Adv. Appl. Prob. **44,** 166–195 (2012)
Printed in Northern Ireland

THE COUPON COLLECTOR'S PROBLEM REVISITED: ASYMPTOTICS OF THE VARIANCE

ARISTIDES V. DOUMAS * ** AND
VASSILIS G. PAPANICOLAOU,* *** *National Technical University of Athens*

Abstract

We develop techniques for computing the asymptotics of the first and second moments of the number T_N of coupons that a collector has to buy in order to find all N existing different coupons as $N \to \infty$. The probabilities (occurring frequencies) of the coupons can be quite arbitrary. From these asymptotics we obtain the leading behavior of the variance $V[T_N]$ of T_N (see Theorems 3.1 and 4.4). Then, we combine our results with the general limit theorems of Neal in order to derive the limit distribution of T_N (appropriately normalized), which, for a large class of probabilities, turns out to be the standard Gumbel distribution. We also give various illustrative examples.

Keywords: Coupon collector's problem; higher asymptotics; limit distribution

2010 Mathematics Subject Classification: Primary 60F05; 60F99

1. Introduction

1.1. Preliminaries

Consider a population whose members are of N different *types* (e.g. colors). For $1 \le j \le N$, we denote by p_j the probability that a member of the population is of type j. The members of the population are sampled independently with replacement and their types are recorded. The so-called *coupon collector problem* (CCP) deals with questions arising in the above procedure. Some key quantities are the moments of the number T_N of trials it takes until all N types are detected (at least once). The coupon collector problem (in its simplest form) has appeared in Feller's classical work [8] and has attracted the attention of various researchers since it has found many applications in several areas of science (computer science—search algorithms, mathematical programming, optimization, learning processes, engineering, ecology, as well as linguistics—see, e.g. [3] and [9]).

It is convenient to introduce the events A_j^k, $1 \le j \le N$, that the type j is *not* detected until trial k (included). Then

$$P\{T_N \ge k\} = P\{A_1^{k-1} \cup \cdots \cup A_N^{k-1}\}, \qquad k = 1, 2, \ldots .$$

By invoking the inclusion–exclusion principle we obtain

$$P\{T_N \ge k\} = \sum_{\substack{J \subset \{1,\ldots,N\} \\ J \ne \varnothing}} (-1)^{|J|-1} \bigg[1 - \bigg(\sum_{j \in J} p_j\bigg)\bigg]^{k-1}, \qquad k = 1, 2, \ldots, \tag{1.1}$$

Received 5 May 2011; revision received 5 September 2011.
* Postal address: Department of Mathematics, National Technical University of Athens, Zografou Campus, 157 80 Athens, Greece.
** Email address: aris.doumas@hotmail.com
*** Email address: papanico@math.ntua.gr

https://doi.org/10.1239/aap/1331216649 Published online by Cambridge University Press

where the sum extends over the $2^N - 1$ nonempty subsets J of $\{1, \dots, N\}$, while $|J|$ denotes the cardinality of J. For $z \in \mathbb{C}$, $|z| \geq 1$, we set

$$G(z) := \mathrm{E}[z^{-T_N}] = 1 + (z^{-1} - 1) \sum_{k=1}^{\infty} z^{-(k-1)} \, \mathrm{P}\{T_N \geq k\}$$

(the second equality follows by partial summation). Using (1.1), we obtain

$$G(z) = 1 + (z-1) \sum_{\substack{J \subset \{1,\dots,N\} \\ J \neq \varnothing}} \frac{(-1)^{|J|}}{z - 1 + \sum_{j \in J} p_j}.$$

Since $\mathrm{E}[T_N] = -\lim_{z \to 1^+} G'(z)$ and $\mathrm{E}[T_N(T_N+1)] = \lim_{z \to 1^+} G''(z)$, we arrive at the well-known formulae (see, e.g. [13, p. 347])

$$\mathrm{E}[T_N] = \sum_{\substack{J \subset \{1,\dots,N\} \\ J \neq \varnothing}} \frac{(-1)^{|J|-1}}{\sum_{j \in J} p_j} = \sum_{m=1}^{N} (-1)^{m-1} \sum_{1 \leq j_1 < \cdots < j_m \leq N} \frac{1}{p_{j_1} + \cdots + p_{j_m}} \tag{1.2}$$

and

$$\mathrm{E}[T_N] = \int_0^{\infty} \left[1 - \prod_{j=1}^{N} (1 - \mathrm{e}^{-p_j t}) \right] \mathrm{d}t = \int_0^1 \left[1 - \prod_{j=1}^{N} (1 - x^{p_j}) \right] \frac{\mathrm{d}x}{x}, \tag{1.3}$$

as well as the formulae

$$\begin{aligned} \mathrm{E}[T_N(T_N+1)] &= 2 \sum_{\substack{J \subset \{1,\dots,N\} \\ J \neq \varnothing}} \frac{(-1)^{|J|-1}}{(\sum_{j \in J} p_j)^2} \\ &= 2 \sum_{m=1}^{N} (-1)^{m-1} \sum_{1 \leq j_1 < \cdots < j_m \leq N} \frac{1}{(p_{j_1} + \cdots + p_{j_m})^2} \end{aligned} \tag{1.4}$$

and

$$\begin{aligned} \mathrm{E}[T_N(T_N+1)] &= 2 \int_0^{\infty} \left[1 - \prod_{j=1}^{N} (1 - \mathrm{e}^{-p_j t}) \right] t \, \mathrm{d}t \\ &= -2 \int_0^1 \left[1 - \prod_{j=1}^{N} (1 - x^{p_j}) \right] \frac{\ln x}{x} \, \mathrm{d}x. \end{aligned} \tag{1.5}$$

Of course,

$$V[T_N] = \mathrm{E}[T_N(T_N+1)] - \mathrm{E}[T_N] - \mathrm{E}[T_N]^2. \tag{1.6}$$

1.2. The case of equal probabilities

Naturally, the simplest case regarding the previous formulae occurs when we take

$$p_1 = \cdots = p_N = \frac{1}{N}. \tag{1.7}$$

It is well known that, under (1.7), (1.2) becomes

$$\mathrm{E}[T_N] = NH_N, \quad \text{where} \quad H_N = \sum_{m=1}^{N} \frac{1}{m}. \tag{1.8}$$

This case, apart from its simplicity, has the property that among all sequences, it is the one with the smallest moments of T_N (see, e.g. [10]). A nice computer simulation of the CCP in the case of equal probabilities is available from http://www-stat.stanford.edu/~susan/surpise/Collector.html.

Conjecture 1.1. *The variance $V[T_N]$ takes its minimum value when all the p_js are equal.*

The results of the present paper (see Theorems 3.1 and 4.4) confirm that, for a large class of probabilities, $V[T_N]$ is actually minimized in the case of equal probabilities, as N becomes sufficiently large. Additional positive evidence for the conjecture comes from the asymptotic formula for the variance given in [6].

Under (1.7), (1.4) and (1.5) become

$$\mathrm{E}[T_N(T_N+1)] = -2\int_0^1 \left[1-(1-x^{1/N})^N\right]\frac{\ln x}{x}\,\mathrm{d}x = 2N^2 \sum_{m=1}^{N} \binom{N}{m} \frac{(-1)^{m-1}}{m^2}. \tag{1.9}$$

Substituting $u = 1 - x^{1/N}$ in the integral of (1.9) and evaluating the resulting integral we also obtain

$$\mathrm{E}[T_N(T_N+1)] = 2N^2 \sum_{m=1}^{N} \frac{H_m}{m} = N^2\left(H_N^2 + \sum_{m=1}^{N} \frac{1}{m^2}\right).$$

From (1.8) and (1.4), we can easily obtain the full asymptotic expansions of $\mathrm{E}[T_N]$, $\mathrm{E}[T_N(T_N+1)]$, and, hence, of $V[T_N]$. In particular, we have

$$\mathrm{E}[T_N] = N\ln N + \gamma N + \frac{1}{2} + O\left(\frac{1}{N}\right)$$

($\gamma = 0.5772\ldots$ is Euler's constant),

$$\mathrm{E}[T_N(T_N+1)] = N^2\left[(\ln N)^2 + 2\gamma \ln N + \gamma^2 + \frac{\pi^2}{6} + O\left(\frac{\ln N}{N}\right)\right],$$

and

$$V[T_N] = \frac{\pi^2}{6}N^2 - N\ln N - (\gamma+1)N + O\left(\frac{\ln N}{N}\right). \tag{1.10}$$

Note that (1.10) is in accordance with the known results (see, e.g. [6]). The coefficient $\pi^2/6$ in the leading order of $V[T_N]$ (refers to the Gumbel distribution and) persists in a large class of cases (see Theorem 4.4 and (5.6)).

1.3. Large N asymptotics

When N is large, it is not clear what information we can obtain from (1.2)–(1.3) and (1.4)–(1.5) for $\mathrm{E}[T_N]$, $\mathrm{E}[T_N(T_N+1)]$, and $V[T_N]$. For this reason, there is a need to develop efficient ways for deriving asymptotics as $N \to \infty$.

As in [4], let $\alpha = \{a_j\}_{j=1}^{\infty}$ be a sequence of strictly positive numbers. Then, for each integer $N > 0$, we can create a probability measure $\pi_N = \{p_1, \dots, p_N\}$ on the set of types $\{1, \dots, N\}$ by taking

$$p_j = \frac{a_j}{A_N}, \quad \text{where} \quad A_N = \sum_{j=1}^{N} a_j. \tag{1.11}$$

Note that p_j depends on α and N; thus, given α, it makes sense to consider the asymptotic behavior of $\mathrm{E}[T_N]$, $\mathrm{E}[T_N(T_N+1)]$, and $V[T_N]$ as $N \to \infty$.

Motivated by (1.2), we introduce the notation (as in [4])

$$E_N(\alpha) := \sum_{\substack{J \subset \{1,\dots,N\} \\ J \neq \varnothing}} \frac{(-1)^{|J|-1}}{\sum_{j \in J} a_j} = \sum_{k=1}^{N} (-1)^{k-1} \sum_{1 \le j_1 < \cdots < j_k \le N} \frac{1}{a_{j_1} + \cdots + a_{j_k}}. \tag{1.12}$$

Then, as in (1.3), we have

$$E_N(\alpha) = \int_0^{\infty} \left[1 - \prod_{j=1}^{N} (1 - \mathrm{e}^{-a_j t})\right] \mathrm{d}t = \int_0^1 \left[1 - \prod_{j=1}^{N} (1 - x^{a_j})\right] \frac{\mathrm{d}x}{x}. \tag{1.13}$$

If $s\alpha = \{s a_j\}_{j=1}^{\infty}$, (1.12) immediately gives $E_N(s\alpha) = s^{-1} E_N(\alpha)$ and, hence, in view of (1.2) and (1.11),

$$\mathrm{E}[T_N] = E_N(A_N^{-1}\alpha) = A_N E_N(\alpha). \tag{1.14}$$

Likewise, motivated by (1.4), in order to analyze $\mathrm{E}[T_N(T_N+1)]$, we introduce

$$Q_N(\alpha) := 2 \sum_{\substack{J \subset \{1,\dots,N\} \\ J \neq \varnothing}} \frac{(-1)^{|J|-1}}{(\sum_{j \in J} a_j)^2} = 2 \sum_{m=1}^{N} (-1)^{m-1} \sum_{1 \le j_1 < \cdots < j_m \le N} \frac{1}{(a_{j_1} + \cdots + a_{j_m})^2}. \tag{1.15}$$

Then, as in (1.5), we have

$$Q_N(\alpha) = 2 \int_0^{\infty} \left[1 - \prod_{j=1}^{N} (1 - \mathrm{e}^{-a_j t})\right] t \, \mathrm{d}t = -2 \int_0^1 \left[1 - \prod_{j=1}^{N} (1 - x^{a_j})\right] \frac{\ln x}{x} \, \mathrm{d}x. \tag{1.16}$$

From (1.15), it immediately follows that $Q_N(s\alpha) = s^{-2} Q_N(\alpha)$; hence,

$$\mathrm{E}[T_N(T_N+1)] = Q_N(A_N^{-1}\alpha) = A_N^2 Q_N(\alpha). \tag{1.17}$$

As noted in [4] for $\mathrm{E}[T_N]$, the problem of estimating $\mathrm{E}[T_N(T_N+1)]$ as $N \to \infty$ can be treated as two separate problems, namely estimating A_N^2 (i.e. A_N) and estimating $Q_N(\alpha)$. Our analysis focuses on estimating $Q_N(\alpha)$. The estimation of A_N will 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. [2, Chapter 6]), or even summation by parts.

The rest of the paper is organized as follows. In Section 2 we discuss a key feature, namely that the sequence α which produces the p_js can be of two (mutually exclusive) kinds. Section 3 deals with the sequences of the first kind. In Section 4 we consider a large class of sequences belonging to the second kind. Here the computations are much more involved. After presenting the detailed asymptotics of $\mathrm{E}[T_N]$ and $\mathrm{E}[T_N(T_N+1)]$ in Theorems 4.2 and 4.3, respectively, we finally give the (leading) asymptotic behavior of $V[T_N]$ in Theorem 4.4. In an earlier work [6] of Brayton (doctoral thesis under N. Levinson) an asymptotic formula for $V[T_N]$ was found under very restrictive assumptions on α. In particular, the probabilities p_j considered in [6] must satisfy

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

Our results complement the results of [6], since they concern quite general sequences for which the above ratio $\lambda(N)$ is not bounded. In particular, we cover some important families of distributions (e.g. linear and Zipf). Then, in Section 5 we use our asymptotic formulae for $\mathrm{E}[T_N]$ and $V[T_N]$ in the limit theorems of Neal [12] and obtain limiting distributions concerning T_N (see formulae (5.6) and (5.7)). Section 6 contains various examples. Finally, the proofs of certain technical theorems and lemmas are given in Appendix A.

2. The dichotomy

For convenience, we set

$$f_N^\alpha(x) := \prod_{j=1}^{N}(1-x^{a_j}), \qquad 0 \le x \le 1.$$

Obviously, (i) $f_N^\alpha(0)=1$ and $f_N^\alpha(1)=0$; (ii) $f_N^\alpha(x)$ is monotone decreasing in x; and (iii) $f_{N+1}^\alpha(x) \le f_N^\alpha(x)$. In particular,

$$\lim_N f_N^\alpha(x) = \prod_{j=1}^{\infty}(1-x^{a_j}) \quad \text{exists.}$$

Thus, applying the monotone convergence theorem to (1.13) and (1.16), we respectively obtain

$$L_1(\alpha) := \lim_N \mathrm{E}_N(\alpha) = \int_0^1 \left[1-\prod_{j=1}^{\infty}(1-x^{a_j})\right]\frac{\mathrm{d}x}{x} \tag{2.1}$$

and

$$L_2(\alpha) := \lim_N Q_N(\alpha) = -2\int_0^1 \left[1-\prod_{j=1}^{\infty}(1-x^{a_j})\right]\frac{\ln x}{x}\,\mathrm{d}x. \tag{2.2}$$

Note that $L_1(\alpha), L_2(\alpha) > 0$ for any α (since, for every $x \in (0,1)$, $f_N^\alpha(x) < 1$ and decreases with N). However, we may have $L_1(\alpha) = \infty$ and/or $L_2(\alpha) = \infty$. In fact, as we will see (in Remark 2.1 below), $L_1(\alpha) = \infty$ if and only if $L_2(\alpha) = \infty$.

Theorem 2.1. *We have $L_2(\alpha) < \infty$ if and only if there exists a $\xi \in (0,1)$ such that*

$$\sum_{j=1}^{\infty} \xi^{a_j} < \infty. \tag{2.3}$$

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 $0 \le b_j \le 1$ for all j. If $\sum_{j=1}^{\infty} b_j < \infty$ then

$$\sum_{j=1}^{\infty} b_j - \sum_{1 \le l < j} b_l b_j \le 1 - \prod_{j=1}^{\infty}(1 - b_j) \le \sum_{j=1}^{\infty} b_j. \tag{2.4}$$

Proof of Theorem 2.1. Assume that there exists a $\xi \in (0, 1)$ such that (2.3) is true. Then, by (2.2) and (2.4), we have

$$\begin{aligned} L_2(\alpha) &\le -2\int_0^{\xi}\left[\sum_{j=1}^{\infty} x^{a_j}\right]\ln x\frac{\mathrm{d}x}{x} - 2\int_{\xi}^{1}\left[1 - \prod_{j=1}^{\infty}(1 - x^{a_j})\right]\ln x\frac{\mathrm{d}x}{x} \\ &\le -2\int_0^{\xi}\left[\sum_{j=1}^{\infty} x^{a_j - 1}\right]\ln(x)\,\mathrm{d}x + \ln^2\xi \end{aligned}$$

or

$$L_2(\alpha) \le -2\sum_{j=1}^{\infty}\left(\frac{1}{a_j}\xi^{a_j}\ln\xi - \frac{1}{a_j^2}\xi^{a_j}\right) + \ln^2\xi.$$

Now, (2.3) implies that $\xi^{a_j} \to 0$; hence, $a_j \to \infty$. Therefore, $\min_j\{a_j\} = a_{j_0} > 0$. Thus,

$$L_2(\alpha) \le 2\frac{1}{a_{j_0}^2}\sum_{j=1}^{\infty}\xi^{a_j} - 2\frac{\ln\xi}{a_{j_0}}\sum_{j=1}^{\infty}\xi^{a_j} + \ln^2\xi < \infty.$$

Conversely, if

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

then, by a well-known property of infinite products (see, e.g. [14, p. 300]),

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

and, hence, (2.2) yields $L_2(\alpha) = -2\int_0^1(\ln x/x)\,\mathrm{d}x = \infty$.

Remark 2.1. It has been shown in [4] that $L_1(\alpha) < \infty$ if and only if there exists a $\xi \in (0, 1)$ such that $\sum_{j=1}^{\infty}\xi^{a_j} < \infty$. Thus, $L_2(\alpha) < \infty$ if and only if $L_1(\alpha) < \infty$.

To summarize, we have the following dichotomy:

$$0 < L_1(\alpha),\ L_2(\alpha) < \infty \quad \text{or} \quad L_1(\alpha) = L_2(\alpha) = \infty.$$

Remark 2.2. Consider the error term defined by

$$\Delta_N := L_2(\alpha) - Q_N(\alpha).$$

Then, by (1.16) and (2.2), we have

$$\Delta_N \le -2\int_0^1\left(\sum_{j=N+1}^{\infty} x^{a_j}\right)\frac{\ln x}{x}\,\mathrm{d}x = 2\sum_{j=N+1}^{\infty}\frac{1}{a_j^2}.$$

https://doi.org/10.1239/aap/1331216649 Published online by Cambridge University Press

3. Second moment and variance I: $L_i(\alpha) < \infty$

Let A_N and $L_i(\alpha)$ be as in (1.11), (2.1), and (2.2), respectively. We note that, by Theorem 2.1, $L_i(\alpha) < \infty$ implies that $\lim_j a_j = \infty$ (hence, $\lim_N A_N = \infty$).

Theorem 3.1. *If $L_i(\alpha) < \infty$, $i \in \{1, 2\}$, then, as $N \to \infty$,*

$$\mathrm{E}[T_N(T_N+1)] = A_N^2 L_2(\alpha)[1+o(1)]$$

and

$$V[T_N] = A_N^2[L_2(\alpha) - L_1(\alpha)^2] + o(A_N^2). \tag{3.1}$$

Proof. We know (see [4]) that

$$\mathrm{E}[T_N] = A_N L_1(\alpha)[1+o(1)] \quad \text{as } N \to \infty.$$

Thus, the formulae of the theorem follow immediately from (1.17), (2.1), (2.2), and (1.6).

Since $V[T_N] > 0$, (3.1) implies that

$$L_2(\alpha) - L_1(\alpha)^2 \geq 0.$$

However, in order that (3.1) is exact, we need to exclude the possibility that $L_2(\alpha) = L_1(\alpha)^2$.

Theorem 3.2. *We have*

$$L_2(\alpha) - L_1(\alpha)^2 > 0.$$

Proof. Set $F(x) = 1 - \prod_{j=1}^{\infty}(1 - x^{a_j})$, $x \in [0, 1]$. Then, clearly, F is increasing on $[0, 1]$, with $F(0) = 0$ and $F(1) = 1$; hence, F is a probability distribution function of some nontrivial (since $L_1(\alpha), L_2(\alpha) < \infty$) random variable X taking values in $[0, 1]$. In view of (2.1) and (2.2), we need to prove that

$$-2\int_0^1 \frac{\ln x}{x} F(x)\,\mathrm{d}x > \left[\int_0^1 \frac{F(x)}{x}\,\mathrm{d}x\right]^2. \tag{3.2}$$

Integration by parts leave us only to validate that

$$\mathrm{E}_F[\ln(X)^2] = \int_0^1 \ln(x)^2\,\mathrm{d}F(x) > \left[\int_0^1 \ln(x)\,\mathrm{d}F(x)\right]^2 = \mathrm{E}_F[\ln X]^2,$$

which is true by Jensen's inequality; thus, (3.2) is established.

4. Second moment and variance II: $L_i(\alpha) = \infty$

As we will see, this case is much more challenging.

4.1. Leading behavior of the second moment

By Theorem 2.1, $L_i(\alpha) = \infty$, $i \in \{1, 2\}$, is equivalent to

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

https://doi.org/10.1239/aap/1331216649 Published online by Cambridge University Press

For our further analysis, we follow [4] and write a_j in the form

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

where

$$f(x) > 0 \quad \text{and} \quad f'(x) > 0. \tag{4.1}$$

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

(C1) $f(x) \to \infty$,

(C2) $f'(x)/f(x) \to 0$,

(C3) $(f''(x)/f'(x))/(f'(x)/f(x)) = O(1)$,

(C4) $f'''(x)f(x)^2/f'(x)^3 = O(1)$

(in [4] the conditions on $f(x)$ were slightly weaker). These conditions are satisfied by a variety of commonly used functions. For example,

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

or various convex combinations of products of such functions.

Remark 4.1. (a) From condition (C2) we have

$$\lim_{x\to\infty} \frac{f(x+1)}{f(x)} = 1. \tag{4.2}$$

This can be justified by considering the function $g(x) = \ln f(x)$ and applying the mean value theorem.

(b) Condition (C3) together with (C1) and (C2) implies that

$$\frac{\ln f'(x)}{\ln f(x)} = O(1). \tag{4.3}$$

For typographical convenience, we set

$$F(x) := -f(x)\ln\left(\frac{f'(x)}{f(x)}\right) \tag{4.4}$$

(note that (4.1) and (C2) imply that $F(x) > 0$ for sufficiently large x).

Theorem 4.1. *If $\alpha = \{1/f(j)\}_{j=1}^{\infty}$, where f satisfies (4.1) and (C1)–(C4), then*

$$Q_N(\alpha) \sim f(N)^2 \ln\left(\frac{f(N)}{f'(N)}\right)^2 = F(N)^2 \quad \text{as } N \to \infty \tag{4.5}$$

(where $\gamma_N \sim \delta_N$ means, as usual, $\gamma_N/\delta_N \to 1$).

https://doi.org/10.1239/aap/1331216649 Published online by Cambridge University Press

The proof is an adaptation of the proof given in [4] for the leading asymptotics of $\mathrm{E}_N(\alpha)$. See Appendix A.

Using Theorem 3.1 in (1.17), we obtain

$$\mathrm{E}[T_N(T_N+1)] \sim A_N^2 f(N)^2 \ln\left(\frac{f(N)}{f'(N)}\right)^2 \quad \text{as } N \to \infty. \tag{4.6}$$

Note that

$$A_N f(N) = \frac{1}{p_N} = \frac{1}{\min_{1\le j\le N}\{p_j\}}.$$

4.2. More terms in the asymptotic behavior of $\mathrm{E}[T_N]$

It was shown in [4] that

$$\mathrm{E}[T_N] \sim A_N f(N) \ln\left(\frac{f(N)}{f'(N)}\right) \quad \text{as } N \to \infty. \tag{4.7}$$

If we substitute (4.6) and (4.7) into (1.6), it is clear that we do not have enough information to find the leading asymptotics of $V[T_N]$. Thus, we need more terms in the expansions of $\mathrm{E}[T_N]$ and $\mathrm{E}[T_N(T_N+1)]$. Starting from (1.13), we rewrite $\mathrm{E}_N(\alpha)$ as

$$\begin{aligned} E_N(\alpha) = F(N)\Bigg(1 - \int_0^1 \exp\Bigg[\sum_{j=1}^N \ln(1-\mathrm{e}^{-F(N)s/f(j)})\Bigg]\mathrm{d}s \\ + \int_1^\infty \Bigg\{1-\exp\Bigg[\sum_{j=1}^N \ln(1-\mathrm{e}^{-F(N)s/f(j)})\Bigg]\Bigg\}\mathrm{d}s\Bigg). \end{aligned} \tag{4.8}$$

Set

$$I_1(N) = \int_0^1 \exp\Bigg[\sum_{j=1}^N \ln(1-\mathrm{e}^{-F(N)s/f(j)})\Bigg]\mathrm{d}s \tag{4.9}$$

and

$$I_2(N) = \int_1^\infty \Bigg\{1-\exp\Bigg[\sum_{j=1}^N \ln(1-\mathrm{e}^{-F(N)s/f(j)})\Bigg]\Bigg\}\mathrm{d}s. \tag{4.10}$$

We know (see [4]) that

$$I_1(N) = o(1) \quad \text{and} \quad I_2(N) = o(1). \tag{4.11}$$

In order to analyze the above integrals more deeply, we need the following lemma.

Lemma 4.1. *Set*

$$J_m(N) := \int_1^N f(x)^m \mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x, \qquad m \ge 0.$$

Then, under (C1)–(C4) and (4.4), we have

$$J_m(N) = \frac{f(N)^{m+2}}{sF(N)f'(N)}\mathrm{e}^{-F(N)s/f(N)}\left[1 + O\left(\frac{f(N)}{F(N)}\right)\right] \quad \text{as } N \to \infty,$$

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

Proof. See Appendix A.

https://doi.org/10.1239/aap/1331216649 Published online by Cambridge University Press

We will also need the second term in the asymptotics of the integral $J_m(N)$.

Corollary 4.1. *If $J_m(N)$ is as in Lemma 4.1 then, as $N \to \infty$,*

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

where

$$\omega(N) := -2 + \frac{f''(N)/f'(N)}{f'(N)/f(N)}. \tag{4.12}$$

Again, the asymptotics are uniform in $s \in [s_0, \infty)$ for any $s_0 > 0$.

Proof. See Appendix A.

4.2.1. *The Integral $I_1(N)$.* Regarding the integral in (4.9), given any $\varepsilon \in (0, 1)$, we have

$$I_1(N) = \int_0^{1-\varepsilon} \exp\left[\sum_{j=1}^{N} \ln(1 - e^{-F(N)s/f(j)})\right] ds + \int_{1-\varepsilon}^{1} \exp\left[\sum_{j=1}^{N} \ln(1 - e^{-F(N)s/f(j)})\right] ds. \tag{4.13}$$

For the first integral in (4.13), we have

$$\begin{aligned} I_{11}(N) &:= \int_0^{1-\varepsilon} \exp\left[\sum_{j=1}^{N} \ln(1 - e^{-F(N)s/f(j)})\right] ds \\ &< (1-\varepsilon) \exp\left[\sum_{j=1}^{N} \ln(1 - e^{-F(N)(1-\varepsilon)/f(j)})\right] \\ &< \exp\left[\sum_{j=1}^{N} \ln(1 - e^{-F(N)(1-\varepsilon)f(j)})\right] \\ &< \exp\left(-\sum_{j=1}^{N} e^{-F(N)(1-\varepsilon)/f(j)}\right), \end{aligned}$$

since $\ln(1-x) < -x$ for $0 < x < 1$. Thus, from (A.6) in Appendix A we obtain

$$I_{11}(N) < \exp\left[-\int_1^N e^{-F(N)(1-\varepsilon)/f(x)}\, dx\right]. \tag{4.14}$$

Applying Lemma 4.1 for $m = 0$ we arrive at

$$I_{11}(N) < \exp\left[-\frac{f(N)^2}{(1-\varepsilon)F(N)f'(N)} e^{-F(N)(1-\varepsilon)/f(N)} \left(1 + M_1 \frac{f(N)}{F(N)}\right)\right], \tag{4.15}$$

https://doi.org/10.1239/aap/1331216649 Published online by Cambridge University Press

where M_1 is a positive constant. Using (4.4), i.e. the definition of F, we have

$$I_{11}(N) < \exp\biggl[-\frac{1}{1-\varepsilon}\frac{(f(N)/f'(N))^{\varepsilon}}{\ln(f(N)/f'(N))}\biggl(1+M_1\frac{1}{\ln(f(N)/f'(N))}\biggr)\biggr]. \tag{4.16}$$

Since $f'(x)/f(x) \to 0$ and $\varepsilon \in (0, 1)$, we have

$$I_{11}(N) < \biggl[\frac{1}{\ln(f(N)/f'(N))}\biggr]^3 \tag{4.17}$$

for sufficiently large N.

Our next task is to compute a few terms of the asymptotic expansion of the second integral in (4.13). For convenience, we set

$$B(N; s) := \sum_{j=1}^{N} \ln(1 - \mathrm{e}^{-F(N)s/f(j)}). \tag{4.18}$$

Since

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

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

$$B(N; s) = \sum_{j=1}^{N} [-\mathrm{e}^{-F(N)s/f(j)} + O(\mathrm{e}^{-2F(N)s/f(j)})]. \tag{4.19}$$

From the comparison of sums and integrals, i.e. (A.6), (4.19) yields

$$B(N; s) = -\biggl[\int_1^N \mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x + O(\mathrm{e}^{-F(N)s/f(N+1)})\biggr] + \sum_{j=1}^{N} O(\mathrm{e}^{-2F(N)s/f(j)}).$$

The above formula, together with Corollary 4.1 for $m = 0$, gives

$$\begin{aligned} B(N; s) = &-\frac{f(N)^2}{sF(N)f'(N)}\mathrm{e}^{-F(N)s/f(N)} \\ &- \omega(N)\frac{f(N)^3}{s^2F(N)^2f'(N)}\mathrm{e}^{-F(N)s/f(N)}\biggl[1 + O\biggl(\frac{f(N)}{F(N)}\biggr)\biggr] \\ &+ O(\mathrm{e}^{-F(N)s/f(N+1)} + N\mathrm{e}^{-2F(N)s/f(N)}). \end{aligned}$$

Using (4.2) the above yields

$$\begin{aligned} B(N; s) = &-\frac{f(N)^2}{sF(N)f'(N)}\mathrm{e}^{-F(N)s/f(N)} \\ &- \omega(N)\frac{f(N)^3}{s^2F(N)^2f'(N)}\mathrm{e}^{-F(N)s/f(N)}\biggl[1 + O\biggl(\frac{f(N)}{F(N)}\biggr)\biggr]. \end{aligned}$$

Hence,

$$I_{12}(N) := \int_{1-\varepsilon}^{1} \mathrm{e}^{B(N;s)}\,\mathrm{d}s$$
$$= \int_{1-\varepsilon}^{1} \exp\bigg[-\frac{f(N)^2}{sF(N)f'(N)}\mathrm{e}^{-F(N)s/f(N)} - \omega(N)\frac{f(N)^3}{s^2F(N)^2f'(N)}\mathrm{e}^{-F(N)s/f(N)}\bigg[1+O\bigg(\frac{f(N)}{F(N)}\bigg)\bigg]\bigg]\,\mathrm{d}s$$

as $N \to \infty$. Using the definition of F and substituting $s = 1-t$, the above expression becomes

$$I_{12}(N) = \int_0^{\varepsilon} \exp\bigg[-\frac{1}{1-t}\frac{(f(N)/f'(N))^t}{\ln(f(N)/f'(N))} - \omega(N)\frac{1}{(1-t)^2}\frac{(f(N)/f'(N))^t}{\ln(f(N)/f'(N))^2}\bigg[1+O\bigg(\frac{1}{\ln(f(N)/f'(N))}\bigg)\bigg]\bigg]\,\mathrm{d}t.$$

For typographical convenience, we set

$$A := \frac{f(N)}{f'(N)} \tag{4.20}$$

(note that $A \to \infty$ as $N \to \infty$). Thus,

$$I_{12}(N) = \int_0^{\varepsilon} \exp\bigg[-\frac{A^t}{\ln A}\bigg(\sum_{n=0}^{\infty} t^n\bigg) - \omega(N)\frac{A^t}{\ln^2 A}\bigg(\sum_{n=1}^{\infty} nt^{n-1}\bigg)\bigg[1+O\bigg(\frac{1}{\ln A}\bigg)\bigg]\bigg]\,\mathrm{d}t. \tag{4.21}$$

Substituting $u = A^t/\ln A$ into the above integral, (4.21) yields

$$I_{12}(N) = \int_{1/\ln A}^{A^{\varepsilon}/\ln A} \exp\bigg[-\frac{u}{1-\ln u/\ln A - \ln(\ln A)/\ln A} - \frac{\omega(N)}{\ln A}\frac{u}{(1-\ln u/\ln A - \ln(\ln A)/\ln A)^2} \times \bigg[1+O\bigg(\frac{1}{\ln A}\bigg)\bigg]\bigg]\frac{\mathrm{d}u}{u\ln A}.$$

If

$$\delta := \frac{1}{\ln A} = \frac{1}{\ln(f(N)/f'(N))} = \frac{f(N)}{F(N)} \tag{4.22}$$

(hence, $A \to \infty$ implies that $\delta \to 0^+$), the above integral becomes

$$I_{12} = \delta\int_{\delta}^{\delta\exp(\varepsilon/\delta)} \exp\bigg(-\frac{u}{1-\delta\ln u+\delta\ln\delta} - \frac{\omega(N)u\delta}{(1-\delta\ln u+\delta\ln\delta)^2}(1+O(\delta))\bigg)\frac{\mathrm{d}u}{u}.$$

Thus,

$$I_{12} = \delta\int_{\delta}^{1/\sqrt{\delta}} \exp\bigg[-\frac{u}{1-\delta\ln u+\delta\ln\delta} - \frac{\omega(N)u\delta}{(1-\delta\ln u+\delta\ln\delta)^2}(1+O(\delta))\bigg]\frac{\mathrm{d}u}{u} + \delta\int_{1/\sqrt{\delta}}^{\delta\exp(\varepsilon/\delta)} \exp\bigg[-\frac{u}{1-\delta\ln u+\delta\ln\delta} - \frac{\omega(N)u\delta}{(1-\delta\ln u+\delta\ln\delta)^2}(1+O(\delta))\bigg]\frac{\mathrm{d}u}{u}. \tag{4.23}$$

https://doi.org/10.1239/aap/1331216649 Published online by Cambridge University Press

First we deal with the second integral in (4.23) and obtain an upper bound as follows:

$$\begin{aligned}
&\int_{1/\sqrt{\delta}}^{\delta\exp(\varepsilon/\delta)} \exp\left[-\frac{u}{1-\delta\ln u+\delta\ln\delta} - \frac{\omega(N)u\delta}{(1-\delta\ln u+\delta\ln\delta)^2}\,(1+O(\delta))\right]\frac{\mathrm{d}u}{u}\\
&\quad= \int_{1/\sqrt{\delta}}^{\delta\exp(\varepsilon/\delta)} \exp\left[-\frac{u}{1-\delta\ln u+\delta\ln\delta}\left[1+\frac{\omega(N)}{1-\delta\ln u+\delta\ln\delta}\,\delta(1+O(\delta))\right]\right]\frac{\mathrm{d}u}{u}\\
&\quad\le \int_{1/\sqrt{\delta}}^{\delta\exp(\varepsilon/\delta)} \exp\left[-\frac{u(1+O(\delta))}{1-\delta\ln(1/\sqrt{\delta})+\delta\ln\delta}\right]\frac{\mathrm{d}u}{1/\sqrt{\delta}}\\
&\quad= O(\sqrt{\delta}\mathrm{e}^{-1/\sqrt{\delta}}).
\end{aligned}\tag{4.24}$$

Denote the first integral in (4.23) as

$$K_1(\delta) := \int_{\delta}^{1/\sqrt{\delta}} \exp\left[-\frac{u}{1-\delta\ln u+\delta\ln\delta} - \frac{u\;\omega(N)\delta}{(1-\delta\ln u+\delta\ln\delta)^2}(1+O(\delta))\right]\frac{\mathrm{d}u}{u}.$$

Since, for $|x|<1$, $(1-x)^{-2}=\sum_{n=1}^{\infty} nx^{n-1}$,

$$\begin{aligned}
K_1(\delta) &= \int_{\delta}^{1/\sqrt{\delta}} \exp\left[-u\sum_{n=0}^{\infty}\left(\delta\ln\frac{u}{\delta}\right)^{n} - u\omega(N)\delta(1+O(\delta))\sum_{n=1}^{\infty} n\left(\delta\ln\frac{u}{\delta}\right)^{n-1}\right]\frac{\mathrm{d}u}{u}\\
&= \int_{\delta}^{1/\sqrt{\delta}} \frac{\mathrm{e}^{-u}}{u}\exp\left[-u\sum_{n=1}^{\infty}\left(\delta\ln\frac{u}{\delta}\right)^{n}\right]\\
&\qquad\qquad\times\exp\left[-u\omega(N)\delta(1+O(\delta))\sum_{n=1}^{\infty} n\left(\delta\ln\frac{u}{\delta}\right)^{n-1}\right]\mathrm{d}u.
\end{aligned}$$

Next we expand the exponentials and obtain

$$\begin{aligned}
K_1(\delta) = \int_{\delta}^{1/\sqrt{\delta}} \frac{\mathrm{e}^{-u}}{u}&\left\{1-u\sum_{n=1}^{\infty}\left(\delta\ln\frac{u}{\delta}\right)^{n} + O\left(u\sum_{n=1}^{\infty}\left(\delta\ln\frac{u}{\delta}\right)^{n}\right)^{2}\right\}\\
&\times\left\{1-\omega(N)u\delta(1+O(\delta))\sum_{n=1}^{\infty}\left(\delta\ln\frac{u}{\delta}\right)^{n-1}\right.\\
&\qquad\left.+\,O\left(\omega(N)u\delta(1+O(\delta))\sum_{n=1}^{\infty}\left(\delta\ln\frac{u}{\delta}\right)^{n-1}\right)^{2}\right\}\mathrm{d}u
\end{aligned}$$

(since $\mathrm{e}^{x}=1+x+O(x^2)$ as $x\to 0$). Hence,

$$\begin{aligned}
K_1(\delta) &= \int_{\delta}^{1/\sqrt{\delta}} \frac{\mathrm{e}^{-u}}{u}\left[1-u\left(\delta\ln\frac{u}{\delta}+\omega(N)\delta(1+O(\delta))\right)+u^2 O\left(\delta^2\ln^2\frac{u}{\delta}\right)\right]\mathrm{d}u\\
&= \int_{\delta}^{\infty} \frac{\mathrm{e}^{-u}}{u}\left[1-u\left(\delta\ln\frac{u}{\delta}+\omega(N)\delta(1+O(\delta))\right)+u^2 O\left(\delta^2\ln^2\frac{u}{\delta}\right)\right]\mathrm{d}u\\
&\quad- \int_{1/\sqrt{\delta}}^{\infty} \frac{\mathrm{e}^{-u}}{u}\left[1-u\left(\delta\ln\frac{u}{\delta}+\omega(N)\delta(1+O(\delta))\right)+u^2 O\left(\delta^2\ln^2\frac{u}{\delta}\right)\right]\mathrm{d}u.
\end{aligned}$$

However,

$$\begin{aligned}
&\int_{1/\sqrt{\delta}}^{\infty} \frac{\mathrm{e}^{-u}}{u}\left[1 - u\left(\delta \ln \frac{u}{\delta} + \omega(N)\delta(1 + O(\delta))\right) + u^2 O\left(\delta^2 \ln^2 \frac{u}{\delta}\right)\right] \mathrm{d}u \\
&\quad \le \int_{1/\sqrt{\delta}}^{\infty} \frac{\mathrm{e}^{-u}}{1/\sqrt{\delta}}\left(1 - \frac{1}{\sqrt{\delta}}\delta \ln\left(\frac{1/\sqrt{\delta}}{\delta}\right) - \omega(N)\delta(1 + O(\delta))\right) \mathrm{d}u \\
&\qquad + \int_{1/\sqrt{\delta}}^{\infty} u\mathrm{e}^{-u} O\left(\delta^2 \ln^2 \frac{u}{\delta}\right) \mathrm{d}u \\
&\quad = O(\sqrt{\delta}\mathrm{e}^{-1/\sqrt{\delta}}) \quad \text{as } \delta \to 0^+.
\end{aligned} \tag{4.25}$$

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

$$I_{12}(N) = \delta \int_{\delta}^{\infty} \frac{\mathrm{e}^{-u}}{u}\left[1 - u\left(\delta \ln \frac{u}{\delta} + \omega(N)\delta(1 + O(\delta))\right) + u^2 O\left(\delta^2 \ln^2 \frac{u}{\delta}\right)\right] \mathrm{d}u \tag{4.26}$$

as $\delta \to 0^+$. To continue, we need the following lemmas.

Lemma 4.2. *For the exponential integral,*

$$\mathrm{E}(x) := \int_x^{\infty} \frac{\mathrm{e}^{-t}}{t}\, \mathrm{d}t,$$

we have the asymptotic expansion

$$\mathrm{E}(x) \sim -\ln x - \gamma + x - \tfrac{1}{4}x^2 + \tfrac{1}{18}x^3 - \cdots$$

as $x \to 0^+$. *Here* $\gamma = 0.5772\ldots$ *is Euler's constant.*

Proof. See [2, p. 252].

Lemma 4.3. *For the integral*

$$G(x) := \int_x^{\infty} \ln t \mathrm{e}^{-t}\, \mathrm{d}t,$$

we have the asymptotic expansion, as $x \to 0^+$,

$$G(x) \sim -\gamma - x\ln x + x + \tfrac{1}{2}x^2 \ln x - \tfrac{1}{2}x^2 - \tfrac{1}{6}x^3 \ln x + \tfrac{1}{6}x^3 + \tfrac{1}{24}x^4 \ln x - \tfrac{1}{24}x^4 + \cdots .$$

Proof. See Appendix A.

Applying Lemmas 4.2 and 4.3 to (4.26) we obtain

$$I_{12} = -\delta \ln \delta - \gamma\delta + \delta^2 \ln \delta \mathrm{e}^{-\delta} + (1 + \gamma)\delta^2 - \delta^2 \omega(N)\mathrm{e}^{-\delta} + O(\delta^3 \ln^2 \delta). \tag{4.27}$$

Since $\mathrm{e}^{-\delta} = 1 + O(\delta)$ as $\delta \to 0^+$, (4.27) yields

$$I_{12} = -\delta \ln \delta - \gamma\delta + \delta^2 \ln \delta + [1 + \gamma - \omega(N)]\delta^2 + O(\delta^3 \ln^2 \delta). \tag{4.28}$$

Note that the error term in (4.28) dominates the terms of (4.24) and (4.25).

https://doi.org/10.1239/aap/1331216649 Published online by Cambridge University Press

4.2.2. *The Integral* $I_2(N)$. Our next goal is to compute the asymptotic behavior of $I_2(N)$. Here we will follow a different approach.

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

$$-(1+\vartheta)x < \ln(1-x) < -(1-\vartheta)x \tag{4.29}$$

and

$$(1-\vartheta)x < 1 - \mathrm{e}^{-x} < (1+\vartheta)x. \tag{4.30}$$

For $j = 1, \dots, N$ and $s \geq 1$, we use the definition of F and the fact that f is increasing to obtain

$$0 < x = \mathrm{e}^{-F(N)s/f(j)} \leq \mathrm{e}^{-F(N)s/f(N)} \leq \mathrm{e}^{-F(N)/f(N)} = \frac{f'(N)}{f(N)} \to 0 \quad \text{as } N \to \infty.$$

Hence, for a given $\vartheta \in (0, 1)$, there is $N_0 = N_0(\vartheta)$ such that, for $N \geq N_0$, (4.29) yields

$$-(1+\vartheta)\mathrm{e}^{-F(N)s/f(j)} < \ln(1 - \mathrm{e}^{-F(N)s/f(j)}) < -(1-\vartheta)\mathrm{e}^{-F(N)s/f(j)}, \qquad j = 1, \dots, N.$$

By summing over j and using (4.18) we get

$$-(1+\vartheta)\sum_{j=1}^{N} \mathrm{e}^{-F(N)s/f(j)} < B(s; N) < -(1-\vartheta)\sum_{j=1}^{N} \mathrm{e}^{-F(N)s/f(j)}.$$

From (A.6) in Appendix A (i.e. the comparison of sums and integrals), we arrive at

$$\begin{aligned} -(1+\vartheta)&\bigg[\mathrm{e}^{-F(N)s/f(N+1)} + \int_1^N \mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x\bigg] \\ &< B(s; N) \\ &< -(1-\vartheta)\bigg[\mathrm{e}^{-F(N)s/f(N+1)} + \int_1^N \mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x\bigg]. \end{aligned} \tag{4.31}$$

Now, from (4.11) we have $B(s; N) \to 0$ as $N \to \infty$, uniformly in $s \in [1, \infty)$. Thus, for given $\vartheta > 0$, there exists $N_1 = N_1(\vartheta)$ such that, for $N \geq N_1$, (4.30) gives

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

Therefore (see (4.10) and (4.18)),

$$-(1-\vartheta)\int_1^\infty B(s; N)\,\mathrm{d}s < I_2(N) < -(1+\vartheta)\int_1^\infty B(s; N)\,\mathrm{d}s.$$

Using the bounds of $B(s; N)$ given in (4.31) in the above formula, we find that, for all $N \geq N_2 = \max\{N_0, N_1\}$,

$$\begin{aligned} (1-\vartheta)^2 &\int_1^\infty\int_1^N \mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x\,\mathrm{d}s - (1-\vartheta)^2\int_1^\infty \mathrm{e}^{-F(N)s/f(N+1)}\,\mathrm{d}s \\ &< I_2(N) \\ &< (1+\vartheta)^2\int_1^\infty\int_1^N \mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x\,\mathrm{d}s + (1+\vartheta)^2\int_1^\infty \mathrm{e}^{-F(N)s/f(N+1)}\,\mathrm{d}s. \end{aligned} \tag{4.32}$$

https://doi.org/10.1239/aap/1331216649 Published online by Cambridge University Press

Now,

$$\int_1^\infty \int_1^N e^{-F(N)s/f(x)}\,dx\,ds = \frac{1}{F(N)} \int_1^N f(x) e^{-F(N)/f(x)}\,dx$$
$$= \left(\frac{f(N)}{F(N)}\right)^2 \left[1 + O\left(\frac{f(N)}{F(N)}\right)\right],$$

where the last equality follows by applying Lemma 4.1 for $m = 1$. Furthermore, by (C1)–(C4) and (4.2), it is straightforward to see that

$$\int_1^\infty e^{-F(N)s/f(N+1)}\,ds = \frac{f(N+1)}{F(N)} e^{-F(N)/f(N+1)} = o\left(\frac{f(N)^2}{F(N)^2}\right).$$

Since $\vartheta \in (0, 1)$ is arbitrary, (4.32) implies that

$$I_2(N) = \left(\frac{f(N)}{F(N)}\right)^2 \left[1 + O\left(\frac{f(N)}{F(N)}\right)\right] \quad \text{as } N \to \infty.$$

Again, using the definition of F and (4.22), we obtain

$$I_2(N) = \delta^2(1 + O(\delta)) \quad \text{as } \delta \to 0^+. \tag{4.33}$$

We are therefore ready to present the following result.

Theorem 4.2. *Let δ be as defined in (4.22) (hence, $\delta \to 0^+$ as $N \to \infty$), and let $\omega(N)$ be as given in (4.12). Then (γ is, as usual, Euler's constant)*

$$E_N(\alpha) = f(N)\left[\frac{1}{\delta} + \ln\delta + \gamma - \delta\ln\delta + (\omega(N) - \gamma)\delta + O(\delta^2 \ln^2 \delta)\right].$$

Proof. The result follows immediately by combining (4.17), (4.28), (4.33), (4.9), and (4.10) with (4.8).

From (1.14) we have (as $\delta \to 0^+$)

$$\mathrm{E}[T_N] = A_N f(N)\left[\frac{1}{\delta} + \ln\delta + \gamma - \delta\ln\delta + (\omega(N) - \gamma)\delta + O(\delta^2 \ln^2 \delta)\right]. \tag{4.34}$$

We mention again that in [4] the leading behavior of $\mathrm{E}[T_N]$ was given. Formula (4.34) is an improvement.

4.3. More asymptotics for $\mathrm{E}[T_N(T_N + 1)]$

Here we will follow a similar approach as in Subsection 4.2, in order to find the first few terms in the asymptotic expansion of $\mathrm{E}[T_N(T_N + 1)]$, so that the leading behavior of $V[T_N]$ can be eventually calculated. Expand $Q_N(\alpha)$ as

$$Q_N(\alpha) = 2F(N)^2\left[\frac{1}{2} - \int_0^1 \exp\left[\sum_{j=1}^N \ln(1 - e^{-F(N)s/f(j)})\right] s\,ds\right]$$
$$+ 2F(N)^2 \int_1^\infty \left[1 - \exp\left[\sum_{j=1}^N \ln(1 - e^{-F(N)s/f(j)})\right]\right] s\,ds. \tag{4.35}$$

https://doi.org/10.1239/aap/1331216649 Published online by Cambridge University Press

Set

$$I_3(N) = \int_0^1 \left[\exp\left[\sum_{j=1}^{N} \ln(1 - e^{-F(N)s/f(j)})\right]\right] s\,ds \tag{4.36}$$

and

$$I_4(N) = \int_1^\infty \left[1 - \exp\left[\sum_{j=1}^{N} \ln(1 - e^{-F(N)s/f(j)})\right]\right] s\,ds. \tag{4.37}$$

From (A.4) and (A.8), we know that

$$I_3(N) = o(1) \quad \text{and} \quad I_4(N) = o(1).$$

4.3.1. *The Integral* $I_3(N)$. For $\varepsilon \in (0, 1)$, we write the integral $I_3(N)$ given in (4.36) as

$$I_3(N) = I_{31}(N) + I_{32}(N), \tag{4.38}$$

where

$$I_{31}(N) = \int_0^{1-\varepsilon} \left[\exp\left[\sum_{j=1}^{N} \ln(1 - e^{-F(N)s/f(j)})\right]\right] s\,ds \tag{4.39}$$

and

$$I_{32}(N) = \int_{1-\varepsilon}^1 \left[\exp\left[\sum_{j=1}^{N} \ln(1 - e^{-F(N)s/f(j)})\right]\right] s\,ds.$$

For $I_{31}(N)$ given in (4.39), as in (4.14), we have

$$I_{31}(N) < \exp\left[-\int_1^N e^{-F(N)(1-\varepsilon)/f(x)}\,dx\right].$$

Applying Lemma 4.1 for $m = 0$ and using the definition of F (in the same manner as we did for (4.15) and (4.16)), we arrive at

$$I_{31}(N) < \left(\frac{1}{\ln(f(N)/f'(N))}\right)^3 \tag{4.40}$$

for sufficiently large N.

Our next task is to compute the asymptotics of $I_{32}(N)$. We can treat $I_{32}(N)$ as we treated $I_{12}(N)$ in Subsubsection 4.2.1. We obtain

$$I_{32} = \int_0^\varepsilon (1-t)\exp\left[-\frac{1}{1-t}\frac{A^t}{\ln A} - \omega(N)\frac{1}{(1-t)^2}\frac{A^t}{\ln^2 A}\left[1 + O\left(\frac{1}{\ln A}\right)\right]\right] dt, \tag{4.41}$$

where A is given in (4.20). Again, substituting $u = A^t/\ln A$ and invoking (4.21), (4.41) yields

$$\begin{aligned} I_{32} = I_{12} - \int_{1/\ln A}^{A^\varepsilon/\ln A} &\frac{\ln u + \ln(\ln A)}{\ln A} \exp\left[-\frac{u}{1 - \ln u/\ln A - \ln(\ln A)/\ln A}\right] \\ &\times \exp\left[-\frac{\omega(N)}{\ln A}\frac{u}{(1 - \ln u/\ln A - \ln(\ln A)/\ln A)^2}\right. \\ &\left.\times \left[1 + O\left(\frac{1}{\ln A}\right)\right]\right]\frac{du}{u \ln A}. \end{aligned} \tag{4.42}$$

https://doi.org/10.1239/aap/1331216649 Published online by Cambridge University Press

Again, using the notation $\delta = 1/\ln A$, (4.42) yields

$$\begin{aligned}
I_{32} &= (1+\delta\ln\delta)I_{12} \\
&\quad - \delta^2 \int_{\delta}^{\delta\exp(\varepsilon/\delta)} \frac{\ln u}{u} \exp\biggl[-\frac{u}{1-\delta\ln u+\delta\ln\delta} - \frac{\omega(N)u\delta(1+O(\delta))}{(1-\delta\ln u+\delta\ln\delta)^2}\biggr]\,\mathrm{d}u \\
&= (1+\delta\ln\delta)I_{12} \\
&\quad - \delta^2 \int_{\delta}^{1/\sqrt{\delta}} \frac{\ln u}{u} \exp\biggl[-\frac{u}{1-\delta\ln u+\delta\ln\delta} - \frac{\omega(N)u\delta(1+O(\delta))}{(1-\delta\ln u+\delta\ln\delta)^2}\biggr]\,\mathrm{d}u \\
&\quad - \delta^2 \int_{1/\sqrt{\delta}}^{\delta\exp(\varepsilon/\delta)} \frac{\ln u}{u} \exp\biggl[-\frac{u}{1-\delta\ln u+\delta\ln\delta} - \frac{\omega(N)u\delta(1+O(\delta))}{(1-\delta\ln u+\delta\ln\delta)^2}\biggr]\,\mathrm{d}u. \qquad (4.43)
\end{aligned}$$

First we deal with the second integral in (4.43) and get an upper bound as follows:

$$\begin{aligned}
&\int_{1/\sqrt{\delta}}^{\delta\exp(\varepsilon/\delta)} \frac{\ln u}{u} \exp\biggl[-\frac{u}{1-\delta\ln u+\delta\ln\delta} - \frac{\omega(N)u\delta}{(1-\delta\ln u+\delta\ln\delta)^2}\,(1+O(\delta))\biggr]\,\mathrm{d}u \\
&\quad = \int_{1/\sqrt{\delta}}^{\delta\exp(\varepsilon/\delta)} \frac{\ln u}{u} \exp\biggl[-\frac{u}{1-\delta\ln u+\delta\ln\delta} \\
&\qquad\qquad \times \biggl(1+\frac{\omega(N)}{1-\delta\ln u+\delta\ln\delta}\,\delta(1+O(\delta))\biggr)\biggr]\,\mathrm{d}u \\
&\quad \le \int_{1/\sqrt{\delta}}^{\delta\exp(\varepsilon/\delta)} \exp\biggl[-\frac{u\ (1+O(\delta))}{1-\delta\ln(1/\sqrt{\delta})+\delta\ln\delta}\biggr] \frac{\ln(1/\sqrt{\delta})\,\mathrm{d}u}{1/\sqrt{\delta}} \\
&\quad = O(\sqrt{\delta}\ln\delta\,\mathrm{e}^{-1/\sqrt{\delta}}). \qquad (4.44)
\end{aligned}$$

The first integral of (4.43) is

$$\begin{aligned}
K_2(\delta) &:= \int_{\delta}^{1/\sqrt{\delta}} \exp\biggl[-\frac{u}{1-\delta\ln u+\delta\ln\delta} - \frac{\omega(N)u\delta}{(1-\delta\ln u+\delta\ln\delta)^2}(1+O(\delta))\biggr]\frac{\ln u}{u}\,\mathrm{d}u \\
&= \int_{\delta}^{1/\sqrt{\delta}} \exp\biggl[-u\sum_{n=0}^{\infty}\Bigl(\delta\ln\frac{u}{\delta}\Bigr)^n - \omega(N)u\delta(1+O(\delta))\sum_{n=1}^{\infty} n\Bigl(\delta\ln\frac{u}{\delta}\Bigr)^{n-1}\biggr]\frac{\ln u}{u}\,\mathrm{d}u \\
&= \int_{\delta}^{1/\sqrt{\delta}} \mathrm{e}^{-u} \exp\biggl[-u\sum_{n=1}^{\infty}\Bigl(\delta\ln\frac{u}{\delta}\Bigr)^n\biggr] \\
&\qquad \times \exp\biggl[-\omega(N)u\delta(1+O(\delta))\sum_{n=1}^{\infty} n\Bigl(\delta\ln\frac{u}{\delta}\Bigr)^{n-1}\biggr]\frac{\ln u}{u}\,\mathrm{d}u.
\end{aligned}$$

We expand the exponentials above and obtain

$$\begin{aligned}
K_2(\delta) = \int_{\delta}^{1/\sqrt{\delta}} \frac{\mathrm{e}^{-u}\ln u}{u} &\biggl\{1 - u\sum_{n=1}^{\infty}\Bigl(\delta\ln\frac{u}{\delta}\Bigr)^n + O\biggl(u\sum_{n=1}^{\infty}\Bigl(\delta\ln\frac{u}{\delta}\Bigr)^n\biggr)^2\biggr\} \\
&\times \biggl\{1 - \omega(N)u\delta(1+O(\delta))\sum_{n=1}^{\infty}\Bigl(\delta\ln\frac{u}{\delta}\Bigr)^{n-1} \\
&\qquad + O\biggl(\omega(N)u\delta(1+O(\delta))\sum_{n=1}^{\infty}\Bigl(\delta\ln\frac{u}{\delta}\Bigr)^{n-1}\biggr)^2\biggr\}\,\mathrm{d}u
\end{aligned}$$

https://doi.org/10.1239/aap/1331216649 Published online by Cambridge University Press

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

Using exactly the same bounds as in Subsubsection 4.2.1 (see (4.25)), we obtain (as $\delta \to 0^+$)

$$\int_{1/\sqrt{\delta}}^{\infty} \frac{e^{-u}\ln u}{u}\left[1 - u\left(\delta \ln\frac{u}{\delta} + \omega(N)\delta(1+O(\delta))\right) + u^2 O\left(\delta^2 \ln^2\frac{u}{\delta}\right)\right] du$$
$$= O(\sqrt{\delta}\ln\delta e^{-1/\sqrt{\delta}}). \tag{4.45}$$

Hence, we can replace the upper limit of the integral $K_2(\delta)$ by ∞. Thus,

$$I_{32} = (1+\delta\ln\delta)I_{12}$$
$$- \delta^2 \int_{\delta}^{\infty} \frac{e^{-u}\ln u}{u}\left[1 - u\left(\delta \ln\frac{u}{\delta} + \omega(N)\delta(1+O(\delta))\right) + u^2 O\left(\delta^2 \ln^2\frac{u}{\delta}\right)\right] du \tag{4.46}$$

as $\delta \to 0^+$.

We now need two additional lemmas in the spirit of Lemmas 4.2 and 4.3. Their proofs are omitted since they are similar to the proof of Lemma 4.3.

Lemma 4.4. *For the integral*

$$L(x) := \int_x^{\infty} \frac{e^{-t}}{t}\ln t\, dt,$$

we have the asymptotic expansion, as $x \to 0^+$,

$$L(x) \sim -\frac{1}{2}\ln^2 x + \frac{1}{2}\left(\gamma^2 + \frac{\pi^2}{6}\right) + x\ln x - x - \frac{1}{4}x^2\ln x + \frac{1}{8}x^2 + \frac{1}{18}x^3\ln x - \frac{1}{54}x^3 + \cdots.$$

We mention that, in the proof of Lemma 4.4 we need to compute the quantity

$$C = \lim_{x\to 0^+}\left(\int_x^{\infty} \ln t \frac{e^{-t}}{t}\, dt + \frac{1}{2}\ln^2 x\right).$$

Integration by parts yields (see [5, p. 213])

$$2C = \int_0^{\infty} e^{-t}\ln^2 t\, dt = \Gamma''(1) = \gamma^2 + \frac{\pi^2}{6}.$$

Lemma 4.5. *For the integral*

$$M(x) := \int_x^{\infty} e^{-t}\ln^2 t\, dt,$$

we have the asymptotic expansion, as $x \to 0^+$,

$$M(x) \sim \left(\gamma^2 + \frac{\pi^2}{6}\right) - x\ln^2 x + 2x\ln x - 2x - \frac{1}{2}x^2\ln x + \frac{1}{4}x^2 + \frac{1}{9}x^3\ln x - \frac{1}{27}x^3 + \cdots.$$

Applying Lemmas 4.2, 4.3, 4.4, and 4.5 and using (4.28), I_{32} of (4.46) becomes

$$I_{32} = (1 + \delta \ln \delta) I_{12} - \delta^2 \biggl[-\frac{1}{2} \ln^2 \delta + \frac{1}{2} \biggl(\gamma^2 + \frac{\pi^2}{6} \biggr) + O(\delta \ln \delta) \biggr],$$

and by invoking (4.28) we arrive at

$$I_{32} = -\delta \ln \delta - \gamma \delta - \frac{1}{2} \delta^2 \ln^2 \delta + (1 - \gamma) \delta^2 \ln \delta + \biggl[1 + \gamma - \omega(N) - \frac{1}{2} \biggl(\gamma^2 + \frac{\pi^2}{6} \biggr) \biggr] \delta^2 + O(\delta^3 \ln^2 \delta). \tag{4.47}$$

Note that the error term in (4.47) dominates the terms of (4.44) and (4.45).

4.3.2. *The Integral $I_4(N)$*. In a similar way as in Subsubsection 4.2.2 (compare with (4.32)), for $\vartheta \in (0, 1)$, we have

$$\begin{aligned} &(1 - \vartheta)^2 \int_1^\infty \biggl(\int_1^N \mathrm{e}^{-F(N)s/f(x)} \, \mathrm{d}x \biggr) s \, \mathrm{d}s - (1 - \vartheta)^2 \int_1^\infty s \mathrm{e}^{-F(N)s/f(N+1)} \, \mathrm{d}s \\ &\quad < I_4(N) \\ &\quad < (1 + \vartheta)^2 \int_1^\infty \biggl(\int_1^N \mathrm{e}^{-F(N)s/f(x)} \, \mathrm{d}x \biggr) s \, \mathrm{d}s + (1 + \vartheta)^2 \int_1^\infty s \mathrm{e}^{-F(N)s/f(N+1)} \, \mathrm{d}s. \end{aligned} \tag{4.48}$$

Invoking (4.4) and applying Lemma 4.1 for $m = 1$ and $m = 2$, we obtain

$$\begin{aligned} &\int_1^\infty \biggl(\int_1^N \mathrm{e}^{-F(N)s/f(x)} \, \mathrm{d}x \biggr) s \, \mathrm{d}s \\ &\qquad = \frac{1}{F(N)} \int_1^N f(x) \mathrm{e}^{-F(N)/f(x)} \, \mathrm{d}x + \frac{1}{F(N)^2} \int_1^N f^2(x) \mathrm{e}^{-F(N)/f(x)} \, \mathrm{d}x \\ &\qquad = \biggl\{ \biggl(\frac{f(N)}{F(N)} \biggr)^2 + \biggl(\frac{f(N)}{F(N)} \biggr)^3 \biggr\} \biggl[1 + O\biggl(\frac{f(N)}{F(N)} \biggr) \biggr]. \end{aligned}$$

Furthermore, using (C1)–(C4), (4.2), the definition of F, and (4.22), we can easily check that

$$\begin{aligned} \int_1^\infty s \mathrm{e}^{-F(N)s/f(N+1)} \, \mathrm{d}s &= \frac{f(N+1)}{F(N)} \mathrm{e}^{-F(N)/f(N+1)} + \biggl[\frac{f(N+1)}{F(N)} \biggr]^2 \mathrm{e}^{-F(N)/f(N+1)} \\ &= o\biggl(\frac{f(N)^2}{F(N)^2} \biggr). \end{aligned}$$

Since $\vartheta \in (0, 1)$ is arbitrary, (4.48) implies that

$$I_4(N) = \biggl\{ \biggl(\frac{f(N)}{F(N)} \biggr)^2 + \biggl(\frac{f(N)}{F(N)} \biggr)^3 \biggr\} \biggl[1 + O\biggl(\frac{f(N)}{F(N)} \biggr) \biggr] \quad \text{as } N \to \infty.$$

Again, using the definition of F and (4.22), we obtain

$$I_4(N) = \delta^2 + \delta^3 + O(\delta^4) \quad \text{as } \delta \to 0^+. \tag{4.49}$$

We are now ready to present the following result.

https://doi.org/10.1239/aap/1331216649 Published online by Cambridge University Press

Theorem 4.3. *Let δ be as defined in (4.22) (hence, $\delta \to 0^+$ as $N \to \infty$), and let $\omega(N)$ be as given in (4.12). Then*

$$Q_N(\alpha) = f(N)^2 \left\{ \frac{1}{\delta^2} + \frac{2\ln\delta}{\delta} + \frac{2\gamma}{\delta} + \ln^2\delta + 2(\gamma - 1)\ln\delta + \left(2\omega(N) + \gamma^2 + \frac{\pi^2}{6} - 2\gamma\right) + O(\delta\ln^2\delta) \right\}.$$

Proof. The result follows immediately upon combining (4.36), (4.37), (4.38), (4.40), (4.47), and (4.49) with (4.35).

It follows (see (1.17)) that, as $\delta \to 0^+$, we have

$$\mathrm{E}[T_N(T_N+1)] = A_N^2 f(N)^2 \left\{ \frac{1}{\delta^2} + \frac{2\ln\delta}{\delta} + \frac{2\gamma}{\delta} + \ln^2\delta + 2(\gamma - 1)\ln\delta + \left(2\omega(N) + \gamma^2 + \frac{\pi^2}{6} - 2\gamma\right) + O(\delta\ln^2\delta) \right\}. \tag{4.50}$$

4.4. Conclusion: asymptotics of $V[T_N]$

We are now ready for our main result regarding the variance.

Theorem 4.4. *Let $\alpha = \{a_j\}_{j=1}^{\infty} = \{1/f(j)\}_{j=1}^{\infty}$, where f satisfies (4.1) and (C1)–(C4) (hence, $L_i(\alpha) = \infty$). Then, as $N \to \infty$, we have*

$$V[T_N] \sim \frac{\pi^2}{6}\, A_N^2\, f(N)^2 = \frac{\pi^2}{6}\frac{1}{p_N^2} = \frac{\pi^2}{6}\frac{1}{\min_{1\le j\le N}\{p_j\}^2},$$

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

Proof. From formulae (4.34) and (4.50) we obtain

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

In view of (1.6), in order to complete the proof, it only remains to show that

$$\frac{\mathrm{E}[T_N]}{A_N^2 f(N)^2} \to 0 \quad \text{as } N \to \infty. \tag{4.51}$$

From (4.34), (4.22), and (4.4), we have

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

Owing to the above, (4.51) is equivalent to

$$\frac{\ln f(N) - \ln f'(N)}{A_N f(N)} \to 0 \quad \text{as } N \to \infty. \tag{4.52}$$

Using (C1) (namely, $f(N) \to \infty$) and (4.3), we easily obtain the validity of (4.52), completing the proof of the theorem.

Remark 4.2. 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 $N \to \infty$, we have

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

5. Limit distributions

Neal [12] established two general limit theorems regarding T_N, where $\pi_N = \{p_1^N, p_2^N, \dots, p_N^N\}$, $N = 1, 2, \dots$, are arbitrary (sub)probability measures, not necessarily of the form (1.11).

Theorem 5.1. ([12, Theorem 2.1].) *Suppose that there exist sequences $\{b_N\}$ and $\{k_N\}$ such that $k_N/b_N \to 0$ as $N \to \infty$ and that, for $y \in \mathbb{R}$,*

$$S_N(y) := \sum_{j=1}^{N} \exp[-p_j^N(b_N + yk_N)] \to g(y) \quad \text{as } N \to \infty \tag{5.1}$$

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

$$\frac{T_N - b_N}{k_N} \xrightarrow{\mathrm{D}} Y \quad \text{as } N \to \infty, \tag{5.2}$$

where Y has distribution function $F(y) = \mathrm{P}\{Y \le y\} = \mathrm{e}^{-g(y)},\ y \in \mathbb{R}$.

Theorem 5.2. ([12, Theorem 2.2].) *Suppose that there exists a sequence $\{k_N\}$ such that, for $y \in \mathbb{R}^+$,*

$$\sum_{j=1}^{N} \exp[-p_j^N yk_N] \to \hat{g}(y) \quad \text{as } N \to \infty \tag{5.3}$$

for a nonincreasing function $\hat{g}(\cdot)$ with $\hat{g}(y) \to \infty$ as $y \to 0$ and $\hat{g}(y) \to 0$ as $y \to \infty$. Furthermore, suppose that there exists a function $h(\cdot)$ such that, for all $y \in \mathbb{R}^+$,

$$\prod_{j=1}^{N} (1 - \exp[-p_j^N yk_N]) \to h(y) \quad \text{as } N \to \infty.$$

Then (5.3) ensures that $h(y) \to 0$ as $y \to 0$ and $h(y) \to 1$ as $y \to \infty$, and

$$\frac{T_N}{k_N} \xrightarrow{\mathrm{D}} \hat{Y} \quad \text{as } N \to \infty,$$

where $\hat{Y}$ has distribution function $\hat{F}(y) = \mathrm{P}\{\hat{Y} \le y\} = h(y),\ y \in \mathbb{R}^+$.

Theorems 5.1 and 5.2 do not indicate how to choose the sequences $\{b_N\}$ and $\{k_N\}$. Here our asymptotic formulae can help.

Case 5.1. Conclusion (5.2) of Theorem 5.1 suggests that, as $N \to \infty$,

$$b_N \sim \mathrm{E}[T_N] \quad \text{and} \quad k_N \sim c\sqrt{V[T_N]} \quad \text{for some } c \neq 0.$$

We remind the reader that in the present work the coupon probabilities p_j^N, $1 \le j \le N$, $N = 1, 2, \dots$, are taken as

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

If $a_j = 1/f(j)$, where $f(x)$ satisfies (C1)–(C4), then, in view of (C1)–(C4), the asymptotic formula (4.34), together with Theorem 4.4, leads to the choices

$$b_N = A_N f(N)[\rho(N) - \ln \rho(N)] \quad \text{and} \quad k_N = A_N f(N), \tag{5.4}$$

where

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

(note that, as $N \to \infty$, $\rho(N) \to \infty$, and, hence, $k_N/b_N \to 0$ as required). Then, $S_N(y)$ in (5.1) becomes

$$S_N(y) := \sum_{j=1}^{N} \exp\left[-\frac{f(N)}{f(j)}[\rho(N) - \ln \rho(N) + y]\right].$$

Since $S_N(y) - S_{N-1}(y) = \exp[\rho(N) - \ln \rho(N) + y] \to 0$ and f is increasing, we have

$$S_N(y) \sim I_N(y) := \int_1^N \exp\left[-\frac{f(N)}{f(x)}[\rho(N) - \ln \rho(N) + y]\right] \mathrm{d}x \quad \text{as } N \to \infty.$$

Integration by parts gives

$$I_N(y) = \left[\frac{1}{M}\frac{f(N)^2}{f'(N)} \exp\left[-\frac{M}{f(x)}\right]\right]_{x=1}^{N} - \frac{1}{M}\int_1^N \left[\frac{f(x)^2}{f'(x)}\right]' \exp\left[-\frac{M}{f(x)}\right] \mathrm{d}x, \tag{5.5}$$

where, for typographical convenience, we have set

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

The integral on the right-hand side of (5.5) is $o(I_N(y))$. Hence,

$$I_N(y) \sim \frac{f(N)}{f'(N)} \frac{\exp[-\rho(N) + \ln \rho(N) - y]}{\rho(N) - \ln \rho(N) + y} \sim \mathrm{e}^{-y}.$$

It follows that $S_N(y) \to \mathrm{e}^{-y}$. Therefore, Theorem 5.1 implies that, for all $y \in \mathbb{R}$,

$$\mathrm{P}\left\{\frac{T_N - b_N}{k_N} \le y\right\} \to \exp(\mathrm{e}^{-y}) \quad \text{as } N \to \infty, \tag{5.6}$$

where $\{b_N\}$ and $\{k_N\}$ are given by (5.4). Note that the limiting distribution in (5.6) is the so-called standard Gumbel, independently of the choice of $f(x)$. In fact, the same limit distribution also arises for various other choices of coupon probabilities, including the case of equal p_j^Ns (see, e.g. [3], [7, p. 142], or [11]).

https://doi.org/10.1239/aap/1331216649 Published online by Cambridge University Press

Case 5.2. Regarding Theorem 5.2, we can see that the suggestions here are that, as $N \to \infty$,

$$\frac{\mathrm{E}[T_N]}{\sqrt{V[T_N]}} \to c_1 \in \mathbb{R} \quad \text{and} \quad k_N \sim c_2\sqrt{V[T_N]} \quad \text{for some } c_2 > 0.$$

For $p_j^N = a_j/A_N$, with $\{a_j\}$ satisfying (2.3) for some $\xi \in (0, 1)$, Theorem 3.1 indicates that the right choice for k_N is

$$k_N = A_N.$$

Then, Theorem 5.2 easily implies that, as $N \to \infty$,

$$\mathrm{P}\left\{\frac{T_N}{A_N} \le y\right\} \to \prod_{j=1}^{\infty}(1 - \mathrm{e}^{-a_j y}). \tag{5.7}$$

Note that here the limiting distribution depends on the choice of the sequence $\{a_j\}$.

Finally, let us mention that the dichotomy is, again, observed here: Case 5.1 versus Case 5.2. Note that in the first case we have $\mathrm{E}[T_N]/\sqrt{V[T_N]} \to \infty$, while in the second case we have $\mathrm{E}[T_N]/\sqrt{V[T_N]} \to c_1 \in \mathbb{R}$.

6. Examples

In this section we give several examples that illustrate the results of the previous sections.

Example 6.1. Let $a_j = j^p$, where $p > 0$. In this case (see Theorem 2.1)

$$L_1(\alpha) = \int_0^1 \left[1 - \prod_{j=1}^{\infty}(1 - x^{j^p})\right]\frac{\mathrm{d}x}{x} < \infty$$

and

$$L_2(\alpha) = (-2)\int_0^1 \left[1 - \prod_{j=1}^{\infty}(1 - x^{j^p})\right]\frac{\ln x}{x}\,\mathrm{d}x < \infty.$$

Hence, Theorem 3.1 gives

$$V[T_N] = \frac{N^{2(p+1)}}{(p+1)^2}(L_{2,p} - L_{1,p}^2)[1 + o(1)].$$

The case $p = 1$ is known as the *linear* case, and it is of particular interest. From the celebrated pentagonal-number formula of Euler (see, e.g. [1, p. 312]),

$$\prod_{j=1}^{\infty}(1 - x^j) = 1 + \sum_{k=1}^{\infty}(-1)^k[x^{\omega(k)} + x^{\omega(-k)}], \qquad \omega(k) = \frac{3k^2 - k}{2},\ k = 0, \pm1, \pm2, \dots,$$

we can compute

$$L_1(\alpha) = \sum_{k=1}^{\infty}\frac{12(-1)^{k+1}}{9k^2 - 1} = \frac{4\pi\sqrt{3}}{3} - 6, \qquad L_2(\alpha) = \sum_{k\in\mathbb{Z}^*}\frac{2(-1)^{k+1}}{\omega(k)^2} = 4(54 - 8\pi\sqrt{3} - \pi^2),$$

https://doi.org/10.1239/aap/1331216649 Published online by Cambridge University Press

where $\mathbb{Z}^* = \mathbb{Z} \setminus \{0\}$. Finally,

$$V[T_N] = \left(45 - 4\pi\sqrt{3} - \frac{7\pi^2}{3}\right) N^2(N+1)^2[1 + O(N^{-\lambda})] \quad \text{for any } \lambda \in (0,1),$$

where the error estimate can be found by exploiting the fact (see the proof of Theorem 14.3 of [1]) that

$$\left|\prod_{j=1}^{N}(1-x^j) - 1 - \sum_{k=1}^{N}(-1)^k[x^{\omega(k)} + x^{\omega(-k)}]\right| \le N x^{N+1}.$$

Example 6.2. Let $a_j = \mathrm{e}^{pj}$ and $b_j = \mathrm{e}^{-pj}$, $p > 0$. For the sequence $\alpha = \{a_j\}_{j=0}^{\infty}$, we have $L_i(\alpha) < \infty$, $i \in \{1, 2\}$. It follows that

$$V[T_N] = \left(\frac{\mathrm{e}^{p(N+1)}}{\mathrm{e}^p - 1}\right)^2 (L_2(\alpha) - L_1(\alpha)^2) + O(\mathrm{e}^{pN}). \tag{6.1}$$

The special case $a_j = 2^j$ (i.e. $p = \ln 2$) is of particular interest. We have

$$\phi(x) := \prod_{j=0}^{\infty}(1 - x^{2^j}) = \sum_{k=0}^{\infty}(-1)^{\delta(k)} x^k = 1 - \sum_{n=0}^{\infty}(1-x)^n x^{2^n},$$

where $\delta(k)$ is the number of 1s in the binary expansion of k (the last equality follows from the observation that $\phi(x) = (1-x)\phi(x^2)$). Then (2.1) and (2.2) give

$$L_1(\alpha) = \sum_{k=1}^{\infty} \frac{(-1)^{\delta(k)-1}}{k} = \sum_{n=0}^{\infty} \frac{n!\,(2^n - 1)!}{(n + 2^n)!}$$

(the second series converges extremely rapidly) and

$$\begin{aligned} L_2(\alpha) &= 2\sum_{k=1}^{\infty} \frac{(-1)^{\delta(k)-1}}{k^2} \\ &= 2\sum_{n=0}^{\infty}\sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{(k+2^n)^2} \\ &= 2\sum_{n=0}^{\infty} \frac{n!\,(2^n-1)!}{(n+2^n)!}(H_{n+2^n} - H_{2^n - 1}), \end{aligned}$$

where, as usual, $H_m = \sum_{k=1}^{m} 1/k$. The last two series above converge extremely rapidly.

Let us now discuss the sequence $\beta = \{b_j\}_{j=0}^{\infty}$. Here we have $L_i(\beta) = \infty$. Furthermore, $f(x) = \mathrm{e}^{px}$ does not satisfy (C2); thus, Theorems 4.1–4.4 cannot be applied. However, the sequences α and β produce the same coupon probabilities. This follows from the fact that, for each N, if we let $c_N = \mathrm{e}^{pN}$ then $\{a_j : 0 \le j \le N\} = \{c_N b_j : 0 \le j \le N\}$, i.e. the elements of the two truncated sequences are proportional to each other. Hence, regarding β, the asymptotics of $V[T_N]$ are also given by (6.1). Note that Theorem 4.4 catches the order of magnitude of $V[T_N]$ modulo a constant factor.

https://doi.org/10.1239/aap/1331216649 Published online by Cambridge University Press

Example 6.3. Let $a_j = 1/j^p$, $p > 0$ (note that $p = 1$ corresponds to the so-called *Zipf* distribution—see [4] and [9] for results regarding $\mathrm{E}[T_N]$). Here $L_1(\alpha) = L_2(\alpha) = \infty$, and, furthermore, $f(x) = x^p$ satisfies the (C1)–(C4); hence, we can apply Theorem 4.4 to obtain

$$V[T_N] \sim \frac{\pi^2}{6}\frac{N^2}{(1-p)^2} \quad \text{if } 0 < p < 1, \qquad V[T_N] \sim \frac{\pi^2}{6}\zeta(p)^2 N^{2p} \quad \text{if } p > 1,$$

where $\zeta(\cdot)$ is the zeta function. As for $p = 1$ (the Zipf case),

$$V[T_N] \sim \frac{\pi^2}{6} N^2 \ln^2 N.$$

Example 6.4. Let $a_j = j!$. Here $L_1(\alpha), L_2(\alpha) < \infty$. Also, Stirling's formula implies that $A_N \sim N!$. Hence, Theorem 3.1 yields

$$V[T_N] \sim (L_2(\alpha) - L_1(\alpha)^2)(N!)^2 \quad \text{as } N \to \infty.$$

Appendix A

Here we give the proofs of certain technical theorems and lemmas appearing in Section 4.

Proof of Theorem 4.1. We can write (1.16) as

$$\begin{aligned} Q_N(\alpha) &= F(N)^2 Q_N(F(N)\alpha) \\ &= 2F(N)^2 \int_0^1 \biggl[1 - \exp\biggl[\sum_{j=1}^N \ln(1 - \mathrm{e}^{-F(N)s/f(j)})\biggr]\biggr] s\,\mathrm{d}s \\ &\quad + 2F(N)^2 \int_1^\infty \biggl[1 - \exp\biggl[\sum_{j=1}^N \ln(1 - \mathrm{e}^{-F(N)s/f(j)})\biggr]\biggr] s\,\mathrm{d}s, \end{aligned} \tag{A.1}$$

where F is defined by formula (4.4). It has been established in [4] that, under conditions (C1)–(C4),

$$\lim_N \sum_{j=1}^N \ln(1 - \mathrm{e}^{-F(N)s/f(j)}) = \begin{cases} -\infty & \text{if } s < 1, \\ 0 & \text{if } s \geq 1, \end{cases} \tag{A.2}$$

and also that

$$\int_1^N \mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x \sim \frac{1}{s\ln[f(N)/f'(N)]}\biggl[\frac{f(N)}{f'(N)}\biggr]^{1-s}. \tag{A.3}$$

Applying the bounded convergence theorem to the first integral of (A.1) yields (in view of (A.2))

$$Q_N(\alpha) = 2F(N)^2\biggl[\frac{1}{2}+o(1)\biggr] + 2F(N)^2 \int_1^\infty \biggl[1-\exp\biggl[\sum_{j=1}^N \ln(1-\mathrm{e}^{-F(N)s/f(j)})\biggr]\biggr] s\,\mathrm{d}s. \tag{A.4}$$

Next, we want to estimate the integral appearing in the above formula. We begin by noting that, by the dominated convergence theorem (since $f(N)/f'(N) \to \infty$),

$$\lim_N \int_1^\infty \biggl[1 - \exp\biggl[-\frac{(f(N)/f'(N))^{1-s}}{s\ln(f(N)/f'(N))}\biggr]\biggr] s\,\mathrm{d}s = 0.$$

https://doi.org/10.1239/aap/1331216649 Published online by Cambridge University Press

Using (A.3), this implies that

$$\lim_N \int_1^\infty \left[1 - \exp\left[-\int_1^N \mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x\right]\right] s\,\mathrm{d}s = 0. \tag{A.5}$$

Since f is increasing, we have

$$\begin{aligned}\int_1^N \mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x &\leq \sum_{j=1}^N \mathrm{e}^{-F(N)s/f(j)} \\ &\leq \int_1^{N+1} \mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x \\ &\leq \int_1^N \mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x + \mathrm{e}^{-F(N)s/f(N+1)}. \end{aligned}\tag{A.6}$$

From the above inequalities, it follows that

$$\begin{aligned}1 - \exp\left[-\int_1^N \mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x\right] &\leq 1 - \exp\left[-\sum_{j=1}^N \mathrm{e}^{-F(N)s/f(j)}\right] \\ &\leq 1 - \exp\left[-\int_1^N \mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x + \mathrm{e}^{-F(N)s/f(N+1)}\right]. \end{aligned}\tag{A.7}$$

However, by (A.3),

$$\lim_N \int_1^N \mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x = \begin{cases}\infty & \text{if } s < 1, \\ 0 & \text{if } s \geq 1.\end{cases}$$

Hence, by taking limits in (A.7) and using (4.2) and (A.5), we obtain

$$\lim_N \int_1^\infty \left[1 - \exp\left[\sum_{j=1}^N \ln(1 - \mathrm{e}^{-F(N)s/f(j)})\right]\right] s\,\mathrm{d}s = 0. \tag{A.8}$$

Finally, by the definition of $F(N)$ and the Taylor expansion of $\ln(1-x)$ as $x \to 0$, (A.4) yields

$$Q_N(\alpha) \sim F(N)^2 = f(N)^2 \ln\left(\frac{f(N)}{f'(N)}\right)^2 \quad \text{as } N \to \infty,$$

completing the proof.

Proof of Lemma 4.1. Integration by parts gives

$$\begin{aligned}\int_1^N f(x)^m \mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x = &\left[\frac{f(x)^{m+2}\mathrm{e}^{-F(N)s/f(x)}}{sF(N)f'(x)}\right]_{x=1}^N \\ &- \int_1^N \frac{\mathrm{e}^{-F(N)s/f(x)}}{sF(N)}\left[\frac{f(x)^{m+2}}{f'(x)}\right]'\,\mathrm{d}x.\end{aligned}$$

https://doi.org/10.1239/aap/1331216649 Published online by Cambridge University Press

Now,

$$\int_1^N \frac{\mathrm{e}^{-F(N)s/f(x)}}{sF(N)}\left[\frac{f(x)^{m+2}}{f'(x)}\right]' \mathrm{d}x = \frac{m+2}{s}\int_1^N \frac{f(x)^{m+1}}{F(N)}\mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x - \frac{1}{s}\int_1^N \frac{f''(x)/f'(x)}{f'(x)/f(x)}\,\frac{f(x)^{m+1}}{F(N)}\mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x. \tag{A.9}$$

Since f is increasing, we have

$$\begin{aligned}\int_1^N \frac{f(x)^{m+1}}{F(N)}\mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x &\le \frac{f(N)}{F(N)}\int_1^N f(x)^m\mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x\\ &= \frac{f(N)}{F(N)}J_m(N)\\ &= o(J_m(N)).\end{aligned}$$

From (C1)–(C4) we also have

$$\int_1^N \frac{f''(x)/f'(x)}{f'(x)/f(x)}\,\frac{f(x)^{m+1}}{F(N)}\mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x = \left[\int_1^N \frac{f(x)^{m+1}}{F(N)}\mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x\right]O(1),$$

which completes the proof.

Proof of Corollary 4.1. Integration by parts in (A.9) gives

$$\begin{aligned}&\frac{m+2}{s}\int_1^N \frac{f(x)^{m+1}}{F(N)}\mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x - \frac{1}{s}\int_1^N \frac{f''(x)/f'(x)}{f'(x)/f(x)}\,\frac{f(x)^{m+1}}{F(N)}\mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x\\ &\quad= (m+2)\left[\frac{f(x)^{m+3}\mathrm{e}^{-F(N)s/f(x)}}{s^2F(N)^2f'(x)}\right]_{x=1}^N - (m+2)\int_1^N \frac{\mathrm{e}^{-F(N)s/f(x)}}{s^2F(N)^2}\left[\frac{f(x)^{m+3}}{f'(x)}\right]'\mathrm{d}x\\ &\qquad - \left[\frac{f(x)^{m+3}\mathrm{e}^{-F(N)s/f(x)}}{s^2F(N)^2f'(x)}\,\frac{f''(x)/f'(x)}{f'(x)/f(x)}\right]_{x=1}^N\\ &\qquad + \int_1^N \frac{\mathrm{e}^{-F(N)s/f(x)}}{s^2F(N)^2}\left[\frac{f(x)^{m+3}}{f'(x)}\,\frac{f''(x)/f'(x)}{f'(x)/f(x)}\right]'\mathrm{d}x.\end{aligned}$$

Now,

$$\begin{aligned}\int_1^N \frac{\mathrm{e}^{-F(N)s/f(x)}}{s^2F(N)^2}\left[\frac{f(x)^{m+3}}{f'(x)}\right]'\mathrm{d}x &= (m+3)\int_1^N \frac{f(x)^{m+2}}{s^2F(N)^2}\mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x\\ &\quad - \int_1^N \frac{f''(x)/f'(x)}{f'(x)/f(x)}\,\frac{f(x)^{m+2}}{s^2F(N)^2}\mathrm{e}^{-F(N)s/f(x)}\,\mathrm{d}x.\end{aligned}$$

Using the assumption that f is increasing and applying (C3) we obtain

$$\int_1^N \frac{\mathrm{e}^{-F(N)s/f(x)}}{s^2F(N)^2}\left[\frac{f(x)^{m+3}}{f'(x)}\right]'\mathrm{d}x = O\left(\frac{f(N)^2}{F(N)^2}J_m(N)\right) = o(J_m(N)).$$

https://doi.org/10.1239/aap/1331216649 Published online by Cambridge University Press

Also,

$$\int_1^N \frac{\mathrm{e}^{-F(N)s/f(x)}}{s^2 F(N)^2}\left[\frac{f(x)^{m+3}}{f'(x)}\,\frac{f''(x)/f'(x)}{f'(x)/f(x)}\right]' \mathrm{d}x$$
$$= \int_1^N \frac{\mathrm{e}^{-F(N)s/f(x)}}{s^2 F(N)^2}\,\frac{f''(x)/f'(x)}{f'(x)/f(x)}\left[\frac{f(x)^{m+3}}{f'(x)}\right]' \mathrm{d}x$$
$$+ \int_1^N \frac{\mathrm{e}^{-F(N)s/f(x)}}{s^2 F(N)^2}\,\frac{f(x)^{m+3}}{f'(x)}\left[\frac{f''(x)/f'(x)}{f'(x)/f(x)}\right]' \mathrm{d}x,$$

and

$$\int_1^N \frac{\mathrm{e}^{-F(N)s/f(x)}}{s^2 F(N)^2}\,\frac{f(x)^{m+3}}{f'(x)}\left[\frac{f''(x)/f'(x)}{f'(x)/f(x)}\right]' \mathrm{d}x$$
$$= \int_1^N \frac{f(x)^2 f'''(x)}{(f'(x))^3} f(x)^{m+2}\frac{\mathrm{e}^{-F(N)s/f(x)}}{s^2 F(N)^2}\,\mathrm{d}x$$
$$+ \int_1^N \frac{f''(x)/f'(x)}{f'(x)/f(x)} f(x)^{m+2}\frac{\mathrm{e}^{-F(N)s/f(x)}}{s^2 F(N)^2}\,\mathrm{d}x$$
$$- 2\int_1^N \left(\frac{f''(x)/f'(x)}{f'(x)/f(x)}\right)^2 f(x)^{m+2}\frac{\mathrm{e}^{-F(N)s/f(x)}}{s^2 F(N)^2}\,\mathrm{d}x.$$

Hence, from (C1)–(C4) we have

$$\int_1^N \frac{\mathrm{e}^{-F(N)s/f(x)}}{s^2 F(N)^2}\,\frac{f(x)^{m+3}}{f'(x)}\left[\frac{f''(x)/f'(x)}{f'(x)/f(x)}\right]' \mathrm{d}x = O\left(\frac{f(N)^2}{F(N)^2} J_m(N)\right) = o(J_m(N)).$$

Thus,

$$J_m(N) = \frac{f(N)^{m+2}}{sF(N)f'(N)}\mathrm{e}^{-F(N)s/f(N)s} + \omega(N)\frac{f(N)^{m+3}}{s^2F(N)^2 f'(N)}\mathrm{e}^{-F(N)s/f(N)} + o(J_m(N)),$$

and the proof is completed by invoking Lemma 4.1 (note that from (C1)–(C4) we immediately obtain $\omega(N) = O(1)$ as $N \to \infty$).

Proof of Lemma 4.3. Since

$$\frac{\mathrm{d}G(x)}{\mathrm{d}x} = -\ln x\mathrm{e}^{-x} = -\ln x\left(1 - x + \frac{1}{2}x^2 - \frac{1}{6}x^3 + \cdots\right),$$

we have

$$G(x) \sim C_1 - x\ln x + x + \tfrac{1}{2}x^2 \ln x - \tfrac{1}{2}x^2 - \tfrac{1}{6}x^3 \ln x + \tfrac{1}{6}x^3 + \tfrac{1}{24}x^4 \ln x - \tfrac{1}{24}x^4 + \cdots, \quad \text{(A.10)}$$

where C_1 is a constant. Next we compute C_1. From (A.10) we see that

$$C_1 = \int_0^\infty \ln(t)\mathrm{e}^{-t}\,\mathrm{d}t = \Gamma'(1) = -\gamma$$

(see [5, p. 213]), and the proof is completed.

Acknowledgements

The authors wish to thank the anonymous referee for reading the manuscript carefully, for making various constructive comments and suggestions, and for bringing Neal's paper [12] to their attention. This work was partially supported by a Π.E.B.E. Grant of the National Technical University of Athens.

References

[1] Apostol, T. M. (1976). *Introduction to Analytic Number Theory*. Springer, New York.
[2] Bender, C. M. and Orszag, S. A. (1999). *Advanced Mathematical Methods for Scientists and Engineers*. Springer, New York.
[3] Boneh, A. and Hofri, M. (1997). The coupon-collector problem revisited—a survey of engineering problems and computational methods. *Commun. Statist. Stoch. Models* **13,** 39–66.
[4] Boneh, S. and Papanicolaou, V. G. (1996). General asymptotic estimates for the coupon collector problem. *J. Comput. Appl. Math.* **67,** 277–289.
[5] Boros, G. and Moll, V. (2004). *Irresistible Integrals*. Cambridge University Press.
[6] Brayton, R. K. (1963). On the asymptotic behavior of the number of trials necessary to complete a set with random selection. *J. Math. Anal. Appl.* **7,** 31–61.
[7] Durrett, R. (2005). *Probability: Theory and Examples*, 3rd edn. Cambridge University Press.
[8] Feller, W. (1966). *An Introduction to Probability Theory and Its Applications*, Vol. I, John Wiley, New York.
[9] Flajolet, P., Gardy, D. and Thimonier, L. (1992). Birthday paradox, coupon collectors, caching algorithms and self-organizing search. *Discrete Appl. Math.* **39,** 207–229.
[10] Hildebrand, M. V. (1993). The birthday problem. *Amer. Math. Monthly* **100,** 643.
[11] Holst, L., Kennedy, J. E. and Quine, M. P. (1988). Rates of Poisson convergence for some coverage and urn problems using coupling. *J. Appl. Prob.* **25,** 717–724.
[12] Neal, P. (2008). The generalised coupon collector problem. *J. Appl. Prob.* **45,** 621–629.
[13] Ross, S. (2006). *A First Course in Probability*, 7th edn. Pearson Prentice Hall.
[14] Rudin, W. (1987). *Real and Complex Analysis*. McGraw-Hill, New York.