Convergence of point processes associated with coupon collector's and Dixie cup problems

Andrii Ilienko*

Abstract

We prove that, in the coupon collector's problem, the point processes given by the times of rth arrivals for coupons of each type, centered and normalized in a proper way, converge toward a non-homogeneous Poisson point process. This result is then used to derive some generalizations and infinite-dimensional extensions of classical limit theorems on the topic.

Keywords: coupon collector's problem; Dixie cup problem; point processes; Poisson convergence; poissonization.
AMS MSC 2010: 60G55; 60F17.
Submitted to ECP on April 29, 2019, final version accepted on August 21, 2019.

1 Introduction

The coupon collector's problem (CCP), as well as its generalization known as the Dixie cup problem (DCP), belong to the classics of combinatorial probability. Their statements are as follows: a person collects coupons, each of which belongs to one of n different types. The coupons arrive one by one at discrete times, the type of each coupon being equiprobable and independent of types of preceding ones. Let $T_c^{(n)}$ stand for the (random) number of coupons a person needs to collect in order to assemble $c \in \mathbb{N}$ complete collections. The most typical questions concern asymptotics of $\mathbb{E}T_c^{(n)}$ and distributional limit theorems for $T_c^{(n)}$ themselves as $n \to \infty$. Sometimes, the case $c = 1$ refers to CCP while $c \geq 2$ to DCP. It should be noted that the terminology is not well established in the literature: sometimes both problems are attributed as CCP or, on the contrary, as DCP. The above terminology follows [12] and [9].

CCP, DCP and their further generalizations have a long history, going back to de Moivre, Euler and Laplace. Since the 60s of the past century, there has appeared an extensive literature on the topic. In particular, we recall here a classical result by Erdős and Rényi [7]:

$$\mathbb{E}T_c^{(n)} = n \ln n + (c-1)n \ln \ln n + (\gamma - \ln(c-1)!)n + o(n), \tag{1.1}$$

$$\lim_{n \to \infty} \mathbb{P}\left\{ \frac{T_c^{(n)}}{n} - \ln n - (c-1)\ln \ln n < x \right\} = \exp\left\{ -\frac{e^{-x}}{(c-1)!} \right\}, \tag{1.2}$$

with $\gamma = -\Gamma'(1)$ standing for the Euler–Mascheroni constant.

*Igor Sikorsky Kyiv Polytechnic Institute, Ukraine. E-mail: ilienko@matan.kpi.ua

Subsequently, the theory was developed and generalized in different directions: non-equal probabilities of coupon types (plenty of literature — [19], [5], [1], to cite just a few), various random sceneries ([13], [6], [9]), collecting pairs ([18], [10]), and so on. A nice and quite elementary introduction to the topic is given in [8].

In his seminal paper, Holst [12] proposed a fruitful poissonization idea which allowed to prove limit results like (1.1), (1.2) avoiding intricate combinatorial calculations. In a very recent paper by Glavaš and Mladenović [10], the connections between CCP and Poisson processes were shown to be even more tight. As a matter of fact, it was proved that the point processes given by the times of first arrivals for coupons of each type, centered and normalized in a proper way, converge toward a non-homogeneous Poisson point process as $n \to \infty$. The above convergence is, as usual, understood as the distributional one in the space of all locally finite point measures, endowed with the vague topology. The proof is based on rather delicate combinatorial arguments. As for the DCP, the authors do not consider the corresponding results, confining themselves to just pointing out that, within the framework of their methods, the relevant formulations and proofs would require much more technical details.

Inspired by this paper, the present note pursues a threefold objective. Firstly, we generalize the above result to the case of DCP. Secondly, to this end, we develop a specific approach involving a poissonization technique in the spirit of [12] and some coupling-based depoissonization procedure. This allows for avoiding sophisticated combinatorial machinery used in [10]. Thirdly, we demonstrate the power of this result from the applications point of view. It can be used to easily derive some generalizations and infinite-dimensional extensions of classical limit theorems on the topic.

2 Preliminaries and notation

Let $Y_{i,r}^{(n)}$, $i \in \mathbb{N}_n = \{1, \ldots, n\}$, $r, n \in \mathbb{N}$, stand for the time the rth coupon of type i arrives. So, $Y_{i,r}^{(n)} \sim \texttt{NegBin}\left(r, \frac{1}{n}\right)$, where by $\texttt{NegBin}$ we mean that version of the negative binomial distribution which counts trials up to (and including) the rth success:

$$\mathbb{P}\{Y_{i,r}^{(n)} = k\} = \binom{k-1}{r-1}\left(\frac{1}{n}\right)^r\left(1 - \frac{1}{n}\right)^{k-r}, \quad k \geq r.$$

For fixed n and r, the random variables $Y_{i,r}^{(n)}$, $i \in \mathbb{N}_n$, are identically distributed but not independent, since $Y_{i,r}^{(n)} \neq Y_{i',r}^{(n)}$ for $i \neq i'$.

In order to cope with this dependency, following [12], we consider a poissonized scheme. That is, we assume that coupons arrive at random times with independent $\texttt{Exp}(1)$-distributed intervals E_j, $j \in \mathbb{N}$. More formally, we introduce the unit-rate independently marked Poisson point process $\Xi^{(n)} = \sum_{k=1}^{\infty} \delta_{(X_k, M_k)}$ with the uniform on $\mathbb{N}_n$ mark distribution: $\mathbb{P}\{M_k = i\} = \frac{1}{n}$, $i \in \mathbb{N}_n$. Here, $X_k = \sum_{j=1}^{k} E_j$ stand for the arrival times, M_k for the types of arriving coupons, and δ_u for the Dirac measure $\mathbb{1}\{u \in \cdot\}$. Hence, by Theorem 5.8 in [16],

$$\Xi_i^{(n)} = \sum_{\substack{(X_k, M_k) \in \\ \text{supp}\, \Xi^{(n)}}} \delta_{X_k} \mathbb{1}\{M_k = i\} = \Xi^{(n)}(\cdot \times \{i\}), \quad i \in \mathbb{N}_n,$$

are independent $\frac{1}{n}$-rate Poisson point processes. The process $\Xi_i^{(n)}$ describes arrivals of coupons of ith type. In this setting, the random variables $Y_{i,r}^{(n)}$ introduced at the beginning of the section admit the following representation:

$$Y_{i,r}^{(n)} = \min\left\{m \geq r : \sum_{k=1}^{m} \mathbb{1}\{M_k = i\} = r\right\}. \tag{2.1}$$

Let $Z_{i,r}^{(n)}$, $i \in \mathbb{N}_n$, $r, n \in \mathbb{N}$, stand for the time the rth coupon of type i arrives in the above poissonized scheme. So, $Z_{i,r}^{(n)}$ are independent gamma-distributed random variables, $Z_{i,r}^{(n)} \sim \Gamma(r, \frac{1}{n})$. For any fixed n, the sequences $(Y_{i,r}^{(n)})$ and $(Z_{i,r}^{(n)})$ are now given on a common probability space and coupled by

$$Z_{i,r}^{(n)} = \sum_{j=1}^{Y_{i,r}^{(n)}} E_j, \quad i \in \mathbb{N}_n, r \in \mathbb{N}. \tag{2.2}$$

Moreover, $(Y_{i,r}^{(n)}, i \in \mathbb{N}_n, r \in \mathbb{N})$ is independent of $(E_j, j \in \mathbb{N})$, since, by (2.1), the former sequence is determined solely by marks M_k.

Let us denote

$$\psi_r^{(n)}(x) = \frac{x}{n} - \ln n - (r-1)\ln\ln n, \quad x \in \mathbb{R}. \tag{2.3}$$

The main object of our study is the (centered and normalized by means of $\psi_r^{(n)}$) point process of rth arrivals of different types:

$$\xi_r^{(n)} = \sum_{i=1}^{n} \delta_{\psi_r^{(n)}\left(Y_{i,r}^{(n)}\right)}. \tag{2.4}$$

In what follows, we will also need the counterpart of $\xi_r^{(n)}$ in the poissonized setting:

$$\eta_r^{(n)} = \sum_{i=1}^{n} \delta_{\psi_r^{(n)}\left(Z_{i,r}^{(n)}\right)}.$$

3 Main result

Before proceeding to the main result, we recall some basic definitions related to convergence of point processes (see [21], [22], or [14] for details). Let $M_p(\mathbb{R})$ denote the space of all locally finite point measures on $\mathbb{R}$. For $\mu, \mu_1, \mu_2, \ldots \in M_p(\mathbb{R})$, μ_n are said to converge vaguely to μ (denoted as $\mu_n \xrightarrow{v} \mu$) if $\int_{\mathbb{R}} f \, d\mu_n \to \int_{\mathbb{R}} f \, d\mu$ for each continuous compactly supported test function $f \colon \mathbb{R} \to [0, +\infty)$. The set $M_p(\mathbb{R})$, endowed with the corresponding topology, can be metrized as a complete separable metric space. This setting allows to consider the distributional convergence of point processes $\xi, \xi_1, \xi_2 \ldots$, denoted as $\xi_n \xrightarrow{vd} \xi$. The main result of this note, Theorem 3.1 below, asserts that the point processes $\xi_r^{(n)}$ converge in this sense toward a non-homogeneous Poisson process.

Theorem 3.1. *Let ξ_r be the Poisson point process on $\mathbb{R}$ with intensity measure $\lambda_r(dx) = \frac{1}{(r-1)!} e^{-x} \, dx$. Then $\xi_r^{(n)} \xrightarrow{vd} \xi_r$ as $n \to \infty$.*

Remark 3.2. The limiting process ξ_r allows for a simple interpretation. Let ξ be a stationary unit-rate Poisson point process restricted to $(0, +\infty)$, and put

$$h(x) = -\ln(r-1)! - \ln x, \quad x > 0. \tag{3.1}$$

Then, $\xi_r \stackrel{d}{=} \xi \circ h^{-1}$. In other words, $\xi_r \stackrel{d}{=} \sum_{x \in \text{supp}\,\xi} \delta_{h(x)}$.

Indeed, by the mapping theorem (see, e.g., Theorem 5.1 in [16]), $\xi \circ h^{-1}$ is a Poisson process with intensity measure of the form $\text{Leb} \circ h^{-1}$, where Leb stands for the Lebesgue measure. Since, for any $[a, b] \subset \mathbb{R}$,

$$\left(\text{Leb} \circ h^{-1}\right)[a, b] = \text{Leb}\left[\frac{e^{-b}}{(r-1)!}, \frac{e^{-a}}{(r-1)!}\right] = \int_a^b \frac{1}{(r-1)!} e^{-x} \, dx = \lambda_r[a, b],$$

the result follows.

We divide the proof of Theorem 3.1 into several steps. First, we prove a similar result in the poissonized setting.

Lemma 3.3. *We have* $\eta_r^{(n)} \xrightarrow{vd} \xi_r$ *as* $n \to \infty$.

Proof. Since $Z_{i,r}^{(n)} \sim \Gamma\big(r, \frac{1}{n}\big)$, the density $f_{i,r}^{(n)}$ of $Z_{i,r}^{(n)}$ is

$$f_{i,r}^{(n)}(x) = \frac{1}{n^r(r-1)!}x^{r-1}\mathrm{e}^{-\frac{x}{n}}\mathbb{1}\{x \geq 0\}.$$

So, the density $\tilde{f}_{i,r}^{(n)}$ of $\psi_r^{(n)}\big(Z_{i,r}^{(n)}\big)$ takes the form

$$\begin{aligned}
\tilde{f}_{i,r}^{(n)}(x) &= nf_{i,r}^{(n)}(nx + n\ln n + (r-1)n\ln\ln n) \\
&= \frac{1}{n(r-1)!}\mathrm{e}^{-x}\Big(1 + (r-1)\frac{\ln\ln n}{\ln n} + \frac{x}{\ln n}\Big)^{r-1}\mathbb{1}\{x \geq -\ln n - (r-1)\ln\ln n\}.
\end{aligned}$$

Hence, for any $x \in \mathbb{R}$,

$$\lim_{n\to\infty} n\tilde{f}_{i,r}^{(n)}(x) = \frac{1}{(r-1)!}\mathrm{e}^{-x},$$

and, moreover, this convergence is uniform over bounded sets. Then, for each such Borel set B, as $n \to \infty$,

$$n\int_B \tilde{f}_{i,r}^{(n)}(x)\,\mathrm{d}x \to \int_B \frac{1}{(r-1)!}\mathrm{e}^{-x}\,\mathrm{d}x,$$

and so, by Proposition 3.12 in [21],

$$n\mathbb{P}\{\psi_r^{(n)}\big(Z_{i,r}^{(n)}\big) \in \cdot\} \xrightarrow{v} \lambda_r(\cdot) \quad \text{on } \mathbb{R}.$$

Taking into account the independence of $Z_{i,r}^{(n)}$ for different i, the well-known fact on the convergence of binomial point processes toward a Poisson one (see, e.g., the warm-up in the proof of Proposition 3.21 in [21]) delivers the claim. $\qquad\square$

In the next stage, we will need some depoissonization procedure in order to turn $\eta_r^{(n)} \xrightarrow{vd} \xi_r$ into $\xi_r^{(n)} \xrightarrow{vd} \xi_r$. Such depoissonization techniques usually involve bounds on distances between random elements in poissonized and depoissonized settings (see, e.g., Lemma 1.4 in [20] for random variables and, as its application, Theorem 3.2 in [17] for point processes). Since $\xi_r^{(n)}$ and $\eta_r^{(n)}$ are related by (2.2), we may use this idea and rate how close both processes are.

Lemma 3.4. *For any segment* $[a,b] \subset \mathbb{R}$ *and any* $n \in \mathbb{N}$, $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P}\big\{\xi_r^{(n)}[a,b] &\neq \eta_r^{(n)}[a,b]\big\} \\
&\leq c_r\varepsilon^{-4}n^{-1} + \mathbb{P}\big\{\eta_r^{(n)}[a-\varepsilon, a+\varepsilon] \geq 1\big\} + \mathbb{P}\big\{\eta_r^{(n)}[b-\varepsilon, b+\varepsilon] \geq 1\big\} \quad (3.2)
\end{aligned}$$

with some $c_r > 0$.

Proof. The idea of (3.2) is pretty simple. Roughly speaking, there may be two reasons for $\xi_r^{(n)}[a,b] \neq \eta_r^{(n)}[a,b]$: either some point of $\xi_r^{(n)}$ deviated far away from the corresponding point of $\eta_r^{(n)}$, or there are points of $\eta_r^{(n)}$ close enough to the boundary of $[a,b]$. The first term on the right-hand side of (3.2) is responsible for the first reason while the rest for the second one. We proceed to the implementation.

Fix an $\varepsilon > 0$. Then, by (2.3) and (2.2),

$$\mathbb{P}\big\{\big|\psi_r^{(n)}\big(Z_{i,r}^{(n)}\big) - \psi_r^{(n)}\big(Y_{i,r}^{(n)}\big)\big| > \varepsilon\big\} = \mathbb{P}\big\{\big|Z_{i,r}^{(n)} - Y_{i,r}^{(n)}\big| > \varepsilon n\big\} = \mathbb{P}\bigg\{\bigg|\sum_{j=1}^{Y_{i,r}^{(n)}}(E_j - 1)\bigg| > \varepsilon n\bigg\}$$

for each $i \in \mathbb{N}_n$. So, by Markov inequality, the latter does not exceed

$$\Delta_{\varepsilon,n} = (\varepsilon n)^{-4} \, \mathbb{E} \left(\sum_{j=1}^{Y_{i,r}^{(n)}} (E_j - 1) \right)^4.$$

Since $E_j - 1$ are centered i.i.d., $\Delta_{\varepsilon,n}$ can be easily calculated by a standard conditioning argument: we have

$$\Delta_{\varepsilon,n} = (\varepsilon n)^{-4} \big(m_4 \, \mathbb{E} Y_{1,r}^{(n)} + m_2^2 \, \mathbb{E} Y_{1,r}^{(n)} (Y_{1,r}^{(n)} - 1) \big),$$

with m_k standing for $\mathbb{E}(E_1 - 1)^k$. As $Y_{1,r}^{(n)} \sim \mathtt{NegBin}\big(r, \frac{1}{n}\big)$, straightforward calculations lead to the bound

$$\Delta_{\varepsilon,n} \leq c_r \varepsilon^{-4} n^{-2}, \quad n \in \mathbb{N},$$

with some $c_r > 0$. Now we can finally bound the probability that some point of $\xi_r^{(n)}$ is far away from the corresponding point of $\eta_r^{(n)}$: by subadditivity,

$$\mathbb{P}\big\{ \big| \psi_r^{(n)} \big(Z_{i,r}^{(n)}\big) - \psi_r^{(n)} \big(Y_{i,r}^{(n)}\big) \big| > \varepsilon \text{ for some } i \in \mathbb{N}_n \big\} \leq n \Delta_{\varepsilon,n} \leq c_r \varepsilon^{-4} n^{-1}.$$

So, (3.2) follows from the reasoning at the beginning of the proof. $\qquad\square$

In order to deal with the last two terms on the right-hand side of (3.2), we will need an easy technical lemma on the distributional convergence of point processes.

Lemma 3.5. *Let* $X^{(n)}$, $n \in \mathbb{N}$, *and* X *be point processes on* $\mathbb{R}$ *such that* $X^{(n)} \xrightarrow{vd} X$, *and* X *has a diffuse intensity measure. If* $(I_n, n \in \mathbb{N})$ *is a decreasing sequence of intervals such that* $I_n \downarrow I$ *for some interval* I, *then*

$$\mathbb{P}\{X^{(n)}(I_n) \geq 1\} \to \mathbb{P}\{X(I) \geq 1\}.$$

Proof. The proof is based on an application of the Skorokhod coupling (see, e.g., [22], p. 41). With the latter in mind, we may assume that $\big(X^{(n)}, n \in \mathbb{N}\big)$ and X are given on a common probability space, and $X^{(n)} \xrightarrow{v} X$ a.s. So,

$$\big| X^{(n)}(I_n) - X(I) \big| \leq \big| X^{(n)}(I) - X(I) \big| + \big| X^{(n)}(I_n) - X^{(n)}(I) \big|. \tag{3.3}$$

Since the intensity measure of X is assumed diffuse, $X(\partial I) = 0$ a.s, and the first term on the right-hand side of (3.3) a.s. vanishes as $n \to \infty$ due to the vague convergence. To deal with the second term, let us fix $N \in \mathbb{N}$. For all $n \geq N$,

$$\big| X^{(n)}(I_n) - X^{(n)}(I) \big| = X^{(n)}(I_n \setminus I) \leq X^{(n)}(I_N \setminus I),$$

and the right-hand side converges a.s. toward $X(I_N \setminus I)$. In other words,

$$\limsup_{n \to \infty} \big| X^{(n)}(I_n) - X^{(n)}(I) \big| \leq X(I_N \setminus I) \quad \text{a.s.}$$

Letting $N \to \infty$ proves that the second term on the right-hand side of (3.3) a.s. vanishes as $n \to \infty$ too. Hence, $X^{(n)}(I_n) \to X(I)$ a.s., and so $X^{(n)}(I_n) \xrightarrow{d} X(I)$, which clearly implies the claim. $\qquad\square$

We may now proceed to the final part of the proof.

Proof of Theorem 3.1. In (3.2), let us take $\varepsilon = n^{-\frac{1}{5}}$. Then,

$$\mathbb{P}\big\{ \xi_r^{(n)}[a,b] \neq \eta_r^{(n)}[a,b] \big\}$$
$$\leq c_r n^{-\frac{1}{5}} + \mathbb{P}\big\{ \eta_r^{(n)} \big[a - n^{-\frac{1}{5}}, a + n^{-\frac{1}{5}}\big] \geq 1 \big\} + \mathbb{P}\big\{ \eta_r^{(n)} \big[b - n^{-\frac{1}{5}}, b + n^{-\frac{1}{5}}\big] \geq 1 \big\}.$$

Lemmas 3.3 and 3.5 imply that, as $n \to \infty$,

$$\mathbb{P}\{\eta_r^{(n)}\big[a - n^{-\frac{1}{5}}, a + n^{-\frac{1}{5}}\big] \geq 1\} \to \mathbb{P}\{\xi_r\{a\} \geq 1\} = 0,$$

and the same holds for b. Together with the foregoing inequality, we have

$$\mathbb{P}\{\xi_r^{(n)}[a,b] \neq \eta_r^{(n)}[a,b]\} \to 0, \quad n \to \infty. \tag{3.4}$$

Let $\mathcal{U}$ stand for the ring of finite unions of bounded closed segments in $\mathbb{R}$. For each $U = [a_1, b_1] \cup \ldots \cup [a_l, b_l] \in \mathcal{U}$ and $k \in \mathbb{N}$, (3.4) implies

$$\big|\mathbb{P}\{\xi_r^{(n)}(U) = k\} - \mathbb{P}\{\eta_r^{(n)}(U) = k\}\big| \leq \mathbb{P}\{\xi_r^{(n)}(U) \neq \eta_r^{(n)}(U)\}$$

$$\leq \sum_{i=1}^{l} \mathbb{P}\{\xi_r^{(n)}[a_i, b_i] \neq \eta_r^{(n)}[a_i, b_i]\} \to 0, \quad n \to \infty.$$

Moreover, by Lemma 3.3,

$$\lim_{n \to \infty} \mathbb{P}\{\eta_r^{(n)}(U) = k\} = \mathbb{P}\{\xi_r(U) = k\}.$$

So, the last two formulas imply

$$\lim_{n \to \infty} \mathbb{P}\{\xi_r^{(n)}(U) = k\} = \mathbb{P}\{\xi_r(U) = k\}.$$

Notice that the process ξ_r is simple, since it is a Poisson process with diffuse intensity measure (see, e.g., Proposition 6.9 in [16]). Hence, $\xi_r^{(n)} \xrightarrow{vd} \xi_r$ on $\mathbb{R}$ by Theorem 4.15 in [14]. $\qquad \square$

In the sequel, we will need the following remark.

Remark 3.6. Theorem 3.1 remains true if we consider $\xi_r^{(n)}$ and ξ_r as point processes on the semi-compactified real axis $\mathbb{R} \cup \{+\infty\}$, endowed with some relevant metric, say, $d(x, y) = |\mathrm{e}^{-y} - \mathrm{e}^{-x}|$. The proof follows along the same lines.

4 Implications of the main result

From Theorem 3.1, we may easily deduce a number of known limit results which were often originally proved by direct complicated calculations. Moreover, this approach allows to obtain some far-reaching generalizations and infinite-dimensional extensions of those results. Finally, an application of Theorem 3.1 often makes it possible to clarify some related surprising phenomena. As an example, consider $T_{r,m}^{(n)}$, $m \leq n - 1$, the first time when some $n - m$ (unspecified) of the n coupon types have already arrived at least r times each, and put $T_{r,m}^{(n)} = 0$ for $m \geq n$. Various limit theorems for the case $r = 1$ were studied in [3], [24], [11], see also §1.2 in [15]. In particular, Theorem 4 in [3] asserts that

$$\ln 2n - \frac{T_{1,m}^{(n)}}{n} \xrightarrow{d} \ln Q, \quad n \to \infty, \tag{4.1}$$

where Q follows χ_{2m+2}^2, the χ^2-distribution with $2m + 2$ degrees of freedom. Theorem 4.1 below gives both the generalization for any $r \in \mathbb{N}$ and the infinite-dimensional extension in the sense of distributional convergence in $\mathbb{R}^\infty$, and also clarifies why a χ^2-distribution appears.

Consider the random elements $V_r^{(n)}$ and V_r in $\mathbb{R}^\infty$ given by

$$V_r^{(n)} = \big(\psi_r^{(n)}\big(T_{r,m}^{(n)}\big), \, m \in \mathbb{N} \cup \{0\}\big),$$

$$V_r = \big(-\ln(r-1)! - \ln \sum_{j=1}^{m+1} E_j, \, m \in \mathbb{N} \cup \{0\}\big),$$

where $\psi_r^{(n)}$ is defined by (2.3), and E_j, $j \in \mathbb{N}$, are i.i.d. Exp(1).

Theorem 4.1. *We have* $V_r^{(n)} \xrightarrow{d} V_r$ *in* $\mathbb{R}^\infty$ *as* $n \to \infty$.

Before proving the theorem, we make a couple of important remarks.

Remark 4.2. Restricting attention only to one-dimensional projections, we obtain

$$\frac{T_{r,m}^{(n)}}{n} - \ln n - (r-1) \ln \ln n \xrightarrow{d} -\ln(r-1)! - \ln S_{m+1}, \quad n \to \infty, \tag{4.2}$$

where $S_{m+1} = \sum_{j=1}^{m+1} E_j \sim \Gamma(m+1, 1)$. Confining ourselves only to the case $r = 1$, we get

$$\ln 2n - \frac{T_{1,m}^{(n)}}{n} \xrightarrow{d} \ln(2S_{m+1}), \quad n \to \infty.$$

Finally, noting that $2S_{m+1} \sim \chi_{2m+2}^2$ leads to (4.1).

Remark 4.3. Now allowing in (4.2) for arbitrary $r \in \mathbb{N}$ but setting $m = 0$, we can easily deduce (1.2), the limit theorem by Erdős and Rényi.

We now turn to the proof.

Proof of Theorem 4.1. We will need some additional notation. Fix $m \in \mathbb{N} \cup \{0\}$, and let

$$L^m \colon M_p(\mathbb{R} \cup \{+\infty\}) \to (\mathbb{R} \cup \{-\infty, +\infty\})^{m+1}$$

map $\sum \delta_{x_i}$ into the vector of its "last-but-j" th points (possibly infinite), $0 \le j \le m$. In other words, $L^m(\mu) = (L_j(\mu), 0 \le j \le m)$, where

$$L_j(\mu) = \inf\{x \in \mathbb{R} \colon \mu(x, +\infty] = j\}.$$

Further, we denote by $\mu|_K$ the point measure μ restricted to the compact set $K \subset \mathbb{R} \cup \{+\infty\}$.

By Theorem 3.1 in the form of Remark 3.6, using the Skorokhod coupling, we may consider $(\xi_r^{(n)}, n \in \mathbb{N})$ and ξ_r on a common probability space and assume that $\xi_r^{(n)} \xrightarrow{v} \xi_r$ on $\mathbb{R} \cup \{+\infty\}$ a.s. Hence, by Proposition 3.13 in [21], $L^m(\xi_r^{(n)}|_{[a,+\infty]}) \to L^m(\xi_r|_{[a,+\infty]})$ a.s. for any $a \in \mathbb{R}$. This implies $L^m(\xi_r^{(n)}) \to L^m(\xi_r)$ a.s., and so

$$L^m(\xi_r^{(n)}) \xrightarrow{d} L^m(\xi_r). \tag{4.3}$$

Note that, by (2.4),

$$L^m(\xi_r^{(n)}) = (L_j(\xi_r^{(n)}), 0 \le j \le m) = (\psi_r^{(n)}(T_{r,j}^{(n)}), 0 \le j \le m),$$

and is thus just the projection of $V_r^{(n)}$ onto the first $m+1$ coordinates (from 0th to m^{th}). On the other hand, by Remark 3.2 and the i.i.d. property of inter-arrival times for ξ, we come to

$$L^m(\xi_r) = (L_j(\xi_r), 0 \le j \le m) = \left(-\ln(r-1)! - \ln \sum_{k=1}^{j+1} E_j, 0 \le j \le m\right),$$

which is, similarly, the projection of V_r onto the first $m+1$ coordinates. Hence, (4.3) proves the finite-dimensional convergence $V_r^{(n)} \xrightarrow{fd} V_r$ in $\mathbb{R}^\infty$ as $n \to \infty$. To complete the proof, it only remains to recall that in $\mathbb{R}^\infty$ the notions of finite-dimensional convergence and convergence in distribution are equivalent (see, e.g., [22], pp. 53–54). $\square$

We now consider another application of Theorem 3.1, namely, to "rare" coupon types. Recall that an arriving coupon is assumed to belong to any of the n types with the same probability $\frac{1}{n}$. But, due to random factors, some coupon types will need a long time until they arrive for the rth time. Taking (1.2) into account, we will call a type $i \in \mathbb{N}_n$ x-rare, $x \in \mathbb{R}$, if

$$Y_{i,r}^{(n)} \geq nx + n \ln n + (r-1)n \ln \ln n. \tag{4.4}$$

Denote by $C_r^{(n)}(x)$ the number of x-rare types:

$$C_r^{(n)}(x) = \sum_{i=1}^{n} \mathbb{1}\big\{Y_{i,r}^{(n)} \geq nx + n \ln n + (r-1)n \ln \ln n\big\}.$$

Below we state and prove a functional limit theorem for $C_r^{(n)} = \big(C_r^{(n)}(x), x \in \mathbb{R}\big)$ in the Skorokhod J_1-topology.

Let $N = \big(N(t), t \geq 0\big)$ be a homogeneous unit-rate Poisson process, considered not as a random point measure, but classically, as a Lévy process with Poisson increments. Actually, $N(t) = \xi(0,t]$, where ξ is introduced in Remark 3.2. Also, let $N_r(x) = N\big(\frac{e^{-x}}{(r-1)!}\big)$, $x \in \mathbb{R}$.

Theorem 4.4. *We have $C_r^{(n)} \xrightarrow{d} N_r$ in $D(\mathbb{R})$, endowed with the J_1-topology, as $n \to \infty$.*

Remark 4.5. Strictly speaking, the processes $C_r^{(n)}$ and N_r are càglàd, not càdlàg, and we should have considered $D_{\text{left}}(\mathbb{R})$, the space of left-continuous functions with finite right limits, instead of $D(\mathbb{R})$. But as the J_1-topology may be introduced on $D_{\text{left}}(\mathbb{R})$ in the same way as on $D(\mathbb{R})$, we will close our eyes to these differences (cf. Remark 3.2 on p. 58 in [22]).

Remark 4.6. For fixed x, the distributional convergence of $C_r^{(n)}(x)$ to $N_r(x)$, as well as its rate in terms of total variation distance, were known before as a result of Stein-Chen method for Poisson approximation (see, e.g., Chapter 6 in [2] and Example 4.34 in [23]). For related finite-dimensional results see also Theorem 1 on p. 172 in [15] and Theorem 2 in [4].

Proof of Theorem 4.4. Let $h^{\leftarrow}(x) = \frac{e^{-x}}{(r-1)!}$, $x \in \mathbb{R}$, be the inverse function to h given by (3.1). Theorem 3.1 and Remark 3.2 imply that, by the continuous mapping theorem, $\xi_r^{(n)} \circ (h^{\leftarrow})^{-1} \xrightarrow{vd} \xi$. (The intuitively obvious continuity of $T: M_p(\mathbb{R}) \to M_p\big((0, +\infty)\big)$ given by $T(\mu) = \mu \circ (h^{\leftarrow})^{-1}$ follows from Proposition 3.18 in [21].) According to Theorem 4.20 in [14], the distributional convergence of random point measures on $(0, +\infty)$ in the vague topology is equivalent to that of the associated cumulative processes in the J_1-topology. So, $\xi_r^{(n)}\big((h^{\leftarrow})^{-1}(0, \cdot]\big) \to N(\cdot)$ in the latter sense, and thus, by transfer, $\xi_r^{(n)}[\cdot, +\infty) \to N_r(\cdot)$. It remains to note that, by (2.4), (2.3), and (4.4), $\xi_r^{(n)}[x, +\infty) = C_r^{(n)}(x)$ for any $x \in \mathbb{R}$. $\qquad\square$

References

[1] E. Anceaume, Y. Busnel, and B. Sericola. New results on a generalized coupon collector problem using Markov chains. *J. Appl. Probab.*, 52(2):405–418, 2015. MR-3372083

[2] A. D. Barbour, L. Holst, and S. Janson. *Poisson approximation.* Oxford: Clarendon Press, 1992. MR-1163825

[3] L. E. Baum and P. Billingsley. Asymptotic distributions for the coupon collector's problem. *Ann. Math. Statist.*, 36(6):1835–1839, 1965. MR-0182039

[4] Yu. V. Bolotnikov. The convergence of the variables $\mu_r(n)$ to Gaussian and Poisson processes in the classical ball problem. *Theory Probab. Appl.*, 13:39–51, 1968. MR-0226705

[5] A. V. Doumas and V. G. Papanicolaou. Asymptotics of the rising moments for the coupon collector's problem. *Electron. J. Probab.*, 18:15, 2013. MR-3040551

[6] A. V. Doumas and V. G. Papanicolaou. Some new aspects for the random coupon collector's problem. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 11(1):197–208, 2014. MR-3225973

[7] P. Erdős and A. Rényi. On a classical problem of probability theory. *Publ. Math. Inst. Hung. Acad. Sci., Ser. A*, 6:215–220, 1961. MR-0150807

[8] M. Ferrante and M. Saltalamacchia. The coupon collector's problem. *Materials matemàtics*, pages 01–35, 2014.

[9] J. C. Fu and W.-C. Lee. On coupon collector's and Dixie cup problems under fixed and random sample size sampling schemes. *Ann. Inst. Stat. Math.*, 69(5):1129–1139, 2017. MR-3689165

[10] L. Glavaš and P. Mladenović. New limit results related to the coupon collector's problem. *Stud. Sci. Math. Hung.*, 55(1):115–140, 2018. MR-3803572

[11] L. Holst. On sequential occupancy problems. *J. Appl. Probab.*, 18:435–442, 1981. MR-0611787

[12] L. Holst. On birthday, collectors', occupancy and other classical urn problems. *Int. Stat. Rev.*, 54:15–27, 1986. MR-0959649

[13] L. Holst. Extreme value distributions for random coupon collector and birthday problems. *Extremes*, 4(2):129–145, 2001. MR-1893869

[14] O. Kallenberg. *Random measures, theory and applications.* Cham: Springer, 2017. MR-3642325

[15] V. F. Kolchin, B. A. Sevast'yanov, and V. P. Chistyakov. *Random allocations.* New York etc.: John Wiley & Sons, 1978. MR-0471016

[16] G. Last and M. Penrose. *Lectures on the Poisson process.* Cambridge: Cambridge University Press, 2017. MR-3791470

[17] M. Mayer and I. Molchanov. Limit theorems for the diameter of a random sample in the unit ball. *Extremes*, 10(3):129–150, 2007. MR-2394205

[18] P. Mladenović. Limit distributions for the problem of collecting pairs. *Bernoulli*, 14(2):419–439, 2008. MR-2544095

[19] P. Neal. The generalised coupon collector problem. *J. Appl. Probab.*, 45(3):621–629, 2008. MR-2455173

[20] M. Penrose. *Random geometric graphs.* Oxford: Oxford University Press, 2003. MR-1986198

[21] S. I. Resnick. *Extreme values, regular variation, and point processes.* New York etc.: Springer-Verlag, 1987. MR-0900810

[22] S. I. Resnick. *Heavy-tail phenomena. Probabilistic and statistical modeling.* New York, NY: Springer, 2007. MR-2271424

[23] N. Ross. Fundamentals of Stein's method. *Probab. Surv.*, 8:210–293, 2011. MR-2861132

[24] E. Samuel-Cahn. Asymptotic distributions for occupancy and waiting time problems with positive probability of falling through the cells. *Ann. Probab.*, 2(3):515–521, 1974. MR-0365668

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: http://www.imstat.org/publications/open.htm