

## A COLLECTOR'S PROBLEM WITH RENEWAL ARRIVAL PROCESSES

OFFER KELLA,\* *The Hebrew University of Jerusalem*

WOLFGANG STADJE,\*\* *University of Osnabrück*

### Abstract

We study a collector's problem with  $K$  renewal arrival processes for different type items, where the objective is to collect complete sets. In particular, we derive the asymptotic distribution of the sequence of interarrival times between set completions.

*Keywords:* Collector's problem; complete set; renewal processes; superposition; asymptotic distribution; mixture

2000 Mathematics Subject Classification: Primary 60K05  
Secondary 60F05

### 1. Introduction

The original collector's problem deals with the probability distributions arising when collecting complete sets from sequentially arriving items of  $K$  different types. For recent work on this classical problem and its ramifications, we refer the reader to [1], [4], and [6]; for surveys on discrete-time models (i.e. new items are obtained at every time instant  $t = 1, 2, \dots$ ), see [2] and [5] for more recent contributions and [7] for older work.

In this paper we tackle the situation when the arrival times of the items are random, which to the authors' knowledge has not been studied before. In principle, we can consider a general marked point process on  $(0, \infty)$  with one of  $K$  different marks assigned to each point. The times of interest to the collector are the arrival times of points completing a set of  $K$  points with different marks. Each completed set is immediately put aside and the waiting for the completion of the next set begins. In this paper we will study the sequence of intercompletion times in the special case that, for every mark  $k \in \{1, \dots, K\}$ , the arrival times of items marked  $k$  form a renewal stream. We will suppose that the  $K$  renewal processes are independent, but our results will also be seen to hold under a slightly weaker assumption.

Technically, let  $\{X_n^k \mid n \geq 1\}$ ,  $k = 1, \dots, K$ , be  $K$  independent sequences of independent and identically distributed (i.i.d.) positive random variables with distribution functions  $F_1, \dots, F_K$ . Let  $S_0^k = 0$  and  $S_n^k = \sum_{i=1}^n X_i^k$  for  $n \geq 1$ . In our setting,  $S_1^k, S_2^k, \dots$  are the arrival times of items of type  $k$ . Now define  $M_n = \max_{1 \leq k \leq K} S_n^k$ ; clearly,  $M_n$  is the time at which the  $n$ th complete set can be formed.

In the corresponding discrete-time models, the basic collector's problem is to determine the expected number of items arriving before obtaining a complete set. In our situation we can start by deriving the expected value of the time  $T$  of the first completion and the expected number

---

Received 3 April 2008; revision received 9 July 2008.

\* Postal address: Department of Statistics, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem, 91905, Israel. Email address: offer.kella@huji.ac.il

\*\* Postal address: Department of Mathematics and Computer Science, University of Osnabrück, 49069 Osnabrück, Germany. Email address: wolfgang@mathematik.uni-osnabrueck.de

of items of type  $k$  obtained before  $T$ . Of course,  $T = \max_{1 \leq k \leq K} X_1^k$ , so that

$$E(T) = \int_0^\infty g \left( 1 - \prod_{k=1}^K F_k(t)g \right) dt.$$

For the expected number of items of a given type that have arrived before  $T$ , we obtain the following result.

**Proposition 1.** *Let  $F_1, \dots, F_K$  be continuous. Let  $N_k(t) = \sum_{n=1}^\infty F_k^{*n}(t)$  be the renewal function associated with  $F_k$  (where  $F_k^{*n}$  denotes the  $n$ -fold convolution of  $F_k$  with itself). Then the expected number of items of type  $k$  obtained before time  $T$ , say  $m_k$ , is given by*

$$m_k = \sum_{\substack{1 \leq j \leq K \\ j \neq k}} \int_0^\infty N_k(x)g \left( \prod_{l \in \{1, \dots, K\} \setminus \{k, j\}} F_l(x)g \right) dF_j(x). \tag{1}$$

*Proof.* The assumption that  $F_1, \dots, F_K$  are continuous excludes ties with probability 1. For any  $k, j \in \{1, \dots, K\}$ , let  $m_{k,j}(x)$  be the conditional expected number of items of type  $k$  obtained before time  $T$  given that  $T = X_1^j = x$ . Then  $m_{k,k}(x) = 0$  and, for  $k \neq j$ , we have

$$\begin{aligned} m_{k,j}(x) &= E(\text{number of items of type } k \text{ obtained before time } x \mid X_1^k < x) \\ &= 1 + \int_0^x \frac{N_k(x-t)}{F_k(x)} dF_k(t) \\ &= 1 + \frac{(N_k * F_k)(x)}{F_k(x)} \\ &= 1 + \frac{N_k(x) - F_k(x)}{F_k(x)} \\ &= \frac{N_k(x)}{F_k(x)}. \end{aligned} \tag{2}$$

Note that  $F_k(x) = 0$  implies that  $N_k(x) = 0$ , and, for those values of  $x$ , we set the right-hand side of (2) equal to 0 (which is obviously the value of  $m_{k,j}(x)$  in this case). As

$$P \left[ \max_{1 \leq l \leq K} X_1^l = X_1^j \in dx \right] = \left( \prod_{l \neq j} F_l(x) \right) dF_j(x),$$

deconditioning in (2) yields

$$m_k = \int_0^\infty \sum_{j: j \neq k} \frac{N_k(x)}{F_k(x)} \left( \prod_{l: l \neq j} F_l(x) \right) dF_j(x),$$

which is tantamount to (1).

The main results of this paper concern the sequence  $Y_n = M_n - M_{n-1}$  of intercompletion times. Of course,  $Y_1, Y_2, \dots$  are neither independent nor identically distributed. We could try to approach  $Y_n$  by embedding  $M_n$  into a complicated high-dimensional Markov chain, say  $Z_n$ . The components of  $Z_n$  are  $M_n$  itself, the index  $\kappa_n$  for which  $\max_{1 \leq k \leq K} S_n^k$  is taken, and, for every  $k \in \{1, \dots, K\} \setminus \{\kappa_n\}$ , the sequence  $(S_{n+j}^k \mathbf{1}_{\{S_{n+j}^k \leq M_n\}})_{j \geq 1}$ , indicating the arrival times of

all the items that are still present just after the first  $n$  complete sets have been removed. We will not follow this idea but use a direct method to determine the asymptotic distribution of  $Y_n$ ; our approach will then be extended to also derive the asymptotic distribution of the tails of the entire sequence  $(Y_{n+N})_{N \geq 1}$  as  $n \rightarrow \infty$ . We will show that if the means  $\mu_1, \dots, \mu_K$  and variances  $\sigma_1^2, \dots, \sigma_K^2$  of  $F_1, \dots, F_K$  are finite and  $\mu_1 = \mu_2 = \dots = \mu_{k_0} > \mu_{k_0+1} \geq \dots \geq \mu_K$ , then  $Y_n$  converges in distribution to the mixture  $\sum_{k=1}^{k_0} \pi_k^{k_0} F_k$ , where the weights  $\pi_k^{k_0}$  can be defined in terms of i.i.d.  $N(0, 1)$  random variables  $Z_1, \dots, Z_{k_0}$  by

$$\pi_k^{k_0} = P[\sigma_k Z_k > \max\{\sigma_\ell Z_\ell \mid 1 \leq \ell \leq k_0, \ell \neq k\}].$$

It is interesting (and maybe counterintuitive) to note that, for  $k_0 = 2$ , the weights are independent of the variances  $\sigma_1^2$  and  $\sigma_2^2$ . Since  $\sigma_1 Z_1 - \sigma_2 Z_2$  is normally distributed with mean 0, its distribution is symmetric around 0 so that

$$\pi_1^2 = P[\sigma_1 Z_1 > \sigma_2 Z_2] = P[\sigma_1 Z_1 - \sigma_2 Z_2 > 0] = \frac{1}{2}$$

and, of course,  $\pi_2^2 = \frac{1}{2}$  also. For  $k_0 > 2$ , the weights do depend on the variances in an intricate way. We will show how to compute the  $\pi_k^{k_0}$  in closed form. The weak limit of the entire tail sequence  $(Y_{n+N})_{N \geq 1}$  as  $n \rightarrow \infty$  is given by the mixture of independent i.i.d. sequences  $(X_1^k, X_2^k, \dots)$  with the same mixing coefficients  $\pi_k^{k_0}$ ,  $k = 1, \dots, k_0$ . We also present an extension to the case of infinite variances, using the theory of stable laws.

### 2. The core result

In the proofs below we need the following general proposition for which no reference could be found. Let ‘ $\xrightarrow{D}$ ’ and ‘ $\xrightarrow{P}$ ’ denote convergence in distribution and in probability, respectively.

**Proposition 2.** *Let  $G$  and  $G_n$ ,  $n \geq 1$ , be distribution functions, and let  $\xi_n$  and  $\eta_n$  ( $n \geq 1$ ) be random variables such that  $\xi_n \sim G_n$  and  $\eta_n \xrightarrow{P} 0$ . If  $G_n \xrightarrow{D} G$  and 0 is a continuity point of  $G$ , then  $\mathbf{1}_{\{\xi_n + \eta_n > 0\}} - \mathbf{1}_{\{\xi_n > 0\}} \xrightarrow{P} 0$ .*

*Proof.* All we need is to show that the probability that  $\mathbf{1}_{\{\xi_n + \eta_n > 0\}} - \mathbf{1}_{\{\xi_n > 0\}}$  is not 0 converges to 0. For any  $\epsilon > 0$  for which  $-\epsilon$  is a continuity point of  $G$ , we have

$$\begin{aligned} P[\xi_n + \eta_n > 0, \xi_n \leq 0] &= P[\eta_n > -\xi_n, \xi_n \leq -\epsilon] + P[\eta_n > -\xi_n, -\epsilon < \xi_n \leq 0] \\ &\leq P[\eta_n > \epsilon] + P[-\epsilon < \xi_n \leq 0] \\ &\rightarrow G(0) - G(-\epsilon) \quad \text{as } n \rightarrow \infty, \end{aligned} \tag{3}$$

and since 0 is a continuity point of  $G$ , the right-hand side converges to 0 as  $\epsilon \downarrow 0$ . Similarly, if  $\epsilon > 0$  is a continuity point of  $G$ ,

$$\begin{aligned} P[\xi_n + \eta_n \leq 0, \xi_n > 0] &= P[\eta_n \leq -\xi_n, \xi_n > 0] \\ &= P[\eta_n \leq -\xi_n, \xi_n \geq \epsilon] + P[\eta_n \leq -\xi_n, 0 < \xi_n \leq \epsilon] \\ &\leq P[\eta_n \leq -\epsilon] + P[0 < \xi_n \leq \epsilon] \\ &\rightarrow G(\epsilon) - G(0) \quad \text{as } n \rightarrow \infty, \end{aligned} \tag{4}$$

and the right-hand side converges to 0 as  $\epsilon \downarrow 0$ .

**Corollary 1.** Let  $G^k$  and  $G_n^k$  be distribution functions for  $k = 1, \dots, K$  and  $n \geq 1$ , and let  $\xi_n^k$  and  $\eta_n^k$  be random variables such that  $\xi_n^k \sim G_n^k \xrightarrow{D} G^k$  and  $\eta_n^k \xrightarrow{P} 0$  as  $n \rightarrow \infty$  for  $k = 1, \dots, K$ . Assume that 0 is a continuity point of  $G^1, \dots, G^K$ . Then

$$\prod_{k=1}^K \mathbf{1}_{\{\xi_n^k + \eta_n^k > 0\}} - \prod_{k=1}^K \mathbf{1}_{\{\xi_n^k > 0\}} \xrightarrow{P} 0. \tag{5}$$

*Proof.* Obviously,

$$\prod_{k=1}^K \mathbf{1}_{\{\xi_n^k + \eta_n^k > 0\}} = \prod_{k=1}^K ((\mathbf{1}_{\{\xi_n^k + \eta_n^k > 0\}} - \mathbf{1}_{\{\xi_n^k > 0\}}) + \mathbf{1}_{\{\xi_n^k > 0\}}). \tag{6}$$

By (6),  $\prod_{k=1}^K \mathbf{1}_{\{\xi_n^k + \eta_n^k > 0\}} - \prod_{k=1}^K \mathbf{1}_{\{\xi_n^k > 0\}}$  can be written as a sum of products, each having at least one term of the form  $\mathbf{1}_{\{\xi_n^k + \eta_n^k > 0\}} - \mathbf{1}_{\{\xi_n^k > 0\}}$  for some  $k \in \{1, \dots, K\}$ . Since this term converges in probability to 0 as  $n \rightarrow \infty$ , it follows that each product and, hence, also the whole sum converges in probability to 0.

Now let  $\{X_n^k \mid n \geq 1\}$ ,  $k = 1, \dots, K$ , be independent sequences of i.i.d. positive random variables with distributions  $F_k$ , and call their means  $\mu_k$  (when finite) and their variances  $\sigma_k^2$  (when finite). Let  $S_0^k = 0$  and, for  $n \geq 1$ ,  $S_n^k = \sum_{i=1}^n X_i^k$ . In our setting  $S_1^k, S_2^k, \dots$  are the arrival times of items of type  $k$ . Finally, we set  $M_n = \max_{1 \leq k \leq K} S_n^k$ , so that  $M_n$  is the time at which the  $n$ th complete set can be formed, and  $Y_n = M_n - M_{n-1}$ . Of course, the intercompletion times  $Y_1, Y_2, \dots$  are neither independent nor identically distributed.

**Theorem 1.** Assume that  $F_1, \dots, F_K$  are continuous distribution functions with finite means  $\mu_1, \dots, \mu_K$  and finite variances  $\sigma_1^2, \dots, \sigma_K^2$ , and that the means are equal:  $\mu_1 = \dots = \mu_K = \mu$ . Let  $Z_1, \dots, Z_K$  be i.i.d.  $N(0, 1)$  random variables that are independent of  $\{X_n^k \mid n \geq 1, k = 1, \dots, K\}$ . Then, with

$$\kappa = \arg \max\{\sigma_k Z_k \mid k = 1, \dots, K\}, \tag{7}$$

we have

$$Y_n \xrightarrow{D} X_1^\kappa = \sum_{k=1}^K X_1^k I_k, \tag{8}$$

where  $I_1, \dots, I_K$  are  $\{0, 1\}$ -valued random variables that are independent of  $X_1^1, \dots, X_1^K$  and have the joint distribution

$$\mathbb{P}[(I_1, \dots, I_K) = e_k] = \mathbb{P}[\sigma_\kappa Z_\kappa > \max\{\sigma_\ell Z_\ell \mid \ell \neq \kappa\}], \tag{9}$$

where  $e_k$  is a unit vector with a 1 in the  $k$ th coordinate and 0 elsewhere.

*Proof.* We first note that

$$M_n = \sum_{k=1}^K S_n^k \mathbf{1}_{\{S_n^k > S_n^\ell; \ell \neq k\}} \quad \text{almost surely (a.s.),} \tag{10}$$

which implies that

$$\begin{aligned} Y_n &= M_n - M_{n-1} \\ &= \sum_{k=1}^K X_n^k \mathbf{1}_{\{S_{n-1}^k > S_{n-1}^\ell; \ell \neq k\}} + \sum_{k=1}^K S_n^k (\mathbf{1}_{\{S_n^k > S_n^\ell; \ell \neq k\}} - \mathbf{1}_{\{S_{n-1}^k > S_{n-1}^\ell; \ell \neq k\}}) \quad \text{a.s.} \end{aligned} \tag{11}$$

If we can show that

$$\mathbf{1}_{\{S_n^k > S_n^\ell; \ell \neq k\}} - \mathbf{1}_{\{S_{n-1}^k > S_{n-1}^\ell; \ell \neq k\}} \xrightarrow{P} 0 \tag{12}$$

then this would immediately imply that

$$\sum_{k=1}^K S_n^k (\mathbf{1}_{\{S_n^k > S_n^\ell; \ell \neq k\}} - \mathbf{1}_{\{S_{n-1}^k > S_{n-1}^\ell; \ell \neq k\}}) \xrightarrow{P} 0. \tag{13}$$

Actually, (12) implies that the probability that the sum in (13) is equal to 0 tends to 1 as  $n \rightarrow \infty$ . Thus, it will be enough to show (12) and that

$$\sum_{k=1}^K X_n^k \mathbf{1}_{\{S_{n-1}^k > S_{n-1}^\ell; \ell \neq k\}} \xrightarrow{D} X_1^\kappa. \tag{14}$$

Now fix  $k$ . Letting

$$\xi_n^\ell = \frac{S_{n-1}^k - S_{n-1}^\ell}{\sqrt{n-1}} \quad \text{and} \quad \eta_n^\ell = \frac{X_n^k - X_n^\ell}{\sqrt{n-1}},$$

we can write

$$\mathbf{1}_{\{S_n^k > S_n^\ell; \ell \neq k\}} - \mathbf{1}_{\{S_{n-1}^k > S_{n-1}^\ell; \ell \neq k\}} = \prod_{\ell \neq k} \mathbf{1}_{\{\xi_n^\ell + \eta_n^\ell > 0\}} - \prod_{\ell \neq k} \mathbf{1}_{\{\xi_n^\ell > 0\}}. \tag{15}$$

Since  $E(X_1^\ell)^2 < \infty$ , we have  $(X_n^\ell)^2/n \rightarrow 0$  a.s. and, thus,  $\eta_n^\ell \xrightarrow{P} 0$  (in fact, even a.s.). Rewriting  $\xi_n^\ell$  as

$$\xi_n^\ell = \sigma_k \frac{S_{n-1}^k - (n-1)\mu}{\sigma_k \sqrt{n-1}} - \sigma_\ell \frac{S_{n-1}^\ell - (n-1)\mu}{\sigma_\ell \sqrt{n-1}}, \tag{16}$$

it can be seen that  $\xi_n^\ell \xrightarrow{D} \sigma_k Z_k - \sigma_\ell Z_\ell \sim N(0, \sigma_k^2 + \sigma_\ell^2)$ . The distribution function of the limit is continuous everywhere, in particular at 0. Thus, the conditions of Corollary 1 are met and, thus, (13) indeed holds.

Next we note that

$$\sum_{k=1}^K X_n^k \mathbf{1}_{\{S_{n-1}^k > S_{n-1}^\ell; \ell \neq k\}} = X_n^{\kappa_n} \quad \text{a.s.}, \tag{17}$$

where  $\kappa_n = \arg \max\{S_{n-1}^k \mid k = 1, \dots, K\}$  is independent of  $(X_n^1, \dots, X_n^K)$ , which in turn is distributed like  $(X_1^1, \dots, X_1^K)$ . Thus, to complete the proof, it suffices to show that  $\kappa_n \xrightarrow{D} \kappa$ . This can be concluded from

$$P[\kappa_n = k] = P[S_{n-1}^k > S_{n-1}^\ell; \ell \neq k] = P[\min\{\xi_n^\ell \mid \ell \neq k\} > 0], \tag{18}$$

where the  $\xi_n^\ell$  are as in (16). Note that  $(\xi_n^\ell)_{\ell \in \{1, \dots, K\} \setminus \{k\}}$  is of the form  $(V_n^k - V_n^\ell)_{\ell \in \{1, \dots, K\} \setminus \{k\}}$ , where  $V_n^1, \dots, V_n^K$  are independent and  $V_n^\ell \xrightarrow{D} \sigma_\ell Z_1$ . Hence,

$$(\xi_n^\ell)_{\ell \in \{1, \dots, K\} \setminus \{k\}} \xrightarrow{D} (\sigma_k Z_k - \sigma_\ell Z_\ell)_{\ell \in \{1, \dots, K\} \setminus \{k\}} \quad \text{as } n \rightarrow \infty.$$

As the function  $h: \mathbb{R}^{K-1} \rightarrow \mathbb{R}$  defined by  $h(u_1, \dots, u_{K-1}) = \min\{u_1, \dots, u_{K-1}\}$  is continuous, it follows that  $\min\{\xi_n^\ell \mid \ell \neq k\}$  converges in distribution to  $\min\{\sigma_k Z_k - \sigma_\ell Z_\ell \mid \ell \neq k\}$ , and since the distribution of the latter is continuous, the right-hand side of (18) converges to  $P[\kappa = k]$ , as desired.

### 3. Generalizations

Theorem 1 can be generalized in several ways.

**Extension 1.** Instead of assuming that the  $K$  renewal processes are independent, it is sufficient to suppose that the  $K$ -vectors  $\{(X_n^1, \dots, X_n^K) \mid n \geq 1\}$  of interarrival times are i.i.d. with a positive definite covariance matrix. In this case the vector  $(V_n^1, \dots, V_n^K)$  converges to a nondegenerate  $K$ -dimensional normal distribution from which the distribution of  $\kappa$  can be derived.

**Extension 2.** The means  $\mu_1, \dots, \mu_K$  do not have to be equal. Let us assume that  $\mu_1 = \dots = \mu_{k_0} > \mu_{k_0+1} \geq \dots \geq \mu_K$ . Write (10) for  $M_n$  as

$$M_n = \sum_{k=1}^{k_0} S_n^k \mathbf{1}_{\{S_n^k > S_n^\ell; \ell \neq k\}} + \sum_{k=k_0+1}^K S_n^k \mathbf{1}_{\{S_n^k > S_n^\ell; \ell \neq k\}} \quad \text{a.s.} \tag{19}$$

By the strong law of large numbers,

$$\lim_{n \rightarrow \infty} \max_{k=k_0+1, \dots, K} \frac{S_n^k}{n} = \mu_{k_0+1} < \lim_{n \rightarrow \infty} \min_{k=1, \dots, k_0} \frac{S_n^k}{n} \quad \text{a.s.}$$

Thus, the second sum in (19) converges to 0 a.s. and can be neglected; the proof can then proceed as in Section 2.

**Extension 3.** Theorem 1 gives the asymptotic distribution of  $Y_n$ . We can even obtain the asymptotic distribution of the tail  $(Y_{n+N})_{N \geq 1}$  of the entire sequence as  $n \rightarrow \infty$ . This distribution turns out to be the mixture obtained from the infinite product distributions of the sequences  $(X_1^k, X_2^k, \dots)$ ,  $k \in \{1, \dots, K\}$ , by choosing  $k$  according to the distribution  $\pi_k^K = P[\sigma_k Z_k > \max\{\sigma_\ell Z_\ell \mid \ell \neq k\}]$ . It suffices to determine the limiting finite-dimensional distributions. We show that, for any  $N \in \mathbb{N}$ ,

$$(Y_n, \dots, Y_{n+N-1}) \xrightarrow{D} (X_1^K, \dots, X_N^K) \quad \text{as } n \rightarrow \infty. \tag{20}$$

It is enough to prove that

$$\lim_{n \rightarrow \infty} P[\kappa_{n+1} = \kappa_{n+2} = \dots = \kappa_{n+N}] = 1. \tag{21}$$

Indeed, since

$$(Y_n, \dots, Y_{n+N-1}) = (X_n^{\kappa_{n+1}}, X_{n+1}^{\kappa_{n+2}}, \dots, X_{n+N-1}^{\kappa_{n+N}}),$$

(20) follows from  $\kappa_n \xrightarrow{D} \kappa$  and (21). Relation (21) in turn is an immediate consequence of

$$\lim_{n \rightarrow \infty} P[\kappa_{n+1} = k, \kappa_{n+N} = l] = 0 \quad \text{for all } k, l \in \{1, \dots, K\}, k \neq l. \tag{22}$$

To prove (22), fix  $k$  and  $l$ ,  $k \neq l$ . Then

$$\begin{aligned} P[\kappa_{n+1} = k, \kappa_{n+N} = l] &\leq P[S_n^k \geq S_n^l, S_{n+N-1}^k \leq S_{n+N}^l] \\ &= P[R_n \geq 0 \geq R_{n+N-1}], \end{aligned}$$

where  $R_n = S_n^k - S_n^l$ . Choose an arbitrary  $\epsilon > 0$ . We have

$$\begin{aligned} P[R_n \geq 0 \geq R_{n+N-1}] &\leq P[0 \leq R_n \leq \epsilon n^{1/2}] + P[R_n > \epsilon n^{1/2}, R_{n+N-1} - R_n < -\epsilon n^{1/2}] \\ &= I_n(\epsilon) + J_n(\epsilon). \end{aligned}$$

By the central limit theorem,

$$\lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} I_n(\epsilon) = 0.$$

Furthermore,  $R_{n+N-1} - R_n$  has the same distribution as  $S_{N-1}^k - S_{N-1}^l$ , which is independent of  $n$ , and, therefore,

$$\lim_{n \rightarrow \infty} J_n(\epsilon) = 0 \quad \text{for every } \epsilon > 0.$$

The latter two relations together imply that

$$\lim_{n \rightarrow \infty} P[R_n \geq 0 \geq R_{n+N-1}] = 0,$$

as remained to be shown.

**Extension 4.** The distribution functions  $F_1, \dots, F_K$  do not have to be continuous. If at least two of the  $F_k$  have discontinuities, it is possible that  $S_n^k = S_n^\ell = \max_l S_n^l$  for some  $n, k, \ell$ . However,

$$\left| M_n - \sum_{k=1}^K S_n^k \mathbf{1}_{\{S_n^k > S_n^\ell; \ell \neq k\}} \right| \leq \sum_{k=1}^K |S_n^k| \mathbf{1}_{\{S_n^k = S_n^\ell \text{ for some } \ell \neq k\}}$$

and, for any fixed indices  $k$  and  $l, k \neq l$ , it is clear that  $\lim_{n \rightarrow \infty} P[S_n^k - S_n^l = 0] = 0$ . Therefore,  $\kappa_n$  is unambiguously defined on a set  $A_n$  satisfying  $\lim_{n \rightarrow \infty} P[A_n] = 0$ . The nonuniqueness of  $\kappa_n$  on  $A_n$  does not disturb the line of reasoning in the proof of Theorem 1. (We may define  $\kappa_n$  in general to be the smallest index  $k \in \{1, \dots, K\}$  for which the maximum of  $S_{n-1}^k$  is attained.)

Summarizing, we obtain the following theorem.

**Theorem 2.** Assume that the  $K$ -vectors  $\{(X_n^1, \dots, X_n^K) \mid n \geq 1\}$  are i.i.d. with a positive definite covariance matrix  $(\sigma_{ij})_{1 \leq i, j \leq K}$  and means  $\mu_k = E(X_n^k)$  satisfying  $\mu_1 = \dots = \mu_{k_0} > \mu_{k_0+1} \geq \dots \geq \mu_K$  for some  $k_0 \in \{1, \dots, K\}$ . Let  $Z_1, \dots, Z_{k_0}$  be jointly normal random variables with zero means and covariance matrix  $(\sigma_{ij})_{1 \leq i, j \leq k_0}$ , which are independent of  $\{X_n^k \mid n \geq 1, k = 1, \dots, k_0\}$ . Then, with

$$\kappa = \arg \max \{ \sigma_{kk}^{1/2} Z_k \mid k = 1, \dots, k_0 \}, \tag{23}$$

we have, for every  $N \in \mathbb{N}$ ,

$$(Y_{n+N})_{N \geq 1} \xrightarrow{D} (X_1^\kappa, X_2^\kappa, X_3^\kappa, \dots). \tag{24}$$

We can also extend the result to the case of infinite variances, using the theory of stable laws. Again, for simplicity, assume that  $F_1, \dots, F_K$  have the same mean  $\mu$ . Fix an arbitrary  $\alpha \in (1, 2)$ . We impose the following condition.

- (C1) Every  $F_k, k = 1, \dots, K$ , belongs to the domain of attraction of a stable law with exponent  $\alpha$ , i.e. there are positive sequences  $(a_n^k)_{n \geq 1}$  such that  $((a_n^k)^{-1}[S_n^k - n\mu])_{n \geq 1}$  converges weakly to a stable distribution with exponent  $\alpha$ .

We use some basic facts from the theory of stable laws; see, e.g. [3, Section XVII.5, pp. 574–581]. It is well known that  $F_k$  (a distribution function on  $(0, \infty)$ ) belongs to the domain of attraction of a stable law with exponent  $\alpha < 2$  if and only if  $1 - F_k(x)$  is regularly varying with exponent  $-\alpha$ , i.e.  $x^\alpha(1 - F_k(x))$  is slowly varying. In this case the constants  $a_n^k$  can be defined implicitly in terms of the *truncated second moment functions*

$$m_k(x) = \int_0^x u^2 dF_k(u)$$

by the relations

$$\lim_{n \rightarrow \infty} \frac{(a_n^k)^2}{nm_k(a_n^k)} = 1. \tag{25}$$

Furthermore, it is known that if (C1) holds then

$$\lim_{x \rightarrow \infty} \frac{x^{2-\alpha} L_k(x)}{m_k(x)} = 1 \tag{26}$$

for some slowly varying function  $L_k(x)$ , and

$$\lim_{x \rightarrow \infty} \frac{x^2(1 - F_k(x))}{m_k(x)} = \frac{2 - \alpha}{\alpha}. \tag{27}$$

We need the following additional conditions that the truncated second moment functions grow proportionally to each other and that the sequences  $a_n^k$  grow fast enough.

(C2) There are constants  $c_1, \dots, c_{K-1} > 0$  such that  $\lim_{x \rightarrow \infty} m_k(x)/m_K(x) = c_k, k = 1, \dots, K$ .

(C3)  $\sum_{n=1}^\infty (1 - F_k(a_n^k)) < \infty$  for some (and then all)  $k$ .

Note that, by (25), condition (C2) implies that

$$\lim_{n \rightarrow \infty} \frac{a_n^k}{a_n^\ell} = \left(\frac{c_k}{c_\ell}\right)^{1/2}. \tag{28}$$

By (25) and (26), it follows that  $a_n^k \sim n^{1/\alpha} L_k(a_n^k)^{1/\alpha}$ , so that condition (C3) is, for example, satisfied if  $E((X_1^k)^\alpha) < \infty$  and  $L_k(x)$  is bounded away from 0.

Now let  $U_1, \dots, U_K$  be i.i.d. random variables with common stable characteristic function

$$\omega(\zeta) = \exp\left\{|\zeta|^\alpha \frac{\Gamma(3 - \alpha)}{\alpha(\alpha - 1)} \left(\cos \frac{\pi\alpha}{2} + i \operatorname{sgn}(\zeta) \sin \frac{\pi\alpha}{2}\right)\right\}, \quad \zeta \in \mathbb{R}.$$

**Theorem 3.** *If (C1)–(C3) hold then*

$$Y_n \xrightarrow{D} \sum_{k=1}^K X_1^k I_k,$$

where the random vector  $(I_1, \dots, I_K)$  takes the values  $e_1, \dots, e_K$ , is independent of  $X_1^1, \dots, X_1^K$ , and has distribution

$$P[(I_1, \dots, I_K) = e_k] = P\left[c_k U_k > \max_{\ell \neq k} c_\ell U_\ell\right], \quad k = 1, \dots, K.$$

*Proof.* It is not difficult to see that the proof of Theorem 1 can be imitated provided that we can show that, for any fixed  $k$ ,

$$\bar{\xi}_n^\ell = \frac{S_{n-1}^k - S_{n-1}^\ell}{a_{n-1}^k} \xrightarrow{D} U_k - \frac{c_l}{c_k} U_\ell \tag{29}$$

and

$$\bar{\eta}_n^\ell = \frac{X_n^k - X_n^\ell}{a_{n-1}^k} \xrightarrow{P} 0. \tag{30}$$

Clearly, by (C1) and (28),

$$\begin{aligned} \bar{\xi}_n^\ell &= \frac{S_{n-1}^k - (n-1)\mu}{a_{n-1}^k} - \frac{a_{n-1}^\ell S_{n-1}^\ell - (n-1)\mu}{a_{n-1}^k a_{n-1}^\ell} \\ &\xrightarrow{D} U_k - \frac{c_l}{c_k} U_\ell, \end{aligned} \tag{31}$$

proving (29). Next, note that (30) would follow from

$$\sum_{n=1}^\infty \mathbb{P}\left[\frac{X_n^k}{a_{n-1}^k} > \epsilon\right] < \infty \quad \text{for all } \epsilon > 0 \tag{32}$$

(which implies that  $X_n^k/a_{n-1}^k \rightarrow 0$  a.s.) together with (28). Since  $\tilde{L}_k(x) = x^\alpha(1 - F_k(x))$  is slowly varying, we have

$$\begin{aligned} 1 - F_k(a_{n-1}^k \epsilon) &= (a_{n-1}^k \epsilon)^{-\alpha} \tilde{L}(a_{n-1}^k \epsilon) \\ &\sim \epsilon^{-\alpha} (a_{n-1}^k)^{-\alpha} \tilde{L}(a_{n-1}^k) \\ &= \epsilon^{-\alpha} (1 - F_k(a_{n-1}^k)). \end{aligned}$$

Thus, the series  $\sum_{n=1}^\infty (1 - F_k(a_{n-1}^k \epsilon))$  converges for every  $\epsilon > 0$  if and only if it converges for  $\epsilon = 1$ . Therefore, (C3) yields (32), and the proof is complete.

**4. The weights of the mixture for the superposition of independent renewal processes**

In the case of independent renewal processes with equal means and finite variances for the interarrival times (which we considered in Theorem 1), the weights of the limiting mixture are given by

$$\pi_k^K = \mathbb{P}[\sigma_k Z_k > \max\{\sigma_\ell Z_\ell \mid \ell \neq k\}].$$

Conditioning on  $Z_k$ , it is easily seen that  $\pi_k^K$  can be written in terms of the density  $\varphi$  and the distribution function  $\Phi$  of  $N(0, 1)$  as follows:

$$\pi_k^K = \int_{-\infty}^\infty \varphi(x) \prod_{\ell \neq k} \Phi\left(\frac{x\sigma_\ell}{\sigma_k}\right) dx, \quad k = 1, \dots, K. \tag{33}$$

These integrals can be computed recursively. Define the auxiliary functions

$$I_{n,l}(u_1, \dots, u_l) = \int_{-\infty}^\infty x^n \varphi(x) \prod_{i=1}^l \Phi(u_i x) dx \tag{34}$$

for  $n, l \in \mathbb{Z}_+$  and  $u_1, \dots, u_l \in \mathbb{R}_+$ . Clearly,

$$\pi_k^K = I_{0, K-1} \left( \frac{\sigma_k}{\sigma_1}, \dots, \frac{\sigma_k}{\sigma_{k-1}}, \frac{\sigma_k}{\sigma_{k+1}}, \dots, \frac{\sigma_k}{\sigma_K} \right). \tag{35}$$

For index pairs  $(n, 0)$ , we have

$$I_{n,0} = E[Z_1^n] = \begin{cases} \frac{n!}{(n/2)! 2^{n/2}}, & n = 0, 2, 4, 6, \dots, \\ 0, & n = 1, 3, 5, \dots \end{cases} \tag{36}$$

Thus, if all the  $I_{n,l}$  can be expressed in terms of the values  $I_{k,0}$ ,  $k \geq 0$ , we obtain explicit formulae for the  $\pi_k^K$ . The next theorem provides a suitable recursion.

**Theorem 4.** For  $n \in \mathbb{Z}_+$ ,  $l \geq 1$ , and  $u_1, \dots, u_l \in \mathbb{R}_+$ , we have

$$\begin{aligned} I_{n,l}(u_1, \dots, u_l) &= \frac{1}{2} I_{n,l-1}(u_2, \dots, u_l) \\ &+ \int_0^{u_1} \frac{1}{\sqrt{2\pi} (1+x^2)^{n/2+1}} I_{n+1,l-1} \left( \frac{u_2}{(1+x^2)^{1/2}}, \dots, \frac{u_l}{(1+x^2)^{1/2}} \right) dx. \end{aligned} \tag{37}$$

*Proof.* Taking the partial derivative with respect to  $u_1$  in (34), using the identities  $\Phi' = \varphi$  and  $\varphi(x)\varphi(u_1x) = (2\pi)^{-1/2} \varphi(x(1+u_1^2)^{1/2})$ , and making the substitution  $y = x(1+u_1^2)^{1/2}$ , we obtain

$$\begin{aligned} \frac{\partial}{\partial u_1} I_{n,l}(u_1, \dots, u_l) &= \int_{-\infty}^{\infty} x^{n+1} \varphi(x) \varphi(u_1x) \prod_{i=2}^l \Phi(u_ix) dx \\ &= \int_{-\infty}^{\infty} x^{n+1} (2\pi)^{-1/2} \varphi(x(1+u_1^2)^{1/2}) \prod_{i=2}^l \Phi(u_ix) dx \\ &= \int_{-\infty}^{\infty} (1+u_1^2)^{-n/2-1} y^{n+1} (2\pi)^{-1/2} \varphi(y) \prod_{i=2}^l \Phi(u_i y (1+u_1^2)^{-1/2}) dy \\ &= (1+u_1^2)^{-n/2-1} (2\pi)^{-1/2} I_{n+1,l-1} \left( \frac{u_2}{(1+u_1^2)^{1/2}}, \dots, \frac{u_l}{(1+u_1^2)^{1/2}} \right). \end{aligned} \tag{38}$$

Integrating (38) yields

$$\begin{aligned} I_{n,l}(u_1, \dots, u_l) - I_{n,l}(0, u_2, \dots, u_l) &= (2\pi)^{-1/2} \int_0^{u_1} (1+x^2)^{-n/2-1} I_{n+1,l-1} \left( \frac{u_2}{(1+x^2)^{1/2}}, \dots, \frac{u_l}{(1+x^2)^{1/2}} \right) dx. \end{aligned} \tag{39}$$

Since  $I_{n,l}(0, u_2, \dots, u_l) = \frac{1}{2} I_{n,l-1}(u_2, \dots, u_l)$ , (37) follows from (39).

Let us compute  $\pi_k^K$  for  $K = 2$  and  $K = 3$ .

**Example 1.** Let  $K = 2$ . We have  $I_{0,0} = 1$  and  $I_{1,0} = 0$  by (36) and, thus,

$$\pi_1^2 = I_{0,1} \left( \frac{\sigma_1}{\sigma_2} \right) = \frac{1}{2} I_{0,0} = \frac{1}{2}$$

and, of course, also

$$\pi_2^2 = \frac{1}{2}.$$

**Example 2.** Let  $K = 3$ . In this case  $\pi_1^3 = I_{0,2}(\sigma_1/\sigma_2, \sigma_1/\sigma_3)$ ,  $\pi_2^3 = I_{0,2}(\sigma_2/\sigma_1, \sigma_2/\sigma_3)$ , and  $\pi_3^3 = I_{0,2}(\sigma_3/\sigma_1, \sigma_3/\sigma_2)$ . To calculate  $I_{0,2}(u_1, u_2)$  using Theorem 3, we need  $I_{0,1}(u_1) \equiv \frac{1}{2}$  and  $I_{1,1}(u_1)$ . Regarding  $I_{1,1}(u_1)$ , we have, by (38),

$$\begin{aligned} I_{1,1}(u_1) &= \frac{1}{2}I_{1,0} + \int_0^{u_1} \frac{1}{\sqrt{2\pi}(1+x^2)^{3/2}} I_{2,0} dx \\ &= \frac{u_1}{\sqrt{2\pi}(1+u_1^2)^{1/2}}, \end{aligned} \quad (40)$$

where we have used  $I_{1,0} = E[Z_1] = 0$ ,  $I_{2,0} = E[Z_1^2] = 1$ , and a simple integral. Next let us derive  $I_{0,2}(u_1, u_2)$ :

$$\begin{aligned} I_{0,2}(u_1, u_2) &= \frac{1}{2}I_{0,1}(u_2) + \int_0^{u_1} \frac{1}{\sqrt{2\pi}(1+x^2)} I_{1,1}\left(\frac{u_2}{(1+x^2)^{1/2}}\right) dx \\ &= \frac{1}{4} + \frac{u_2}{2\pi} \int_0^{u_1} \frac{1}{(1+x^2)(1+x^2+u_2^2)^{1/2}} dx \\ &= \frac{1}{4} + \frac{1}{2\pi} \arctan\left(\frac{u_1 u_2}{(1+u_1^2+u_2^2)^{1/2}}\right), \end{aligned} \quad (41)$$

where we have used the indefinite integral

$$\int \frac{1}{(1+x^2)(1+x^2+u^2)^{1/2}} dx = \frac{1}{u} \arctan\left(\frac{ux}{(1+x^2+u^2)^{1/2}}\right), \quad u > 0.$$

It follows that

$$\pi_k^3 = \frac{1}{4} + \frac{1}{2\pi} \arctan\left(\frac{\sigma_k^2}{(\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2)^{1/2}}\right), \quad k = 1, 2, 3.$$

### Acknowledgements

This research was carried out while Wolfgang Stadje visited the Hebrew University of Jerusalem as a Lady Davis Visiting Professor. Offer Kella was supported in part by grant number 964/06 from the Israel Science Foundation and by the Vigevani Chair in Statistics.

### References

- [1] ADLER, I., OREN, S. AND ROSS, S. (2003). The coupon-collector's problem revisited. *J. Appl. Prob.* **40**, 513–518.
- [2] BONEH, A. AND HOFRI, M. (1997). The coupon-collector problem revisited – a survey of engineering problems and computational methods. *Stoch. Models* **13**, 39–66.
- [3] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd edn. John Wiley, New York.
- [4] FLAJOLET, P., GARDY, D. AND THIMONIER, L. (1992). Birthday paradox, coupon collectors, catching algorithms and self-organizing search. *Discrete Appl. Math.* **39**, 207–229.
- [5] KOBZA, J. E., JACOBSON, S. H. AND VAUGHAN, D. E. (2007). A survey of the coupon collector's problem with random sample sizes. *Methodol. Comput. Appl. Prob.* **9**, 573–584.
- [6] MYERS, A. M. AND WILF, H. S. (2006). Some new aspects of the coupon collector's problem. *SIAM Rev.* **48**, 549–565.
- [7] STADJE, W. (1990). The collector's problem with group drawings. *Adv. Appl. Prob.* **22**, 866–882.