Asymptotic Analysis of the Moments for the Coupon Collector’s Problem

Aristides V. Doumas
(joint work with Vassilis G. Papanicolaou)

National Technical University of Athens, Department of Applied Mathematics

Patra, May 2012
Outline

1. Introduction
   - History of the problem
   - Preliminaries

2. The equally likely case

3. Large $N$ asymptotics

4. The dichotomy

5. The case $L_r(\alpha) < \infty$

6. The case $L_r(\alpha) = \infty$
   - The leading behavior of the moments of $T_N$
   - Asymptotic estimates for the moments of $T_N$ by comparison with known sequences

7. Examples
The Coupon Collector’s Problem

Consider a population whose members are of $N$ different types (e.g. baseball cards).
The Coupon Collector’s Problem

- Consider a population whose members are of $N$ different types (e.g. baseball cards).
- For $1 \leq j \leq N$ we denote by $p_j$ the probability that a member of the population is of type $j$. 
The Coupon Collector’s Problem

- Consider a population whose members are of \( N \) different types (e.g. baseball cards).
- For \( 1 \leq j \leq N \) 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 Coupon Collector’s Problem

- Consider a population whose members are of $N$ different types (e.g. baseball cards).
- For $1 \leq j \leq N$ 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’s Problem” (CCP) deals with questions arising in the above procedure.
History of the Problem

- In particular, CCP pertains to the family of Urn problems. Other classical such problems are birthday, dixie cup or occupancy problems, whose origin can be traced back to De Moivre’s treatise *De Mensura Sortis* of 1712 (see, e.g., Holst, International Statistical Review (1986)).
- CCP (in its simplest form, i.e. the case of equal probabilities) had appeared in W. Feller’s classical work and has attracted the attention of various researchers, since it has found many applications in many areas of science (computer science–search algorithms, mathematical programming, optimization, learning processes, engineering, ecology, as well as linguistics – see, e.g., P. Erdős and A. Rényi, (1961), A. Boneh and M. Hofri (1997), P. Flajolet, D. Gardy and L. Thimonier (1992), A. Torn and A. Zallinskas (1999), and M. Karwan *et al* (1993)).
History of the Problem

- With $T_N$ we denote the number of trials it takes until all $N$ types are detected (at least once).
- Apart from the distribution (see P. Neal (2008), D–V.G. Papanicolaou (2012)), some key quantities are the moments of the number $T_N$.
- For the case of equal sampling probabilities the first and the second moment of $T_N$ are well known.
- Furthermore, asymptotics and limiting results have been obtained by several authors (see for instance S. Boneh and V.G. Papanicolaou (1996), R. Durrett (2005), L. Holst (1986), S. Janson (1983) and D.J. Newman and L. Shepp (1960).
- For unequal probabilities, general asymptotic estimates regarding the first and the second moment, as well as for the variance, have been obtained by several authors (see R.K. Brayton (1963), D–V.G. Papanicolaou (2012)).
Preliminaries

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

$$P \{ T_N \geq k \} = P \left( A_1^{k-1} \cup \ldots \cup A_N^{k-1} \right), \ k = 1, 2, \ldots.$$

By invoking the inclusion-exclusion principle one gets

$$P \{ T_N \geq k \} = \sum_{J \subset \{1, \ldots, N\}, J \neq \emptyset} (-1)^{|J|-1} \left[ 1 - \left( \sum_{j \in J} \rho_j \right) \right]^{k-1}, \ k = 1, 2, \ldots, \tag{1}$$

where the sum extends over all $2^N - 1$ nonempty subsets $J$ of $\{1, \ldots, N\}$, while $|J|$ denotes the cardinality of $J$. 
Preliminaries

For $z \in \mathbb{C}, |z| \geq 1$, we introduce the following moment generating function of $T_N$,

$$G(z) := E \left[ z^{-T_N} \right] = 1 + \left( z^{-1} - 1 \right) \sum_{k=1}^{\infty} z^{-(k-1)} P \{ T_N \geq k \} \quad (2)$$

(the second equality follows by partial summation).

Consequently, by using (1) one arrives at

$$G(z) = 1 + \left( z^{-1} - 1 \right) \sum_{J \subset \{1, \ldots, N\}} (-1)^{|J|-1} \sum_{k=1}^{\infty} z^{-(k-1)} \left[ 1 - \left( \sum_{j \in J} p_j \right) \right]^k$$

hence, by summing the geometric series

$$G(z) = 1 - (z - 1) \sum_{J \subset \{1, \ldots, N\}} \frac{(-1)^{|J|-1}}{z - 1 + \left( \sum_{j \in J} p_j \right)}.$$
Preliminaries

We proceed by noticing that

$$\prod_{j=1}^{N} (1 - e^{-p_j t}) = \sum_{J \subset \{1, \ldots, N\}} (-1)^{|J|} \exp \left(-t \sum_{j \in J} p_j \right). \quad (4)$$

Thus, at least for $\Re\{z\} \geq 1$,

$$\int_0^\infty \left[ 1 - \prod_{j=1}^{N} (1 - e^{-p_j t}) \right] e^{-(z-1)t} \, dt = \sum_{J \subset \{1, \ldots, N\}, J \neq \emptyset} \frac{(-1)^{|J|-1}}{z - 1 + \left( \sum_{j \in J} p_j \right)}.$$  \quad (5)

Finally, by comparing (3) and (5) we get

$$G(z) = 1 - (z - 1) \int_0^\infty \left[ 1 - \prod_{j=1}^{N} (1 - e^{-p_j t}) \right] e^{-(z-1)t} \, dt, \quad (6)$$

or, equivalently, by substituting $x = e^{-t}$ in the integral,
Preliminaries

\[ G(z) = 1 - (z - 1) \int_0^1 \left[ 1 - \prod_{j=1}^N (1 - x^p_j) \right] x^{z-2} dx. \quad (7) \]

Let \( r \geq 1 \), be an integer. We consider the following \( r \)-th moment of \( T_N \)

\[ E \left[ T_N^{(r)} \right] := E \left[ T_N (T_N + 1) (T_N + 2) \cdots (T_N + r - 1) \right]. \]

Observe that,

\[ E \left[ T_N^{(r)} \right] = (-1)^r \lim_{z \to 1^+} G^{(r)}(z). \]
Thus, we arrive at the formulas (H.V. Schelling, *Coupon Collecting for Unequal Probabilities*, Amer. Math. Monthly (1954))

\[
E \left[ T_N^{(r)} \right] = r \sum_{m=1}^{N} (-1)^{m-1} \sum_{1 \leq j_1 < \cdots < j_m \leq N} \frac{1}{(p_{j_1} + \cdots + p_{j_m})^r} \\
= r \int_{0}^{\infty} \left[ 1 - \prod_{j=1}^{N} (1 - e^{-p_j t}) \right] t^{r-1} dt \\
= (-1)^{r-1} r \int_{0}^{1} \left[ 1 - \prod_{j=1}^{N} (1 - x^{p_j}) \right] \ln(x)^{r-1} \frac{dx}{x}. \quad (8)
\]
Outline

1. Introduction
   - History of the problem
   - Preliminaries

2. The equally likely case

3. Large $N$ asymptotics

4. The dichotomy

5. The case $L_r(\alpha) < \infty$

6. The case $L_r(\alpha) = \infty$
   - The leading behavior of the moments of $T_N$
   - Asymptotic estimates for the moments of $T_N$ by comparison with known sequences

7. Examples
The equally likely case

Naturally, the simplest case regarding the previous formulas occurs when one takes

\[ p_1 = \cdots = p_N = \frac{1}{N}. \]  \hspace{1cm} (9)

Actually, this case apart from its simplicity, has the property that among all sequences, it is the one with the smallest moments of \( T_N \). This is a known result (see M.V. Hildebrand (1993), A. Boneh and M. Hofri (1997)). For example, (7) and (8) imply immediately that

\[ G(z) = E \left[ z^{-T_N} \right], \quad \text{where} \quad z \geq 1, \]

attains its maximum value, while \( E \left[ T_N^{(r)} \right] \) attain their minimum values, when all \( p_j \)'s are equal. In particular, we have

\[ G(z) = N! \frac{\Gamma \left( (z - 1) N + 1 \right)}{\Gamma (zN + 1)}, \]
The equally likely case

\[ E \left[ T_N^{(r)} \right] = (-1)^{r-1} r \int_0^1 \left[ 1 - \left(1 - x^{1/N}\right)^N \right] \ln \frac{x}{x} \, dx. \quad (10) \]

Substituting \( u = 1 - x^{1/N} \) in the integral of (10) and after repeated integration by parts one gets

\[ E \left[ T_N^{(r)} \right] = r! \, N^r \sum_{m=1}^{N} \left( \frac{1}{m} \sum_{m_1=1}^{m} \left( \frac{1}{m_1} \sum_{m_2=1}^{m_1} \frac{1}{m_2} \cdots \right) \sum_{m_{r-1}=1}^{m_{r-2}} \frac{1}{m_{r-1}} \right). \quad (11) \]

It seems that formulas for \( E \left[ T_N^{(r)} \right] \) had been first obtained in H.J. Goodwin, On cartophily and motor cars, *Math. Gazette* (1949). Notice that the term cartophily first appeared in G. Maunsell, *Math. Gazette* (1938)!. To continue we consider the recursive formula

\[ \alpha_1^{(N)} = \sum_{m=1}^{N} \frac{1}{m}, \quad \alpha_r^{(N)} = \sum_{m=1}^{N} \frac{1}{m} \alpha_{r-1}^{(m)}. \]
The equally likely case

Foata et al (2001), called these numbers *hyperharmonic* and derived asymptotics using multivariate generating functions. Soon after, Adler et al (2003), gave explicit expression for the hyperharmonic numbers using basic probability arguments. It is remarkable that the study of these numbers appeared in a much more general version of the classical CCP. In particular, we have (see, Foata et al)

\[
\alpha_r^{(N)} \sim \frac{(\ln N)^r}{r!} \quad \text{as} \quad N \to \infty,
\]

hence (11) yields

\[
E \left[ T_N^{(r)} \right] \sim N^r (\ln N)^r \quad \text{as} \quad N \to \infty. \quad (12)
\]
The equally likely case

For fixed $r$, one also has explicit asymptotics for $E \left[ T_N^{(r)} \right]$. If for example, $r = 3$ then, either from Foata et al, or with a little patience and paper after repeated application of Abel partial summation method, one finally arrives at

$$E \left[ T_N^{(3)} \right] = N^3 \left[ \ln^3 N + 3\gamma \ln^2 N + \left( 3\gamma^2 + \frac{\pi^2}{2} \right) \ln N ight.$$

$$+ \left( 2\zeta(3) + \gamma^3 + \frac{\gamma \pi^2}{2} \right) + O \left( \frac{\ln N}{N} \right) \right].$$

A part of this particular case appeared in the American Mathematical Monthly (problem [4582], proposed by M.S. Klamkin (1954); published solution by Leonard Carlitz).
Outline

1. Introduction
   - History of the problem
   - Preliminaries

2. The equally likely case

3. **Large N asymptotics**

4. The dichotomy

5. The case $L_r(\alpha) < \infty$

6. The case $L_r(\alpha) = \infty$
   - The leading behavior of the moments of $T_N$
   - Asymptotic estimates for the moments of $T_N$ by comparison with known sequences

7. Examples
Large $N$ asymptotics

When $N$ is large it is not obvious at all what information one can obtain for $E \left[T_{N}^{(r)}\right]$, from the formula (8). For this reason there is a need to develop efficient ways for deriving asymptotics as $N \to \infty$ (we have already analyzed the very special case of equal probabilities—see formulas (12)–(13).

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 $\pi_N = \{\rho_1, \ldots, \rho_N\}$ on the set $\{1, \ldots, N\}$ by taking

$$p_j = \frac{a_j}{A_N}, \quad \text{where} \quad A_N = \sum_{j=1}^{N} a_j. \quad (14)$$
Large $N$ asymptotics

When $N$ is large it is not obvious at all what information one can obtain for $E \left[ T_N^{(r)} \right]$, from the formula (8). For this reason there is a need to develop efficient ways for deriving asymptotics as $N \to \infty$ (we have already analyzed the very special case of equal probabilities—see formulas (12)–(13).

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 $\{1, ..., N\}$ by taking

$$p_j = \frac{a_j}{A_N}, \quad \text{where} \quad A_N = \sum_{j=1}^{N} a_j. \quad (14)$$

Notice that $p_j$ depends on $\alpha$ and $N$, thus, given $\alpha$, it makes sense to consider the asymptotic behavior of $E \left[ T_N^{(r)} \right]$ as $N \to \infty$. This way of producing sequences of probability measures first appeared in [BP] (1996).
Large $N$ asymptotics

Motivated by (8) we introduce the notation (see [BP])

$$H_N(\alpha; r) := r \sum_{J \subset \{1, \ldots, N\}, J \neq \emptyset} \frac{(-1)^{|J|-1}}{(\sum_{j \in J} a_j)^r} = r \sum_{k=1}^{N} (-1)^{k-1} \sum_{1 \leq j_1 < \cdots < j_k \leq N} \frac{1}{(a_{j_1} + \cdots + a_{j_k})^r}. \tag{15}$$

Thus,

$$H_N(\alpha; r) = r \int_0^\infty \left[ 1 - \prod_{j=1}^{N} \left( 1 - e^{-a_j t} \right) \right] t^{r-1} \, dt$$

$$= (-1)^{r-1} r \int_0^1 \left[ 1 - \prod_{j=1}^{N} \left( 1 - x a_j \right) \right] \ln(x)^{r-1} \frac{dx}{x}. \tag{16}$$

If $s\alpha := \{sa_j\}_{j=1}^\infty$, then (15) gives immediately that

$$H_N(s\alpha; r) = s^{-r} H_N(\alpha; r) \tag{17}$$

and hence, in view of (8) and (14),
\[ E \left[ T_N^{(r)} \right] = A_N^r H_N(\alpha; r). \] (18)

As it was noticed in [BP], and D–Papanicolaou, for \( E \left[ T_N \right] \), the problem of estimating \( E \left[ T_N^{(r)} \right] \) as \( N \to \infty \), can be treated as two separate problems, namely estimating \( A_N^r \) and estimating \( H_N(\alpha; r) \).

- Our analysis focuses on estimating \( H_N(\alpha; r) \).
- The estimation of \( A_N^r \) will be considered an external matter which can be handled by existing powerful methods, such as the Euler-Maclaurin Summation formula, the Laplace method for sums (see, e.g., 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)), or even summation by parts.
Outline

1. Introduction
   - History of the problem
   - Preliminaries

2. The equally likely case

3. Large $N$ asymptotics

4. The dichotomy

5. The case $L_r(\alpha) < \infty$

6. The case $L_r(\alpha) = \infty$
   - The leading behavior of the moments of $T_N$
   - Asymptotic estimates for the moments of $T_N$ by comparison with known sequences

7. Examples
The dichotomy

For convenience, we denote

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

The following properties of the functions \( f_N^\alpha \) are immediate:

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

\[ \lim_{N \to \infty} f_N^\alpha(x) = \prod_{j=1}^{\infty} (1 - x^{a_j}) \text{ exists.} \]
The dichotomy

Thus, by applying the Monotone Convergence Theorem in (16) we get

\[
L_r(\alpha) := \lim_{N} H_N(\alpha; r) = (-1)^{r-1} r \int_{0}^{1} \left[ 1 - \prod_{j=1}^{\infty} (1 - x^{a_j}) \right] \ln(x)^{r-1} \frac{dx}{x}.
\]

(19)

Notice that \(L_r(\alpha) > 0\), for any \(\alpha\) (since, for every \(x \in (0, 1)\), \(f_N^{\alpha}(x) < 1\) and decreases with \(N\)). However, we may have \(L_r(\alpha) = \infty\). In fact as we will see (in Remark 1 below), \(L_r(\alpha) = \infty\) if and only if \(L_1(\alpha) = \infty\).
The dichotomy

Theorem 1

\( L_r(\alpha) < \infty \) if and only if there exist a \( \xi \in (0,1) \) such that

\[
\sum_{j=1}^{\infty} \xi^{a_j} < \infty.
\]  

(20)

Before proving the theorem we recall the following lemma:

Lemma 2

Let \( \{ b_j \}_{j=1}^{\infty} \) be a sequence of real numbers such that

\( 0 \leq b_j \leq 1 \), for all \( j \). If \( \sum_{j=1}^{\infty} b_j < \infty \), then

\[
\sum_{j=1}^{\infty} b_j - \sum_{1 \leq l < j} b_l b_j \leq 1 - \prod_{j=1}^{\infty} (1 - b_j) \leq \sum_{j=1}^{\infty} b_j.
\]
The dichotomy

Assume that there is a $\xi \in (0, 1)$ such that (20) is true. Then, for all positive integers $r$, by (19) and Lemma 2 we have

$$L_r(\alpha) \leq (-1)^{r-1} r \int_0^\xi \left[ \sum_{j=1}^{\infty} x^a_j \right] \ln (x)^{r-1} \frac{dx}{x} +$$

$$+ (-1)^{r-1} r \int_\xi^1 \left[ 1 - \prod_{j=1}^{\infty} (1 - x^a_j) \right] \ln (x)^{r-1} \frac{dx}{x}$$

$$\leq (-1)^{r-1} r \int_0^\xi \left[ \sum_{j=1}^{\infty} x^{a_j-1} \right] \ln (x)^{r-1} \, dx + (-1)^r \ln (\xi)^r. \quad (21)$$
The dichotomy

Using repeated integration integration by parts we evaluate the integral,

$$I_j(\xi; r) := \int_0^{\xi} x^{a_j-1} \ln (x)^{r-1} \, dx =$$

$$\frac{1}{a_j} \xi^{a_j} \sum_{k=0}^{r-1} (-1)^k (r-1)_k \frac{1}{a_j^k} \ln (\xi)^{r-1-k},$$  \quad (22)

where \((r-1)_k = (r-1)! / (r-1 - k)!\) is the falling Pochhammer symbol.
The dichotomy

Next we apply Tonelli’s Theorem and use (22) in (21) and get

\[ L_r(\alpha) \leq (-1)^{r-1} r \left( \sum_{j=1}^{\infty} \frac{1}{a_j} \xi^{a_j} \right) \]

\[ \left( \sum_{k=0}^{r-1} (-1)^k (r-1)_k \frac{1}{a_j^k} \ln^{r-1-k}(\xi) \right) + (-1)^r \ln^r \xi. \]  

(23)

Now, (20) implies that \( \xi^{a_j} \to 0 \), hence \( a_j \to \infty \). Therefore, \( \min_j \{ a_j \} = a_{j_0} > 0 \). Thus,

\[ L_r(\alpha) \leq (-1)^{r-1} r \frac{1}{a_{j_0}} \]

\[ \left( \sum_{j=1}^{\infty} \xi^{a_j} \right) \left( \sum_{k=0}^{r-1} (-1)^k (r-1)_k \frac{1}{a_{j_0}^k} \ln^{r-1-k}(\xi) \right) + (-1)^r \ln^r \xi. \]

Since \( \xi \in (0, 1) \), \( r \) is a fixed positive integer, and (20) is valid, one obtains \( L_r(\alpha) < \infty \).
The dichotomy

Conversely, if $\sum_{j=1}^{\infty} \xi^{a_j} = \infty$, for all $\xi \in (0, 1)$, then, by a well-known property of infinite products (see, e.g. W. Rudin, *Real and Complex Analysis*, McGraw-Hill, New York, (1987)).

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

and hence (19) yields

$$L_r(\alpha) = (-1)^{r-1} \int_0^1 \left( \ln^{r-1}(x)/x \right) dx = \infty.$$
The dichotomy

Remark 1. It has been shown in [BP], that $L_1(\alpha) < \infty$, if and only if there exist a $\xi \in (0, 1)$ such that $\sum_{j=1}^{\infty} \xi^{a_j} < \infty$. Thus, $L_r(\alpha) < \infty$ if and only if $L_1(\alpha) < \infty$. To sum up we have the following dichotomy, simultaneously for all positive integers $r$:

(i) $0 < L_r(\alpha) < \infty$ or (ii) $L_r(\alpha) = \infty$.  \hspace{1cm} (24)

Remark 2. Consider the error term, defined by

$$\Delta_r(N) := L_r(\alpha) - H_N(\alpha; r).$$

Then, for all positive integers $r$ by (16), (19), Tonelli’s Theorem, and repeated integration by parts, we have

$$\Delta_r(N) \leq (-1)^{r-1} r ! \int_{0}^{1} \left( \sum_{j=N+1}^{\infty} x^{a_j} \right) \ln(x)^{r-1} \frac{dx}{x} = r ! \sum_{j=N+1}^{\infty} a_j^{-r}.  \hspace{1cm} (25)$$

Thus, if $\sum_{j=1}^{\infty} a_j^{-r} < \infty$, then (25) can serve as an upper bound for the error $\Delta_r(N)$. 

Outline

1. Introduction
   - History of the problem
   - Preliminaries

2. The equally likely case

3. Large $N$ asymptotics

4. The dichotomy

5. **The case $L_r(\alpha) < \infty$**
   - The leading behavior of the moments of $T_N$
   - Asymptotic estimates for the moments of $T_N$ by comparison with known sequences

6. **The case $L_r(\alpha) = \infty$**

7. Examples
The case $L_r(\alpha) < \infty$

Let $A_N$ and $L_r(\alpha)$ be as in (14) and (19) respectively. We note that, by Theorem 1, $L_r(\alpha) < \infty$ implies that $\lim_j a_j = \infty$ (hence $\lim_N A_N = \infty$).

**Theorem 3**

*If $L_r(\alpha) < \infty$, then as $N \to \infty$,*

$$E \left[ T_N^{(r)} \right] = A_N^r L_r(\alpha) \left[ 1 + o(1) \right],$$  \hspace{1cm} (26)

*Proof.* Since, $E \left[ T_N^{(r)} \right] = A_N^r H_N(\alpha; r)$ and $L_r(\alpha) := \lim_N H_N(\alpha; r) < \infty$, formula (26) follows immediately.
The case $L_r(\alpha) < \infty$

Theorem 3 states that if $L_r(\alpha) < \infty$, then the asymptotics of $E \left[ T_N^{(r)} \right]$, are essentially determined by the asymptotics of $A_N$. As was already mentioned, asymptotic estimates of $A_N$ can be obtained by various known methods. These methods apply to a large variety of cases. Alternatively, one can resort to specific features of $\alpha$. For instance, if $\alpha$ is of the type

$$a_j = e^{jc_j}, \quad \text{where} \quad c_j \uparrow \infty, \quad \text{then as} \quad N \to \infty,$$

(27)

$$A_N = \sum_{j=1}^{N} a_j \sim a_N.$$  

(28)

In words, if a sequence satisfies (27), then in a sum of type (28), the last term dominates all the previous terms. Other examples of such sequences are $a_j = e^{jr}$ with $r > 1$, $a_j = j^r$ and $a_j = j!$ (see Example 5). We now continue with a much more challenging case.
Outline

1. Introduction
   - History of the problem
   - Preliminaries

2. The equally likely case

3. Large $N$ asymptotics

4. The dichotomy

5. The case $L_r(\alpha) < \infty$

6. The case $L_r(\alpha) = \infty$
   - The leading behavior of the moments of $T_N$
   - Asymptotic estimates for the moments of $T_N$ by comparison with known sequences

7. Examples
The leading behavior of the moments of $T_N$

By Theorem 1, $L_r(\alpha) = \infty$ is equivalent to $L_j(\alpha) = \infty$, for all $j = 1, 2, \cdots, r-1$, and also equivalent to $\sum_{j=1}^{\infty} x^{a_j} = \infty$, for all $x \in (0, 1)$. For our further analysis, we follow [BP], and write $a_j$ in the form

$$a_j = \frac{1}{f(j)}, \quad \text{where} \quad f(x) > 0,$$

(29)

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

(i) $f(x) \nearrow \infty$, (ii) $\frac{f'(x)}{f(x)} \searrow 0$, and (iii) $\frac{f''(x)}{f'(x)} \ln \left[ \frac{f'(x)}{f(x)} \right] \to 0$. (30)
The leading behavior of the moments of $T_N$

Conditions (30) are satisfied by a variety of commonly used functions. For example,

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

or various convex combinations of products of such functions.

**Remark 3.** From condition (ii) of (30), one has

$$\lim_{x \to \infty} \frac{f(x + 1)}{f(x)} = 1. \quad (31)$$

This can be justified by considering the function $g(x) = \ln(f(x))$ and applying the Mean Value Theorem.
The leading behavior of the moments of $T_N$

**Theorem 4**

If $\alpha = \{1/f(j)\}_{j=1}^{\infty}$, where $f$ satisfies (29) and (30), then

$$H_N(\alpha; r) \sim f(N)^r \ln \left( \frac{f(N)}{f'(N)} \right)^r, \quad N \to \infty. \quad (32)$$

**Proof.** Set

$$F(x) := -f(x) \ln \left[ \frac{f'(x)}{f(x)} \right]. \quad (33)$$

Notice that (29) and (ii) of (30) imply that $F(x) > 0$, at least for $x$ sufficiently large. Hence, one can write (16) as:

$$H_N(\alpha; r) = F(N)^r H_N[F(N)\alpha; r]$$
The leading behavior of the moments of $T_N$

$$= rF(N)^r \int_0^1 \left[ 1 - \exp \left( \sum_{j=1}^{N} \ln \left( 1 - e^{-\frac{F(N)}{T(u)}} s \right) \right) \right] s^{r-1} ds$$

$$+ rF(N)^r \int_1^\infty \left[ 1 - \exp \left( \sum_{j=1}^{N} \ln \left( 1 - e^{-\frac{F(N)}{T(u)}} s \right) \right) \right] s^{r-1} ds. \quad (34)$$

It has been established in [BP] that,

$$\lim_{N} \sum_{j=1}^{N} \ln \left( 1 - e^{-\frac{F(N)}{T(u)}} s \right) = \left\{ \begin{array}{ll} -\infty, & \text{if } s < 1; \\ 0, & \text{if } s \geq 1. \end{array} \right. \quad (35)$$

and also that

$$\int_1^N e^{-\frac{F(N)}{T(x)}} s dx \sim \frac{1}{s \ln \left[ f(N)/f'(N) \right]} \left[ \frac{f(N)}{f'(N)} \right]^{1-s}. \quad (36)$$

These two results came out under conditions (30).
The leading behavior of the moments of $T_N$

Applying the Bounded Convergence Theorem for the first integral on (34) yields (in view of (35))

$$H_N(\alpha; r) = rF(N)^r \left[ \frac{1}{r} + o(1) \right]$$

$$+ rF(N)^r \int_1^{\infty} \left[ 1 - \exp \left( \sum_{j=1}^N \ln \left( 1 - e^{-\frac{F(N)}{\frac{j}{r}}} s \right) \right) \right] s^{r-1} ds. \quad (37)$$

Next, we want to estimate the integral appearing in the above formula.
The leading behavior of the moments of $T_N$

Idea!
Since

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

and

$$\ln(1-x) \sim -x \text{ as } x \rightarrow 0$$

we can replace

$$\sum_{j=1}^{N} \ln \left( 1 - e^{-\frac{F(N)}{f(j)} s} \right) \text{ by } - \sum_{j=1}^{N} e^{-\frac{F(N)}{f(j)} s}$$

Hence, we can work with the following integral, since $f$ is increasing,

$$\int_{1}^{N} e^{-\frac{F(N)}{f(x)} s} dx \leq \sum_{j=1}^{N} e^{-\frac{F(N)}{f(j)} s} \leq \int_{1}^{N+1} e^{-\frac{F(N)}{f(x)} s} dx$$
The leading behavior of the moments of $T_N$

We begin by noticing that by the Dominated Convergence Theorem (since $f(N)/f'(N) \to \infty$)

$$\lim_N \int_1^\infty \left[ 1 - \exp \left( - \frac{(f(N)/f'(N))^{1-s}}{s \ln (f(N)/f'(N))} \right) \right] s^{r-1} ds = 0.$$  

Using (36) this implies that

$$\lim_N \int_1^\infty \left[ 1 - \exp \left( - \int_1^N \frac{F(N)}{f(x)} s dx \right) \right] s^{r-1} ds = 0. \quad (38)$$

Since $f$ is increasing, we have

$$\int_1^N e^{-\frac{F(N)}{f(x)} s} dx \leq \sum_{j=1}^{N} e^{-\frac{F(N)}{f(j)} s} \leq \int_1^{N+1} e^{-\frac{F(N)}{f(x)} s} dx \leq \int_1^N e^{-\frac{F(N)}{f(x)} s} dx + e^{-\frac{F(N)}{f(N+1)} s}. \quad (39)$$
The leading behavior of the moments of $T_N$

From the above inequalities it follows

$$1 - \exp \left( - \int_1^N e^{-\frac{F(N)}{f(x)} s} \, dx \right) \leq 1 - \exp \left( - \sum_{j=1}^N e^{-\frac{F(N)}{f(j)} s} \right)$$

$$1 - \exp \left( - \int_1^N e^{-\frac{F(N)}{f(x)} s} \, dx \right) \leq 1 - \exp \left( - \int_1^N e^{-\frac{F(N)}{f(x)} s} \, dx + e^{-\frac{F(N)}{f(N+1)} s} \right).$$

(40)

However, by (36)

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

(41)

Hence by taking limits in (40) and using (38) and (31), we get

$$\lim_N \int_1^\infty \left[ 1 - \exp \left( \sum_{j=1}^N \ln \left( 1 - e^{-\frac{F(N)}{f(j)} s} \right) \right) \right] s^{r-1} \, ds = 0.$$

(42)
The leading behavior of the moments of $T_N$

Finally, by the definition of $F(\cdot)$, and the Taylor expansion for the logarithm, namely $\ln(1 - x) \sim -x$ as $x \to 0$, (37) yields

$$H_N(\alpha; r) \sim F(N)^r = f(N)^r \ln \left( \frac{f(N)}{f'(N)} \right)^r, \quad N \to \infty$$

(43)

and the proof is completed.

**Remark 4.** Using Theorem 4 in (18), we get as $N \to \infty$

$$E \left[ T_N^{(r)} \right] \sim A_N^r f(N)^r \ln \left( \frac{f(N)}{f'(N)} \right)^r = \frac{1}{p_N^r} \ln \left( \frac{f(N)}{f'(N)} \right)^r = \frac{1}{\min_{1 \leq j \leq N} \{p_j\}^r} \ln \left( \frac{f(N)}{f'(N)} \right)^r,$$

(44)

where the last equality follows from (14).
Asymptotic estimates for the moments of $T_N$ by comparison with known sequences

In this subsection we will present a theorem that helps us obtain asymptotic estimates by comparison with sequences $\alpha$ for which the asymptotic estimates of $H_N(\alpha; r)$ are known (for instance, via Theorem 4). First, we recall the following notation. Suppose that $\{s_j\}_{j=1}^{\infty}$ and $\{t_j\}_{j=1}^{\infty}$ are two sequences of nonnegative terms. The symbol $s_j \asymp t_j$ means that there are two constants $C_1 > C_2 > 0$ and an integer $j_0 > 0$ such that

$$C_2 t_j \leq s_j \leq C_1 t_j, \quad \text{for all } j \geq j_0,$$

i.e. $s_j = O(t_j)$ and $t_j = O(s_j)$. 
Asymptotic estimates for the moments of $T_N$ by comparison with known sequences

Theorem 5

Let $\alpha = \{a_j\}_{j=1}^{\infty}$ and $\beta = \{b_j\}_{j=1}^{\infty}$ be sequences of strictly positive terms such that $\lim_N H_N(\alpha; r) = \lim_N H_N(\beta; r) = \infty$.

(i) If there exists an $j_0$ such that $a_j = b_j$, for all $j \geq j_0$, then $H_N(\beta; r) - H_N(\alpha; r)$ is bounded,

(ii) if $a_j = O(b_j)$, then $H_N(\beta; r) = O(H_N(\alpha; r))$ as $N \to \infty$,

(iii) if $a_j = o(b_j)$, then $H_N(\beta, r) = o(H_N(\alpha; r))$ as $N \to \infty$,

(iv) if $a_j \sim b_j$, then $H_N(\beta; r) \sim H_N(\alpha; r)$ as $N \to \infty$,

(v) if $a_j \sim b_j$, then $H_N(\beta; r) \sim H_N(\alpha; r)$ as $N \to \infty$. 
Outline

1. Introduction
   - History of the problem
   - Preliminaries
2. The equally likely case
3. Large $N$ asymptotics
4. The dichotomy
5. The case $L_r(\alpha) < \infty$
6. The case $L_r(\alpha) = \infty$
   - The leading behavior of the moments of $T_N$
   - Asymptotic estimates for the moments of $T_N$ by comparison with known sequences
7. Examples
Examples

Example 1. The case \( a_j = 1 \), for all \( j \), has been already discussed in detail in Section 2. This case can also provide us with an application of Theorem 5: If \( \beta = \{b_j\}_{j=1}^{\infty} \) is a sequence such that \( 0 < \lim b_j \leq \lim b_j < \infty \) then, there are two constants \( C_1 > C_2 > 0 \) and an integer \( j_0 > 0 \) such that

\[
C_2 b_j \leq 1 \leq C_1 b_j, \quad \text{for all } j \geq j_0, \quad \text{i.e. } 1 \sim b_j.
\]

Hence, by part (iv) of Theorem 5, \( H_N(\beta; r) \sim \ln^r N \). If, in addition, \( \lim b_j = b \) exists, then \( ba_j \sim b_j \). Hence, by part (v) of Theorem 5, \( H_N(\beta; r) \sim H_N(b\alpha; r) \). Using (17) we get

\[
H_N(\beta; r) \sim b^{-r} \ln^r N.
\]
Examples

Example 2. $a_j = j^p$, where $p > 0$. In this case

$$L_{r,p} = L_r(\alpha) = (-1)^{r-1} r \int_0^1 \left[ 1 - \prod_{j=1}^{\infty} (1 - x^j) \right] \ln^{r-1}(x) \frac{dx}{x},$$

By Theorem 1 and for all positive integers $r$ we have: $L_{r,p} < \infty$.

From the Euler-Maclaurin summation formula we get the full asymptotic expansion of $A_N = \sum_{n=1}^{N} n^p$, exactly as a polynomial in $N$ of degree $p + 1$, if $p$ is a positive integer, i.e.

$$A_N = \sum_{n=1}^{N} n^p = \frac{N^{p+1}}{p+1} \left[ 1 + O \left( \frac{1}{N} \right) \right].$$

Therefore,

$$E \left[ T_N^{(r)} \right] = \frac{N^{r(p+1)}}{(p+1)^r} L_{r,p} \left[ 1 + o(1) \right].$$
Examples

The case $p = 1$ is known as the *linear* case, and it is of particular interest. From Euler’s pentagonal-number formula (*a remarkable proof of this formula is due to F. Franklin (1881), see. e.g., T. Apostol, Introduction to Analytic Number Theory*).

$$\prod_{j=1}^{\infty} (1 - x^j) = 1 + \sum_{k=1}^{\infty} (-1)^k \left[ x^{\omega(k)} + x^{\omega(-k)} \right], \quad \omega(k) = \frac{3k^2 - k}{2}, \quad k = 0, \pm 1, \pm 2, \ldots$$

In that case $L_r$ becomes

$$L_r = (-1)^r r \sum_{k=1}^{\infty} (-1)^k \left[ \int_0^1 x^{\omega(k)-1} \ln(x)^{r-1} \, dx + \int_0^1 x^{\omega(-k)-1} \ln(x)^{r-1} \, dx \right].$$

Repeated integration by parts yields,

$$L_r = 2^r r! \sum_{k=1}^{\infty} (-1)^{k+1} \left[ \frac{1}{(3k^2 - k)^r} + \frac{1}{(3k^2 + k)^r} \right].$$

For example,

$$L_1 = \frac{4\pi \sqrt{3}}{3} - 6 \approx 1.2552, \quad L_2 = 4(54 - 8\pi \sqrt{3} - \pi^2) \approx 2.39684.$$  

As for $L_3$, a numerical computation gives

$$L_3 \approx 6.68903.$$
Example 3. $\alpha_j = e^{pj}, b_j = e^{-pj}, p > 0$. For the sequence $\alpha = \{a_j\}_j^{\infty}$ we have, $L_r(\alpha) < \infty$, $r = \{1, 2, \cdots \}$. Furthermore,

$$\Delta_r(N) = L_r(\alpha) - H_N(\alpha; r) \leq r! \sum_{j=N+1}^{\infty} e^{-rpj} = \frac{r! \ e^{-rp(N+1)}}{1 - e^{-rp}}, \quad (46)$$

$$A_N = \frac{e^{p(N+1)} - 1}{e^p - 1}, \quad \text{and} \quad E \left[ T_N^{(r)} \right] = \left( \frac{e^{p(N+1)}}{e^p - 1} \right)^r L_r(\alpha) + O(1).$$
Examples

In the special case of $a_j = 2^j$ (i.e. $p = \ln 2$), we have

$$
\phi(x) := \prod_{j=0}^{\infty} (1 - x^{2^j}) = \sum_{k=0}^{\infty} (-1)^{\delta(k)} x^k = 1 - \sum_{n=0}^{\infty} (1 - x)^n x^{2^n}, \tag{47}
$$

where $\delta(k)$ is the number of ones in the binary expansion of $k$. The first equality is strait forward. The last equality follows from the observation that $\phi(x) = (1 - x) \phi(x^2)$. Therefore, by (19)

$$
L_r(\alpha) = r! \sum_{k=1}^{\infty} \frac{(-1)^{\delta(k)-1}}{k^r}. \tag{48}
$$

The last of the above series converges (of course), but it is hard to compute. We may express (48) in an equivalent form. By the last equality of (47)

$$
L_r(\alpha) = (-1)^{r-1} r \sum_{n=0}^{\infty} \int_0^1 (1 - x)^n x^{2^n-1} \ln(x)^{r-1} dx.
$$
Examples

Applying binomial formula and repeated integration by parts, the integral above yields

\[ \int_0^1 (1 - x)^n x^{2n-1} \ln(x)^{r-1} dx = \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k (r-1)!}{(k+2^n)^r}. \]

Hence, \( L_r(\alpha) \) of (48) becomes

\[ L_r(\alpha) = r! \sum_{n=0}^{\infty} \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{(k+2^n)^r}. \]

The above series converges extremely rapidly.
Examples

Now, for the sequence $\beta = \{b_j\}_{j=0}^{\infty}$ we have $L(\beta) = \infty$. Furthermore $f(x) = e^{px}$ does not satisfy condition (ii) and of (30), thus Theorem 4 can not be applied.

However, the sequences $\alpha$ and $\beta$ produce the same (!) coupon probabilities.

This follows from the fact that for each $N$, if we let $c_N = e^{pN}$, then $\{a_j : 0 \leq j \leq N\} = \{c_N b_j : 0 \leq j \leq N\}$, i.e. the elements of the two truncated sequences are proportional to each others. In particular $\alpha$ and $\beta$ lead to the same $H_N(\alpha; r)$.
Examples

Example 4. $a_j = 1/j^p$, $p > 0$ (note that $p = 1$ corresponds to the so-called Zipf distribution. In this case Theorem 1 implies $L_r(\alpha) = \infty$. If $f(x) = x^p$, then $f$ satisfies (30) and hence Theorem 4 apply. It is now straightforward to estimate $A_N^r$ (from the Euler-Maclaurin Summation formula) and get

(i) $E \left[ T_N^{(r)} \right] \sim \frac{N^r \ln^r(N)}{(1 - p)^r}$ for $0 < p < 1$,

(ii) $E \left[ T_N^{(r)} \right] \sim N^r \ln^{2r}(N)$ for $p = 1$,

(iii) $E \left[ T_N^{(r)} \right] \sim \zeta(p)^r N^{rp} \ln^r(N)$ for $p > 1$. 
Examples

**Example 5.** \( a_j = j! \). We have \( L_r(\alpha) < \infty \). Here, the Euler-Maclaurin summation formula is not effective for the estimation of \( A_N \). However, using Stirling's formula i.e. \( n! \sim (n/e)^n \sqrt{2\pi n} \) and (28) we have \( A_N \sim N! \).

Hence, by Theorem 1 we get

\[
E \left[ T_N^{(r)} \right] \sim L_r(\alpha) (N!)^r \quad \text{as} \quad N \to \infty.
\]
Introduction

The equally likely case
Large $N$ asymptotics
The dichotomy
The case $L_r(\alpha) < \infty$
The case $L_r(\alpha) = \infty$

references


references


references


references


Introduction  The equally likely case  Large $N$ asymptotics  The dichotomy  The case $L_r(\alpha) < \infty$  The case $L_r(\alpha) = \infty$

references


