The logarithmic Zipf law in a general urn problem

Aristides V. Doumas[1] and Vassilis G. Papanicolaou[2]
Department of Mathematics
National Technical University of Athens
Zografou Campus
157 80 Athens, GREECE
[1]aris.doumas@hotmail.com, adou@math.ntua.gr [2]papanico@math.ntua.gr

Abstract

The origin of power-law behavior (also known variously as Zipf's law) has been a topic of debate in the scientific community for more than a century. Power laws appear widely in physics, biology, earth and planetary sciences, economics and finance, computer science, demography and the social sciences. In a highly cited article, Mark Newman [*Contemporary Physics* **46** (2005) 323–351] reviewed some of the empirical evidence for the existence of power-law forms, however underscored that even though many distributions do not follow a power law, quite often many of the quantities that scientists measure are *close* to a Zipf law, and hence are of importance. In this paper we engage a variant of Zipf's law with a general urn problem. A collector wishes to collect m *complete* sets of N distinct coupons. The draws from the population are considered to be independent and identically distributed with replacement, and the probability that a type-j coupon is drawn is denoted by p_j, $j = 1, 2, \cdots, N$. Let $T_m(N)$ the number of trials needed for this problem. We present the asymptotics for the expectation (five terms plus an error), the second rising moment (six terms plus an error), and the variance of $T_m(N)$ (leading term) as $N \to \infty$, when

$$p_j = \frac{a_j}{\sum_{j=2}^{N+1} a_j}, \quad \text{where} \quad a_j = (\ln j)^{-p}, \ p > 0.$$

Moreover, we prove that $T_m(N)$ (appropriately normalized) converges in distribution to a Gumbel random variable. These "log-Zipf" classes of coupon probabilities are not covered by the existing literature and the present paper comes to fill this gap. In the spirit of a recent paper of ours [*ESAIM: Prob. and Stat.* **20** (2016) 367–399] we enlarge the classes for which the *Dixie cup* problem is solved w.r.t. its moments, variance, distribution.

Keywords. Generalized Zipf law; Urn problems; coupon collector's problem; double Dixie cup problem; Gumbel distribution; Laplace method for integrals - determination of higher order terms; Eulerian logarithmic integral.

This provisional PDF is the accepted version. The article should be cited as: ESAIM: PS, doi: 10.1051/ps/2020011

2010 AMS Mathematics Classification. 60F05; 60F99; 60G70.

1 Panoramic view of the problem and motivation

The coupon collector's problem (CCP) is a classic urn problem of probability theory. It refers to a population whose members are of N different *types*. For $j = 1, 2, \cdots, N$, we denote by p_j the probability that a randomly chosen coupon is of type j, where $p_j > 0$ and $\sum_{j=1}^{N} p_j = 1$. The members of the population are sampled independently *with replacement* and their types are recorded. Naturally, the main object of study is the number $T(N)$ of trials needed until all N types are detected (at least once for the first time). The simple case where all p_j's are equal has a long history. It began with A. De Moivre at the eighteenth century and later with P.S. Laplace (see [16], [5]). In the recent years D.J. Newman and L. Shepp studied the more general problem where the collector's goal is to complete m *sets* of all N existing different coupons (still uniformly distributed), [22]. This problem is known as the double *Dixie cup* problem due to a successful marketing policy of the Dixie Cup Company, (see [18]). Let $T_m(N)$ be the number of trials needed for this case. The main result of [22] was the asymptotics of $E\left[\, T_m(N)\,\right]$ for any fixed m as $N \to \infty$. Soon after, P. Erdős and A. Rényi went a step further and determined the limit distribution of $T_m(N)$, see [12].

For the case of unequal coupon probabilities R.K. Brayton (1963) under a quite restrictive assumption of "nearly equal coupon probabilities" employed the formulae (see, [3])

$$E[T_m(N)] = \int_0^\infty \left\{ 1 - \prod_{j=1}^{N} \left[1 - S_m(p_j t) e^{-p_j t}\right] \right\} dt, \tag{1.1}$$

$$E\left[T_m(N)\left(T_m(N)+1\right)\right] = 2\int_0^\infty \left\{ 1 - \prod_{j=1}^{N} \left[1 - S_m(p_j t) e^{-p_j t}\right] \right\} t dt \tag{1.2}$$

and obtained detailed asymptotics for the expectation $E[T_m(N)]$ and the second rising moment $E\left[T_m(N)\left(T_m(N)+1\right)\right]$. Here and in what follows $S_m(y)$ denotes the m-th partial sum of e^y, namely

$$S_m(y) := 1 + y + \frac{y^2}{2!} + \cdots + \frac{y^{m-1}}{(m-1)!} = \sum_{\ell=0}^{m-1} \frac{y^\ell}{\ell!}\,. \tag{1.3}$$

For the case of unequal coupon probabilities and for $m = 1$, general results have been published in [6] and [7], while for general (however fixed) values of m we refer the reader to our recent work [8].

Our motivation arises from [8], hence we will briefly present its setup. 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 $\pi_N = \{p_1, ..., p_N\}$ on the set of types $\{1, ..., N\}$ by taking

$$p_j = \frac{a_j}{A_N}, \qquad \text{where} \quad A_N = \sum_{j=1}^{N} a_j. \tag{1.4}$$

It follows that

$$E\left[\, T_m(N) \,\right] = A_N \int_0^{\infty} \left[1 - \prod_{j=1}^{N} \left(1 - e^{-a_j t}\, S_m\left(a_j t\right) \right) \right] dt, \tag{1.5}$$

$$E\left[\, T_m(N)\left(T_m(N)+1\right) \,\right] = 2\, A_N^2 \int_0^{\infty} t \left[1 - \prod_{j=1}^{N} \left(1 - e^{-a_j t}\, S_m\left(a_j t\right) \right) \right] dt, \tag{1.6}$$

The sequences $\alpha = \{a_j\}_{j=1}^{\infty}$ were separated as follows:

$$\textbf{(Case I)} \qquad \sum_{j=1}^{\infty} e^{-a_j \xi} < \infty \quad \text{for some } \xi > 0,$$

$$\textbf{(Case II)} \qquad \sum_{j=1}^{\infty} e^{-a_j \xi} = \infty \quad \text{for all } \xi > 0.$$

Regarding Case I, both the expectation and the second (rising) moment of $T_m(N)$ are enough to obtain the leading asymptotics of its variance. As for the distribution of $T_m(N)$, for all $s \in [0, \infty)$ one has

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

Examples of sequences falling in this case are $a_j = j^p$, $p > 0$ (for $p = 1$ we have the so-called *linear case*), $b_j = e^{pj}$, $p > 0$, and $c_j = j!$.
Regarding Case II and as in [6], [7] and, mainly, in [8] the authors wrote a_j as

$$a_j = f(j)^{-1}, \tag{1.7}$$

where

$$f(x) > 0 \qquad \text{and} \qquad f'(x) > 0,$$

and assumed that $f(x)$ possesses three derivatives satisfying the following conditions as $x \to \infty$:

$$
\begin{aligned}
&\text{(i)} \quad f(x) \to \infty, \\
&\text{(ii)} \quad \frac{f'(x)}{f(x)} \to 0, \\
&\text{(iii)} \quad \frac{f''(x)/f'(x)}{f'(x)/f(x)} = O\left(1\right), \\
&\text{(iv)} \quad \frac{f'''(x)\, f(x)^2}{f'(x)^3} = O\left(1\right).
\end{aligned} \tag{1.8}
$$

Then, the asymptotics of the expectation of $T_m(N)$ (up to the fifth term), and the asymptotics of its second rising moment (up to the sixth term) are obtained. These results are needed for the leading asymptotics of the variance $V[\,T_m(N)\,]$ to appear. Finally, they proved that the random variable $T_m(N)$ (under the appropriate normalization) converges in distribution to a Gumbel random variable.

Remark 1. Roughly speaking, $f(\cdot)$ belongs to the class of positive and strictly increasing functions, which grow to ∞ (as $x \to \infty$) *slower than exponentials, but faster than powers of logarithms.*

In particular, (*ii*) is a subexponential condition. Conditions (*iii*) (mainly) and (*iv*) interpret the above remark for the growth of $f(\cdot)$. These conditions are satisfied by a variety of commonly used functions. For example,

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

or various convex combinations of products of such functions. The case

$$
f(x) = x^p,\ p > 0
$$

namely, the case where the coupon probabilities are

$$
p_j = \frac{a_j}{\sum_{j=1}^N a_j}, \quad a_j = \frac{1}{j^p}, \quad p > 0 \tag{1.9}
$$

is the so-called *generalized Zipf distribution*, a surprising law of nature, which has attracted the interest of many researchers, mainly due to its application in computer science and linguistics, physics, biology, earth and planetary sciences, economics and finance, computer science, demography and the social sciences. The literature for the Zipf law is extensive see, e.g. the highly cited article of Mark Newman [21], where he reviewed some of the empirical evidence for the existence of power-law forms, and the references therein. It is also worth mentioning the recent work of Locey and Lennon (*Proc. Natl. Acad. Sci. USA* (2016)) on the applications of power-law in biology. With respect to the CCP the standard Zipf distribution (that is the case where $p = 1$) and when $m = 1$,

the asymptotics of the expectation (leading term) of $T_1(N)$, was first studied by Flajolet *et al*, see [14]. Moreover, an interesting variant of the classic CCP namely, when the family of coupon probabilities is the *mixing* of two subfamilies one of which is the *uniform* family, while the other belongs to the standard *Zipf family* has been studied in [9]. For another mixing type CCP problem we refer the interested reader to [10].
To summarize, we have an answer for the asymptotics of the expectation and the second rising moment of $T_m(N)$, as well as the leading asymptotics of the variance $V[\,T_m(N)\,]$, and its limiting distribution for rich classes of coupon probabilities. The question arises naturally: *can we extend the classes of functions* $f(\cdot)$*? What happens if our functions grows as powers of logarithms?*.

2 Discussion and main results

Here, we address the following:

Problem. What can be said about the moments, the variance, and the distribution of the random variable $T_m(N)$, when $f(x) = \ln x$, or more generally when $f(x) = (\ln x)^p$, $p > 0$? In other words what can be said for the case the coupon probabilities satisfy:

$$p_j = \frac{a_j}{\sum_{j=2}^{N+1} a_j}, \quad \text{where} \;\; a_j = (\ln j)^{-p}\,, \;\; p > 0. \tag{2.1}$$

Remark 2. Formulae (1.9) and (2.1) explain the title of this paper. Consider the case $a_j = (\ln j)^{-p}\,,\; p > 0$. Clearly,

$$\sum_{j=2}^{\infty} e^{-\xi(\ln j)^{-p}} = \infty \quad \text{for all} \quad \xi > 0.$$

Therefore, these sequences fall into Case II of the previous section. However, conditions (*iii*) and (*iv*) of (1.8) are *violated*. In view of (2.1) and (1.5) we get

$$E[\,T_m(N)\,] = \left(\sum_{j=2}^{N+1} (\ln j)^{-p} \right) \times \int_0^\infty \left\{ 1 - \prod_{j=2}^{N+1} \left[1 - S_m\left(t\,(\ln j)^{-p} \right) e^{-t\,(\ln j)^{-p}} \right] \right\} dt. \tag{2.2}$$

Remark 3. Since $E[T_m(N+1)]$ has the same asymptotics as $E[T_m(N)]$, there is no loss of information if we replace $(N+1)$ by N in both the sum and the product in (2.2).

The sum $\sum_{j=2}^{N} (\ln j)^{-p}$ in (2.2) is easy to handle. In fact one may easily obtain its full asymptotic expanssion by using the Euler-Maclaurin summation formula, and hence the associated integral $\int_{j=2}^{N} (\ln x)^{-p}\, dx$, and then repeated integration by parts, (see [1]). In particular, for $p = 1$ we get the so-called *offset logarithmic integral* or *Eulerian logarithmic integral*. We get

$$A_N = \sum_{j=2}^{N} \frac{1}{(\ln j)^p} = \frac{N}{(\ln N)^p} + \frac{p\, N}{(\ln N)^{p+1}} + \frac{p\,(p+1)\; N}{(\ln N)^{p+2}} + O\left(\frac{N}{(\ln N)^{p+3}}\right). \quad (2.3)$$

Remark 4. As we will later see (in the proof of Theorem 2.3), even the *third term* of A_N has a contribution in the asymptotics of the average of $T_m(N)$.
The integral appearing in (2.2) is our main task. Our approach lies in three steps.
Step 1 is a change of variables

$$t = g(N)\, s$$

where

$$\lim_N g(N) = \infty.$$

There are maybe infinite choices for $g(N)$, but a convenient one is

$$g(N) = (\ln N)^{p+1},$$

which makes things simpler by invoking (2.3). Thus,

$$E[\,T_m(N)\,] = \left(N \ln N + p\,N + p\,(p+1)\,\frac{N}{\ln N} + O\left(\frac{N}{(\ln N)^2}\right)\right)$$
$$\times \int_0^\infty \left\{1 - \exp\left(\sum_{j=2}^{N} \ln\left[1 - S_m\left(\frac{(\ln N)^{p+1}}{(\ln j)^p}\, s\right) \exp\left(-\frac{(\ln N)^{p+1}}{(\ln j)^p}\, s\right)\right]\right)\right\} ds. \quad (2.4)$$

Step 2 is to obtain the asymptotics (as $N \to \infty$) of the integral

$$J_k(N) := \int_2^N \exp\left(-\frac{(\ln N)^{p+1}}{(\ln x)^p}\, s\right) \frac{dx}{(\ln x)^{kp}}, \quad k = 0, 1, \ldots, m-1,\ \ p > 0. \quad (2.5)$$

Lemma 2.1

$$J_k(N) = N^{1-s}\, (\ln N)^{-kp} \times \left[\frac{1}{1+ps} + \frac{kp}{(1+ps)^2 \ln N} - \frac{p\,(p+1)\,s}{(1+ps)^3 \ln N}\left(1 + O\left(\frac{1}{\ln N}\right)\right)\right],$$

uniformly in $s \in [s_0, \infty)$, *for any fixed* $s_0 > 0$, where the hidden constant in the error term does not depend on s. All the proofs of the paper are gathered in Section 3. For now, we note that the main tool to estimate the integral of Lemma 2.1 above is the Laplace method for integrals for the determination of higher order terms. Hence,

$$\lim_N \int_2^N \exp\Big(-\frac{(\ln N)^{p+1}}{(\ln x)^p}s\Big) S_1\Big(\frac{(\ln N)^{p+1}}{(\ln j)^p}s\Big)dx = \begin{cases} \infty, & \text{if } s<1,\\ (1+p)^{-1}, & \text{if } s=1,\\ 0, & \text{if } s>1, \end{cases} \tag{2.6}$$

while for $m \geq 2$

$$\lim_N \int_2^N \exp\Big(-\frac{(\ln N)^{p+1}}{(\ln x)^p}s\Big) S_m\Big(\frac{(\ln N)^{p+1}}{(\ln j)^p}s\Big)dx = \begin{cases} \infty, & \text{if } \quad s\leq 1,\\ 0, & \text{if } \quad s>1, \end{cases} \tag{2.7}$$

Now from the comparison of sums and integrals it follows that the limits above are valid, if the integral is replaced by the associated sum. Moreover, from the Taylor expansion for the logarithm, namely $\ln(1-x) \sim -x$ as $x \to 0$, one gets the corresponding limits, e.g. for all $m \geq 2$

$$\lim_N \sum_{j=2}^N \ln\Big[1 - S_m\Big(\frac{(\ln N)^{p+1}}{(\ln j)^p}s\Big)\exp\Big(-\frac{(\ln N)^{p+1}}{(\ln j)^p}s\Big)\Big] = \begin{cases} -\infty, & \text{if } s<1\\ 0, & \text{if } s\geq 1. \end{cases} \tag{2.8}$$

The limit above drives us to Step 3. We actually build on a method proposed recently in [6] and [8] which was based on conditions (1.8). It is remarkable that even though conditions (1.8) are here *violated*, the spirit of this method leads to a solution. We will briefly discuss it here and complete the proof in the next section. Let us denote by $\tilde{E}_m(N;\alpha)$ the integral appearing in (2.4). For any given $\varepsilon \in (0,1)$ one has

$$\tilde{E}_m(N;\alpha) = [\,1+\varepsilon - I_1(N) - I_2(N) + I_3(N)\,], \tag{2.9}$$

where

$$I_1(N) := \int_0^{1-\varepsilon} e^{M_m(N;s)}\,ds, \tag{2.10}$$

$$I_2(N) := \int_{1-\varepsilon}^{1+\varepsilon} e^{M_m(N;s)}\,ds, \tag{2.11}$$

$$I_3(N) := \int_{1+\varepsilon}^{\infty} 1 - e^{M_m(N;s)}\,ds, \tag{2.12}$$

and

$$M_m(N;s) := \sum_{j=2}^N \ln\Big[1 - S_m\Big(\frac{(\ln N)^{p+1}}{(\ln j)^p}s\Big)\exp\Big(-\frac{(\ln N)^{p+1}}{(\ln j)^p}s\Big)\Big]. \tag{2.13}$$

The heart of Step 3 is that $I_3(N)$ and $I_1(N)$ are dominated by the sixth term in the asymptotics of $I_2(N)$ as $N \to \infty$. Intuitively one expects that the main contribution of $\tilde{E}_m(N;\alpha)$ should come from $I_2(N)$ (due to the limit of (2.8). As it turns out $I_2(N)$ is much more important. The analysis of $I_2(N)$ lies in Lemma 2.1(critical contribution), as well as in classical techniques of asymptotic analysis. The computations needed are often quite involved.

Theorem 2.2 ***(Main result)***
Let $T_m(N)$ the number of trials a collector needs to complete m sets of N different types of coupons with replacement. Assume that the coupon probabilities satisfy

$$p_j = \frac{a_j}{\sum_{j=2}^{N} a_j}, \quad \textit{where} \quad a_j = (\ln j)^{-p}\,, \ \ p > 0.$$

Then, for all $y \in \mathbb{R}$ and for all positive integers m we have as $N \to \infty$

$$P\left\{\frac{T_m(N) - N\ln N - (m-1)N\ln\ln N - c(p)N}{N} \leq y\right\} \to e^{-e^{-y}}, \tag{2.14}$$

where, $c(p) := \gamma + p - \ln\Bigg((p+1)(m-1)!\Bigg)$. Hence, the random variable $T_m(N)$ (under the normalization above) converges in distribution to a Gumbel random variable.

To obtain the distribution of the random variable $T_m(N)$ we need information for the asymptotics of its expectation and variance in order to take advantage of a well known and very general limit theorem ([20], see Section 3). It turns out that we need five terms of the asymptotics of the first moment and six terms of the asymptotics of the second moment to have leading asymptotics of the variance $V\,[\,T_m(N)\,]$. We describe these results in the following theorems.

Theorem 2.3 *Let $T_m(N)$ the random variable of Theorem 2. Then, the asymptotics of the average of $T_m(N)$ (as $N \to \infty$) satisfy*

$$\begin{aligned} E\,[\,T_m(N)\,] = N\ln N &+ (m-1)\,N\ln\ln N + [\,p + \gamma - \ln(m-1)! - \ln(p+1)\,]\;N \\ &- (m-1)\left[\frac{p}{p+1} - (m-1) - p\right]\frac{\ln\ln N}{\ln N}\,N \\ &+ \Bigg[p\,(p+1) - p\,\Bigg(\ln(m-1)! + \ln(p+1) - \gamma\Bigg) \\ &\quad - \Bigg(\frac{p}{p+1} - (m-1)\Bigg) \times [\gamma - \ln(m-1)! - \ln(p+1) \\ &\quad - \frac{1}{(p+1)^2}\left(\frac{m-1}{p+1} - \frac{p+1}{p} - 3\left(\frac{p}{p+1}\right)^2\right)\Bigg]\Bigg]\,\frac{N}{\ln N} \\ &+ O\left(\frac{(\ln\ln N)^2}{(\ln N)^2}\,N\right), \end{aligned} \tag{2.15}$$

where γ is the Euler-Mascheroni constant.

Remark 5. Notice that the expected value in (2.15) is slightly bigger than the corresponding expected value for the case of equal coupon probabilities (compare with the results of [12]) due to the term $p-\ln(p+1)$ which is strictly positive for all $p>0$. This is in accordance with the statement: For fixed positive integers m and N, the case of equal probabilities has the property that it is the one with the stochastically smallest $T_m(N)$. This result is due to [19].

Theorem 2.4 *For the second (rising) moment of the random variable $T_m(N)$ we have the following asymptotic expression as $N\to\infty$*

$$\begin{aligned}
E\left[\,T_m(N)\,(T_m(N)+1)\,\right] &= N^2\,(\ln N)^2 + 2\,(m-1)\,N^2 \ln N\,(\ln\ln N) \\
&\quad + 2\left[\,p+\gamma-\ln(m-1)!-\ln(p+1)\,\right]\,N^2\ln N \\
&\quad + (m-1)^2\,N^2\,(\ln\ln N)^2 \\
&\quad - 2\,(m-1)\left(\frac{p}{p+1}-(m-1)-\gamma-2p\right. \\
&\qquad\qquad \left. +\ln(m-1)!+\ln(p+1)\right)N^2\ln\ln N \\
&+ N^2\left[p^2+2p\,(p+1)-2\,(2p+\gamma)\,\Big(\ln(m-1)!+\ln(p+1)\Big)\right. \\
&\qquad +4p\gamma-\Big(\ln(m-1)!+\ln(p+1)\Big)^2+\gamma^2+\frac{\pi^2}{6} \\
&\qquad -2\left(\frac{p}{p+1}-(m-1)\right)\times\left[\gamma-\ln(m-1)!-\ln(p+1)\right. \\
&\qquad \left.\left. -\frac{1}{(p+1)^2}\left(\frac{m-1}{p+1}-\frac{p+1}{p}-3\left(\frac{p}{p+1}\right)^2\right)\right]\right] \\
&+ O\left(\frac{(\ln\ln N)^2}{\ln N}\,N^2\right).
\end{aligned} \tag{2.16}$$

Theorem 2.5 *Let $T_m(N)$ the number of trials a collector needs to complete m sets of N different types of coupons with replacement (m is a fixed positive integer). When the coupon probabilities satisfy*

$$p_j=\frac{a_j}{\sum_{j=2}^{N}a_j}, \quad \textit{where} \quad a_j=(\ln j)^{-p}\,,\ \ p>0$$

we have as $N\to\infty$

$$V\left[\,T_m(N)\,\right]\sim\frac{\pi^2}{6}\,N^2 \tag{2.17}$$

independently of the value of the positive integer m.

2.1 Final comments

The main purpose of this paper is to enlarge the classes of distributions for which we have an answer to the general Dixie cup problem (w.r.t. the average, the

variance and the limiting distribution), by engaging the *logarithmic-Zipf* coupon probabilities. Recall that under the notation $a_j = f(j)^{-1}$ we have information for functions $f(\cdot)$ growing *slower than exponentials but faster than powers of logarithms*, [8].
Since the full asymptotic expansion of $\sum_{j=2}^{N} (\ln j)^{-p}$ is easy to obtain (see (2.2) and (2.3)) our approach is analytic (continuous). We approximate sums by integrals. For example, a key formula is (2.13), which is valid for $m \geq 2$:

$$\lim_N \sum_{j=2}^{N} \ln \left[1 - S_m\left(\frac{(\ln N)^{p+1}}{(\ln j)^p} s \right) \exp\left(-\frac{(\ln N)^{p+1}}{(\ln j)^p} s \right) \right] = \begin{cases} -\infty, & \text{if } s < 1 \\ 0, & \text{if } s \geq 1. \end{cases}$$

The analysis of the corresponding integrals is complicated. We build on a method proposed in previous works of ours even though the original conditions are violated and one would expect that this approach does not *garantee* a path to a solution. A key ingredient which is used here is the application of the Laplace method for the determination of higher order terms, for the asymptotic expansion of integrals. We believe that this method (or something similar) could be valuable for future researchers in order to further *enlarge* the classes of distributions for this problem.
Let us set

$$E\left[T_m(N) \left(T_m(N) + 1 \right) \cdots \left(T_m(N) + r - 1 \right) \right]$$

the r-th rising moment of $T_m(N)$. Following the steps of Theorem 2.3 one easily has the leading term asymptotically, namely

$$E\left[T_m(N) \left(T_m(N) + 1 \right) \cdots \left(T_m(N) + r - 1 \right) \right] \sim N^r \left(\ln N \right)^r, \ N \to \infty.$$

To get deeper in the asymptotics we notice that in view of (2.11) and (3.15) (see Section 3), the key integral for the r rising moment of $T_m(N)$ should be

$$I(N) := \int_{1-\varepsilon}^{1+\varepsilon} s^{r-1} \, e^{M_m(N;s)} \, ds,$$

where

$$M_m(N; s) := \sum_{j=2}^{N} \ln \left[1 - S_m\left(\frac{(\ln N)^{p+1}}{(\ln j)^p} s \right) \exp\left(-\frac{(\ln N)^{p+1}}{(\ln j)^p} s \right) \right].$$

We note that the computations regarding this higher order term asymptotic analysis are very much involved.

3 Proofs

Proof of Lemma 2.1. From (2.5) we easily have

$$J_k(N) = \int_{\ln 2}^{\ln N} \exp\left(-\frac{(\ln N)^{p+1}}{y^p} s \right) \frac{e^y}{y^{kp}} dy.$$

The substitution $y = \left(s^{1/(p+1)} \ln N\right) t$ yields

$$J_k(N) = \frac{s^{\frac{1-kp}{p+1}}}{(\ln N)^{kp-1}} \int_{a_N\, s^{-1/(p+1)}}^{s^{-1/(p+1)}} \exp\left(s^{1/(p+1)} \ln N \left(t - t^{-p}\right) \right) \frac{dt}{t^{kp}}. \tag{3.1}$$

where $a_N = \ln 2/\ln N$. For convenience we set

$$\tilde{J}_k(N) := \int_{a_N\, s^{-1/(p+1)}}^{s^{-1/(p+1)}} \exp\left(s^{1/(p+1)} \ln N \left(t - t^{-p}\right) \right) \frac{dt}{t^{kp}}. \tag{3.2}$$

Now as long as $s \geq s_0 > 0$ for any fixed s_0, we have

$$\lim_N s^{1/(p+1)} \ln N = \infty, \quad \text{for all } p > 0.$$

Moreover, the function

$$\phi(t) := t - t^{-p}$$

is strictly increasing and attains its maximum value at $t_0 = s^{-1/(p+1)}$. In particular, $\phi'(t_0) = 1 + ps > 0$. Hence, only the immediate neighborhood of t_0 contributes to the *full* asymptotic expansion of $\tilde{I}_k(N)$. Let us set

$$h(t) := t^{-kp}.$$

By applying the standard analysis of Laplace integrals one has easily the leading term of the integral of (3.2), see e.g., [4]. One has as $N \to \infty$

$$\tilde{J}_k(N) \sim h(t_0) \frac{\exp\left(s^{1/(p+1)} \ln N \phi(t_0)\right)}{s^{1/(p+1)} \ln N \phi'(t_0)}.$$

By invoking (3.1) we get as $N \to \infty$

$$J_k(N) \sim \left(\frac{1}{1+ps}\right) N^{1-s} (\ln N)^{-kp}.$$

To get higher order terms in the asymptotic expansion of (3.2) (and finally, of (3.1)) we follow [1] (pp. 272-274, alternatively we refer the interested reader to [4], pp. 66-69), and replace $\phi(t)$ by $\phi(t_0) + (t-t_0)\phi'(t_0) + \frac{1}{2}(t-t_0)^2\phi''(t_0)$ and $h(t)$ by $h(t_0) + (t-t_0)h'(t_0) + \frac{1}{2}(t-t_0)^2 h''(t_0)$. Then,

$$\begin{aligned}\tilde{J}_k(N) \sim \int_{t_0-\epsilon}^{t_0} &\left[h(t_0) + (t-t_0)h'(t_0) + \frac{1}{2}(t-t_0)^2 h''(t_0)\right] \\ &\times \exp\left(s^{1/(p+1)} \ln N \left[\phi(t_0) + (t-t_0)\phi'(t_0) + \frac{1}{2}(t-t_0)^2\phi''(t_0)\right]\right) dt.\end{aligned}$$

Next, we use the Taylor expansion of the term

$$\exp\left[s^{1/(p+1)} \ln N \frac{1}{2}(t-t_0)^2 \phi''(t_0)\right],$$

since ϵ may be chosen small enough. Substituting this expansion in the above, then collecting powers of $(t-t_0)$, and finally, extending the range of integration to $(-\infty, t_0]$, yields

$$\tilde{J}_k(N) \sim e^{s^{1/(p+1)}\ln N\,\phi(t_0)} \int_{-\infty}^{t_0} e^{s^{1/(p+1)}\ln N\,(t-t_0)\,\phi'(t_0)}$$
$$\times \left[h(t_0) + (t-t_0)h'(t_0) + \frac{1}{2}(t-t_0)^2 \Big(h''(t_0) + s^{1/p+1}\ln N\, h(t_0)\,\phi''(t_0) \Big) + \cdots \right] dt.$$

and the proof completes the evaluation of the above integral.

Proof of Theorem 2.3. To analyze (2.9) we start from $I_2(N)$ (see (2.11)) and obtain the five first terms in its asymptotic expansion (plus an error). Then we calculate the leading term of $I_3(N)$ and prove that is negligible compared to the sixth term of $I_2(N)$ as $N \to \infty$. Finally, we estimate the leading term of $I_1(N)$, for which we will see that is negligible compared to the leading term of $I_3(N)$. Since

$$S_m(xs)\exp(-xs) \to 0 \quad \text{as } x \to +\infty,$$

meaning that, for all $\epsilon > 0$, there exists x_ϵ such that

$$S_m(xs)\exp(-xs) \le \epsilon, \quad \textit{for all} \quad x \ge x_\epsilon.$$

For all $j \in \{2, \cdots, N\}$ we clearly have

$$(\ln N)^{p+1}/\,(\ln j)^p \ge \ln N.$$

This implies that for $N \ge e^{x_\epsilon}$, one has $(\ln N)^{p+1}/\,(\ln j)^p \ge \ln N \ge x_\epsilon$. Hence,

$$c_j(N) := S_m\left(\frac{(\ln N)^{p+1}}{(\ln j)^p}s\right)\exp\left(-\frac{(\ln N)^{p+1}}{(\ln j)^p}s\right) \le \epsilon$$

that is, for all $j \in \{2, \cdots, N\}$ we have

$$c_j(N) \le \sup_{2\le j\le N} c_j(N) = o(1).$$

Moreover, $\sum_{j=2}^N \log(1-c_j) = -\sum_{j=2}^N c_j + O(c_j^2)$ when $N \to \infty$, if $c_j \to 0$ when $j \to \infty$. It follows from (2.13) that

$$\begin{aligned} M_m(N;s) = &-\sum_{k=0}^{m-1} \frac{(\ln N)^{k(p+1)}\ s^k}{k!} \left(\sum_{j=2}^{N} (\ln j)^{-kp} \exp\left(-\frac{(\ln N)^{p+1}}{(\ln j)^p}s \right) \right) \\ &+ \sum_{j=2}^{N} O\left(e^{-\frac{2(\ln N)^{p+1}}{(\ln j)^p}s} \left[S_m\left(\frac{(\ln N)^{p+1}}{(\ln j)^p}s \right) \right]^2 \right), \end{aligned} \tag{3.3}$$

as long as $s \ge s_0 > 0$. Let

$$h(x) := (\ln x)^{-kp} \exp\left(-\frac{(\ln N)^{p+1}}{(\ln x)^p}s \right), \;\; x \in [2, N], \;\; k = 0, 1, \cdots, m-1, \;\; s > 0.$$

It is easy to check that for sufficiently large N, $h(\cdot)$ is strictly decreasing in $[2, N]$. Hence,

$$\int_2^{N+1} h(x)dx \leq \sum_{j=2}^{N} h(j) \leq h(2) + \int_2^{N} h(x)dx.$$

From the comparison of sums and integrals we have

$$\begin{aligned} M_m(N;s) = -\sum_{k=0}^{m-1} \frac{(\ln N)^{k(p+1)}\ s^k}{k!} \left(\int_2^N (\ln 2)^{-kp} \exp\left(-\frac{(\ln N)^{p+1}}{(\ln x)^p} s \right) dx \right. \\ \left. + O\left(\exp\left(-\frac{(\ln N)^{p+1}}{(\ln 2)^p} s \right) \right) \right) \\ + \sum_{j=1}^{N} O\left(e^{-\frac{2(\ln N)^{p+1}}{(\ln j)^p} s} \left[S_m\left(\frac{(\ln N)^{p+1}}{(\ln j)^p} s \right) \right]^2 \right). \end{aligned} \tag{3.4}$$

Moreover, we have

$$\begin{aligned} &\sum_{j=1}^{N} O\left(e^{-\frac{2(\ln N)^{p+1}}{(\ln j)^p} s} \left[S_m\left(\frac{(\ln N)^{p+1}}{(\ln j)^p} s \right) \right]^2 \right) \\ &= e^{-2s \ln N} \sum_{j=1}^{N} O\left(\left[S_m\left(\frac{(\ln N)^{p+1}}{(\ln j)^p} s \right) \right]^2 \right) \\ &= e^{-2s \ln N} \sum_{j=1}^{N} O\left(\left[\sum_{k=1}^{m-1} \frac{1}{k!} \frac{(\ln N)^{k(p+1)}}{(\ln j)^{kp}} s^k \right]^2 \right). \end{aligned}$$

The main contribution comes from the last term of the inner sum, namely $\frac{1}{k!}\frac{(\ln N)^{k(p+1)}}{(\ln j)^{kp}} s^k$. By expanding the square of the bracket above and using (2.3) we have

$$\begin{aligned} &\sum_{j=1}^{N} O\left(e^{-\frac{2(\ln N)^{p+1}}{(\ln j)^p} s} \left[S_m\left(\frac{(\ln N)^{p+1}}{(\ln j)^p} s \right) \right]^2 \right) \\ =&N^{-2s} \frac{(\ln N)^{2(m-1)(p+1)}}{[(m-1)!]^2} s^{2(m-1)} O\left(\sum_{j=2}^{N} \frac{1}{(\ln j)^{2p(m-1)}} \right) \\ =&\frac{s^{2(m-1)}}{[(m-1)!]^2} N^{1-2s} (\ln N)^{2(m-1)} \left(1 + O\left(\frac{N}{\ln N} \right) \right). \end{aligned} \tag{3.5}$$

Since s is *strictly* positive, we are able to apply Lemma 2.1, and by invoking (3.4) and (3.5), and get

$$M_m(N;s) = -N^{1-s} \sum_{k=0}^{m-1} \frac{(\ln N)^k\ s^k}{k!} \left[\frac{1}{1+ps} + \frac{kp}{(1+ps)^2 \ln N} \right.$$
$$\left. - \frac{p\,(p+1)\,s}{(1+ps)^3 \ln N} \left(1 + O\left(\frac{1}{\ln N} \right) \right) \right]. \tag{3.6}$$

Next, we substitute (3.6) into (2.11)) and apply the change of variables $s = 1-t$. Thus,

$$I_2(N) =$$
$$\int_{-\varepsilon}^{\varepsilon} \exp \left\{ -N^t\,(\ln N)^{m-1}\,(1-t)^{m-1}\,\frac{(1-b)}{(m-1)!} \sum_{n=0}^{\infty} (b\,t)^n + (\ln N)^{m-2}\,\frac{(1-t)^{m-2}}{(m-1)!} \right.$$
$$\times \left[(m-1)\,(1-b) \sum_{n=0}^{\infty} (b\,t)^n + (m-1)\,(1-b)\,(1-t) \sum_{n=1}^{\infty} n b^n t^{n-1} \right.$$
$$\left.\left. - \frac{1-b}{2b}\,(1-t)^2 \sum_{n=2}^{\infty} n\,(n-1)\,b^n t^{n-2} \left(1 + O\left(\frac{1}{\ln N} \right) \right) \right] \right\} dt,$$

where

$$b = \frac{p}{p+1}, \tag{3.7}$$

and we have used that

$$(1-bt)^{-1} = \sum_{n=0}^{\infty} (b\,t)^n\,,\ \ (1-bt)^{-2} = b^{-1} \sum_{n=1}^{\infty} n b^n\, t^{n-1},$$
$$(1-bt)^{-3} = 2b^{-2} \sum_{n=2}^{\infty} n\,(n-1)\,b^n\, t^{n-2},$$

since $\varepsilon \in (0,1)$, $b \in (0,1)$, and $t \in [-\varepsilon, \varepsilon]$. If we change the variables as

$$N^t = u\,\omega_N^{m-1},$$

where

$$\omega_N := (\ln N)^{-1}\,,$$

and apply the binomial theorem, after some careful computations we get

$$I_2(N) = \omega_N \int_{\omega_N^{1-m} \exp(-\varepsilon/\omega_N)}^{\omega_N^{1-m} \exp(\varepsilon/\omega_N)} \exp \left\{ -\frac{(1-b)\,u}{(m-1)!} \left[1 + (b-(m-1))\,\omega_N\, \ln\left(u\omega_N^{m-1}\right) \right.\right.$$
$$\left.\left. + O\left(\omega_N\, \ln\left(u\omega_N^{m-1}\right)\right)^2 \right] \right\}$$
$$\times \exp \left\{ -\frac{\omega_N\, u}{(m-1)!} \left[d_1 + O\left(\omega_N\, \ln\left(u\omega_N^{m-1}\right)\right) \right] \right\} \frac{du}{u},$$

where

$$d_1 = \left(1-b^2\right)(m-1) - \frac{1-b}{b} - 3b^2\,(1-b)\,. \tag{3.8}$$

Notice that, $N \to \infty$ implies $\omega_N \to 0^+$. We **claim** that we can replace the upper limit in the above expression by ∞. Let us rewrite $I_2(N)$ as

$$I_2(N) = \omega_N \left(\int_{\omega_N^{1-m}\exp(-\varepsilon/\omega_N)}^{1/\sqrt{\omega_N}} + \int_{1/\sqrt{\omega_N}}^{\omega_N^{1-m}\exp(\varepsilon/\omega_N)} \right). \tag{3.9}$$

For the second integral of (3.9) we have

$$\begin{aligned}
&\int_{\omega_N^{1-m}\exp(-\varepsilon/\omega_N)}^{\omega_N^{1-m}\exp(\varepsilon/\omega_N)} \exp\left\{-\frac{(1-b)\,u}{(m-1)!}\left[1+(b-(m-1))\,\omega_N\,\ln\left(u\omega_N^{m-1}\right)\right.\right.\\
&\qquad\qquad \left.\left. +O\left(\omega_N\,\ln\left(u\omega_N^{m-1}\right)\right)^2\right]\right\}\\
&\qquad\qquad \times\exp\left\{-\frac{\omega_N\,u}{(m-1)!}\left[d_1+O\left(\omega_N\,\ln\left(u\omega_N^{m-1}\right)\right)\right]\right\}\frac{du}{u}\\
&\le \int_{\omega_N^{1-m}\exp(-\varepsilon/\omega_N)}^{\omega_N^{1-m}\exp(\varepsilon/\omega_N)} \exp\left\{-\frac{(1-b)}{(m-1)!}u\,(1+O(\omega_N))\right\}\frac{du}{1/\sqrt{\omega_N}}\\
&= \sqrt{\omega_N}\int_{\omega_N^{1-m}\exp(-\varepsilon/\omega_N)}^{\omega_N^{1-m}\exp(\varepsilon/\omega_N)} \exp\left\{-\frac{(1-b)}{(m-1)!}u\,(1+O(\omega_N))\right\}du\\
&= O\left(\sqrt{\omega_N}\int_{\omega_N^{1-m}\exp(-\varepsilon/\omega_N)}^{\infty} \exp\left\{-\frac{(1-b)}{(m-1)!}u\,(1+O(\omega_N))\right\}du\right)\\
&= O\left(\sqrt{\omega_N}\;e^{-(1-b)/(m-1)!\sqrt{\omega_N}}\right),
\end{aligned}$$

since $(1-b) \in (0,1)$. Let us denote $I_{21}(\omega_N)$ the first integral of (3.9). We expand the exponentials and get

$$\begin{aligned}
I_{21}(\omega_N) = \int_{\omega_N^{1-m}\exp(-\varepsilon/\omega_N)}^{1/\sqrt{\omega_N}} \frac{e^{-(1-b)u/(m-1)!}}{u}&\left[1-\frac{1-b}{(m-1)!}\right.\\
&\quad\times(b-(m-1))\,u\omega_N\,\ln\left(u\omega_N^{m-1}\right)\\
&\left.-\frac{d_1}{(m-1)!}\,u\,\omega_N\,\left(1+O\left(\omega_N\ln\left(u\,\omega_N^{m-1}\right)\right)\right)\right]du.
\end{aligned}$$

We write the integral above as

$$I_{21}(\omega_N) = \int_{\omega_N^{1-m}\exp(-\varepsilon/\omega_N)}^{\infty} - \int_{1/\sqrt{\omega_N}}^{\infty}. \tag{3.10}$$

Again, the second integral of (3.10) is easily bounded by $O\left(\sqrt{\omega_N}\,e^{-(1-b)/(m-1)!\sqrt{\omega_N}}\right)$ as $\omega_N \to 0^+$, and our **claim is proved**. It is now an easy exercise to evaluate

$I_2(N)$. We have

$$\begin{aligned} I_2(N) = &\varepsilon + (m-1)\,\omega_N \ln \omega_N + [\,\ln (m-1)! + \ln (p+1) - \gamma\,]\,\omega_N \\ &- (m-1)\,(b-(m-1))\,\omega_N^2 \ln \omega_N \\ &+ [(b-(m-1))\,(\gamma - \ln (m-1)! - \ln (p+1) - d_1\,(1-b))]\,\omega_N^2 \\ &+ O\left(\omega_N^3\,(\ln \omega_N)^2\right), \end{aligned} \tag{3.11}$$

(where b and d_1 as defined in (3.7) and (3.8) respectively). Notice that the error term in the above dominates (as $\omega_N \to 0^+$) the previously mentioned term $O\left(\sqrt{\omega_N}\, e^{-(1-b)/(m-1)!\sqrt{\omega_N}}\right)$.
Now, we turn our attention to $I_3(N)$ of (2.12). As we will see the leading term is enough. The idea is that one can replace the integrand of (2.12) with $[\,-M_m(N;s)\,]$ and then by the quantity

$$N_m(N;s) := \sum_{j=2}^{N} \left[S_m\left(\frac{(\ln N)^{p+1}}{(\ln j)^p}s\right) \exp\left(-\frac{(\ln N)^{p+1}}{(\ln j)^p}s\right)\right].$$

For a rigorous approach see [6]. Hence as $N \to \infty$

$$I_3(N) = \int_{1+\varepsilon}^{\infty} N_m(N;s)\,[1 + O\,(N_m(N;s))]\,ds.$$

From the comparison of sums and integrals and Lemma 2.1 one easily arrives at

$$I_3(N) = \sum_{k=0}^{m-1} \frac{(\ln N)^k}{k!} \int_{1+\varepsilon}^{\infty} \frac{s^k N^{1-s}}{1+ps}\,[1 + O\,(\ln N)]\,ds.$$

Substitute $s = 1 - t$ and apply the Lapace method for integrals yields

$$I_3(N) = \frac{(1+\varepsilon)^{m-1}}{(1+p)\,(m-1)!\,\omega_N^{m-2}}\, e^{-\varepsilon/\omega_N} \left[1 + O\left(\frac{1}{\omega_N}\right)\right] \tag{3.12}$$

as $\omega_N \to 0^+$ and as we have set $\omega_N = (\ln N)^{-1}$. The reader now observes that the leading term of $I_3(N)$ is dominated by the sixth term of $I_2(N)$ as $N \to \infty$. We finish our approach by estimating the integral $I_1(N)$ of (2.10). For any given

$\varepsilon \in (0,1)$ it is easy to see that

$$
\begin{aligned}
I_1(N) &= \int_0^{1-\varepsilon} \exp\left[\sum_{j=2}^{N} \ln\left[1 - S_m\left(\frac{(\ln N)^{p+1}}{(\ln j)^p} s\right) \exp\left(-\frac{(\ln N)^{p+1}}{(\ln j)^p} s\right)\right]\right] ds \\
&\le \exp\left[\sum_{j=2}^{N} \ln\left[1 - S_m\,(0) \exp\left(-\frac{(\ln N)^{p+1}}{(\ln j)^p}\,(1-\varepsilon)\right)\right]\right] ds \\
&= \exp\left[\sum_{j=2}^{N} \ln\left[1 - \exp\left(-\frac{(\ln N)^{p+1}}{(\ln j)^p}\,(1-\varepsilon)\right)\right]\right] \\
&\le \exp\left[-\sum_{j=2}^{N} \exp\left(-\frac{(\ln N)^{p+1}}{(\ln j)^p}\,(1-\varepsilon)\right)\right] ds,
\end{aligned}
$$

since $\ln(1-x) \le -x$ for all $x \in [0,1]$. From the comparison for sums and integrals it follows that (as $N \to \infty$)

$$
\sum_{j=2}^{N} e^{-(1-\varepsilon)\frac{(\ln N)^{p+1}}{(\ln j)^p}} \sim \int_{j=2}^{N} e^{-(1-\varepsilon)\frac{(\ln N)^{p+1}}{(\ln x)^p}}\,dx.
$$

Since the quantity $1-\varepsilon$ is strictly positive *it is safe* to apply Lemma 2.1 and have

$$
I_1(N) = \exp\left[-\frac{1}{(1+p\,(1-\varepsilon))}\, e^{\varepsilon/\omega_N}\ (1 + M_1\,\omega_N)\right], \tag{3.13}
$$

where M_1 is a positive constant. The reader now observes that the leading term of $I_1(N)$ is dominated by the sixth term of $I_2(N)$ as $N \to \infty$.
From (3.11), (3.12), and (3.13), Theorem 2.3 follows immediately. It is notable that the *third term* of $A_N = \sum_{j=2}^{N} (\ln j)^{-p}$ contributes to the average of $T_m(N)$.

Proof of Theorem 2.4. From (1.6) and (2.3) we have

$$
\begin{aligned}
&E[\,T_m(N)\,(T_m(N)+1)\,] \\
&= 2\,N^2\Big(\,(\ln N)^2 + 2p\,\ln N + \big(p^2 + 2p\,(p+1)\big) + O\left(\frac{1}{\ln N}\right)\Big) \\
&\times \int_0^{\infty} s\left\{1 - \exp\left(\sum_{j=2}^{N} \ln\left[1 - S_m\left(\frac{(\ln N)^{p+1}}{(\ln j)^p} s\right) \exp\left(-\frac{(\ln N)^{p+1}}{(\ln j)^p} s\right)\right]\right)\right\} ds.
\end{aligned} \tag{3.14}
$$

Let us denote $\tilde{Q}_m(N;\alpha)$ the integral above. Then, for any given $\varepsilon \in (0,1)$ we have

$$
\tilde{Q}_m(N;\alpha) = \left[\,\frac{1}{2} + \varepsilon + \varepsilon^2 - I_4(N) - I_5(N) + I_6(N)\,\right],
$$

Let us denote $\tilde{Q}_m(N;\alpha)$ the integral above. Then, for any given $\varepsilon \in (0,1)$ we have

$$\tilde{Q}_m(N;\alpha) = \left[\frac{1}{2} + \varepsilon + \varepsilon^2 - I_4(N) - I_5(N) + I_6(N) \right],$$

where

$$\begin{aligned} I_4(N) &:= \int_0^{1-\varepsilon} s\, e^{M_m(N;s)}\, ds, \\ I_5(N) &:= \int_{1-\varepsilon}^{1+\varepsilon} s\, e^{M_m(N;s)}\, ds, \\ I_6(N) &:= \int_{1+\varepsilon}^{\infty} s \left[1 - e^{M_m(N;s)}\right] ds, \end{aligned} \tag{3.15}$$

and $M_m(N;s)$ is given in (2.13). If we treat $I_5(N)$ as we treated $I_2(N)$ and with a little patiences and paper, one finally arrives at

$$\begin{aligned} &I_5(N) = \\ &\varepsilon + \frac{\varepsilon^2}{2} + (m-1)\,\omega_N \ln \omega_N + [\ln(m-1)! + \ln(p+1) - \gamma]\,\omega_N - \frac{(m-1)^2}{2}\omega_N^2 \ln^2 \omega_N \\ &\quad + (m-1)\left[(m-1) - \frac{p}{p+1} - \ln(m-1)! - \ln(p+1) + \gamma\right]\omega_N^2 \ln \omega_N \\ &\quad + [(b-(m-1))\,(\gamma - \ln(m-1)! - \ln(p+1) - d_1\,(1-b)) \\ &\qquad - \frac{1}{2}\left(\gamma^2 + \frac{\pi^2}{6}\right) + \gamma\,(\ln(m-1)! + \ln(p+1)) \\ &\qquad + \frac{1}{2}\left(\ln(m-1)! + \ln(p+1)\right)^2\Big]\,\omega_N^2 + O\left(\omega_N^3\,(\ln \omega_N)^2\right), \end{aligned}$$

(where b and d_1 as defined in (3.7) and (3.8) respectively). With similar steps as in Theorem 2.3 one has that $I_4(N)$ and $I_6(N)$ are negligible compared to the *eighth* of $I_5(N)$. Now Theorem 2.4 follows immediately by invoking (3.14).

Proof of Theorem 2.5. The proof follows immediately from the identity

$$V[\,T_m(N)\,] = E[\,T_m(N)\,(T_m(N)+1)\,] - E[\,T_m(N)\,] - E[\,T_m(N)\,]^2$$

by invoking Theorems 2.3 and 2.4.

Proof of Main result. By a well known but very general theorem of P. Neal [20], which we have applied in [6] and extensively in [8], it suffices to find sequences $\{b_N\}$ and $\{k_N\}$ such that $k_N/b_N \to 0$ as $N \to \infty$ and that, for $y \in \mathbb{R}$,

$$\Lambda_N(y\,;m) := \frac{b_N^{m-1}}{(m-1)!}\sum_{j=1}^{N} p_{Nj}^{m-1} \exp\Big(-p_{Nj}\,(b_N + y k_N)\Big) \to g(y), \quad N \to \infty, \tag{3.16}$$

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

$$\frac{T_m(N) - b_N}{k_N} \xrightarrow{D} Y, \qquad N \to \infty,$$

where Y have distribution function

$$F(y) = P\{Y \le y\} = e^{-g(y)}, \qquad y \in \mathbb{R}.$$

As already mentioned the theorem in its original form is very general and without the knowledge of the sequences b_N and k_N can not be applied. Here our asymptotic formulas can help. In particular, we choose

$$b_N = N \ln N + (m-1) N \ln\ln N \qquad \text{and} \qquad k_N = N \tag{3.17}$$

and for all $y \in \mathbb{R}$ we will prove that

$$P\left\{\frac{T_m(N) - N\ln N - (m-1)N\ln\ln N}{N} \le y\right\} \to \exp\left(-\frac{e^{-(y-p)}}{(p+1)(m-1)!}\right) \tag{3.18}$$

as $N \to \infty$, which is equivalent to Main result. Under the choice of (3.17), $\Lambda_N(y\,; m)$ of (3.16) satisfies, as $N \to \infty$,

$$\Lambda_N(y\,; m) \sim \frac{(N\ln N)^{m-1}}{(m-1)!} \sum_{j=2}^{N} \left(\frac{a_j}{A_N}\right)^{m-1} e^{-(a_j/A_N)(N\ln N + (m-1)N\ln\ln N + Ny)} \tag{3.19}$$

where

$$a_j = \frac{1}{(\ln j)^p} \quad \text{and} \quad A_N = \sum_{j=2}^{N} \frac{1}{(\ln j)^p} = \frac{N}{(\ln N)^p} + \frac{pN}{(\ln N)^{p+1}} + O\left(\frac{N}{(\ln N)^{p+2}}\right).$$

Hence, (3.19) yields

$$\Lambda_N(y\,; m) \sim \frac{(\ln N)^{(p+1)(m-1)}}{(m-1)!} S_N(y), \tag{3.20}$$

where

$$S_N(y) := \sum_{j=2}^{N} \frac{1}{(\ln j)^{p(m-1)}} \times \exp\left(-\frac{(\ln N)^p (1 - p/\ln N)(\ln N + (m-1)\ln\ln N + y)}{(\ln j)^p}\right). \tag{3.21}$$

Now,

$$S_N(y) \sim I_N(y) \tag{3.22}$$

where

$$I_N(y) := \int_2^N \frac{1}{(\ln x)^{p(m-1)}} \times \exp\left(-\frac{(\ln N)^p (1 - p/\ln N)(\ln N + (m-1)\ln\ln N + y)}{(\ln x)^p}\right) dx.$$

By substituting $u = \ln x$ in the above integral we get

$$I_N(y) := \int_2^M \frac{1}{u^{p(m-1)}} \exp\left(-\frac{B}{u^p} + u\right) du, \tag{3.23}$$

where for typographical convenience we have set

$$B := \omega_N^{-(p+1)} (1 - p\,\omega_N) \left(1 - \frac{(m-1)\omega_N \ln \omega_N}{M} + y\,\omega_N\right), \quad \omega_N := (\ln N)^{-1}, \tag{3.24}$$

so that $B \to \infty$ and $\omega_N \to 0^+$ as $N \to \infty$. Next, we substitute

$$u = B^{1/(p+1)} t$$

in the integral of (3.23) and obtain

$$I_N(y) \sim B^{1 - \frac{pm}{p+1}} \int_0^\theta \frac{1}{t^{p(m-1)}} \, e^{B^{1/(p+1)} \phi(t)} dt, \tag{3.25}$$

where

$$\theta := \omega_N^{-1} B^{-1/(p+1)} \quad \text{and} \quad \phi(t) := t - \frac{1}{t^p}. \tag{3.26}$$

The integral in the right-hand side of (3.25) can be treated as a Laplace integral [15], where the large parameter is $B^{1/(p+1)}$. Since $\phi(t)$ is strictly increasing, the main contribution to the asymptotics of this integral comes from the endpoint θ (notice that $\theta \sim 1$ as $N \to \infty$). Thus, by applying the standard analysis of Laplace integrals, after some straightforward algebraic manipulations (3.25) becomes

$$I_N(y) \sim M^{-(p+1)(m-1)} \frac{e^{-(y-p)}}{(p+1)}. \tag{3.27}$$

Finally, by combining (3.27) with (3.24), (3.22), and (3.20) we obtain

$$\Lambda_N(y\,; m) \sim \frac{e^{-(y-p)}}{(p+1)(m-1)!} \tag{3.28}$$

and the proof is completed.

Acknowledgements. The authors wish to thank the two anonymous referees for reading the manuscript very carefully and making several valuable comments and suggestions which improved significantly this work.

References

[1] 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.

[2] S. Boneh and V.G. Papanicolaou, General Asymptotic Estimates for the Coupon Collector Problem, *Journal of Computational and Applied Mathematics* **67** (2) (Mar. 1996) 277–289.

[3] R.K. Brayton, On the asymptotic behavior of the number of trials necessary to complete a set with random selection, *Journal of Mathematical Analysis and Applications* **7** (1963) 31–61.

[4] N. G. de Bruijn, *Asymptotic Methods in Analysis, second edition*, North Holland, Amsterdam, 1961.

[5] P. Diaconis and S. Holmes, A Bayesian peek into Feller volume I, *Sankhyā*, Special issue in memory of D. Basu, **64** Ser. A (3, part 2) (2002) 820–841.

[6] A.V. Doumas and V.G. Papanicolaou, The Coupon Collector's Problem Revisited: Asymptotics of the Variance, *Adv. Appl. Prob.* **44** (1) (2012) 166–195.

[7] A.V. Doumas and V.G. Papanicolaou, Asymptotics of the rising moments for the Coupon Collector's Problem, *Electron. J. Probab.* **18** (Article no. 41) (2012) 1–15.

[8] A.V. Doumas and V.G. Papanicolaou, The Coupon Collector's Problem Revisited: Generalizing the Double Dixie Cup Problem of Newman and Shepp, *ESAIM: Prob. and Stat.* **20** (2016) 367–399.

[9] A.V. Doumas and V.G. Papanicolaou, Uniform versus Zipf distribution in a mixing collection process, *Stat. Probab. Lett.* **155** (2019), https://doi.org/10.1016/j.spl.2019.108559.

[10] A.V. Doumas and V.G. Papanicolaou, Sampling from a mixture of different groups of coupons, *https://arxiv.org/abs/1709.04500.*

[11] R. Durrett, *Probability: Theory and Examples*, Third Edition, Duxbury Advanced Series, Brooks/Cole—Thomson Learning. Belmont, CA, USA, 2005.

[12] P. Erdős and A. Rényi, On a classical problem of probability theory, *Magyar. Tud. Akad. Mat. Kutató Int. Közl.*, **6** (1961), 215–220.

[13] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. I & II, John Wiley & Sons, Inc., New York, 1966.

[14] P. Flajolet, D. Gardy and L. Thimonier, Birthday paradox, coupon collectors, caching algorithms and self-organizing search, *Discrete Applied Mathematics* **39** (1992) 207–229.

[15] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, 1st Ed., 2009.

[16] L. Holst, On Birthday, Collectors', Occupancy and other classical Urn problems, *International Statistical Review* **54** (1986) 15–27.

[17] K.L. Locey and J.T. Lennon, Scaling laws predict global microbial diversity, *Proc. Natl. Acad. Sci. USA* (2016) doi:/10.1073/pnas.1521291113

[18] H.M. Mahmoud, *Pólya urn models*, CRC Press, New York, 2008.

[19] A.W. Marshall, I. Olkin, and B. Arnold, *Inequalities: Theory of Majorization and Its Applications*, Springer, 2nd Ed., 2009.

[20] P. Neal, The Generalised Coupon Collector Problem, *J. Appl. Prob.* **45** (2008) 621–629.

[21] M. E. J. Newman, Power laws, Pareto distributions and Zipf's law, *Contemporary Physics* **46** (2005) 323–351.

[22] D.J. Newman and L. Shepp, The double Dixie cup problem, *Amer. Math. Monthly* **67** (1960) 58–61. MR0120672