rising moments for the Coupon Collector's Problem. *Electron. J. Probab.* **18** (2012) 1–15.

- [10] R. Durrett, *Probability: Theory and Examples*, *Duxbury Advanced Series*. Brooks/Cole – Thomson Learning, 3rd edn. Belmont, CA, USA (2005).
- [11] P. Erdős and A. Rényi, On a classical problem of probability theory. *Magyar. Tud. Akad. Mat. Kutató Int. Közl.* **6** (1961) 215–220.
- [12] R. Fagin, Asymptotic miss ratios over independent references. *J. Comput. System Sci.* **14** (1977) 222–250.
- [13] W. Feller, *An Introduction to Probability Theory and Its Applications*. Vols. I and II. John Wiley & Sons, Inc., New York (1966).
- [14] L. Flatto, Limit Theorems for Some Random Variables Associated with Urn Models. *Ann. Prob.* **10** (1982) 927–934.
- [15] L. Holst, On Birthday, Collectors', Occupancy and other classical Urn problems. *Int. Statist. Rev.* **54** (1986) 15–27.
- [16] P.R. Jelenković, Asymptotic approximation of the move-to-front search cost distribution and least-recently-used caching fault probabilities. *Ann. Appl. Prob.* **9** (1999) 430–464.
- [17] N. Kaplan, A generalization of a result of Erdős and Rényi. *J. Appl. Prob.* **14** (1977) 212–216.
- [18] K.L. Locey and J.T. Lennon, Scaling laws predict global microbial diversity. *Proc. of Natl. Acad. Sci. USA* **113** (2016) 5970–5975. DOI:10.1073/pnas.1521291113
- [19] H.M. Mahmoud, *Pólya urn models*. CRC Press, New York (2008).
- [20] P. Neal, The Generalised Coupon Collector Problem. *J. Appl. Prob.* **45** (2008) 621–629.
- [21] D.J. Newman and L. Shepp, The double Dixie cup problem. *Amer. Math. Monthly* **67** (1960) 58–61. MR0120672
- [22] S. Ross, *A First Course in Probability*, 8th edn. Pearson Prentice Hall, Pearson Education, Inc., Upper Saddle River, NJ (2010).
- [23] S. Ross, *Introduction to Probability Models*, 10th edn. Elsevier Inc., Burlington, MA (2010).
- [24] W. Rudin, *Real and Complex Analysis*. McGraw-Hill, New York (1987).
- [25] The Dixie Cup Company History. Available at: <http://sites.lafayette.edu/dixiecollection/> (2016).