Fourth Colloquium on Mathematics and Computer Science DMTCS proc. **AG**, 2006, 215–224

A coupon collector's problem with bonuses

Toshio Nakata[1] and Izumi Kubo[2†]

[1] *Department of Information Education, Fukuoka University of Education, Akama-Bunkyomachi, Munakata, Fukuoka, 811-4192, JAPAN.*
[2] *Department of Environmental Design, Faculty of Environmental Studies, Hiroshima Institute of Technology, 2-1-1 Miyake, Saeki-ku, Hiroshima 731-5193, JAPAN*

In this article, we study a variant of the coupon collector's problem introducing a notion of a *bonus*. Suppose that there are c different types of coupons made up of bonus coupons and ordinary coupons, and that a collector gets every coupon with probability $1/c$ each day. Moreover suppose that every time he gets a bonus coupon he immediately obtains one more coupon. Under this setting, we consider the number of days he needs to collect in order to have at least one of each type. We then give not only the expectation but also the exact distribution represented by a gamma distribution. Moreover we investigate their limits as the Gumbel (double exponential) distribution and the Gauss (normal) distribution.

Keywords: Coupon collector's problem, Gumbel distribution, central limit theorem, randomized algorithm

1 Introduction and results

1.1 The coupon collector's problem

Discrete probability, which often appears in randomized algorithms, plays an important role in theoretical computer science. The coupon collector's problem is one of the most popular topics in discrete probability, as it is simple and useful. There are a lot of standard probability/algorithm textbooks describing it. For example see (8, p.48),(7, p.38),(15, p.55), (17, §2.4.1), (12, Problem 3.3.2), (18, §3.6), (21, p.12 Example 1.4f). The problem is the following:

> The coupon collector's problem (classical version): Suppose that there are c different types of coupons, and that a collector gets every coupon with probability $1/c$ each day. What is the number of days that he needs to collect in order to have at least one of each type?

For $i = 1, \cdots, c$, let X_i^c be the number of waiting days until another new type has been obtained after $i-1$ distinct types have been collected. We define $Y^c = X_1^c + \cdots + X_c^c$, that is, Y^c is the number of days the collector needs to collect in order to have at least one of c types. The expected number of days collecting all c coupons is well-known, that is,

$$E[Y^c] = \sum_{i=1}^{c} E[X_i^c] = \sum_{i=1}^{c} \frac{c}{c-i+1} = c\sum_{i=1}^{c}\frac{1}{i} \sim c\log c, \tag{1}$$

where $a_n \sim b_n$ denotes $\lim_{n\to\infty}\frac{a_n}{b_n} = 1$. In fact, $\{X_i^c\}_{i=1}^c$ are geometrically distributed with parameter $(c-i+1)/c$ respectively. Therefore since $E[X_i^c] = \frac{c}{c-i+1}$, we have Eq. (1). In addition to this, a closed form of the exact distribution denoted by $p^c(n) = P(Y^c = n)$ for $n = 1, 2, \cdots$ is known, that is,

$$p^c(n) = \sum_{m=0}^{c-1}\binom{c-1}{m}(-1)^m\left(1-\frac{m+1}{c}\right)^{n-1} = \begin{cases} \dfrac{c!\left\{ {n-1 \atop c-1} \right\}}{c^n}, & \text{if } n \ge c, \\ 0 & \text{if } n < c, \end{cases} \tag{2}$$

where $\left\{ {n \atop k} \right\}$ is the *Stirling number of the second kind* (see (10; 22; 19)). Moreover some limit theorems are also known. For example, $Y^c/c\log c$ converges to 1 in probability (see (7, p.38, Example 5.3)), that is, for an arbitrary $\eta > 0$ we have

$$\lim_{c\to\infty} P((1-\eta)c\log c \le Y^c \le (1+\eta)c\log c) = 1. \tag{3}$$

†This work was supported by Grant-in-Aid for KAKENHI No. 17540138 from MEXT Japan.

More precisely, we note a *sharp threshold* theorem (see (7, p. 144), (18, Theorem 3.8)). Namely, for any real number x we have

$$\lim_{c\to\infty} P(Y^c \le c(\log c + x)) = e^{-e^{-x}}, \tag{4}$$

$$\lim_{c\to\infty} P(Y^c \ge c(\log c - x)) = 1 - e^{-e^{x}}. \tag{5}$$

Eq. (5) is easily obtained by Eq. (4). The right hand side of Eq. (4) is said to be the *Gumbel (double exponential)* distribution, which appears in extreme value theory (see (5; 13)). The result of Eqs. (4) and (5) is almost like deterministic as a randomized algorithm since it so sharply identifies the threshold value for collecting all coupons (see (18)).

1.1.1 Related topics:

More recently various applied researches concerning the problem are reported. Let us explain a small subset of them. In combinatorial analysis Myers and Wilf (19) studied a variant of the classical coupon collector's problem using a combinatorial property of the Stirling number of the second kind. In computer science Flajolet et al. (9) gave a unified study for the birthday paradox, the coupon collector's problem, caching algorithms, and self-organizing searches which are in the form of an integral representation. Martinez (16) generalized their results focusing on a ratio limit theorem. Gardy (11) surveyed some problems including the coupon collector's problem as *occupancy urn* models from the viewpoint of computer science. There are some other recent applications to computer science. For example, an application to a packet delivery system as a randomized algorithm is described in (17, p. 34). M. Adler et al. (3) and Dimitrov, Plaxton (6) studied load balancing in peer-to-peer networks using a structured coupon collector's problem. I. Adler et al. (1) studied a coalescing particle model which is applicable to population biology. They gave sharp bounds for expected times and asymptotic results. Finally we point out a well-known fact that the coupon collector's problem is considered as a *cover time* of a random walk on a complete graph. For a cover time on any graph there are diverse researches including graph theoretical issues (see (18, §6.5), (4, Chap. 6) and references therein).

1.2 A coupon collector's problem with bonuses

On the classical coupon collector's problem, it seems reasonable to get all types of coupons as soon as possible. Then, in this article, to get these coupons in a shorter time we give a variant of the coupon collector's problem by introducing a notion of a *bonus*. That means the following:

> A coupon collector's problem with bonuses (bonus version): Suppose that there are c coupons made up of k *bonus* coupons and $l = c - k$ ordinary coupons. Whenever a collector gets one of the bonus coupons, he immediately collects one more coupon. In this setting, what is the number of days that he needs to collect in order to have at least one of each type?

Every time the collector gets a bonus coupon, he immediately obtains a new coupon on the same day. Consequently he may collect all k bonuses and one ordinary coupon in a day if he has a stretch of luck. Let $Y^{c,k}$ be the number of days that he has to collect in order to have at least one of each of c types of coupons including k bonuses. By definition, we see $Y^{c,0} = Y^c$. We consider closed forms of the expectation $E[Y^{c,k}]$, corresponding to Eq. (1), and the probability function $p^{c,k}(n)$ corresponding to Eq. (2) respectively, where $p^{c,k}(n) = P(Y^{c,k} = n)$. By definition, $E[Y^{c,k}]$ is smaller than $E[Y^c]$ for any $k \ge 1$. However it seems nontrivial to estimate the difference between $E[Y^{c,k}]$ and $E[Y^c]$.

Before illustrating the closed forms, we prepare a gamma random variable. Let $X_{n,l}$ be a gamma random variable with parameter (n, l), denoted by $X_{n,l} \sim \Gamma(n, l)$, if the probability density function is

$$\frac{l(lx)^{n-1}}{\Gamma(n)} e^{-lx} \quad \text{for } x > 0, \tag{6}$$

where $\Gamma(s)$ is the gamma function, that is, $\Gamma(s) = \int_0^\infty e^{-x} x^{s-1} dx$.

Theorem 1.1 (Exact distribution) *For any integers $k \ge 0$ and $l \ge 1$, consider the coupon collector's problem with k bonuses and l ordinary coupons. Then the exact distribution is*

$$P(Y^{k+l,k} \le n) = E[(1 - e^{-X_{n,l}})^k] P(Y^l \le n) \quad \textit{for } n \ge 1, \tag{7}$$

where $X_{n,l}$ is a gamma random variable with parameter (n, l). The probability function is

$$p^{k+l,k}(n) = \sum_{i=0}^{k}\sum_{j=0}^{l}\binom{k}{i}\binom{l}{j}(-1)^{i+j+1}\frac{i+j}{l+i}\left(\frac{l-j}{l+i}\right)^{n-1} \quad \text{for } n \geq 1. \tag{8}$$

Remark 1.1 *If $l = 1$, then the probability of Eq. (7) is*

$$P(Y^{k+1,k} \leq n) = E[(1-e^{-X_{n,1}})^k] = P(\Delta_n > k) \quad \text{for } n \geq 1,$$

where Δ_n is the inter-record time *between the $(n-1)$-th record and the n-th record appeared in extreme value theory (see (20; 5; 13)).*

Corollary 1.1 (Expectation) *We have the following expectation of $Y^{k+l,k}$:*

$$E[Y^{k+l,k}] = l\sum_{m=1}^{k+l}\frac{1}{m} + \frac{k}{k+l} \sim l\log(k+l). \tag{9}$$

Remark 1.2 *Eq. (9) means*

$$E[Y^{k+l,k}] = \frac{l}{k+l}\cdot\{(k+l)H_{k+l}\} + \frac{k}{k+l}\cdot 1. \tag{10}$$

As a result, Eq. (10) says that $E[Y^{k+l,k}]$ is the dividing point whose ratio is $k : l$ between $E[Y^{k+l,0}] = (k+l)H_{k+l}$ (all $k+l$ ordinary coupons) and $E[Y^{k+0,k}] = 1$ (all k bonus coupons).

Next, we consider the concentration of $Y^{k+l,k}$ near $E[Y^{k+l,k}] \sim l\log(k+l)$ for large k and l. Before observing asymptotic behaviors, we define the following growth rates:

(C1) For $k \geq 0$ and $l \geq 1$ we assume that there exist $0 \leq \alpha \leq \infty$ and $0 \leq \beta \leq \infty$ (including ∞) satisfying

$$\alpha = \lim_{k+l\to\infty}\frac{k}{l}, \tag{11}$$

$$\beta = \lim_{k+l\to\infty}\frac{\log k}{l}, \tag{12}$$

respectively. The notation $k+l \to \infty$ means that at least one of k and l must go to infinity.

Note that $0 \leq \beta \leq \alpha \leq \infty$. Both of them do not be positive and finite. Moreover we see the following:

- If $\alpha, \beta = 0$ then $l \to \infty$ but $0 \leq k$ may be bounded.
- If $\alpha, \beta = \infty$ then $k \to \infty$ but $1 \leq l$ may be bounded.

Unless otherwise noted, we assume the existence of α and β in (C1) throughout the paper.

Proposition 1.1 (Law of Large Numbers) *For any $k \geq 0$ and $l \geq 1$ we have*

$$\lim_{k+l\to\infty}\frac{Y^{k+l,k}}{l\log(k+l)} = 1 \quad \text{in probability.} \tag{13}$$

Remark 1.3 *If $k = 0$, then Eq. (13) is equivalent to Eq. (3).*

Before stating more precise results, we prepare some notations. Let $N_{0,1}$ denote a standard Gauss (normal) random variable, that is, the density function ϕ and the distribution function Φ are

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}, \quad \Phi(x) = P(N_{0,1} \leq x) = \int_{-\infty}^{x}\phi(t)dt$$

respectively.

Theorem 1.2 (Limit distribution) *The limit distributions of $Y^{k+l,k}$ for $k \geq 0$ and $l \geq 1$ are classified by β which is defined by Eq. (12):*

1. *If $\beta = 0$, then we have*

$$\lim_{k+l\to\infty} P(Y^{k+l,k} \le l\{\log(l+k)+x\}) = e^{-e^{-x}}. \tag{14}$$

2. *If $0 < \beta < \infty$, then we have*

$$\lim_{k+l\to\infty} P(Y^{k+l,k} \le l\{\log(l+k)+x\}) = E\left[\exp\left\{-e^{-x-N_{0,1}\sqrt{\beta}}\right\}\right]$$
$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left\{-e^{-x-y\sqrt{\beta}}-\frac{1}{2}y^2\right\}dy.$$

3. *If $\beta = \infty$, then a central limit theorem holds. Namely, for any real number x we have*

$$\lim_{k+l\to\infty} P\left(Y^{k+l,k} \le l\log(k+l) + x\sqrt{l\log(k+l)}\right) = \Phi(x). \tag{15}$$

Remark 1.4 1. *If $k = 0$ then Eq. (14) is equivalent to Eq. (4).*

2. *Put $F(x) = E\left[\exp\left\{-e^{-x-N_{0,1}\sqrt{\beta}}\right\}\right]$. Then we see that F is a distribution function by the Lebesgue bounded convergence theorem.*

Corollary 1.2 *Suppose that α does not exist. Namely, $\underline{\alpha} < \overline{\alpha}$ holds, where $\overline{\alpha} = \limsup_{k+l\to\infty}\frac{k}{l}$ and $\underline{\alpha} = \liminf_{k+l\to\infty}\frac{k}{l}$. Then we have the following:*

- *If $\underline{\alpha} < \overline{\alpha} < \infty$, Eq.* (14) *holds.*
- *If $\underline{\alpha} < \infty$ and $\overline{\alpha} = \infty$, we have the following:*
  - *If $\beta = 0$, then Eq.* (14) *holds.*
  - *If $\beta > 0$, then any limit distribution does not exist.*

Corollary 1.2 is immediately obtained, because of considering adequate subsequences in Theorem 1.2 for $\overline{\alpha}$ and $\underline{\alpha}$ respectively. In a similar fashion, if the limit β does not exist, then any limit distribution does not exist.

2 Proof of the exact distribution (Theorem 1.1)

First, we state a key lemma.

Lemma 2.1 *Under the assumption of Theorem 1.1, we have*

$$P(Y^{k+l,k} \le n) = \sum_{m=0}^{k}\binom{k}{m}(-1)^m\left(\frac{l}{l+m}\right)^n P(Y^l \le n). \tag{16}$$

Proof: We define the following events:

- Let E_i be the event that the bonus coupon i does *not* be collected for $i \in \{1,\cdots,k\}$ by the n-th day.
- Let F be the event that all l types of ordinary coupons are collected by the n-th day.

Then we have

$$\{Y^{k+l,k} \le n\} = \bigcap_{i=1}^{k} E_i^c \cap F.$$

Moreover since $\bigcap_{i=1}^{k} E_i^c$ and F are independent, we obtain

$$P(Y^{k+l,k} \le n) = P\left(\bigcap_{i=1}^{k} E_i^c\right)P(F) = \left(1 - P\left(\bigcup_{i=1}^{k} E_i\right)\right)P(Y^l \le n). \tag{17}$$

By the principle of inclusion-exclusion, it turns out

$$P\left(\bigcup_{i=1}^{k} E_i\right) = \sum_{m=1}^{k} \binom{k}{m} (-1)^{m-1} P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_m}). \tag{18}$$

Now, if we could show

$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_m}) = \left(\frac{l}{m+l}\right)^n \tag{19}$$

for any $1 \le i_1 < i_2 < \cdots < i_m \le k$, then Eq. (16) holds by Eqs. (17) and (18). Hence we check Eq. (19). For n and $i_1, \cdots, i_m$, let G_t be an event that any bonus coupon $i_1, \cdots, i_m$ does not appear on the t-th day. Then we see that $P(G_t) = \frac{l}{m+l}$, because any one of ordinary l coupons is collected earlier than bonus coupons $i_1, \cdots, i_m$. Since $\{G_t\}_{t=1}^n$ are independent,

$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_m}) = P\left(\bigcap_{t=1}^{n} G_t\right) = \prod_{t=1}^{n} P(G_t) = \left(\frac{l}{m+l}\right)^n.$$

□

Proof of Theorem 1.1: First, to check Eq. (7), we claim

$$E\left[(1-e^{-X_{n,l}})^k\right] = \sum_{m=0}^{k} \binom{k}{m} (-1)^m \left(\frac{l}{m+l}\right)^n. \tag{20}$$

In fact, by Eq. (6) we see

$$E\left[(1-e^{-X_{n,l}})^k\right] = \int_0^\infty \frac{l\cdot (lx)^{n-1} e^{-lx}}{\Gamma(n)} (1-e^{-x})^k dx$$
$$= \int_0^\infty \frac{l\cdot (lx)^{n-1} e^{-lx}}{\Gamma(n)} \sum_{m=0}^{k} \binom{k}{m} (-1)^m e^{-mx} dx = \sum_{m=0}^{k} \binom{k}{m} (-1)^m \left(\frac{l}{m+l}\right)^n.$$

Combining Eqs. (16) and (20) we have Eq. (7).

Next, to check Eq. (8), we review the distribution of the classical coupon collector's problem. Namely, for $n \ge c$, we have

$$P(Y^c \le n) = \sum_{m=0}^{c} \binom{c}{m} (-1)^m \left(1 - \frac{m}{c}\right)^n$$

(see (8, IV.2 (2.3)), (10, p.80) and (22, p.157)). Using Eq. (16), we have

$$P(Y^{k+l,k} \le n) = \sum_{i=0}^{k} \sum_{j=0}^{l} \binom{k}{i} \binom{l}{j} (-1)^{i+j} \left(\frac{l-j}{l+i}\right)^n. \tag{21}$$

Hence we obtain Eq. (8). □

Proof of Corollary 1.1: Since $P(Y^{k+l,k} \ge 0) = 1$, by Eq. (21) we see

$$E[Y^{k+l,k}] = \sum_{n=0}^{\infty} P(Y^{k+l,k} > n) = \sum_{n=0}^{\infty} \sum_{(i,j)\ne(0,0)} \binom{k}{i} \binom{l}{j} (-1)^{i+j+1} \left(\frac{l-j}{l+i}\right)^n,$$

where the summation of the last term is in

$$\{(i,j) : 0 \le i \le k, 0 \le j \le l\} \setminus \{(0,0)\}.$$

Therefore we have

$$\begin{aligned}
E[Y^{k+l,k}] &= \sum_{(i,j)\neq(0,0)} \binom{k}{i}\binom{l}{j}(-1)^{i+j+1}\frac{l+i}{i+j} = \sum_{m=1}^{k+l}\sum_{i=0}^{k}\binom{k}{i}\binom{l}{m-i}(-1)^{m+1}\frac{l+i}{m} \\
&= \sum_{m=1}^{k+l}\frac{l}{m}\sum_{i=0}^{k}\binom{k}{i}\binom{l}{m-i}(-1)^{m+1} + \sum_{m=1}^{k+l}\frac{1}{m}\sum_{i=0}^{k} i\binom{k}{i}\binom{l}{m-i}(-1)^{m+1} \\
&\overset{(*)}{=} \sum_{m=1}^{k+l}\frac{l}{m}(-1)^{m+1}\binom{k+l}{m} + \sum_{m=1}^{k+l} k\binom{k-1+l}{m-1}(-1)^{m+1}\frac{1}{m} \\
&= l\sum_{m=1}^{k+l}\binom{k+l}{m}(-1)^{m+1}\frac{1}{m} + k\sum_{i=0}^{k+l-1}\binom{k-1+l}{i}(-1)^{i}\frac{1}{i+1}.
\end{aligned}$$

To check the equality of $(*)$, we used

$$\sum_{i=0}^{k}\binom{k}{i}\binom{l}{m-i} = \binom{k+l}{m} \quad \text{and} \quad \sum_{i=0}^{k} i\binom{k}{i}\binom{l}{m-i} = \sum_{j=0}^{k-1} k\binom{k-1}{j}\binom{l}{m-1-j} = k\binom{k-1+l}{m-1}$$

(see (8, II.12 (12.9))). It is known that

$$\sum_{m=1}^{n}\binom{n}{m}(-1)^{m+1}\frac{1}{m} = H_n, \quad \sum_{i=0}^{n}\binom{n}{i}(-1)^{i}\frac{1}{i+1} = \frac{1}{n+1}$$

(see (8, II.12 (12.19))). Thus we obtain

$$E[Y^{k+l,k}] = lH_{k+l} + \frac{k}{k+l}.$$

Hence we have Eq. (9). □

3 Proof of limit distributions

To prove Proposition 1.1 and Theorem 1.2, we prepare three lemmas. First, we approximately put the expectation with respect to a gamma distribution into the expectation with respect to the Gauss distribution.

Lemma 3.1 ((20, Lemma 1)) *For any $\epsilon > 0$, there exist a real number $a > 0$ and an integer $n_0 > 0$ such that for $n \geq n_0$, $k \geq 0$ and $l \geq 1$*

$$\left| E[(1-e^{-X_{n,l}})^k] - E\left[\left(1-e^{-\frac{n}{l}-N_{0,1}\sqrt{\frac{n}{l^2}}}\right)^k 1_{\{-a\leq N_{0,1}\leq a\}}\right]\right| < \epsilon, \tag{22}$$

where 1_A denotes the indicator function, that is, $1_A(\omega) = \begin{cases} 1, \textit{if } \omega \in A \\ 0, \textit{otherwise.}\end{cases}$

Proof: The proof is essentially due to (20, Lemma 1). Actually, if $l = 1$ then it is exactly the same statement. The detail of the proof is omitted. □

Using Lemma 3.1, we can prove Proposition 1.1. However the proof is also omitted. Next, we consider $P(Y^{k+l,k} \leq n)$ by estimating $P(Y^l \leq n)$ and $E[(1-e^{-X_{n,l}})^k]$ in virtue of Eq. (7).

Lemma 3.2 *For $k \geq 0, l \geq 1$ and any real x, we have*

$$\lim_{k+l\to\infty} P(Y^l \leq l(\log(l+k)+x)) = \begin{cases} e^{-\frac{e^{-x}}{1+\alpha}}, & \textit{if } 0 \leq \alpha < \infty, \\ 1, & \textit{if } \alpha = \infty. \end{cases} \tag{23}$$

Proof: Note that

$$P(Y^l \leq l(\log(l+k)+x)) = P\left(Y^l \leq l\left\{\log l + x + \log\left(1+\frac{k}{l}\right)\right\}\right).$$

By Eqs. (4) and (11), we obtain

$$\lim_{k+l\to\infty} P(Y^l \le l(\log(l+k)+x)) = e^{-e^{-x-\log(1+\alpha)}} = e^{-\frac{e^{-x}}{1+\alpha}}$$

if $\alpha < \infty$. In the case of $\alpha = \infty$, we see

$$\lim_{k+l\to\infty} P(Y^l \le l(\log(l+k)+x)) = \lim_{k+l\to+\infty} e^{-e^{-C(k,l,x)}} = 1,$$

where $C(k,l,x)$ denotes some sequence satisfying $\lim_{k+l\to\infty} C(k,l,x) = \infty$. □

By Lemma 3.1, we should estimate $E[(1-e^{-\frac{n}{l}-N_{0,1}\sqrt{\frac{n}{l^2}}})^k]$ instead of $E[(1-e^{-X_{n,l}})^k]$.

Lemma 3.3 *For $k \ge 0, l \ge 1$ and any real x, put $n = n(k,l,x) = l(\log(l+k)+x)$. Then $E[(1-e^{-X_{n,l}})^k]$ has the following limit:*

- *If $\alpha < \infty$, then we have*
$$\lim_{k+l\to\infty} E[(1-e^{-X_{n,l}})^k] = e^{-\frac{\alpha e^{-x}}{\alpha+1}}. \tag{24}$$
- *If $\alpha = \infty$, then two claims follow.*
  - *If $\beta = 0$, then we have*
$$\lim_{k+l\to\infty} E[(1-e^{-X_{n,l}})^k] = e^{-e^{-x}}. \tag{25}$$
  - *If $0 < \beta < \infty$, then we have*
$$\begin{aligned}\lim_{k+l\to\infty} E[(1-e^{-X_{n,l}})^k] &= E\left[\exp\left\{-e^{-x-N_{0,1}\sqrt{\beta}}\right\}\right] \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\left\{-e^{-x-y\sqrt{\beta}-\frac{1}{2}y^2}\right\} dy. \end{aligned}\tag{26}$$

Proof: By Lemma 3.1, we consider $E[(1-e^{-\frac{n}{l}-N_{0,1}\sqrt{\frac{n}{l^2}}})^k]$ instead of $E[(1-e^{-X_{n,l}})^k]$. Letting y be a realization of $N_{0,1}$, we deal with the integrand in the expectation, that is,

$$(1-e^{-\frac{n}{l}-y\sqrt{\frac{n}{l^2}}})^k = \left(1 - \frac{e^{-x}}{k+l}e^{-y\sqrt{\frac{\log(k+l)+x}{l}}}\right)^k. \tag{27}$$

Then we have the following claims:

- The case of $\alpha < \infty$:
  Since $\lim_{k+l\to\infty} k/l = \alpha < \infty$, we deduce
$$\lim_{k+l\to\infty} \sqrt{\frac{\log(k+l)+x}{l}} = 0. \tag{28}$$
  We see $k \sim \alpha l$ if l is sufficiently large. By Eq. (28) we obtain
$$\lim_{k+l\to\infty} (1-e^{-\frac{n}{l}-y\sqrt{\frac{n}{l^2}}})^k = \lim_{l\to\infty}\left(1-\frac{e^{-x}}{(\alpha+1)l}\right)^{\alpha l} = e^{-\frac{\alpha e^{-x}}{\alpha+1}}.$$
  Therefore by the Lebesgue bounded convergence theorem, we conclude
$$\lim_{k+l\to\infty} E[(1-e^{-X_{n,l}})^k] = \lim_{k+l\to\infty} E[(1-e^{-\frac{n}{l}-N_{0,1}\sqrt{\frac{n}{l^2}}})^k] = e^{-\frac{\alpha e^{-x}}{\alpha+1}}.$$
- The case of $\alpha = \infty$:

– If $\beta = 0$, then Eq. (28) also holds. In this case, there exists a sequence $\{c_l\}$ such that $\lim_{l\to\infty} \frac{\log c_l}{l} = 0$, $\lim_{l\to\infty} c_l = \infty$, and $k = lc_l$. Therefore using Eq. (27) we obtain

$$\lim_{k+l\to\infty} (1 - e^{-\frac{n}{l} - y\sqrt{\frac{n}{l^2}}})^k = \lim_{l\to\infty} \left(1 - \frac{e^{-x}}{(c_l+1)l}\right)^{c_l l} = e^{-e^{-x}}.$$

Hence applying again the theorem, we conclude

$$\lim_{k+l\to\infty} E[(1 - e^{-X_{n,l}})^k] = \lim_{k+l\to\infty} E[(1 - e^{-\frac{n}{l} - N_{0,1}\sqrt{\frac{n}{l^2}}})^k] = e^{-e^{-x}}.$$

– If $0 < \beta < \infty$, that is $\log k \sim \beta l$, then the limit of Eq. (28) is not 0 but $\sqrt{\beta}$ for any x. Therefore the limit of Eq. (27) is $\exp\left\{-e^{-x-y\sqrt{\beta}}\right\}$. Hence applying again the theorem, we have

$$\lim_{k+l\to\infty} E[(1 - e^{-X_{n,l}})^k] = E\left[\exp\left\{-e^{-x-N_{0,1}\sqrt{\beta}}\right\}\right].$$

□

Remark 3.1 *If $\alpha = \infty$ and $\beta = \infty$, then we see that $\lim_{k+l\to\infty} E[(1 - e^{-X_{n,l}})^k]$ is degenerate for $n = l(\log(l+k) + x)$. Hence we use $n = l\log(l+k) + x\sqrt{l\log(l+k)}$ in Item (iii) of Theorem 1.2.*

Proof of Theorem 1.2. Using Eq. (7), we combine Lemmas 3.2 and 3.3.

- Item (i): Since $\beta = 0$, the case includes both $\alpha < \infty$ and $\alpha = \infty$.
  - If $\alpha < \infty$, then by Eqs. (23) and (24) we have

$$\lim_{k+l\to\infty} P(Y^{k+l,k} \le l\{\log(l+k) + x\}) = e^{-\frac{\alpha e^{-x}}{\alpha+1}} \times e^{-\frac{e^{-x}}{\alpha+1}} = e^{-e^{-x}}.$$

  - If $\alpha = \infty$, then by Eqs. (23) and (25) we have

$$\lim_{k+l\to\infty} P(Y^{k+l,k} \le l\{\log(l+k) + x\}) = 1 \times e^{-e^{-x}}.$$

- Item (ii): Since $0 < \beta < \infty$, we see $\alpha = \infty$. By Eqs. (23) and (26) we have

$$\lim_{k+l\to\infty} P(Y^{k+l,k} \le l\{\log(l+k) + x\}) = 1 \times E\left[\exp\left\{-e^{-x-N_{0,1}\sqrt{\beta}}\right\}\right].$$

- Item (iii): Since $\beta = \infty$, we see $\alpha = \infty$ and $l = o(\log k)$. For k, l and any real x, put

$$n = n(k,l,x) = l\log(k+l) - x\sqrt{l\log(k+l)}. \tag{29}$$

By Eq. (7) and $\alpha = \infty$ in Lemma 3.2, we have for n in Eq. (29)

$$\lim_{k+l\to\infty} P(Y^{k+l,k} \le l\log(k+l) - x\sqrt{l\log(k+l)}) = \lim_{k+l\to\infty} E[(1 - e^{-X_{n,l}})^k].$$

Applying Lemma 3.1, we consider $E\left[\left(1 - e^{-\frac{n}{l} - N_{0,1}\sqrt{\frac{n}{l^2}}}\right)^k\right]$ instead of $E[(1 - e^{-X_{n,l}})^k]$ for a sufficiently large n defined by Eq. (29). Hence for any $\epsilon > 0$, we have for sufficiently large numbers a, k

$$\left| P(Y^{k+l,k} \le l\log(k+l) - x\sqrt{l\log(k+l)}) - \int_{-a}^{+a} \left(1 - e^{-\frac{n}{l} - y\sqrt{\frac{n}{l^2}}}\right)^k \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy \right| < \epsilon.$$

If we can show that

$$\lim_{k+l\to\infty} \{1 - e^{-n/l - y\sqrt{n}/l}\}^k = \begin{cases} 0, & \text{if } x > y \\ 1, & \text{if } x < y \end{cases} \tag{30}$$

for $y \in (-a, a)$, then by the Lebesgue bounded convergence theorem we have

$$\lim_{k+l\to\infty} P(Y^{k+l,k} \le l\log(k+l) - x\sqrt{l\log(k+l)}) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{y^2}{2}}dy = 1 - \Phi(x).$$

Hence we conclude the desired result, namely Eq. (15). We check Eq. (30). To obtain Eq. (30), it is sufficient to show

$$\begin{aligned}\lim_{k+l\to\infty} \log\{1 - e^{-n/l - y\sqrt{n}/l}\}^k &= -\lim_{k+l\to\infty} \exp\left\{-\frac{n}{l} - y\frac{\sqrt{n}}{l} + \log k\right\} \\ &= \begin{cases} -\infty, & \text{if } x > y \\ 0, & \text{if } x < y. \end{cases}\end{aligned} \tag{31}$$

In fact, the limit of the exponential part of Eq. (31) is

$$\begin{aligned}&\lim_{k+l\to\infty}\left\{-\frac{n}{l} - y\frac{\sqrt{n}}{l} + \log k\right\} \\ =\ &\lim_{k+l\to\infty}\left\{\log\frac{k}{k+l} + x\sqrt{\frac{\log(k+l)}{l}} - y\sqrt{\frac{\log(k+l)}{l} - x\sqrt{\frac{\log(k+l)}{l^3}}}\right\} \\ \overset{(*)}{=}\ &\begin{cases} +\infty, & \text{if } x > y, \\ -\infty, & \text{if } x < y. \end{cases}\end{aligned}$$

Since $l = o(\log k)$, we have

$$\lim_{k+l\to\infty} \frac{\log(k+l)}{l} = \infty.$$

Therefore the equality of $(*)$ holds. Hence Eq. (31) follows.

4 Conclusion

In this article, we study a random collecting time of a coupon collector's problem with bonuses. We give not only the expectation but also the exact distribution. As a result, the distribution function has a representation that the effect of bonus coupons and the one of ordinary coupons are specifically separated (see Eq. (7)). Moreover the limit distributions are classified by the growth order of the number of bonus coupons compared to the one of ordinary coupons.

- If the growth order is smaller than the exponential order, then the limit distribution is Gumbel.
- If the growth order is greater than the exponential order, then the limit distribution is Gauss.
- If the growth order is exponential, then the limit distribution is mixed by the Gumbel distribution and the Gauss distribution.

We will study the distribution more precisely by investigating the record-time appeared in extreme value theory in our future works.

References

[1] I. Adler, H. Ahn, R. Karp, and S. Ross, Coalescing Times for IID Random Variables with Applications to Population Biology, *Random Struct. Alg.*, **23**, No. 2, 155–166, 2003.

[2] I. Adler, S. Oren, and S. Ross, The coupon-collector's problem revised, *J. Appl. Probab.*, **40**, No. 2, 513–518, 2003.

[3] M. Adler, E. Halperin, R. Karp, and V. Vazirani, A stochastic process on the hypercube with applications to peer-to-peer networks, *Proc. STOC 2003*, 575–584, 2003.

[4] D. Aldous and J. Fill, *Reversible Markov Chains and Random Walks on Graphs*, to appear http://www.stat.berkeley.edu/users/aldous/RWG/book.html

[5] B. Arnold, N. Balakrishnan, and H. Nagaraja, *Records*, Wiley, 1998.

[6] N. Dimitrov and C. Plaxton, Optimal Cover Time for a Graph-Based Coupon Collector Process, *Proc. ICALP 2005, LNCS* **3580**, 702-716, 2005.

[7] R. Durrett, *Probability: Theory and Examples, Second Edition*, Duxbury Press, 1996.

[8] W. Feller, *An Introduction to Probability Theory and Its Applications, vol. 1,* New York: Wiley, 1968.

[9] P. Flajolet, D. Gardy, and L. Thimonier, Birthday paradox, coupon collectors, caching algorithms and self-organizing search, *Discrete Appl. Math.*, **39**, no. 3, 207–229, 1992.

[10] P. Flajolet and R. Sedgewick, *Analytic combinatorics*, Symbolic Combinatorics, http://algo.inria.fr/flajolet/Publications/books.html, 2002.

[11] D. Gardy, Occupancy urn models in the analysis of algorithms, Special issue on lattice path combinatorics and applications (Vienna, 1998), *J. Statist. Plann. Inference,* **101**, no. 1-2, 95–105, 2002.

[12] G. Grimmett and D. Stirzaker, *One Thousand Exercises in Probability,* Oxford, 1 edition, 2002.

[13] E. J. Gumbel, *Statistics of Extremes*, Dover Publications, 2004.

[14] S. Karlin and H. Taylor, *A First Course in Stochastic Processes, (2nd ed.)*, Academic Press, New York, 1975.

[15] T. Lindvall, *Lectures on the coupling method*, Wiley, 1992.

[16] S. Martinez, Some bounds on the coupon collector problem, *Random Structures Algorithms*, **25**, no. 2, 208–226, 2004.

[17] M. Mitzenmacher and E. Upfal, *Probability and Computing*, Cambridge University Press, 2005.

[18] R. Motwani and P. Raghavan, *Randomized Algorithms*, Cambridge University Press, New York, 1995.

[19] A. Myers and H. Wilf, Some new aspects of the coupon collector's problem, *SIAM J. Discrete Math.*, **17**, no. 1, 1–17, 2003.

[20] M. Neuts, Waiting times between record observations, *J. Appl. Prob.*, **4**, 206–208, 1967.

[21] S. Ross, *Probability models for computer science*, Harcourt/Academic Press, 2002.

[22] H. Wilf, *generatingfunctionology*, Academic Press, 1990.