MAXIMAL COUNTS IN THE STOPPED OCCUPANCY PROBLEM

ALEXANDER GNEDIN, SVANTE JANSON AND YAAKOV MALINOVSKY

Abstract. We revisit a version of the classic occupancy scheme, where balls are thrown until almost all boxes receive a given number of balls. Special cases are widely known as coupon-collectors and dixie cup problems. We show that as the number of boxes tends to infinity, the distribution of the maximal occupancy count does not converge, but can be approximated by a convolution of two Gumbel distributions, with the approximating distribution having oscillations close to periodic on a logarithmic scale. We pursue two approaches: one relies on lattice point processes obtained by poissonisation of the number of balls and boxes, and the other employs interpolation of the multiset of occupancy counts to a point process on reals. This way we gain considerable insight in known asymptotics obtained previously by mostly analytic tools. Further results concern the moments of maximal occupancy counts and ties for the maximum.

1. Introduction

In the classic sequential occupancy scheme balls are thrown independently in n boxes, with each ball landing with equal probability $1/n$ in each box. The allied waiting time problems concern the distribution of the random number of trials required to satisfy specified occupancy conditions for boxes [28]. Technically, the random variables in focus here are stopping times that terminate the allocation process by way of an adapted nonanticipating rule. From the early days of probability theory such questions attracted much attention and were studied under different guises. For instance, a problem treated in de Moivre's seminal treatise [11] asks one 'to find in how many trials [a gambler] may with equal chance [i.e., probability about 1/2] undertake with a pair of common dice to throw all the doublets'. A better known modern textbook example is the coupon collector's problem (CCP), which in terms of the occupancy scheme deals with the number of balls thrown until no empty boxes are left. The model where the balls are thrown until every box contains more than m balls is sometimes called the dixie cup problem [18].

In the present paper we are not primarily interested in the stopping time itself, i.e., the required number of balls, but rather in the configuration when the stopping occurs. More precisely, we study the largest number of balls in a box at that time, and also the second largest, and so on. Ivchenko in a series of papers [20; 21; 22] studied a wide range of statistics of the sequential occupancy scheme terminated by a stopping time; a summary of some of his work is found in [23]. The present note is inspired by his result on the maximum box occupancy count M_n observed when the stopping occurs as in the CCP or according to a more general criterion (this appeared as a special case of [21, Theorem 9] and [22, Theorem 2]). The distribution of M_n does not converge as $n \to \infty$ even after suitable normalisation; instead the approximating asymptotic distribution oscillates on a logarithmic scale. This is a common and well understood phenomenon when considering asymptotics of discrete random variables that do not require scaling due to bounded variances. The asymptotic result can formally be stated either as convergence in distribution of suitable subsequences, or (equivalently) as approximation in total variation distance with some family of random variables. A simple but typical example

Date: 25 June 2025.

of this kind is the distribution of the maximum of n i.i.d. geometric random variables (see for instance [26, Example 4.3], also see Lemma 5.5 in the sequel for a general framework).

In this paper we employ the familiar device of embedding the occupancy scheme in a Poisson process, to link directly properties of M_n with the extreme value theory [14; 43]. This way we identify the asymptotic distribution with a discretised convolution of two Gumbel distributions (Theorem 4.4). The benefits of the poissonisation of the number of balls in the occupancy problems have been long known [15; 31; 36]. A novel element in our proof is the use of 'bi-poissonisation', which amounts to also replacing fixed number of boxes n by random, thus achieving independence of multiplicities of occupancy counts for each given time. Moreover, the time evolution of the occupancy model as a whole becomes similar to the processes familiar from studies in population dynamics and queueing theory, which sheds a new light on the processes of small counts (Section 7). A competing approach we pursue here (Section 6) relies on interpolating the multiset of integer occupancy counts to a point process on $\mathbb{R}$, then showing that it can be approximated by a Poisson process with a suitable (exponential) rate, with all points rounded to integers. Extending the result about M_n we show that a similar approximation holds for any given number of maximal order statistics of the box occupancy counts (Theorem 5.6).

We also discuss (Section 9) the asymptotics of moments for M_n and other stopped maximal order statistics of occupancy counts. Complementing the previous studies [19; 21; 22; 36; 45], we obtain these indirectly by means of new exponential tail estimates, which might be of independent interest for other contexts related to the extreme value theory.

Sampling from a discrete distribution outputs repeated values. Criteria are known to ensure that ties for the maximum do not vanish asymptotically [5]. In particular, for the geometric source distribution the probability of such a tie is known to undergo tiny fluctuations on the logarithmic scale [9; 35]. In this direction, we examine (Section 10) the multiplicity of M_n, and show that the fluctuations turn even smaller due to a smoothing effect caused by random stopping.

Following [21; 22], we consider the occupancy process stopped when there remain only ℓ boxes that contain at most m balls each. Thus the case $m = 0$ and $\ell = 0$ corresponds to the CCP. The integer parameters $m \geqslant 0$ and $\ell \geqslant 0$ will be fixed throughout the paper; many variables below depend on them, but this will not be shown in the notation. (The parameters ℓ in [21] and m in [22] can be allowed to vary with n, but we will only focus on the case of fixed values where the results turn out to be most interesting.)

1.1. **Notation.** For common probability distributions we use the self-explaining notation, for instance Poisson(t), Geometric(α). For identity and convergence in distribution we write $\overset{\mathrm{d}}{=}$ and $\overset{\mathrm{d}}{\to}$, respectively. The total variation distance between the distributions of random elements X and Y is denoted $d_{\mathrm{TV}}(X, Y)$. The symbols $\lfloor\ \rfloor, \lceil\ \rceil$ and $\{\ \}$ denote, respectively, the floor, the ceiling and the fractional part functions. For shorthand we set $L := \log n$ and $\log L := \log\log n$. Unspecified limits and asymptotic relations such as $o(1)$, $f_n \sim g_n$ (meaning $f_n/g_n \to 1$, or equivalently $f_n = g_n(1 + o(1))$) and $f_n \asymp g_n$ (meaning that $f_n = O(g_n)$ and $g_n = O(f_n)$) all presume $n \to \infty$. We say that some limit holds uniformly in $x = o(f_n)$ if it holds uniformly for all x with $|x| \leqslant g_n$ for any function $g_n = o(f_n)$; the uniformity in $x = O(f_n)$ is understood similarly. $O_{\mathrm{p}}(1)$ means bounded in probability (i.e., tight), and $o_{\mathrm{p}}(1)$ means convergence to 0 in probability. The term 'with high probability' (w.h.p.) will mean that a certain event has probability converging to one as $n \to \infty$, and 'almost surely' (a.s.) will mean an event of probability one.

2. The poissonised occupancy scheme

We will be dealing with a standard continuous-time version of the discrete sequential occupancy scheme. Let $(\Pi_i(t), t \geqslant 0), i \in \mathbb{Z}_{>0}$, be i.i.d. replicas of a Poisson counting process $(\Pi(t), t \geqslant 0)$ with unit rate. We interpret $\Pi_i(t)$ as the number of balls allocated into box i by time t; this box occupancy count appears in the literature under different names such as score or load, to mention a few.

The problem with n boxes concerns the occupancy counts $\Pi_i(t)$ for $i \in [n]$, where $[n] := \{1, \dots, n\}$. The aggregate arrival process to the batch of n boxes is Poisson with rate n. Given such an arrival occurring at time t, the ball is allocated into each of the n boxes with the same probability $1/n$, independently of the past allocations. Thus the random sequence of states of the set of n boxes follows the same dynamics as in the classic discrete-time occupancy scheme studied in detail in [36]. But keep in mind that the time scale differs from the discrete case in that the mean number of balls dropped in t units of continuous time is nt.

The advantage of the poissonised model is the exact independence among the boxes, whereas in the discrete-time occupancy scheme the independence only holds asymptotically for large number of balls [31; 36]. Moreover, the adopted setting with infinitely many Π_i's allows one to consistently define the occupancy schemes for all n on the same probability space.

For point and cumulative probabilities of the Poisson distribution we use the notation

$$p_r(t) = e^{-t}\frac{t^r}{r!}, \ P_r(t) = \sum_{i=0}^{r} p_i(t), \ \overline{P}_r(t) = 1 - P_r(t), \tag{2.1}$$

where $r \in \mathbb{Z}_{\geqslant 0}, t \geqslant 0$. These are further related via the standard formulas

$$\overline{P}_r(t) = p_r(t) \sum_{i=1}^{\infty} \frac{t^i}{(r+1)\cdots(r+i)}, \tag{2.2}$$

$$P_r(t) = \int_t^{\infty} p_r(s)\mathrm{d}s, \tag{2.3}$$

where (2.3) connects Gamma and Poisson distributions. If $t, r \to \infty$ so that $\limsup t/r < 1$, then from (2.2) follows the asymptotics

$$\overline{P}_r(t) \sim \frac{t}{r-t}\, p_r(t). \tag{2.4}$$

Let, as in the Introduction, $m \geqslant 0$ and $\ell \geqslant 0$ be fixed integers. Define τ_n to be the first time when there remain only ℓ out of n boxes that contain at most m balls each, that is

$$\tau_n := \min\{t \geqslant 0 : \Pi_i(t) > m \text{ for all but } \ell \text{ indices } i \in [n]\} \tag{2.5}$$

(we assume $n > \ell$ to ensure $0 < \tau_n < \infty$ a.s.). In particular, for $m = 0, \ell = 0$ this is the time when each of the n boxes becomes occupied by at least one ball, which is the termination condition in the CCP. The case $\ell = 0, m \geqslant 1$ aligns with the dixie cup problem.

Since the Poisson distribution is discrete, repetitions among the occupancy counts $\Pi_i(t)$ occur with positive probability. Let $M_{n,1}(t) \geqslant \cdots \geqslant M_{n,n}(t)$ be the nonincreasing sequence of order statistics of $\Pi_1(t), \dots, \Pi_n(t)$. The order statistics capture the allocation of some random number of indistinguishable balls into n indistinguishable boxes. From a combinatorial viewpoint, this can be regarded as a generalised partition of an integer into n parts which can be zero. The same data can be equivalently encoded in the sequence of *multiplicities*

$$\mu_{n,r}(t) := \#\{i \in [n] : M_{n,i}(t) = r\}, \ r \in \mathbb{Z}_{\geqslant 0}, \tag{2.6}$$

that count repetitions among the box occupancy counts. In particular, $\mu_{n,0}(t)$ is the number of empty boxes. By independence among the boxes

$$\mu_{n,r}(t) \stackrel{\mathrm{d}}{=} \mathrm{Binomial}(n, p_r(t)), \qquad r \in \mathbb{Z}_{\geqslant 0}, \tag{2.7}$$

and the joint distribution of the multiplicities is multinomial with infinitely many classes. The study of multiplicities for large number of balls and boxes is a central theme in the occupancy problems [36].

The equivalence of the two descriptions of the allocation of balls is established by the relations

$$M_{n,i}(t) \leqslant k \Longleftrightarrow \sum_{j=1}^{\infty} \mu_{n,k+j}(t) \leqslant i-1 \tag{2.8}$$

for $i \in [n], k \in \mathbb{Z}_{\geqslant 0}$. Specifically, for the largest box occupancy count we have

$$M_{n,1}(t) = \max\{r : \mu_{n,r}(t) > 0\}. \tag{2.9}$$

With these notations, the number of balls thrown by time t has the threefold representation

$$\sum_{i=1}^{n} \Pi_i(t) = \sum_{i=1}^{n} M_{n,i}(t) = \sum_{r=0}^{\infty} r\mu_{n,r}(t), \tag{2.10}$$

Note also that the random variables of our primary interest are the largest few occupancy numbers when we stop (which are the same in the discrete-time and continuous-time models), that is $M_{n,i}(\tau_n)$'s with i less than some bound not depending on n. We therefore use the shorthand notation

$$M_{n,i} := M_{n,i}(\tau_n). \tag{2.11}$$

We also sometimes use the even shorter notation

$$M_n := M_{n,1} = M_{n,1}(\tau_n) \tag{2.12}$$

for the maximum box occupancy count when the allocation is stopped.

3. Preliminaries on point processes

Throughout we will be exploiting basic facts on Poisson and related point processes as found in many excellent texts [30; 33; 37; 43]. In this section we remind the bare minimum, also using this opportunity to introduce constructions needed in later sections.

A point process H on $\mathbb{R}$ (or a more general Polish space) is a random locally finite Borel measure with values in $\mathbb{Z}_{\geqslant 0}$. Such point process can be represented as a (finite or infinite) sum

$$\mathrm{H} = \sum_i \delta_{\eta_i} \tag{3.1}$$

of Dirac masses at random points η_i. We will sometimes with a minor abuse of notation write this as $\mathrm{H} = \{\eta_i\}$, thus identifying the point process with the multiset $\{\eta_i\}$ of its points, that is its atoms endowed with some multiplicities. (We prefer the measure-theoretic term 'atom' to make difference with nonrandom points of the background space.) The intensity measure of H is the function that assigns $\mathbb{E}[\mathrm{H}(A)]$ to each Borel set A. We call H simple if the multiplicity of each atom is one almost surely.

Recall that a Poisson point process is a point process H where $\mathrm{H}(A)$ (the number of points in A) has a Poisson distribution for every Borel set A, and these numbers for different, disjoint sets are independent.

3.1. **Transformations of Poisson processes.** The unit rate Poisson process Π on $\mathbb{R}_+$ is a simple point process whose atoms are representable as sums $\eta_i = E_1 + \cdots + E_i$, where the terms are independent with standard exponential distribution, so the ith atom has Gamma$(i,1)$ (aka Erlang) distribution satisfying

$$\mathbb{P}[\eta_i \leqslant t] = \overline{P}_{i-1}(t), \qquad t \geqslant 0. \tag{3.2}$$

The intensity measure of Π is the Lebesgue measure on the halfline.

Inhomogeneous Poisson point processes are uniquely characterised by their intensity measures. Such processes on $\mathbb{R}$ can be constructed from Π by the measure-theoretic pushforward $\Pi \circ f^{-1}$, which is implemented through transporting the atoms with function f; thus the multiset $\{\eta_i\}$ is mapped to $\{f(\eta_i)\}$. The intensity measure of $\Pi \circ f^{-1}$ is the pushforward of the Lebesgue measure by f.

An important role in the extreme-value theory is played by the *exponential* Poisson process Ξ on $\mathbb{R}$, obtained as pushforward of Π by $f(t) = -\log t$, thus with the intensity measure

$$\mathbb{E}\,[\Xi(\mathrm{d}x)] = e^{-x}\mathrm{d}x, \ x \in \mathbb{R}. \tag{3.3}$$

(The name is not common but has been used in the literature, see [7, Section 6.2.2].) The atoms of Ξ comprise a decreasing to $-\infty$ sequence of random variables $\xi_i := -\log \eta_i$, with distribution, by (3.2),

$$\mathbb{P}[\xi_i \leqslant x] = P_{i-1}(e^{-x}), \ x \in \mathbb{R}, \tag{3.4}$$

and thus with density, see (2.3),

$$e^{-x} p_{i-1}(e^{-x}), \ x \in \mathbb{R}. \tag{3.5}$$

In particular, the largest atom ξ_1 of Ξ has the (standard) Gumbel distribution

$$\mathbb{P}[\xi_1 \leqslant x] = e^{-e^{-x}}, \ x \in \mathbb{R}. \tag{3.6}$$

The best known instance $i = 1$, as well as the distributions (3.4) with $i > 1$, were introduced in [17], which justifies the notation Gumbel(i) for the distribution (3.4) of ξ_i. In this nomenclature the (standard) Gumbel distribution becomes Gumbel(1). For each i there is an associated scale-location family of distributions.

We define $\Xi_b := \Xi + b, \ b \in \mathbb{R}$, to be the translation of Ξ with atoms $\xi_i + b$; these have shifted Gumbel(i) distributions

$$\mathbb{P}[\xi_i + b \leqslant x] = P_{i-1}(e^{-x+b}), \ x \in \mathbb{R}. \tag{3.7}$$

Thus, $\Xi_b = \{\xi_i + b\}$ is a Poisson process with intensity measure $e^{b-x}\mathrm{d}x$, $x \in \mathbb{R}$.

3.2. **Lattice point processes.** For a point process with nonatomic intensity measure the probability of an atom occurring at fixed location is zero. In contrast to that, a point process on the integer lattice $\mathbb{Z}$ is in essence a two-sided random sequence of multiplicities at integer locations. In particular, a lattice Poisson point process is identifiable with an array of independent Poisson random variables with given parameters.

We denote by $\Xi_b^{\uparrow}$ the lattice counterpart of Ξ_b obtained by applying the ceiling function:

$$\Xi_b^{\uparrow} := \{\lceil \xi_i + b \rceil\}. \tag{3.8}$$

We choose rounding up (rather than down) to have the intensity measures agreeing on semi-closed intervals $(-\infty, r]$, $r \in \mathbb{Z}$, thus forcing the distribution functions of the respective atoms to coincide at integers:

$$\mathbb{P}[\lceil \xi_i + b \rceil \leqslant r] = \mathbb{P}[\xi_i + b \leqslant r] = P_{i-1}(e^{-r+b}), \ r \in \mathbb{Z}. \tag{3.9}$$

In the case $i = 1, b = 0$ we obtain a natural discrete analogue of the Gumbel distribution. The intensity measure of $\Xi_b^\uparrow$ is supported by $\mathbb{Z}$, with masses comprising a two-sided geometric sequence

$$\mathbb{E}[\Xi_b^\uparrow(\{r\})] = \mathbb{E}\left[\Xi_b(r-1, r]\right] = \int_{r-1}^{r} e^{b-x}\,\mathrm{d}x = e^{-r+b}(e-1), \tag{3.10}$$

hence with the right tail

$$\mathbb{E}[\Xi_b^\uparrow(r,\infty]] = \mathbb{E}[\Xi_b^\uparrow[r+1,\infty]] = \mathbb{E}\left[\Xi_b(r,\infty]\right] = e^{-r+b}. \tag{3.11}$$

3.3. **The marking theorem and the Poisson shift.** In plain terms, a basic version of the marking theorem says that if η_i's are random atoms of some Poisson point process H and random 'marks' ζ_i's are i.i.d. independent of H, then the pairs (η_i, ζ_i) define a bivariate Poisson process in a product space. This proves very useful to construct other Poisson processes as transforms $\{f(\eta_i, \zeta_i)\}$. In particular, if H is on $\mathbb{R}$ and ζ_i's are real-valued, then the pairwise sums $\eta_i + \zeta_i$ are atoms defining another Poisson process, whose intensity measure is the convolution of the intensity measure of H and the distribution of ζ_1.

For a general point process H on $\mathbb{R}$ and $h \geqslant 0$ we define its *Poisson shift* $T_h \circ \mathrm{H}$ as the above operation with atom-wise adding of independent $\zeta_i \stackrel{\mathrm{d}}{=} \mathrm{Poisson}(h)$. Note that if H has k atoms at the same location x, then each of them contributes to $T_h \circ \mathrm{H}$ a unit mass at x shifted by an independent Poisson variable.

Notably, on the exponential Poisson process Ξ the Poisson shift acts in distribution like a deterministic translation

$$T_h \circ \Xi \stackrel{\mathrm{d}}{=} \Xi_{(e-1)h}. \tag{3.12}$$

The proof follows by a simple calculation found in [7, p. 153] and the formula $\mathbb{E}[e^{\zeta_1}] = e^{h(e-1)}$. Applying T_h to the lattice process $\Xi^\uparrow$ yields a distributional copy of $\Xi^\uparrow_{(e-1)h}$, as is clear from (3.12) since ζ_i's are integer-valued.

The *Poisson flow* with initial state H is the Markov measure-valued process $(T_h \circ \mathrm{H},\ h \geqslant 0)$, where each atom independently of the others undergoes unit jumps to the right at the unit rate. If H is a Poisson point process then so is also every $T_h \circ \mathrm{H}$. See the recent monograph [13] for the general theory of measure-valued processes.

3.4. **Mixed binomial point processes.** A Poisson point process with finite intensity measure has a random sum representation

$$\mathrm{H} = \sum_{i=1}^{N} \delta_{\eta_i}, \tag{3.13}$$

where the random variables $\eta_1, \eta_2, \ldots$ are i.i.d., and N is an independent from η_i's Poisson random variable whose parameter is equal to the total mass of the intensity measure.

The latter form is an instance of the more general *mixed binomial* point process (MBPP) [30], where N is allowed to have arbitrary distribution on $\mathbb{Z}_{\geqslant 0}$. Conditionally on $N = n$, such a MBPP is just a scatter of n i.i.d. random points. A subprocess obtained by restricting H to a Borel set A is again a MBPP

$$\mathrm{H}|_A = \sum_{i=1}^{N} \mathbf{1}\{\eta_i \in A\}\delta_{\eta_i} \stackrel{\mathrm{d}}{=} \sum_{j=1}^{\widehat{N}} \delta_{\widehat{\eta}_j}, \tag{3.14}$$

where $\widehat{N}$ has a mixed binomial distribution Binomial(N, α) with $\alpha = \mathbb{P}[\eta_1 \in A]$, and $\widehat{\eta}_j$'s are i.i.d. with distribution $\mathbb{P}[\widehat{\eta}_1 \in B] = \mathbb{P}[\eta_1 \in B \mid \eta_1 \in A]$. If N has a Poisson (respectively, binomial) distribution then also $\widehat{N}$ has a Poisson (respectively, binomial) distribution.

For another MBPP

$$\mathrm{H}' = \sum_{i=1}^{N'} \delta_{\eta_i}, \tag{3.15}$$

that only differs from H by the distribution of the total count, presuming N, N', η_i's defined on the same probability space, we will have $\{N = N'\} = \{\mathrm{H} = \mathrm{H}'\}$. Therefore, for the total variation distance we have the identity

$$d_{\mathrm{TV}}(\mathrm{H}, \mathrm{H}') = d_{\mathrm{TV}}(N, N'), \tag{3.16}$$

which follows from the definition of the distance as the infimum of the non-coincidence probability $\mathbb{P}[\mathrm{H} \neq \mathrm{H}']$ taken over all couplings.

3.5. **Weak convergence.** Weak convergence (convergence in distribution) of point processes is defined using the 'vague topology' in the space of locally finite measures; this means roughly convergence in the weak topology of restrictions on compact subsets. Note that convergence in the weak topology on $\mathbb{R}$ is defined only for a.s. finite point processes with an a.s. finite limit process, which is a situation not applicable here. Instead, the point processes treated in this paper are point processes H on $\mathbb{R}$ such that $\mathrm{H}(x, \infty) < \infty$ a.s. for every $x \in \mathbb{R}$; for such point processes we may use the representation $\mathrm{H} = \sum_{i=1}^{\infty} \delta_{\eta_i}$ with $\eta_1 \geqslant \eta_2 \geqslant \ldots$, setting formally $\eta_i = -\infty$ if the sequence of atoms is finite. Note that $\eta_i \to -\infty$ as $i \to \infty$ since point processes are assumed to be locally finite, that is finite on every compact set.

A technical note is that a point process on $\mathbb{R}$ such that $\mathrm{H}(x, \infty) < \infty$ a.s. for every $x \in \mathbb{R}$, also can be regarded as a point process on $(-\infty, \infty]$; we will not actually put any atoms at $+\infty$, but the inclusion of $+\infty$ in the background space increases the family of compact sets (for example, $[x, \infty]$ becomes compact), which makes the vague topology stronger. We will use this one-point closure and note that convergence of point processes in the vague topology on $(-\infty, \infty]$ is equivalent to the finite dimensional convergence of the ordered sequence of atoms, as shown by the following lemma (cf. [25, Lemma 4.4] with a trivial change of variables $(-\infty, \infty] \to (0, \infty]$).

Lemma 3.1. *Suppose that* H_n, $1 \leqslant n \leqslant \infty$, *are point processes on* $(-\infty, \infty]$, *and write* $\mathrm{H}_n = \{\eta_{n,i}\}_{i=1}^{N_n}$ *with* $\eta_{n,1} \geqslant \eta_{n,2} \geqslant \ldots$ *and* $0 \leqslant N_n \leqslant \infty$. *If some* $N_n < \infty$, *define further* $\eta_{n,i} = -\infty$ *for* $i > N_n$. *Then* $\mathrm{H}_n \xrightarrow{\mathrm{d}} \mathrm{H}_\infty$, *in the vague topology for measures on* $(-\infty, \infty]$, *if and only if* $(\eta_{n,1}, \eta_{n,2}, \ldots) \xrightarrow{\mathrm{d}} (\eta_{\infty,1}, \eta_{\infty,2}, \ldots)$ *in the standard sense that all finite dimensional distributions converge.* □

4. Stopped maximum

Recall the notation $M_n := M_{n,1}(\tau_n)$ for the maximum box occupancy count observed as the allocation is stopped at time τ_n. The distribution of M_n was studied by Ivchenko in the framework of the classic discrete-time occupancy scheme. We here study its distribution using the poissonised continuous-time scheme.

4.1. **Stopping time.** Let $m \geqslant 0$ and $\ell \geqslant 0$ be fixed, and consider the stopping time τ_n in (2.5), which can be alternatively defined through the multiplicities as

$$\tau_n = \min\{t : \mu_{n,0}(t) + \cdots + \mu_{n,m}(t) = \ell\}, \tag{4.1}$$

where we assume $n > \ell$.

First, the distribution of the stopping time follows readily by identifying τ_n with the $(\ell+1)$st last time when one of n boxes receives its $(m+1)$st ball. The distribution of the $(m+1)$st arrival time to a particular box is Gamma$(m+1,1)$, whence by independence among the boxes

$$\mathbb{P}[\tau_n \in \mathrm{d}t] = n\binom{n-1}{\ell}(P_m(t))^{\ell}(\overline{P}_m(t))^{n-\ell-1}p_m(t)\mathrm{d}t,\ t \geqslant 0. \tag{4.2}$$

In greater detail, the event defining τ_n occurs when one box receives its $(m+1)$st ball, ℓ boxes contain at most m balls each, and the remaining $n-\ell-1$ boxes contain at least $m+1$ balls each. By exchangeability of the boxes, the distribution of M_n conditional on τ_n will not change if we also condition on the indices of the boxes that satisfy the said constraints.

This implies the following fact noticed in [21, Lemma 1].

Fact 4.1. *Conditioned on $\tau_n = t$, the subsequence of $n-\ell-1$ stopped occupancy numbers of boxes that received at least $m+1$ balls (strictly) before time t is i.i.d. and independent of the complementing subsequence (also i.i.d.) of ℓ boxes with at most m balls. Moreover, both subsequences have truncated* Poisson(t) *distributions: the first one on $\{m+1, m+2, \ldots\}$ and the second on $\{0,\ldots,m\}$.*

For $t \to \infty$ the truncated Poisson distribution on $\{0,\ldots,m\}$ converges to the Dirac measure δ_m, because $p_{r+1}(t)/p_r(t) = t/(r+1) \to \infty$. Since $\tau_n \xrightarrow{\mathrm{p}} \infty$, w.h.p. the number of boxes with exactly m balls immediately before stopping is $\ell+1$ and no box contains lesser number of balls, whence by monotonicity in the stopping condition in (4.1)

$$\tau_n = \sup\{t : \mu_{n,m}(t) = \ell+1\} \qquad \text{w.h.p.} \tag{4.3}$$

(where $\sup\varnothing := \infty$), which gives yet another, asymptotic, interpretation of τ_n in terms of the sole multiplicity $\mu_{n,m}(t)$. The relation (4.3) holds a.s. for $m=0$.

We emphasise (4.3) to match the temporal domain we need here with a classification of asymptotic regimes for the growing number of balls in [36]. In the terminology of this book, the range of τ_n is the *right m-domain*, which for the poissonised model can be characterised by the properties $\mu_{n,m}(t) = O_{\mathrm{p}}(1)$, $\mu_{n,r}(t) = o_{\mathrm{p}}(1)$ for $r<m$, and $\mu_{n,r}(t) \xrightarrow{\mathrm{p}} \infty$ for $r>m$.

4.2. **Approximating the distribution of M_n.** We turn to the stopped maximum. By the virtue of the adopted stopping condition we cannot have $M_n < m+1$, and for values $r \geqslant m+1$ conditioning on the indices yields

$$\mathbb{P}[M_n \leqslant r \,|\, \tau_n = t] =$$
$$\mathbb{P}[M_n \leqslant r \,|\, \Pi_n(t-) < \Pi_n(t) = m+1 \text{ and } \Pi_i(t) \leqslant m < \Pi_j(t) \text{ for } 1 \leqslant i \leqslant \ell < j \leqslant n-1].$$

Since the event defined by the condition entails $M_n = \max(\Pi_{\ell+1}(t),\ldots,\Pi_{n-1}(t))$ (if $n > \ell+1$), by independence the last formula becomes

$$\mathbb{P}[M_n \leqslant r \,|\, \tau_n = t] = \prod_{j=\ell+1}^{n-1}\mathbb{P}[\Pi_j(t) \leqslant r \,|\, \Pi_j(t) \geqslant m+1] = \left(1 - \frac{\overline{P}_r(t)}{\overline{P}_m(t)}\right)^{n-\ell-1}. \tag{4.4}$$

Integrating out the stopping time we obtain the unconditional distribution of the stopped maximum in the form of a mixture

$$\mathbb{P}[M_n \leqslant r] = \int_0^\infty \left(1 - \frac{\overline{P}_r(t)}{\overline{P}_m(t)}\right)^{n-\ell-1}\mathbb{P}[\tau_n \in \mathrm{d}t], \qquad r \geqslant m+1. \tag{4.5}$$

Finding the asymptotics of (4.5) requires approximating both the mixing distribution and the integrand.

The first part is a customary task from the extreme-value theory, which we include for completeness. The stopping time τ_n has the same distribution as the $(\ell+1)$st order statistic for n i.i.d. Gamma$(m+1,1)$ random variables. The constant

$$\alpha_n := L + m\log L - \log m! \tag{4.6}$$

is a well known $o(1)$ approximation to the upper $1/n$-quantile of Gamma$(m+1,1)$, see [14, p. 156]. With this centering, we have for any fixed $s \in \mathbb{R}$,

$$P_m(\alpha_n + s) \sim \frac{L^m}{m!} e^{-\alpha_n - s} = \frac{1}{n} e^{-s} \tag{4.7}$$

and thus Binomial$\big(n, P_m(\alpha_n+s)\big) \xrightarrow{\mathrm{d}}$ Poisson(e^{-s}). Hence,

$$\mathbb{P}[\tau_n \leqslant \alpha_n + s] = \mathbb{P}\left[\text{Binomial}(n, P_m(\alpha_n+s)) \leqslant \ell\right] \to P_\ell(e^{-s}), \tag{4.8}$$

which means that we have weak convergence

$$\tau_n - \alpha_n \xrightarrow{\mathrm{d}} \tau \tag{4.9}$$

to a random variable τ with $\mathbb{P}[\tau \leqslant s] = P_\ell(e^{-s})$ and thus by (3.4) $\tau \overset{\mathrm{d}}{=} \xi_{\ell+1}$ in the notation there; in other words τ has Gumbel$(\ell+1)$ distribution with density (3.5), that is

$$\mathbb{P}[\tau \in \mathrm{d}s] = e^{-s} p_\ell(e^{-s}) = \frac{1}{\ell!} \exp\big(-(\ell+1)s - e^{-s}\big), \qquad s \in \mathbb{R}. \tag{4.10}$$

For $\ell = 0$ the limit distribution of τ_n is standard Gumbel, which is a well known result in the context of CCP and the dixie cup problems [18]. Comparing with Section 3.1, we see that τ can be realised as the $(\ell+1)$st largest atom of the exponential Poisson process Ξ.

Identifying a proper norming for the integrand (hence M_n) is a much more delicate matter, requiring a bivariate approximation of the Poisson probabilities. To that end, we introduce

$$a_n := eL + \left((e-1)m - \frac{1}{2}\right)\log L - \log\left((e-1)m!^{e-1}\sqrt{2\pi e}\right), \tag{4.11}$$

where the choice of the constant term is the matter of convenience. Set further

$$b_n := \lfloor a_n \rfloor, \qquad c_n := \{a_n\} \tag{4.12}$$

so $a_n = b_n + c_n$ is the decomposition in integer and fractional parts. The next lemma is our main technical tool, giving asymptotics of Poisson probabilities in a vicinity of α_n and b_n. For the time being we may ignore the connection $L = \log n$ and just treat L as a large parameter.

Lemma 4.2. *Let $t = t(L,u) \geqslant 0$ and integer $r = r(L,v) \geqslant 0$ for large enough L be given by*

$$t := L + m\log L + u, \tag{4.13}$$

$$r := eL + \left((e-1)m - \frac{1}{2}\right)\log L + v. \tag{4.14}$$

Then, as $L \to \infty$,

$$\overline{P}_r(t) \sim \frac{1}{e-1} p_r(t), \tag{4.15}$$

$$\log p_r(t) = -L + (e-1)u - v - \log\sqrt{2\pi e} - \frac{1}{2}\log\left(1 + \frac{v}{eL}\right) - \frac{(eu-v)^2}{2(eL+v)} + O\left(\frac{\log L}{L^{1/2}}\right) \tag{4.16}$$

uniformly in u, v within the range

$$\frac{u}{L} + 1 \geqslant \varepsilon, \tag{4.17}$$

$$|eu - v| = O(L^{1/2}) \tag{4.18}$$

for any fixed $\varepsilon > 0$.

Proof. It follows readily from the assumptions (4.17)–(4.18) that $t, r \to \infty$ with $L = O(t \wedge r)$ and $t/r \to 1/e$. This gives the asymptotics in (4.15) by the virtue of (2.4).

For (4.16), Stirling's formula yields

$$\log p_r(t) = -t + r \log\left(\frac{et}{r}\right) - \log\sqrt{2\pi} - \frac{1}{2}\log r + O\left(r^{-1}\right). \tag{4.19}$$

To work out the second term of (4.19), write $et - r = eu - v + (m + 1/2)\log L$ and note that for large L the assumptions (4.17)–(4.18) further imply $(et - r)/r = O(L^{-1/2})$ along with

$$eL + v > eL + eu + O(L^{1/2}) > \varepsilon eL + O(L^{1/2}) > \varepsilon L. \tag{4.20}$$

With all this in hand we calculate by expanding the logarithm

$$\begin{aligned} r\log\left(\frac{et}{r}\right) &= et - r - \frac{(et-r)^2}{2r} + O\left(L^{-1/2}\right) \\ &= eu - v + (m+1/2)\log L - \frac{(eu-v)^2}{2(eL+v)} + O\left(L^{-1/2}\log L\right). \end{aligned} \tag{4.21}$$

Another logarithm expansion gives

$$\log r = \log(eL) + \log\left(1 + \frac{v}{eL}\right) + O(L^{-1}\log L). \tag{4.22}$$

Plugging the last two formulas in (4.19), we obtain (4.16) by careful bookkeeping. □

Applying (4.15) with $L = \log n$ results in the tail asymptotics

$$\overline{P}_{b_n+k}(\alpha_n + s) \quad \sim \quad n^{-1}e^{(e-1)s+c_n-k}, \tag{4.23}$$

uniformly in s and k satisfying $s \geqslant -(1-\varepsilon)L$, and $|es - k| = o(L^{1/2})$, and $k = o(L)$, which hold in particular if s, k assume values in a bounded range. For times $t = \alpha_n + s$ in (4.5) we have $P_m(t) \to 0$ locally uniformly in s. Making use of (4.23) we approximate the integrand to arrive at the following theorem.

Theorem 4.3. *For every fixed* $k \in \mathbb{Z}$, *as* $n \to \infty$,

$$\begin{aligned} \mathbb{P}[M_n - b_n \leqslant k] &= \int_{-\infty}^{\infty} p_\ell(e^{-s})e^{-s}\exp\left(-e^{(e-1)s+c_n-k}\right)\mathrm{d}s + o(1) \\ &= \frac{1}{\ell!}\int_{-\infty}^{\infty}\exp\left(-e^{-s} - (\ell+1)s - e^{(e-1)s+c_n-k}\right)\mathrm{d}s + o(1). \end{aligned} \tag{4.24}$$

□

A peculiar feature of this result (which is a common phenomenon for asymptotics of discrete random variables that do not require scaling) is that the distribution defined by (4.24) has, asymptotically, periodic fluctuations on the $\log n$ scale. Thus, weak convergence of the centered stopped maximum only holds along certain subsequences (n_j) and there are different possible limit distributions; more precisely, we see from (4.24) (or (4.26) below) that convergence in distribution holds for any subsequence such that c_n converges. In [21], equation (4.5) was manipulated using the change of variable $z = nP_{m-1}(t)$, which lead to a representation equivalent to (4.24) via the substitution $z = e^{-s}$.

The following equivalent reformulation of Theorem 4.3 in terms of the total variation distance is new to our knowledge.

Theorem 4.4. *Let b_n and c_n be given by* (4.11)–(4.12)*. Then*

$$\lim_{n\to\infty} d_{\mathrm{TV}}(M_n - b_n, Z_n) = 0 \tag{4.25}$$

with

$$Z_n := \lceil \xi_1 + (e-1)\tau + c_n \rceil, \tag{4.26}$$

where ξ_1 is standard Gumbel-distributed and independent of $\tau \overset{\mathrm{d}}{=} \mathrm{Gumbel}(\ell+1)$ given by (4.10)*.*

Note that the distribution of Z_n depends on n through the constant $0 \leqslant c_n \leqslant 1$, which is the source of the oscillations. Recall also that the variables M_n depend on the fixed parameters $m \geqslant 0$ and $\ell \geqslant 0$; in Theorem 4.4, b_n and c_n depend on m, and τ depends on ℓ. In the special case $\ell = 0$ (CCP and dixie cup problem), ξ_1 and τ are two independent standard Gumbel variables.

Proof. Using (4.10) and (3.6), we can write the first integral in (4.24) as

$$\begin{aligned}\int_{-\infty}^{\infty} \exp\bigl(-e^{(e-1)s+c_n-k}\bigr)\,\mathbb{P}\left[\tau \in \mathrm{d}s\right] &= \int_{-\infty}^{\infty} \mathbb{P}\left[\xi_1 \leqslant -(e-1)s - c_n + k\right]\mathbb{P}\left[\tau \in \mathrm{d}s\right] \\ &= \mathbb{P}\left[\xi_1 \leqslant -(e-1)\tau - c_n + k\right] = \mathbb{P}\left[\xi_1 + (e-1)\tau + c_n \leqslant k\right] \\ &= \mathbb{P}\left[Z_n \leqslant k\right],\end{aligned} \tag{4.27}$$

using also (4.26). Hence (4.24) yields

$$\mathbb{P}[M_n - b_n \leqslant k] = \mathbb{P}[Z_n \leqslant k] + o(1), \qquad \text{for every } k \in \mathbb{Z}. \tag{4.28}$$

It is easy to see, using that the sequence (c_n) is bounded, that (4.28) is equivalent to (4.25), see [26, Lemma 4.1] (or Lemma 5.5 in what follows). □

We have stated Theorem 4.4 with (roughly) centered variables $M_n - b_n$. The result can also be stated without centering as follows.

Theorem 4.5. *Let a_n be given by* (4.11)*. Then*

$$\lim_{n\to\infty} d_{\mathrm{TV}}(M_n, \lceil \xi_1 + (e-1)\tau + a_n \rceil) = 0, \tag{4.29}$$

where ξ_1 is standard Gumbel-distributed and independent of $\tau \overset{\mathrm{d}}{=} \mathrm{Gumbel}(\ell+1)$ given by (4.10)*.*

Proof. We have $d_{\mathrm{TV}}(M_n - b_n, Z_n) = d_{\mathrm{TV}}(M_n, Z_n + b_n)$, and since b_n is an integer,

$$Z_n + b_n = \bigl\lceil \xi_1 + (e-1)\tau + c_n + b_n \bigr\rceil = \bigl\lceil \xi_1 + (e-1)\tau + a_n \bigr\rceil, \tag{4.30}$$

using (4.12). Thus (4.29) is an immediate consequence of Theorem 4.4. □

5. Maximal occupancy counts at fixed and random times

For the classic occupancy scheme the multivariate asymptotics of order statistics were explored in [19; 45] by the method of moments. The approach in [36] relies on poissonisation and a conditioning relation connecting to the multinomial distribution. This work revealed that the maximum occupancy count may exhibit different asymptotic behaviours depending on how the number of balls compares to n. Parallel studies in the extreme-value theory confirmed that the maximum of n Poisson variables with fixed parameter is asymptotically degenerate [1] and later focused on an instance of the triangular scheme, where a continuous Gumbel limit exists [2]. In this section we contribute to the past development by treating the occupancy scheme in the framework of point processes and introducing a simple tool to deal with dependence among the multiplicities.

As in the previous section, we deal with the time range $t = \alpha_n + O(1)$ (corresponding to the right m-domain recalled in Section 4.1), that is $t = \alpha_n + s$, where α_n is defined by (4.6) and s is allowed to vary in a large but bounded range. With the norming suggested by Lemma 4.2, we introduce a lattice point process M_n^s with n atoms

$$M_{n,i}(\alpha_n + s) - b_n,\ i \in [n], \tag{5.1}$$

each corresponding to some box occupancy count $\Pi_j(\alpha_n + s)$, $j \in [n]$. Representing via multiplicities,

$$\mathrm{M}_n^s := \sum_{k=-\infty}^{\infty} \mu_{n,b_n+k}(\alpha_n + s)\delta_k. \tag{5.2}$$

Intuitively, the point process M_n^s captures a few maximal occupancy counts, which are rare among the n boxes for the temporal regime in focus. An obstacle on the way of approximating M_n^s for large n is the dependence among the multiplicities, which persists in the poissonised scheme through the identity $\sum_r \mu_{n,r}(t) = n$. Our strategy to circumvent the dependence makes use of replacing the fixed number of boxes with a random number to pass to a Poisson point process $\widehat{\mathrm{M}}_n^s$, then compare $\widehat{\mathrm{M}}_n^s$ with a suitable exponential Poisson process Ξ_b.

5.1. **Bi-poissonisation.** We randomise the poissonised occupancy scheme by introducing an auxiliary unit-rate Poisson point process $\mathrm{N} = \{\theta_i\}$ with atoms $\theta_1 < \theta_2 < \ldots$, independent of $\Pi_1, \Pi_2, \ldots$. We associate with θ_i the ith box, and treat the whole arrival process Π_i to this box as a random mark attached to θ_i. Geometrically, the now bivariate point data can be plotted in the positive quadrant of the (θ, t)-plane, by first erecting a vertical line at each site θ_i of the θ-axis, then populating line i with the atoms of Π_i. (The resulting planar point process has the intensity measure $\mathrm{d}\theta\mathrm{d}t$ but is not Poisson.) In this picture, the bi-poissonised occupancy scheme with size parameter n corresponds to the configuration of atoms in a vertical strip of width n.

For shorthand we write $N = \mathrm{N}[0, n]$ whenever $n \geqslant 0$ appears in formulas as fixed parameter (which need not be integer). In the bi-poissonised occupancy scheme the box occupancy counts at time $t \geqslant 0$ are

$$\Pi_1(t), \ldots, \Pi_N(t), \tag{5.3}$$

where the number of boxes is random, $N \overset{\mathrm{d}}{=} \mathrm{Poisson}(n)$; in particular with probability e^{-n} the number of boxes is zero. For varying n the models are consistent, so that for $n' < n$ the model with a smaller number of boxes is obtained from the larger model by discarding some number of boxes $\mathrm{N}(n', n]$. The number of balls allocated by time t, equal to $\sum_{i=1}^N \Pi_i(t)$, has a compound Poisson distribution with p.g.f. $z \mapsto \exp\big(n(e^{t(z-1)} - 1)\big)$.

We denote by $\widehat{M_{n,i}}(t), \widehat{\mu}_{n,r}(t), \widehat{\tau}_n$ the bi-poissonised counterparts of the fixed-n random variables and set them equal to zero in the event N takes a small value and they are not defined in a natural way. For instance, in the event $N < \ell + 1$ we set $\widehat{\tau}_n = 0$ for the stopping time which, as before, terminates the allocation process as soon as the number of boxes with at most m balls becomes ℓ. Such conventions do not impact the envisaged distributional asymptotics.

The technical benefit of poissonising the number of boxes relates to the marking theorem, which for many applications can be used in the following transparent form. Let $\mathrm{P}_1, \ldots, \mathrm{P}_k$ $(1 \leqslant k \leqslant \infty)$ be a collection of mutually exclusive properties of a countable subset of $[0, \infty]$, such that a unit-rate Poisson point process possesses one of them almost surely. Let N_j be the set of atoms θ_i whose Π_i satisfies P_j. Then $\mathrm{N}_1, \ldots, \mathrm{N}_k$ are independent Poisson processes.

This applied, immediately gives that the bi-poissonised multiplicities are independent in r (for fixed n, t) and satisfy

$$\widehat{\mu}_{n,r}(t) \overset{\mathrm{d}}{=} \mathrm{Poisson}(np_r(t)),\ r \in \mathbb{Z}_{\geqslant 0}. \tag{5.4}$$

A further consequence concerns independence in the time domain. For $r \geqslant 1$ define *the rth arrival* point process $\widehat{\mathrm{R}}_{n,r}$ to be the set of rth atoms of Π_i's for $i \leqslant N$. That is, instant t accomodates an atom of $\widehat{\mathrm{R}}_{n,r}$ if one of N boxes receives its rth ball at time t. (In the occupancy problems these times are sometimes called r-records, see [12] and references therein.)

Proposition 5.1. *Each point process $\widehat{\mathrm{R}}_{n,r}, r \in \mathbb{Z}_{>0}$, is Poisson, with the intensity measure*

$$\mathbb{E}[\widehat{\mathrm{R}}_{n,r}(\mathrm{d}t)] = np_{r-1}(t)\mathrm{d}t,\ \ t \geqslant 0, \tag{5.5}$$

or, equivalently,

$$\mathbb{E}[\widehat{\mathrm{R}}_{n,r}[t,\infty]] = nP_{r-1}(t),\ \ t \geqslant 0. \tag{5.6}$$

Proof. For θ_i an atom of N and η_i the rth arrival in Π_i, we have that the pairs (θ_i, η_i) comprise a bivariate Poisson process. Projecting these for $i \leqslant N$ yields $\widehat{\mathrm{R}}_{n,r}$, which is thus Poisson. □

The proposition bears some similarity with the r-records (rank r observations) in the sense of extreme-value theory [43, Section 4.6]. The major difference, however, is that the $\widehat{\mathrm{R}}_{n,r}$'s are not independent for different values of r.

The bi-poissonised stopping time $\widehat{\tau}_n$ identifies as the $(\ell+1)$st rightmost atom of $\widehat{\mathrm{R}}_{n,m+1}$, hence Proposition 5.1 yields an explicit formula for the distribution

$$\mathbb{P}[\widehat{\tau}_n \leqslant t] = P_\ell(nP_m(t)), \tag{5.7}$$

analogous to (4.2). This can be manipulated to show that $\widehat{\tau}_n - \alpha_n \xrightarrow{\mathrm{d}} \tau$, where $\tau \stackrel{\mathrm{d}}{=} \mathrm{Gumbel}(\ell+1)$ as in the fixed-n model. (Alternatively, condition on N to prove $\widehat{\tau}_n - \alpha_N \xrightarrow{\mathrm{d}} \tau$ and use $|\alpha_N - \alpha_n| = o_p(1)$.) This becomes most elementary in the CCP case $\ell = 0, m = 0$, where the centering constant is $\alpha_n = L$ and the distribution function coincides *exactly* with the standard Gumbel distribution on $(-L, \infty)$:

$$\mathbb{P}[\widehat{\tau}_n - L \leqslant s] = \exp(-ne^{-s-L}) = e^{-e^{-s}},\ \ s \geqslant -L, \tag{5.8}$$

which entails that the value $\widehat{\tau}_n = 0$ has positive probability e^{-n}.

5.2. **Approximating M_n^s by $\widehat{\mathrm{M}}_n^s$.** By the marking theorem the multiplicities $\widehat{\mu}_{n,r}(t), r \in \mathbb{Z}_{\geqslant 0}$, for each fixed t are independent, as said above, which makes

$$\widehat{\mathrm{M}}_n^s := \sum_{k=-\infty}^{\infty} \widehat{\mu}_{n,b_n+k}(\alpha_n + s)\delta_k \tag{5.9}$$

a lattice Poisson point process, with mean multiplicity

$$np_{b_n+k}(\alpha_n + s) \tag{5.10}$$

(equal to zero for $k < -b_n$ or $s \leqslant -\alpha_n$) at site $k \in \mathbb{Z}$. We proceed with estimating the proximity of M_n^s to $\widehat{\mathrm{M}}_n^s$ restricted to a region containing the few largest atoms with high probability.

For b_n as in (4.12) and $r_n < b_n$ yet to be chosen, consider $b_n - r_n$ as a truncation level for the occupancy counts. The number of boxes $i \leqslant n$ with $\Pi_i(\alpha_n + s) > b_n - r_n$ is a binomial random variable X with parameters n and $\overline{P}_{b_n - r_n}(\alpha_n + s)$. Likewise the number of boxes $i \leqslant N(n)$ satisfying this condition is a Poisson random variable Y with the same mean $\mathbb{E}[X] = \mathbb{E}[Y]$. A criterion for selecting the truncation level is that the mean number of overshoots goes to ∞ but the overshoot probability for any particular box i approaches 0. Choosing $r_n \sim \alpha L$ with some $0 < \alpha < e - 1$, Stirling's formula gives the asymptotics

$$-\log(\overline{P}_{b_n - r_n}(\alpha_n + s)) \sim \beta L \tag{5.11}$$

with some $0 < \beta < 1$ depending on α. This entails the desired $n\overline{P}_{b_n - r_n}(\alpha_n + s) \to \infty$ and $\overline{P}_{b_n - r_n}(\alpha_n + s) \to 0$, to enable application of Prohorov's Poisson-binomial bound (see e.g. [4, p. 2, and also (1.6) and (1.23)] and the references there) that becomes

$$d_{\mathrm{TV}}(X, Y) = O(n^{-\beta'}), \tag{5.12}$$

for any $\beta' < \beta$, in fact locally uniformly in s. Appealing to the identity (3.16) for MBPP's this translates as

$$d_{\mathrm{TV}}(\mathrm{M}^s_n|_{[-r_n,\infty]}\,, \widehat{\mathrm{M}}^s_n|_{[-r_n,\infty]}) = O(n^{-\beta'}) \tag{5.13}$$

(where r_n grows logarithmically as above). For fixed positive integer K, the events

$$\{X < K\} = \{M_{n,K}(\alpha_n + s) \leqslant b_n - r_n\} = \{M_{n,K}(\alpha_n + s) - b_n \leqslant -r_n\}, \tag{5.14}$$

$$\{Y < K\} = \{\widehat{M}_{n,K}(\alpha_n + s) \leqslant b_n - r_n\} = \{\widehat{M}_{n,K}(\alpha_n + s) - b_n \leqslant -r_n\} \tag{5.15}$$

have much smaller probabilities, also as a consequence of (5.11), and by a coupling argument we can force the K top atoms of M^s_n and $\widehat{\mathrm{M}}^s_n$ to coincide. From this, we obtain that, for any fixed K,

$$d_{\mathrm{TV}}\big((M_{n,i}(\alpha_n + s))_{i=1}^K, (\widehat{M}_{n,i}(\alpha_n + s))_{i=1}^K\big) = O(n^{-\beta'}). \tag{5.16}$$

We remark that it is possible to improve on the above by letting K grow as a small power of n, this way extending the approximation to cover some intermediate order statistics.

5.3. **Coupling by the index of box.** An alternative coupling of a fixed number of maximal occupancy counts in the fixed-n and bi-poissonised models only uses symmetry and not distribution of the counts. For fixed $t > 0$ and $n > K$ there exist K distinct boxes whose ordered occupancy counts are $M_{n,1}(t), \dots, M_{n,K}(t)$. If $M_{n,K} > M_{n,K+1}$ the set of such boxes is uniquely determined, otherwise we choose at random a suitable number from the boxes with $M_{n,K}$ balls. The labels of thus selected K boxes is a random sample from $[n]$; let I be the largest index in the sample. For $N \overset{\mathrm{d}}{=} \mathrm{Poisson}(n)$, if $N > K$ define $\widehat{I}$ similarly, leaving the index undefined otherwise. For $\varepsilon > 0$, if the event $A := \{I \vee \widehat{I} \leqslant n - n\varepsilon < N < n + n\varepsilon\}$ occurs then we can couple in such a way that

$$(M_{n,i}(t))_{i=1}^K \text{ and } (\widehat{M}_{n,i}(t))_{i=1}^K \tag{5.17}$$

are equal. Using exchangeability we obtain estimates (at least if $n\varepsilon$ is an integer)

$$\mathbb{P}[I > n - n\varepsilon] \leqslant K\varepsilon, \qquad \mathbb{P}[|N - n| < n\varepsilon, \widehat{I} > n - n\varepsilon] \leqslant \frac{2K\varepsilon}{1 - \varepsilon}. \tag{5.18}$$

Choosing $\varepsilon \sim 2\sqrt{L/n}$, these dominate the known tail bound $\mathbb{P}[|N - n| \geqslant n\varepsilon] \leqslant 2e^{-n\varepsilon^2/3}$ (following from Bennet's inequality, see also [24, Corollary 2.3 and Remark 2.6]); hence the total variation distance between the vectors in (5.17) does not exceed $\mathbb{P}[A^c] < 7K\sqrt{L}n^{-1/2}$ for n not too small.

5.4. **Approximating $\widehat{\mathrm{M}}^s_n$ by an exponential process.** By passing to the Poisson point process $\widehat{\mathrm{M}}^s_n$ the approximation problem is reduced to asymptotics of the intensity measure. For Poisson random variables X and Y, we have that $d_{\mathrm{TV}}(X, Y) \leqslant |\mathbb{E}[X] - \mathbb{E}[Y]|$. From this the total variation distance between lattice Poisson processes does not exceed the ℓ_1-distance between their intensity measures, seen as sequences of point masses at integer locations. (Stronger bounds exist, as follows from [27, Theorem 2.2(i) and its proof], but we do not need them.)

The intensity measure of $\Xi^{\uparrow}_b$ has masses growing exponentially fast in the negative direction. This forces us to impose a less generous truncation than in Section 5.2 to keep the quality of approximation high.

Proposition 5.2. *For $r'_n \sim \beta \log L$ with any $0 < \beta < 1/2$, locally uniformly in s,*

$$d_{\mathrm{TV}}(\widehat{\mathrm{M}}^s_n|_{[-r'_n,\infty]}\,,\ \Xi^{\uparrow}_{(e-1)s+c_n}|_{[-r'_n,\infty]}) \to\ 0. \tag{5.19}$$

Proof. We first restrict both Poisson processes to the interval $(-r'_n\,, L^{1/4})$. Within this range we approximate the point masses of the intensity measure of $\Xi^{\uparrow}_{(e-1)s+c_n}$ by their $\widehat{\mathrm{M}}^s_n$-counterparts; recall that these intensities are given by (3.10) and (5.10). To that end, Lemma 4.2 is applied with $u = O(1)$ and v varying in the range from $-r'_n + O(1)$ to $L^{1/4} + O(1)$. A simple calculation, using (4.6), (4.11) and (4.12), shows that the pointwise *relative* approximation error for the point masses is then given by the last three terms in (4.16), which altogether are estimated as $cL^{-1/2}\log L$ for some $c > 0$, as one easily checks. Since for the exponential process

$$\mathbb{E}\,[\Xi^{\uparrow}_b(-r'_n,\infty]] \asymp e^{r'_n} \tag{5.20}$$

(locally uniformly in b), the ℓ_1-distance between the mean measures on $(-r'_n\,, L^{1/4})$ is of the order $O(e^{r'_n}L^{-1/2}\log L)$, which is in fact $o(1)$ by our choice of r'_n. By this very token and (4.15), for both processes the total mean measure of $[L^{1/4},\infty]$ is $O(\exp(-L^{1/4}))$, which makes a negligible contribution to the total variation distance in (5.19). □

Clearly, (5.19) combined with (5.13) yields the desired approximation

$$d_{\mathrm{TV}}(\mathrm{M}^s_n|_{[-r'_n,\infty]}\,,\ \Xi^{\uparrow}_{(e-1)s+c_n}|_{[-r'_n,\infty]}) \to\ 0. \tag{5.21}$$

Obviously from (5.20), for any fixed K, for both processes the K largest atoms exceed $-r'_n$ w.h.p., whence with the notation of Section 3 we obtain as a corollary

Theorem 5.3. *For every $K \in \mathbb{Z}_{>0}$, as $n \to \infty$ locally uniformly in $s \in \mathbb{R}$*

$$d_{\mathrm{TV}}\big((M_{n,i}(\alpha_n + s) - b_n)_{i=1}^K\,,\ (\lceil \xi_i + (e-1)s + c_n\rceil)_{i=1}^K\big) \to 0, \tag{5.22}$$

where $\xi_i \stackrel{\mathrm{d}}{=} \mathrm{Gumbel}(i)$ is the decreasing sequence of atoms of Ξ.

Remark 5.4. (The need for a condition for convergence in distribution.) The more general asymptotic regime $t_n \sim \alpha \log n$ was addressed in [36, Ch. II, Section 6, Theorem 2]. Adjusted to the poissonised model, the cited result claims a weak convergence of $M_{n,1}(t) - r(n,t)$ provided an (integer) centering constant $r = r(n,t)$ is chosen so that $np_r(t) \to \lambda$, for some $\lambda \in (0,\infty)$. However, for the generic sequence (t_n) with such asymptotics the required centering need not exist, as exemplified by Lemma 4.2 in the case $\alpha = 1$. Thus in essense the weak convergence necessitates an additional constraint on (t_n) to avoid oscillations of certain Poisson probabilities. △

Oscillatory asymptotics for a sequence of distributions involve closeness to a set of accumulation points of the sequence. Leaving aside the periodicity patterns, this fits in the following broad scenario. Let (S,d) be a metric space, and let $\gamma \mapsto H_\gamma$ be a continuous map from a compact interval (or, more generally, some compact metric space) J into S.

Lemma 5.5. *Let (G_n) be a sequence in S and let $\rho(n)$ be a sequence in J. Then the following are equivalent.*

(i) *For any subsequence (n_j) and $\gamma \in J$ such that $\rho(n_j) \to \gamma$ as $j \to \infty$, we have $G_{n_j} \to H_\gamma$.*
(ii) *As $n \to \infty$,*

$$d(G_n, H_{\rho(n)}) \to 0. \tag{5.23}$$

Furthermore, if these hold, then (G_n) is relatively compact.

Proof. (ii)⇒(i). If (ii) holds and $\rho(n_j) \to \gamma$, then $d(G_{n_j}, H_{\rho(n_j)}) \to 0$ by (ii), and $d(H_{\rho(n_j)}, H_\gamma) \to 0$ by the assumption on H. Consequently, $d(G_{n_j}, H_\gamma) \to 0$.

(i)⇒(ii). Consider any subsequence (n_j). Since the set J is compact, we may select a subsubsequence (n'_j) such that $n'_j \to \gamma$ for some $\gamma \in J$, and thus $G_{n'_j} \to H_\gamma$ by the assumption (i). Furthermore, $H_{\rho(n'_j)} \to H_\gamma$ by continuity. Consequently, as $j \to \infty$,

$$d(G_{n'_j}, H_{\rho(n'_j)}) \leqslant d(G_{n'_j}, H_\gamma) + d(H_\gamma, H_{\rho(n'_j)}) \to 0. \tag{5.24}$$

Hence every subsequence has a subsubsequence for which (5.23) holds; as is well-known, this implies (5.23) for the full sequence.

Finally, if (i) and (ii) hold, then for any subsequence (n_j) we may select a subsubsequence (n'_j) such that $\rho(n'_j)$ converges, which by (i) implies that $(G_{n'_j})$ converges. Hence, (G_n) is relatively compact. □

Applying the lemma to the setting of Theorem 5.3, we take for (S, d) the space of probability distributions on $\mathbb{Z}^K$ with $d = d_{\mathrm{TV}}$, $J := [0, 1]$, and $\rho(n) := c_n$. Then (5.22) is an instance of (5.23), and (i) in Lemma 5.5 describes subsequential limits in distribution, for subsequences with $c_n \to \gamma$ for some $\gamma \in [0, 1]$.

5.5. **Stopped maxima.** We have now all prerequisites to derive one of our main results, a multivariate extension of Theorems 4.3 and 4.4, from Fact 4.1. The latter says that conditionally on $\tau_n = t$, the $n - \ell - 1$ box occupancy counts exceeding m at this time, excluding the box that gets the ball at τ_n, are i.i.d. with the truncated Poisson distribution

$$\mathbb{P}[\Pi_i(t) = r \mid \Pi_i(t) \geqslant m + 1] = \frac{p_r(t)}{\overline{P}_m(t)}, \qquad r \geqslant m + 1. \tag{5.25}$$

Recall also the shorthand notation $M_{n,i} = M_{n,i}(\tau_n)$ for the box occupancies when the allocation is stopped.

Theorem 5.6. *For every $K \in \mathbb{Z}_{>0}$, as $n \to \infty$,*

$$d_{\mathrm{TV}}\left((M_{n,i} - b_n)_{i=1}^K, (\lceil \xi_i + (e-1)\tau + c_n \rceil)_{i=1}^K\right) \to 0, \tag{5.26}$$

where $\tau \overset{\mathrm{d}}{=} \mathrm{Gumbel}(\ell + 1)$ is independent of Ξ.

Proof. For $t = \alpha_n + O(1)$ we have $P_m(t) = O(n^{-1})$, therefore the asymptotics in Lemma 4.2 hold for the truncated Poisson distribution in (5.25) as well; we may also replace n by $n-\ell-1$ without changing the result. In view of this, it follows from the fact mentioned before the theorem that the approximation in Theorem 5.3 remains valid also for $(M_{n,i}(\tau_n) - b_n)_{i=1}^K$ conditioned on $\tau_n - \alpha_n = s$, locally uniformly in s. In other words, conditionally on τ_n and with $\Xi = \{\xi_i\}$ independent of τ_n,

$$d_{\mathrm{TV}}\left((M_{n,i}(\tau_n) - b_n)_{i=1}^K, (\lceil \xi_i + (e-1)(\tau_n - \alpha_n) + c_n \rceil)_{i=1}^K\right) \to 0, \tag{5.27}$$

uniformly for $\tau_n - \alpha_n$ in a compact set. Recall that $\tau_n - \alpha_n \overset{\mathrm{d}}{\to} \tau$. In particular, $\tau_n - \alpha_n$ is tight, and thus it follows that (5.27) holds also unconditionally.

If $c_{n_j} \to c_0$ for some subsequence (n_j) then furthermore, along the subsequence,

$$(\lceil \xi_i + (e-1)(\tau_n - \alpha_n) + c_n \rceil)_{i=1}^K \overset{\mathrm{d}}{\to} (\lceil \xi_i + (e-1)\tau + c_0 \rceil)_{i=1}^K \tag{5.28}$$

by the mapping theorem [8, Theorem 5.1] (since $x \mapsto \lceil x \rceil$ a.s. is continuous at $\xi_i + (e-1)\tau + c_0$); since the random variables in (5.28) take values in the countable set $\mathbb{Z}^K$, it follows by Scheffé's lemma that (5.28) holds also in total variation. Hence, (5.27) implies that, along (n_j),

$$d_{\mathrm{TV}}\left((M_{n,i}(\tau_n) - b_n)_{i=1}^K, (\lceil \xi_i + (e-1)\tau + c_0 \rceil)_{i=1}^K\right) \to 0. \tag{5.29}$$

Finally, Lemma 5.5 enables us to pass from the convergence of subsequences to the claimed approximation (5.26). □

5.6. **Equivalent formulations.** By analogy with Theorem 4.5, we may formulate Theorem 5.6 in an equivalent way for the non-centered variables $M_{n,i} = M_{n,i}(\tau_n)$.

Theorem 5.7. *For every $K \in \mathbb{Z}_{>0}$, as $n \to \infty$,*

$$d_{\mathrm{TV}}\left((M_{n,i})_{i=1}^{K}\,,\,(\lceil \xi_i + (e-1)\tau + a_n\rceil)_{i=1}^{K}\right) \to 0, \tag{5.30}$$

where $\tau \overset{\mathrm{d}}{=} \mathrm{Gumbel}(\ell+1)$ is independent of $\Xi = \{\xi_i\}$.

As emphasised in Section 3.5, it is technically convenient to regard the point processes in focus as point processes on $(-\infty,\infty]$, although they never have an atom at ∞. Let $\mathcal{N}(-\infty,\infty]$ be the space of locally finite integer-valued measures on $(-\infty,\infty]$, and regard a point process on $(-\infty,\infty]$ as a random element of $\mathcal{N}(-\infty,\infty]$. The space $\mathcal{N}(-\infty,\infty]$ is equipped with the vague topology, which is metrisable (and Polish) [6, Proposition 3.1]. Using this framework we can state Theorem 5.6 in the following equivalent form that involves a limit of the entire set $\{\Pi_i(\tau_n)\}_1^n = \{M_{n,i}\}_1^n$ regarded as a point process.

Theorem 5.8. *Let d be any metric on the space $\mathcal{N}(-\infty,\infty]$ that induces the vague topology. Then*

$$d\big(\{M_{n,i} - b_n\}_{i=1}^{n}\,,\,\{\lceil \xi_i + (e-1)\tau + c_n)\rceil\}_{i=1}^{\infty}\big) \to 0. \tag{5.31}$$

Proof. Consider a subsequence (n_j) such that $c_{n_j} \to c_0$ for some $c_0 \in [0,1]$. Then, by the virtue of Lemma 3.1, we obtain from (5.29) that, along the subsequence,

$$d\left(\{M_{n,i} - b_n\}_{i=1}^{n}\,,\,\left\{\lceil \xi_i + (e-1)\tau + c_0\rceil\right\}_{i=1}^{\infty}\right) \to 0. \tag{5.32}$$

The result now follows from Lemma 5.5. □

At last, instead of looking at few rightmost atoms, we may restrict our point processes to a vicinity of ∞. This leads by the virtue of Lemma 3.1 (and Lemma 5.5) to the following equivalent version.

Theorem 5.9. *For every $r \in \mathbb{Z}$*

$$d_{\mathrm{TV}}\big(\mathrm{M}_n^{\tau_n - \alpha_n}|_{[r,\infty]}, \Xi^{\uparrow}_{(e-1)\tau + c_n}|_{[r,\infty]}\big) \to 0. \tag{5.33}$$

6. Proof of the main result by interpolation

Theorem 5.8 shows that the lattice point process of stopped counts $\{\Pi_i(\tau_n)\}_{i=1}^n$ may be approximated by the exponential Poisson process Ξ that is shifted and then has all atoms rounded to integers. We present here an alternative proof based on the idea of interpolation of the lattice process to $\mathbb{R}$, that amounts to artificially adding the 'missing' fractional parts to the atoms and then showing convergence to an exponential Poisson process on $\mathbb{R}$. By this approach, the oscillations are revealed only at the final stages of the argument.

The shift operation on Ξ from Section 3.1 makes sense for arbitrary point process on $\mathbb{R}$. Thus, for a point process $\mathrm{H} = \{\eta_i\}$ and a real number b, we let $\mathrm{H} \pm b$ denote the shifted processes

$$\mathrm{H} + b := \{\eta_i + b\}, \qquad \mathrm{H} - b := \{\eta_i - b\} \tag{6.1}$$

obtained by translating each atom the same way.

For convenience of notation, let $W_t \overset{\mathrm{d}}{=} \mathrm{Poisson}(t)$, and let W'_t and W''_t have the truncated (conditioned) distributions $W'_t \overset{\mathrm{d}}{=} (W_t \mid W_t \geqslant m+1)$ and $W''_t \overset{\mathrm{d}}{=} (W_t \mid W_t \leqslant m)$. By Fact 4.1, if we condition on $\tau_n = t$, then the occupancy numbers $\{\Pi_i(\tau_n)\}_{i=1}^n$ are given by

$$\{\Pi'_i(t)\}_{i=1}^{n-\ell-1} \cup \{\Pi''_j(t)\}_{j=1}^{\ell} \cup \{m+1\}, \tag{6.2}$$

where all random variables are independent and have the distributions $\Pi'_i(t) \overset{\mathrm{d}}{=} W'_t$ and $\Pi''_j(t) \overset{\mathrm{d}}{=} W''_t$. The last $\ell+1$ of the numbers in (6.2) are $\leqslant m+1$ and may be ignored, as will be seen below, so only the $n-\ell-1$ numbers $\Pi'_i(t)$ are important asymptotically.

Let $E \overset{\mathrm{d}}{=} \mathrm{Exponential}(1)$ and consider $R \overset{\mathrm{d}}{=} (E \mid E < 1)$; thus R is a random variable in $[0,1)$ with the distribution function

$$\mathbb{P}[R \leqslant x] = \frac{1-e^{-x}}{1-e^{-1}}, \qquad 0 \leqslant x \leqslant 1. \tag{6.3}$$

Let R_i ($i \geqslant 1$) be independent copies of R, also independent of all other variables. We define, for a given $t = t_n$, the modified variables

$$\widetilde{\Pi}'_i := \Pi'_i(t_n) + R_i \tag{6.4}$$

and note that we are back to integer counts via $\Pi'_i(t_n) = \lfloor \widetilde{\Pi}'_i \rfloor$.

Let s_n and x_n be any bounded sequences of real numbers, and consider only n that are so large that $\log n + s_n \geqslant 0$. Let, recalling (4.6), (4.11), and $L := \log n$,

$$t_n := \alpha_n + s_n, \tag{6.5}$$
$$y_n := a_n + x_n, \tag{6.6}$$
$$k_n := \lfloor y_n \rfloor, \tag{6.7}$$
$$x'_n := k_n - a_n. \tag{6.8}$$

Note that

$$x_n - x'_n = y_n - k_n = \{y_n\} \in [0,1). \tag{6.9}$$

Then, for $i \leqslant n-\ell-1$ we have, with W'_{t_n} as above and independent of R_i, using (6.3),

$$\begin{aligned}
\mathbb{P}[\widetilde{\Pi}'_i > y_n \mid \tau_n = t_n] &= \mathbb{P}[W'_{t_n} + R_i > y_n] \\
&= \mathbb{P}[W'_{t_n} > k_n] + \mathbb{P}[W'_{t_n} = k_n]\,\mathbb{P}[R_i > y_n - k_n] \\
&= \mathbb{P}[W'_{t_n} > k_n] + \mathbb{P}[W'_{t_n} = k_n]\frac{e^{-(y_n-k_n)} - e^{-1}}{1-e^{-1}}.
\end{aligned} \tag{6.10}$$

We have $t_n \to \infty$, and thus $\mathbb{P}[W_{t_n} \leqslant m] \to 0$. Hence, by Lemma 4.2 and simple calculations,

$$\begin{aligned}
\mathbb{P}[W'_{t_n} = k_n] &= \mathbb{P}[W_{t_n} = k_n \mid W_{t_n} > 0] = \mathbb{P}[W_{t_n} = k_n]\bigl(1+o(1)\bigr) \\
&= \frac{e-1}{n} e^{(e-1)s_n - x'_n + o(1)}.
\end{aligned} \tag{6.11}$$

and similarly

$$\mathbb{P}[W'_{t_n} > k_n] = \mathbb{P}[W_{t_n} > k_n]\bigl(1+o(1)\bigr) = \frac{1}{n} e^{(e-1)s_n - x'_n + o(1)}. \tag{6.12}$$

Then (6.10)–(6.12) yield, recalling (6.6) and (6.9),

$$\begin{aligned}
\mathbb{P}[\widetilde{\Pi}'_i > a_n + x_n] &= \frac{1}{n} e^{(e-1)s_n - x'_n + o(1)} + \frac{e-1}{n} e^{(e-1)s_n - x'_n + o(1)} \frac{e^{-(y_n-k_n)} - e^{-1}}{1-e^{-1}} \\
&= \frac{1}{n} e^{(e-1)s_n - x'_n} \Bigl(e^{o(1)} + e^{o(1)}\bigl(e^{1-(x_n-x'_n)} - 1\bigr)\Bigr)
\end{aligned}$$

$$= \frac{1}{n}\Big(e^{(e-1)s_n+1-x_n} + o(1)\Big). \tag{6.13}$$

Let $x \in \mathbb{R}$ and choose $x_n := x + (e-1)s_n + 1$. Furthermore, define

$$\widehat{\Pi}_i(s_n) := \widetilde{\Pi}'_i - a_n - (e-1)s_n - 1 = \Pi'_i(t_n) + R_i - a_n - (e-1)s_n - 1. \tag{6.14}$$

Then (6.13) yields

$$\mathbb{P}[\widehat{\Pi}_i(s_n) > x] = \mathbb{P}[\widetilde{\Pi}'_i > a_n + x_n] = \frac{1}{n}\Big(e^{-x} + o(1)\Big). \tag{6.15}$$

Note that $\widehat{\Pi}_i(s_n)$ depends on the chosen bounded sequence s_n, both directly and through t_n, but the right-hand side of (6.15) does not.

The random variables $\widehat{\Pi}_i(s_n)$ are independent for $1 \leqslant i \leqslant n-\ell-1$. Consequently, if we define the point process

$$\begin{aligned}\widehat{\Xi}'_n[s_n] &:= \big\{\widehat{\Pi}_i(s_n)\big\}_{i=1}^{n-\ell-1} \\ &= \big\{\Pi'_i(t_n) + R_i\big\}_{i=1}^{n-\ell-1} - a_n - (e-1)s_n - 1,\end{aligned} \tag{6.16}$$

then (6.15) shows by the standard Poisson convergence of binomial distributions that, still for any bounded sequence s_n,

$$\widehat{\Xi}'_n[s_n](x,\infty] \xrightarrow{\mathrm{d}} \operatorname{Poisson}\big(e^{-x}\big). \tag{6.17}$$

Since $\Xi(x,\infty] \overset{\mathrm{d}}{=} \operatorname{Poisson}\big(e^{-x}\big)$, cf. (3.3), we thus obtain from (6.17)

$$\widehat{\Xi}'_n[s_n](A) \xrightarrow{\mathrm{d}} \Xi(A) \tag{6.18}$$

for every interval $A = (x,\infty]$.

This is not quite enough to show convergence in distribution in the space $\mathcal{N}(-\infty,\infty]$, but it is not far from it. Let $\mathcal{U}$ be the family of all finite unions $\bigcup_1^k (u_j, v_j]$ with $-\infty < u_j < v_j \leqslant \infty$. For any such set $A \in \mathcal{U}$, we can use (6.15) for $x = u_j$ and v_j, $j = 1, \dots, k$, and conclude that

$$\mathbb{P}[\widehat{\Pi}_i(s_n) \in A] = \frac{1}{n}\big(\mu(A) + o(1)\big), \tag{6.19}$$

where $\mathrm{d}\mu(x) = e^{-x}\,\mathrm{d}x$ is the intensity measure of Ξ, and it follows as above that (6.18) holds for every $A \in \mathcal{U}$. This implies convergence

$$\widehat{\Xi}'_n[s_n] \xrightarrow{\mathrm{d}} \Xi \tag{6.20}$$

in $\mathcal{N}(-\infty,\infty]$, see for example [30, Theorem 4.15]. (In the terminology there, $\mathcal{U}$ is a dissecting ring, and we may take $\mathcal{I} = \mathcal{U}$; both conditions in the theorem follow from (6.18) for $A \in \mathcal{U}$. See also the version in [29, Proposition 16.17].) Alternatively, (6.20) follows easily from (6.15) using [30, Corollary 4.25]; we leave the details to the reader.

Define $S_n := \tau_n - \alpha_n$, and recall from (4.9) that

$$S_n \xrightarrow{\mathrm{d}} \tau, \quad \text{with } \tau \overset{\mathrm{d}}{=} \operatorname{Gumbel}(\ell+1). \tag{6.21}$$

Define also the point processes

$$\widetilde{\Xi}_n := \big\{\Pi_i(\tau_n) + R_i\big\}_{i=1}^{n}, \tag{6.22}$$

$$\widehat{\Xi}_n[s_n] := \widetilde{\Xi}_n - (e-1)S_n - a_n - 1 = \big\{\Pi_i(\tau_n) + R_i - (e-1)S_n - a_n - 1\big\}_{i=1}^{n}. \tag{6.23}$$

Then, by (6.5) and (6.2),

$$\big(\widetilde{\Xi}_n \mid S_n = s_n\big) = \big(\widetilde{\Xi}_n \mid \tau_n = t_n\big) \overset{\mathrm{d}}{=} \{\Pi'_i(t_n) + R_i\}_{i=1}^{n-\ell-1} \cup \{\Pi''_j(t_n) + R_{n-j}\}_{j=0}^{\ell}, \tag{6.24}$$

where we for convenience let $\Pi''_n(t_n) := m+1$. (We have only equality in distribution in (6.24), since the equality involves a harmless relabelling of $R_1, \dots, R_n$.) Hence, also using (6.23) and (6.16),

$$\begin{aligned}\bigl(\widehat{\Xi}_n[s_n] \mid S_n = s_n\bigr) &= \bigl(\widetilde{\Xi}_n \mid S_n = s_n\bigr) - (e-1)s_n - a_n - 1\\ &\overset{\mathrm{d}}{=} \widehat{\Xi}'_n[s_n] \cup \bigl\{\Pi''_j(t_n) + R_{n-j} - (e-1)s_n - a_n - 1\bigr\}_{j=0}^{\ell},\end{aligned} \tag{6.25}$$

For $0 \leqslant j \leqslant \ell$, we have $\Pi''_j(t_n) \leqslant m+1$ and thus, recalling $s_n = O(1)$,

$$\Pi''_j(t_n) + R_{n-j} - (e-1)s_n - a_n - 1 = -a_n + O(1) \to -\infty. \tag{6.26}$$

Hence, if $A \in \mathcal{U}$ is as above, then for large n, the final multiset in (6.25) is disjoint from A, and thus (6.25) shows that (6.18) holds also for $\bigl(\widehat{\Xi}_n[s_n] \mid S_n = s_n\bigr)$, which, as for (6.20) above, implies

$$\bigl(\widehat{\Xi}_n[s_n] \mid S_n = s_n\bigr) \xrightarrow{\mathrm{d}} \Xi \tag{6.27}$$

in $\mathcal{N}(-\infty, \infty]$. (Alternatively, this follows from (6.20), (6.25), and (6.26) using Lemma 3.1.)

We have shown (6.27) for any bounded sequence s_n. Thus Lemma 6.1 below applies and yields

$$\bigl(\widehat{\Xi}_n[s_n], S_n\bigr) \xrightarrow{\mathrm{d}} (\Xi, \tau), \tag{6.28}$$

with Ξ and $\tau \overset{\mathrm{d}}{=} \mathrm{Gumbel}(\ell+1)$ independent. Consequently, by (6.23) and the continuous mapping theorem,

$$\bigl\{\Pi_i(\tau_n) + R_i - a_n - 1\bigr\}_1^n = \widehat{\Xi}_n[s_n] + (e-1)S_n \xrightarrow{\mathrm{d}} \Xi + (e-1)\tau. \tag{6.29}$$

This is our continuous version of Theorems 5.6–5.8, where we have added artificial fractional parts R_i in order to get a nice limit $\Xi + (e-1)\tau$ consisting of the Poisson process Ξ with an independent random shift $(e-1)\tau$.

To obtain the desired conclusions about the occupancy counts it now remains only to remove the fractional parts. Arrange $\widetilde{\Pi}_i := \Pi_i(\tau) + R_i$ in decreasing order as $\widetilde{\Pi}_{(1)} \geqslant \widetilde{\Pi}_{(2)} \geqslant \dots$, and note that then

$$M_{n,i} = \lfloor \widetilde{\Pi}_{(i)} \rfloor. \tag{6.30}$$

Lemma 3.1 shows that (6.29) is equivalent to, with $\Xi = \{\xi_i\}_{i=1}^{\infty}$ as in Section 3.1,

$$\bigl(\widetilde{\Pi}_{(i)} - a_n - 1\bigr)_{i=1}^K \xrightarrow{\mathrm{d}} \bigl(\xi_i + (e-1)\tau\bigr)_{i=1}^K \tag{6.31}$$

for every fixed $K \geqslant 1$.

We write as in (4.12) $a_n = b_n + c_n$ where $b_n := \lfloor a_n \rfloor$ is an integer and $c_n := \{a_n\} \in [0,1)$ is the fractional part. Consider a subsequence such that $c_n \to \gamma$ for some $\gamma \in [0,1]$. Then, along this subsequence, it follows from (6.31) that

$$\bigl(\widetilde{\Pi}_{(i)} - b_n\bigr)_{i=1}^K \xrightarrow{\mathrm{d}} \bigl(\xi_i + (e-1)\tau + 1 + \gamma\bigr)_{i=1}^K. \tag{6.32}$$

The K variables on the right-hand side of (6.32) have continuous distributions, and are thus a.s. not integers; hence, the vector on the right-hand side is a.s. a continuity point of the mapping $F : \mathbb{R}^K \to \mathbb{R}^K$ given by $(z_i)_1^K \mapsto (\lfloor z_i \rfloor)_1^K$. Consequently, by [8, Theorem 5.1], we may apply this mapping F and conclude, using (6.30) and (6.32), that

$$\bigl(M_{n,i} - b_n\bigr)_{i=1}^K = \bigl(\lfloor \widetilde{\Pi}_{(i)} - b_n \rfloor\bigr)_{i=1}^K \xrightarrow{\mathrm{d}} \bigl(\lfloor \xi_i + (e-1)\tau + 1 + \gamma \rfloor\bigr)_{i=1}^K = \bigl(\lceil \xi_i + (e-1)\tau + \gamma \rceil\bigr)_{i=1}^K. \tag{6.33}$$

This yields Theorem 5.6 by Lemma 5.5, again taking $d = d_{\mathrm{TV}}$ and letting G_n and H_γ be distributions of the random vectors on left and right sides of (6.33).

Theorems 5.7 and 5.8 follow easily as in Section 5.5.

6.1. **A general lemma.** We used in the proof above the following simple lemma on joint convergence using conditional distributions. We admit this may belong to the folklore, and give a detailed proof since we are not aware of any explicit reference.

Lemma 6.1. *Let (X_n, Y_n), $n \geqslant 1$, be a sequence of pairs of random variables taking values in $\mathcal{X} \times \mathcal{Y}$ for some Polish spaces $\mathcal{X}$ and $\mathcal{Y}$. Let X be a random variable in $\mathcal{X}$ and suppose that there exists regular conditional distributions $\mathcal{L}(X_n \mid Y_n = y)$, $y \in \mathcal{Y}$, such that, as $n \to \infty$, for any convergent sequence $y_n \to y$ in $\mathcal{Y}$,*

$$(X_n \mid Y_n = y_n) \xrightarrow{\mathrm{d}} X. \tag{6.34}$$

Suppose further that $Y_n \xrightarrow{\mathrm{d}} Y$ as $n \to \infty$, for some random variable Y in $\mathcal{Y}$. Assume, as we may, that X and Y are independent. Then, as $n \to \infty$,

$$(X_n, Y_n) \xrightarrow{\mathrm{d}} (X, Y). \tag{6.35}$$

The assumption (6.34) means that $\mathcal{L}(X_n \mid Y_n = y_n)$ converges to the distribution $\mathcal{L}(X)$. Note that the limit distribution does not depend on y.

Proof. Let $A \subseteq \mathcal{X}$ and $B \subseteq \mathcal{Y}$ be such that $\mathbb{P}[X \in \partial A] = 0$ and $\mathbb{P}[Y \in \partial B] = 0$ (i.e., these are continuity sets for X and Y). We have

$$\mathbb{P}[(X_n, Y_n) \in A \times B] = \mathbb{E}\left[\mathbb{P}[X_n \in A \mid Y_n] \mathbf{1}\{Y_n \in B\}\right], \tag{6.36}$$

where we use the regular conditional distributions in the assumption. By the Skorohod coupling theorem [29, Theorem 4.30], we may assume that $Y_n \to Y$ almost surely. Then (6.34) implies $\mathbb{P}[X_n \in A \mid Y_n] \to \mathbb{P}[X \in A]$ a.s. (since A is an X-continuity set), and $Y_n \to Y$ implies $\mathbf{1}\{Y_n \in B\} \to \mathbf{1}\{Y \in B\}$ a.s. (since B is a Y-continuity set). Consequently, (6.36) implies by the dominated convergence theorem that

$$\begin{aligned}\mathbb{P}[(X_n, Y_n) \in A \times B] &\to \mathbb{E}\left[\mathbb{P}(X \in A) \mathbf{1}\{Y \in B\}\right] = \mathbb{P}[X \in A]\,\mathbb{P}[Y \in B] \\ &= \mathbb{P}[(X, Y) \in A \times B]. \end{aligned}\tag{6.37}$$

This implies (6.35) by [8, Theorem 3.1]. □

7. Dynamical aspects

7.1. **Small counts.** The number of empty boxes decreases each time a box receives its first ball. Functional limit theorems for this process in the setting of the discrete-time occupancy scheme were first obtained by Sevastyanov [44] through asymptotic analysis of the multivariate p.g.f. of the finite-dimensional distributions. For the regime of interest here, Theorem 5 of the cited paper showed (in a minor disguise) the convergence to an exponential Poisson process (see also [36, Ch. 4 Section 5]). In the context of CCP an equivalent result was proved quite recently by another method for the process of first arrivals (see [39, Theorem 4.3.38] and references therein), although the connection with [44] was apparently overlooked. Ilienko [18, Theorem 3.1] used the poissonised scheme to identify the Poisson limits for the processes of r-th arrivals.

We aim next to demonstrate, in the framework of the bi-poissonised occupancy scheme, how the time evolution of small counts connects to the processes of rth arrivals.

Recall from Proposition 5.1 that the pre-limit processes $\widehat{\mathrm{R}}_{n,r}$ of r-arrivals ($r \geqslant 1$) are nonhomogeneous Poisson point processes with intensity measure $np_{r-1}(t)\mathrm{d}t,\ t \geqslant 0$. Let

$$\alpha_{n,r} := L + (r-1)\log L - \log(r-1)! \tag{7.1}$$

which is (4.6) with $m = r-1$. Employing (2.3) and (4.7) to control the intensity measure, we obtain for every fixed $r \geqslant 1$ convergence of Poisson point processes in the form of their counting functions

$$(\widehat{\mathrm{R}}_{n,r}(\alpha_{n,r}+s,\infty],\ s \in \mathbb{R}) \xrightarrow{\mathrm{d}} (\Xi(s,\infty],\ s \in \mathbb{R}). \tag{7.2}$$

In fact, it is easy to see that on every fixed interval $(a,\infty]$, the intensities converge in L^1, and thus the intensity measures in total variation, and (7.2) follows. Taking the de-poissonisation of the numbers of balls and boxes for granted, (7.2) recovers the cited results from [18; 39].

We stress that the r-arrival processes have no common asymptotic time scale, meaning that $\widehat{\mathrm{R}}_{n,r'}(\alpha_{n,r}+s,\infty]$ for $r' < r$ converges in probability to 0, and for $r' > r$ converges in probability to ∞. These relations are just features of the right $(r-1)$-domain in terms of [36].

Sevastyanov's result on convergence of the process of empty boxes is equivalent to the $r = 1$ instance of (7.2) by the virtue of identity

$$N - \widehat{\mu}_{n,0}(t) = \widehat{\mathrm{R}}_{n,1}(t), \tag{7.3}$$

which holds pathwise a.s. for all $t \geqslant 0$, and just says that the number of empty boxes decreases each time some box receives its first ball. In this formula $N \overset{\mathrm{d}}{=} \mathrm{Poisson}(n)$, $\widehat{\mu}_{n,0}(t) \overset{\mathrm{d}}{=} \mathrm{Poisson}(ne^{-t})$, with $N - \widehat{\mu}_{n,0}(t)$ and $\widehat{\mu}_{n,0}(t)$ being independent for each fixed t. The number of empty boxes $(\widehat{\mu}_{n,0}(t),\ t \geqslant 0)$ is a pure-death process with unit death rate per capita, and $(\widehat{\mathrm{R}}_{n,1}(t),\ t \geqslant 0)$ is a Poisson process with the exponential jump rate ne^{-t}.

To generalise for all $r \geqslant 1$, we observe that

$$\widehat{\mu}_{n,r-1}(t) + \sum_{k=0}^{r-2} \widehat{\mu}_{n,k}(t) = \widehat{\mathrm{R}}_{n,r}(t,\infty],\ t \geqslant 0. \tag{7.4}$$

The term $\widehat{\mu}_{n,r-1}(t)$ is placed separately to emphasise that the remaining sum is asymptotically negligible for $t = \alpha_{n,r} + O(1)$ (which corresponds to the right $(r-1)$-domain of [36]). We center and use (7.2), which, using that the separated term in (7.4) dominates the rest, yields (again in the Skorohod space $D(-\infty,\infty]$):

Proposition 7.1. *For $r \in \mathbb{Z}_{>0}$, as $n \to \infty$,*

$$(\widehat{\mu}_{n,r-1}(\alpha_{n,r}+s),\ s \in \mathbb{R}) \xrightarrow{\mathrm{d}} (\Xi(s,\infty],\ s \in \mathbb{R}). \tag{7.5}$$

Notably, although $(\widehat{\mu}_{n,r}(t),\ t \geqslant 0)$ for $r > 0$ is a nonmonotonic birth-death process with upward jumps occuring at rate $np_{r-1}(t)$ and downward at rate one per capita, in the limit there are only downward jumps at times of a nonhomogeneous Poisson process. This feature of the occupancy scheme in the right $(r-1)$-domain is new to our knowledge.

7.2. **Parallels with queueing theory.** For a more comprehensive picture of the dynamics, the bi-poissonised model should be considered as a whole. Representing the occupancy counts as a random measure on the integer lattice, in the form

$$\sum_{r=0}^{\infty} \widehat{\mu}_{n,r}(t)\delta_r = \sum_{i=1}^{N} \delta_{\widehat{M}_{n,i}}, \tag{7.6}$$

and letting the time $t \geqslant 0$ vary defines a Markov process with values in the space of point measures $\mathcal{N}(-\infty, \infty]$, with the random starting state $N\delta_0$ (compare with $n\delta_0$ for the fixed-n poissonised scheme) and transitions driven by a Poisson flow

$$\sum_{r=0}^{\infty} \widehat{\mu}_{n,r}(t+h)\delta_r \stackrel{\mathrm{d}}{=} T_h \circ \sum_{r=0}^{\infty} \widehat{\mu}_{n,r}(t)\delta_r. \tag{7.7}$$

In intuitive terms, we may think of a box as a particle jumping at unit probability rate, independently of other particles, from one site on the lattice to the next on its right, with each such move caused by a new ball added to the box. Site 0 has a distinguished role as a source emitting particles that proceed to enter site 1 at epochs of the Poisson process $\widehat{\mathrm{R}}_{n,1}$ with nonhomogeneous rate ne^{-t}. The configuration on $\mathbb{Z}_{>0}$ (representing the multiplicities of nonzero box counts) evolves like an infinite network of infinite-server queues connected in series [38] (also known as a tandem of $\mathrm{M}_t/\mathrm{M}/\infty$ queues). In this context, the facts like independence of multiplicities in (5.4) and features of the arrival processes in Proposition 5.1 appear as specialisation of properties of open networks with Poisson inputs [32]. Other way round, our large n results on the occupancy problem for times $t = \alpha_{n,r} + O(1)$ (as well as results for other temporal regimes [36]) admit transparent interpretation in terms of a heavy-traffic approximation for the series of infinite-server queues with exponential input rate ne^{-t}.

A similar representation has been used to study a continuous-time growth process of random permutations [16]. In that model, like in other combinatorial processes related to tandem networks, the focus is on convergence to stationarity of the configuration of particles on any fixed finite set of sites, see [3] for examples.

8. Asymptotic independence of extreme occupancy counts

8.1. **Asymptotic independence.** The bi-poissonised multiplicities $\widehat{\mu}_{n,r}(t)$'s for different r are only independent for every fixed t (and n) but not as processes: e.g. the larger the number of boxes with r balls, the higher is the likelihood that the number of boxes with $r+1$ balls will increase in the nearest time. Nevertheless, Theorem 5.6 suggests that the independence of Ξ and τ stems from some kind of weak dependence of a vector of K maximal and a vector of K minimal occupancy counts for times $t = \alpha_n + O(1)$. With the latter shown, we will be in position to approximate the joint distribution of the measure-valued process $(\widehat{\mathrm{M}}_n^s,\ s \in \mathbb{R})$ taken together with $\widehat{\tau}_n - \alpha_n$, thus putting (5.26) (through the bi-poissonised version of the result) in light of the theorem on continuity of compositions [46].

To introduce the appropriate independence concept formally, consider a bivariate sequence of random elements (X_n, Y_n) taking values in a product Polish space $\mathcal{X} \times \mathcal{Y}$. We say that X_n and Y_n are *asymptotically independent* if for independent X_n' and Y_n' with $X_n' \stackrel{\mathrm{d}}{=} X_n$, $Y_n' \stackrel{\mathrm{d}}{=} Y_n$ it holds that

$$\lim_{n\to\infty} d_{\mathrm{TV}}((X_n, Y_n), (X_n', Y_n')) = 0. \tag{8.1}$$

(Then every subsequential weak limit will be a product measure.) See [10, Condition AI-4]) for this and weaker forms of asymptotic independence.

We are interested in times around the instant $L = \log n$. The covering interval with endpoints

$$t_0 := \frac{2}{3}L, \qquad t_1 := L + (K-1)\log L \tag{8.2}$$

will serve our purpose. The processes in the next lemma are to be considered as random elements of the space of cadlag functions endowed with the Skorohod topology.

Lemma 8.1. *For $K > 1$ and t_0, t_1 given by* (8.2) *the K-variate extreme-value processes*

$$((\widehat{M}_{n,n-i+1}(t))_{i=1}^K,\ t \in [t_0, t_1]) \text{ and } ((\widehat{M}_{n,i}(t))_{i=1}^K,\ t \in [t_0, t_1]) \tag{8.3}$$

are asymptotically independent.

Proof. For the time being let us regard boxes with at most K balls as 'small' and the others as 'big'. Accordingly, we split the sequence of multiplicities into two blocks

$$S(t) = (\hat{\mu}_{n,r}(t))_{r=0}^K, \qquad B(t) = (\hat{\mu}_{n,r}(t))_{r=K+1}^\infty, \tag{8.4}$$

which for every fixed t are independent, and stem from two complementary collections of boxes.

The block $S(t)$ of small box multiplicities is a Markov process, whose lifetime until absorption at zero is $L + K \log L + O_{\mathrm{p}}(1)$, in consequence of the discussion around the centering constant (4.6) (now with K assuming the role of m). The lifetime exceeds t_1 by $\log L + O_{\mathrm{p}}(1)$, hence at time t_1 w.h.p. the number of small boxes is at least K and, by monotonicity, the K minimal occupancy counts for all $t \leqslant t_1$ are due to small boxes. So we are reduced to show that the process of small box multiplicities on $[t_0, t_1]$ is asymptotically independent of the K maximal box occupancy counts.

To that end, for times $t \geqslant t_0$ we further decompose the process of big boxes as

$$B(t) = B'(t) + B''(t), \tag{8.5}$$

where $B'(t)$ is the sequence of multiplicities representing occupancy counts of those big boxes that contained more than K balls already at time t_0, and $B''(t)$ appears due to the increase of the content of small boxes. Independence of the blocks at t_0 and the Poisson flow dynamics entail that the processes $(S(t),\ t \geqslant t_0)$ and $(B'(t),\ t \geqslant t_0)$ are independent. It remains to show that for the range $t \in [t_0, t_1]$, the nonzero multiplicities in $B(t)$ that account for the K maximal box occupancy counts coincide w.h.p. with their counterparts in $B'(t)$. That is to say, we assert that boxes small at time t_0 are unlikely to overtake the largest ones at later stages up to time t_1.

Indeed, by Theorem 5.3 (with $m = 0$) and (5.16), for every $\varepsilon > 0$ w.h.p.

$$(e - \varepsilon)L < \widehat{M}_{n,K}(L) \leqslant \widehat{M}_{n,1}(L) < (e + \varepsilon)L, \tag{8.6}$$

where we recall $L = \log n$. Hence, by monotonicity for $t \geqslant t_0$, also

$$\frac{2}{3}(e - \varepsilon)L < \widehat{M}_{n^{2/3},K}(t_0) \leqslant \widehat{M}_{n,K}(t_0) \leqslant \widehat{M}_{n,K}(t). \tag{8.7}$$

On the other hand, the total number $\sum_{r=0}^K \hat{\mu}_{n,r}(t_0)$ of small boxes existing at time t_0 has a Poisson distribution with mean

$$nP_K(t_0) \leqslant ne^{-t_0} t_0^K < n^{1/3} L^K. \tag{8.8}$$

Hence it satisfies $\sum_{r=0}^K \hat{\mu}_{n,r}(t_0) < n^{1/2}$ w.h.p., which implies that the maximum number of balls any of these boxes can contain at a later time t (i.e., the index of the largest nonzero component of $B''(t)$) for $t_0 \leqslant t \leqslant t_1 < t_0 + L/2$ does not exceed $K + J$, where $J \overset{\mathrm{d}}{=} \widehat{M}_{n^{1/2},1}(L/2)$. By (8.6)

$$\widehat{M}_{n^{1/2},1}(L/2) < (e + \varepsilon)\frac{L}{2} < \frac{2}{3}(e - \varepsilon)L \tag{8.9}$$

w.h.p. for $\varepsilon < e/7$. Comparing with (8.7) yields $K + J < \widehat{M}_{n,K}(t_0)$ w.h.p., which shows the claim above that w.h.p. for $t \in [t_0, t_1]$ the K largest box occupancy counts are not represented. by $B''(t)$. Thus, in (8.3) the K-variate minimal process coincides w.h.p. with the K minimal counts contributing to S, and the maximal process coincides w.h.p. with the K maximal counts contributing to B', where S and B' are independent. The proof is completed by appealing to [10, Proposition 3] which ensures the asserted asymptotic independence. $\square$

Corollary 8.2. $\widehat{\tau}_n$ and $((\widehat{M}_{n,i}(t))_{i=1}^K,\ t \geqslant 0)$ *are asymptotically independent for each* $K \in \mathbb{Z}_{>0}$.

Proof. We may assume $K \geqslant m+2$ to have w.h.p. $t_0 < \widehat{\tau}_n < t_1$, for the bounds defined in (8.2). The truncated stopping time $\tau_n' = (\widehat{\tau}_n \vee t_0) \wedge t_1$ is adapted to the minimal process in (8.3), hence asymptotically independent of the maximal process. Since $\tau_n' = \widehat{\tau}_n$ w.h.p. we can apply [10, Proposition 3] again. $\square$

8.2. **Stopped maxima via continuity of compositions.** The value of a random process at a random time is sometimes referred to as composition. The continuity of compositions theorems connect convergence of such evaluations with the convergence of underlying processes and times. We sketch the ingredients needed for an alternative proof of Theorem 5.6 following this thread.

Firstly, we have observed a weak convergence of $\tau_n - \alpha_n$ to some random variable τ. As a next step, with a minor extra effort Theorem 5.3 extends to a functional approximation result in the sense of Lemma 5.5. In particular, for n running along a subsequence of integers with $c_n \to c_0 \in [0,1]$, the functional convergence

$$(\mathrm{M}_n^s, s \in \mathbb{R}) \xrightarrow{\mathrm{d}} (\Xi^{\uparrow}_{(e-1)s+c_0}, s \in \mathbb{R}), \tag{8.10}$$

follows from the marginal convergence for each fixed $s = s_0$ and the fact that both processes are driven by the same Poisson flow. Convergence (8.10) and the asymptotic independence in Corollary 8.2 allow one to control the joint distribution to show that

$$(\tau_n - \alpha_n, (\mathrm{M}_n^s,\ s \in \mathbb{R})) \xrightarrow{\mathrm{d}} \left(\tau, (\Xi^{\uparrow}_{(e-1)s+c_0},\ s \in \mathbb{R})\right), \tag{8.11}$$

where τ and the limit process are independent. The composition theorem from [46, Corollary 13.3.2, p. 433] now applies to yield convergence of the stopped point process

$$\mathrm{M}_n^{\tau_n - \alpha_n} \xrightarrow{\mathrm{d}} \Xi^{\uparrow}_{(e-1)\tau+c_0} \tag{8.12}$$

along the subsequence. Finally, the full extent of Theorem 5.6 with oscillatory asymptotics obtains by Lemma 5.5.

A version of the following formula for the joint distribution of stopped maximum and its multiplicity was stated in [21, Theorem 12] without proof.

Corollary 8.3. *For fixed* $k \in \mathbb{Z}$, $j \in \mathbb{Z}_{>0}$ *and* $M_n := M_{n,1}(\tau_n)$,

$$\mathbb{P}[M_n - b_n = k,\ \mu_{n,b_n+k}(\tau_n) = j] = \int_{-\infty}^{\infty} p_0\left(e^{(e-1)s+c_n-k}\right) p_j\left((e-1)e^{(e-1)s+c_n-k}\right) e^{-s} p_\ell(e^{-s})\mathrm{d}s + o(1). \tag{8.13}$$

Proof. We need to compute the analogous probability for the stopped approximating process $\Xi^{\uparrow}_{(e-1)\tau+c_n}$. Recalling the intensity measure (3.10)–(3.11) we obtain $\mathbb{P}[\Xi^{\uparrow}_b[k+1,\infty] = 0] = p_0(e^{b-k})$ and $\mathbb{P}[\Xi^{\uparrow}_b(\{k\}) = j] = p_j((e-1)e^{b-k})$, for the events which determine a j-fold rightmost atom at location k. Conditionally on $\tau = s$, we multiply these probabilities while setting $b = (e-1)s + c_n$, then integrate in s over the Gumbel$(\ell+1)$ density of τ given in (3.5). $\square$

9. Exponential tail estimates and moments

We proceed with uniform in (large) n exponential tail estimates for the maximal stopped occupancy counts in the fixed-n poissonised scheme. Apparently the underlying light-tail phenomenon has not been given due attention in the literature. We take therefore first a wider view on maximal order statistics, complementing the established theory found in [14; 42].

9.1. **The general setting.** Consider a sequence of distribution functions F_n on $\mathbb{R}_+$, and let $X_{n,1} \geqslant \cdots \geqslant X_{n,n}$ be an ordered i.i.d. sample from F_n. Suppose it is possible to choose an approximate upper $1/n$ quantile, i.e. to find x_n satisfying

$$c_- n^{-1} \leqslant \overline{F}_n(x_n) \leqslant c_+ n^{-1}, \tag{9.1}$$

where $\overline{F}_n := 1 - F_n$ and c_-, c_+ are some positive constants. If F_n is continuous then, of course, a $1/n$ quantile can be chosen exactly. Assuming that

$$\frac{\overline{F}_n(x+y)}{\overline{F}_n(x)} \leqslant Ce^{-cy} \quad \text{for} \quad \theta x_n \leqslant x \leqslant x_n \quad \text{and} \quad y \geqslant 0, \tag{9.2}$$

with some positive constants c, C, and $\theta \in (0,1)$, we wish to conclude on a similar tail estimate for the centered statistic $X_{n,i} - x_n$ with fixed i.

Lemma 9.1. *Under the assumptions* (9.1) *and* (9.2)*, for $y \geqslant 0$ and fixed $i \geqslant 1$,*

$$\mathbb{P}[X_{n,i} - x_n > y] \leqslant (c_+C)e^{-cy}, \tag{9.3}$$

$$\mathbb{P}[X_{n,i} - x_n \leqslant -y] \leqslant C_2 e^{-c_2 e^{c(1-\theta)y}}, \tag{9.4}$$

with some constant $C_2 > 0$ and $c_2 = c_-/(2C)$.

Proof. For the right tail estimate we only need the upper bound in (9.1) and that (9.2) holds with $x = x_n$. Granted that, we have for $y > 0$

$$\begin{aligned} \mathbb{P}[X_{n,i} - x_n > y] &\leqslant \mathbb{P}[X_{n,1} - x_n > y] = 1 - F_n^n(x_n + y) \leqslant n\overline{F}_n(x_n + y) \\ &\leqslant n\overline{F}_n(x_n)Ce^{-cy} \leqslant (c_+C)e^{-cy}. \end{aligned} \tag{9.5}$$

The left tail requires more effort. For $0 \leqslant y \leqslant x_n$ write the exact formula

$$\mathbb{P}[X_{n,i} - x_n \leqslant -y] = \sum_{j=0}^{i-1} \binom{n}{j} \overline{F}_n^j(x_n - y) F_n^{n-j}(x_n - y). \tag{9.6}$$

To bound this sum from the above we recall that the binomial distribution is stochastically increasing as the success probability increases; therefore it is enough to estimate $\overline{F}_n(x_n - y)$ from below. Inverting (9.2), we find

$$\frac{\overline{F}_n(x_n - y)}{\overline{F}_n(x_n)} \geqslant C^{-1}e^{cy}, \qquad 0 \leqslant y \leqslant (1-\theta)x_n. \tag{9.7}$$

Hence, noting for $0 \leqslant y \leqslant (1-\theta)x_n$ that $C^{-1}e^{cy}\overline{F}_n(x_n) \leqslant 1$ by (9.7), we obtain for the binomial sum in (9.6) an upper bound

$$\begin{aligned} &\sum_{j=0}^{i-1} \binom{n}{j} \bigl(C^{-1}e^{cy}\overline{F}_n(x_n)\bigr)^j \bigl(1 - C^{-1}e^{cy}\overline{F}_n(x_n)\bigr)^{n-j} \\ &\qquad \leqslant \sum_{j=0}^{i-1} \frac{1}{C^j j!} \bigl(n\overline{F}_n(x_n)\bigr)^j e^{cjy} \exp\bigl(-(n-j)\overline{F}_n(x_n)C^{-1}e^{cy}\bigr) \\ &\qquad \leqslant \sum_{j=0}^{i-1} \frac{c_1^j e^j}{j!} \exp\left(cjy - 2c_2e^{cy}\right) \leqslant C_1 \exp\left(ciy - 2c_2e^{cy}\right) \leqslant C_2 e^{-c_2e^{cy}} \end{aligned} \tag{9.8}$$

where $c_1 = c_+/C$, $c_2 = c_-/(2C)$, and C_1, C_2 are some constants; the final inequality holds since $\sup(az - e^{cz}) < \infty$ for every $a, c > 0$. This implies (9.4) for $0 \leqslant y \leqslant (1-\theta)x_n$. In particular, the bound (9.8) is valid for the cutoff $y = (1-\theta)x_n$. Therefore, for the remaining range $(1-\theta)x_n \leqslant y \leqslant x_n$ we have

$$\mathbb{P}[X_{n,i} - x_n \leqslant -y] \leqslant \mathbb{P}[X_{n,i} - x_n \leqslant -(1-\theta)x_n] \leqslant C_2 e^{-c_2 e^{c(1-\theta)x_n}} \leqslant C_2 e^{-c_2 e^{c(1-\theta)y}}. \tag{9.9}$$

Consequently, combining (9.8) and (9.9), we obtain (9.4). □

The striking asymmetry between the right and left tails is partly explained by a similar behaviour of the maximal point in the exponential process Ξ, which has the Gumbel distribution (3.6). Even so the above estimates do not presume approximability or convergence of $X_{n,i}$'s in distribution. Replacing the double exponent in (9.4) by a weaker exponential bound, we have

$$\mathbb{P}[|X_{n,i} - x_n| > y] \leqslant C_0 e^{-cy} \tag{9.10}$$

with suitable $C_0 > 0$. This two-sided estimate will be sufficient for our purposes, but see Lemma 9.2 below.

9.2. **Gamma and Poisson examples.** We illustrate the obtained tail bounds for maximal order statistics in two examples relevant to our stopped occupancy problem.

If F is Gamma$(m+1,1)$ with $m \geqslant 0$, then the hazard rate $h(x) = F'(x)/\overline{F}(x)$ is increasing to 1. From this, (9.2) holds in the form

$$\frac{\overline{F}(x+y)}{\overline{F}(x)} = \exp\left(-\int_x^{x+y} h(z)\mathrm{d}z\right) \leqslant e^{-h(x)y} \leqslant e^{-(1-\varepsilon)y}, \qquad \varepsilon > 0, \tag{9.11}$$

for large enough x.

For another example, suppose F_n is Poisson(t_n) with $t_n \sim L = \log n$. The Poisson distribution also has an increasing hazard rate (as being log-concave); thus for $x_n \sim eL$, and hence $t_n/x_n \to e^{-1}$, (2.4) gives

$$\frac{\overline{P}_{x_n+k}(t_n)}{\overline{P}_{x_n}(t_n)} \leqslant \left(\frac{\overline{P}_{x_n+1}(t_n)}{\overline{P}_{x_n}(t_n)}\right)^k \leqslant e^{-(1-\varepsilon)k}, \qquad k > 0, \tag{9.12}$$

for large enough n, similarly to the Gamma example above.

9.3. **Tail estimates for stopped maximal occupancy counts.** We combine the Gamma and Poisson bounds to obtain tail estimates for $M_{n,i}(\tau_n)$. The idea comes from the property of the Poisson distribution in Lemma 4.2, which tells us that for $t \sim L$ an increment u of the parameter is compensated by about $v = (e-1)u$ change of the quantile.

Throughout the subsection, k is a nonnegative integer; $C_1, C_2, \ldots$ and $c_1, c_2, \ldots$ are strictly positive constants that may disagree with those in Section 9.1.

Lemma 9.2. *For any fixed $i \geqslant 1$ and $\varepsilon > 0$, there exist constants $C_\cdot$ and $c_\cdot$ (that may depend on i and ε) such that for all $n \geqslant i$ and $k \geqslant 0$,*

$$\mathbb{P}[M_{n,i}(\tau_n) - b_n > k] \leqslant C_1 e^{-(1-\varepsilon)k/e}, \tag{9.13}$$

$$\mathbb{P}[M_{n,i}(\tau_n) - b_n \leqslant -k] \leqslant C_2 e^{-c_1 e^{c_2 k}} \leqslant C_3 e^{-k}. \tag{9.14}$$

Proof. The estimates (9.13)–(9.14) are more or less trivial for each fixed n, so we may assume that n is large when needed.

We consider first τ_n. Recall the realisation of τ_n as a maximal order statistic from Gamma$(m+1,1)$, and note that α_n is an approximate upper $1/n$ quantile, see (4.6)–(4.7). Taking $x_n = \alpha_n$, we see from (9.11) that (at least for large n) (9.2) holds with $c = 1-\varepsilon$ and $\theta = \varepsilon$. Hence, Lemma 9.1 yields

$$\mathbb{P}[\tau_n - \alpha_n > k/e] \leqslant C_4 e^{-(1-\varepsilon)k/e}, \tag{9.15}$$

$$\mathbb{P}[\tau_n - \alpha_n < -k/e] \leqslant C_5 e^{-c_3 e^{(1-2\varepsilon)k/e}}. \tag{9.16}$$

Consider now $M_{n,i}(\tau_n)$. For the right tail, the event $M_{n,i}(\tau_n) - b_n > k$ implies that either $\tau_n - \alpha_n > k/e$, or $\tau_n - \alpha_n \leqslant k/e$ and then $M_{n,i}(\alpha_n + k/e) - b_n > k$. Thus with the account of $M_{n,i} \leqslant M_{n,1}$,

$$\mathbb{P}[M_{n,i}(\tau_n) - b_n > k] \leqslant \mathbb{P}[\tau_n - \alpha_n > k/e] + \mathbb{P}[M_{n,1}(\alpha_n + k/e) - b_n > k]. \tag{9.17}$$

For the first part on the right-hand side, we apply (9.15). For the second part, we have a bound

$$\mathbb{P}[M_{n,1}(\alpha_n + k/e) - b_n > k] = 1 - (P_{b_n+k}(\alpha_n + k/e))^n \leqslant n\overline{P}_{b_n+k}(\alpha_n + k/e) \leqslant C_2 e^{-k/e}, \tag{9.18}$$

where the last inequality follows from Lemma 4.2 with $t = \alpha_n + k/e$ and $r = b_n + k$ and thus $u = k/e+O(1)$ and $v = k+O(1)$, by discarding some negligible or negative terms in (4.15)–(4.16). This proves (9.13).

For the left tail, the event $M_{n,i}(\tau_n)-b_n \leqslant -k$ implies that either $\tau_n-\alpha_n < -k/e$, or otherwise $\tau_n - \alpha_n \geqslant -k/e$ and then $M_{n,i}(\alpha_n - k/e) - b_n \leqslant M_{n,i}(\tau_n) - b_n \leqslant -k$. Splitting this way yields

$$\mathbb{P}[M_{n,i}(\tau_n) - b_n \leqslant -k] \leqslant \mathbb{P}[\tau_n - \alpha_n < -k/e] + \mathbb{P}[M_{n,i}(\alpha_n - k/e) - b_n \leqslant -k], \tag{9.19}$$

where the first part is estimated using (9.16). To bound the left tail of $M_{n,i}(\alpha_n - k/e)$, we only need to take care of k within the range $k \leqslant b_n$, since otherwise $\mathbb{P}[M_{n,i}(\alpha_n - k/e) - b_n \leqslant -k] = 0$. We consider first $k \leqslant (1-\varepsilon)b_n$ using Lemma 4.2 with $t = \alpha_n - k/e$ and $r = b_n - k$, and thus $u = -k/e + O(1)$ and $v = -k + O(1)$; note that in this range

$$\frac{u}{L} \geqslant -\frac{(1-\varepsilon)b_n}{eL} + o(1) \sim -(1-\varepsilon) \tag{9.20}$$

and thus (4.17) holds. The right-hand side of (4.16) becomes $-L + k/e + O(1)$, and thus (4.15) yields

$$\overline{P}_{b_n-k}(\alpha_n - k/e) \geqslant c_4 n^{-1} e^{k/e}. \tag{9.21}$$

The event $M_{n,i}(\alpha_n - k/e) - b_n \leqslant -k$ holds when less than i of the occupancy counts $\Pi_j(\alpha_n - k/e), j \in [n]$, are greater than $b_n - k$. We may thus argue as in (9.6) and (9.8) (with $\overline{F}_n(x_n - y)$ replaced by $\overline{P}_{b_n-k}(\alpha_n - k/e)$) and obtain from (9.21)

$$\mathbb{P}[M_{n,i}(\alpha_n - k/e) - b_n \leqslant -k] \leqslant C_6 e^{ik/e - c_4 e^{k/e}}. \tag{9.22}$$

The first inequality in (9.14) follows from (9.19), (9.16), and (9.22), under our assumption $k_n \leqslant (1-\varepsilon)b_n$. The remaining range $(1-\varepsilon)b_n \leqslant k \leqslant b_n$ is dealt with as in (9.9).

Finally, the second inequality in (9.14) is trivial. □

The proof shows (replacing ε by $\varepsilon/3$) that we may take c_2 as $(1-\varepsilon)/e$.

9.4. **Moments of the stopped maximal occupancy counts: approximability.** Complementing Theorem 5.6, we assert that the analogous result also holds for the mean and higher moments, in the natural sense.

Theorem 9.3. *For fixed $i, k \in \mathbb{Z}_{>0}$, as $n \to \infty$,*

$$\mathbb{E}[(M_{n,i}(\tau_n) - b_n)^k] = \mathbb{E}[\lceil \xi_i + (e-1)\tau + c_n \rceil^k] + o(1). \tag{9.23}$$

Proof. The exponential tail bounds in Lemma 9.2 imply that the sequence $(M_{n,i}(\tau_n) - b_n)^k$ is uniformly integrable. If the fractional parts c_{n_j} converge to some c_0 along a subsequence (n_j), then Theorem 5.6 ensures a weak convergence

$$M_{n_j,i}(\tau_{n_j}) - b_{n_j} \xrightarrow{\mathrm{d}} \lceil \xi_i + (e-1)\tau + c_0 \rceil, \tag{9.24}$$

which together with the uniform integrability imply the convergence of all moments along (n_j). The assertion now readily follows from the fact that every infinite set of positive integers contains such (n_j) (see Lemma 5.5). $\square$

9.5. **Computing the asymptotic moments.** Extending notation (4.26), denote the approximating variable

$$Z_{n,j} := \lceil \xi_j + (e-1)\tau + c_n \rceil, \tag{9.25}$$

where $\xi_j \overset{\mathrm{d}}{=} \mathrm{Gumbel}(j)$ and $\tau \overset{\mathrm{d}}{=} \mathrm{Gumbel}(\ell+1)$ are independent. These have the familiar characteristic functions, which are easily shown from e.g. the density (3.5),

$$\mathbb{E}[e^{\mathrm{i}x\xi_j}] = \frac{\Gamma(j-\mathrm{i}x)}{(j-1)!}, \qquad \mathbb{E}[e^{\mathrm{i}x\tau}] = \frac{\Gamma(\ell+1-\mathrm{i}x)}{\ell!} \tag{9.26}$$

and expected values

$$\mathbb{E}[\xi_j] = \gamma - H_{j-1}, \qquad \mathbb{E}[\tau] = \gamma - H_\ell, \tag{9.27}$$

where $\gamma \doteq 0.57721$ is the Euler constant and $H_k := \sum_{j=1}^k 1/j$ (so $H_0 := 0$).

To evaluate the mean of (9.25) we may apply [26, Theorem 2.3], which asserts that for a continuous random variable X with characteristic function φ

$$\mathbb{E}\lceil X \rceil = \mathbb{E}(X) + \frac{1}{2} + \sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{\varphi(2\pi k)}{2\pi\mathrm{i}k}, \tag{9.28}$$

provided $\varphi(x) = O(|x|^{-\varepsilon})$ for $x \to \pm\infty$. For (9.25) this condition is readily justified using the functional recursion

$$\Gamma(k-\mathrm{i}x) = (-\mathrm{i}x)_k\, \Gamma(-\mathrm{i}x), \tag{9.29}$$

where $(x)_\ell$ denotes the Pochhammer factorial, together with the reflection formula [40, 5.5.1 and 5.5.3], which yield (see also [40, 5.11.9]),

$$|\Gamma(\mathrm{i}x)|^2 = \frac{\pi}{x\sinh \pi x} \sim \frac{2\pi}{|x|}e^{-\pi|x|}, \quad x \to \pm\infty. \tag{9.30}$$

Applying (9.28) and (9.26)–(9.27), for $i \geqslant 1$ and $\ell \geqslant 0$,

$$\begin{aligned}\mathbb{E}[Z_{n,j}] &= \gamma e - H_{j-1} - (e-1)H_\ell + \frac{1}{2} + c_n \\ &\quad + \sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{\Gamma(j-2\pi\mathrm{i}k)\Gamma(\ell+1-2\pi(e-1)\mathrm{i}k)}{(j-1)!\,\ell!} \cdot \frac{e^{2\pi\mathrm{i}kc_n}}{2\pi\mathrm{i}k}.\end{aligned} \tag{9.31}$$

Consequently, (9.23) yields, recalling (4.12),

$$\begin{aligned}\mathbb{E}\lceil M_{n,j}\rceil &= b_n + \mathbb{E}\, Z_{n,j} + o(1) \\ &= a_n + \gamma e - H_{j-1} - (e-1)H_\ell + \frac{1}{2} \\ &\quad + \sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{\Gamma(j-2\pi\mathrm{i}k)\Gamma(\ell+1-2\pi(e-1)\mathrm{i}k)}{(j-1)!\,\ell!} \cdot \frac{e^{2\pi\mathrm{i}kc_n}}{2\pi\mathrm{i}k} + o(1).\end{aligned} \tag{9.32}$$

The sum in (9.31) and (9.32) is a Fourier series with small and rapidly decreasing coefficients, as is seen from (9.29)–(9.30).

For example, in the case $j = 1, \ell = 0$ (CCP and dixie cup problems) the coefficients have asymptotics

$$(2\pi k)^{-1}\big|\Gamma(1-(e-1)2\pi\mathrm{i}k)\Gamma(1-2\pi\mathrm{i}k)\big| = \frac{\pi(e-1)^{1/2}}{(\sinh(2\pi^2 k)\sinh(2(e-1)\pi^2 k))^{1/2}} \tag{9.33}$$

$$\sim 2\pi(e-1)^{1/2}e^{-e\pi^2 k}.$$

The terms with $k=\pm1$ are the largest by far and have the absolute value about $0.18379\cdot 10^{-10}$; hence the sum in (9.31) makes to the mean a small, oscillating on the $\log n$ scale contribution, whose amplitude does not exceed $4\cdot 10^{-11}$.

The Fourier coefficients increase with j and ℓ, though remain small. For instance, for $\ell=0$ the coefficient for $k=\pm1$ is bounded in absolute value, for all $j\geqslant 1$, by (since $|\mathbb{E}\,[e^{2\pi \mathrm{i}k\xi_j}]|\leqslant 1$)

$$(e-1)\big|\Gamma(-\mathrm{i}(e-1)2\pi)\big| \doteq 0.56552\cdot 10^{-7}. \tag{9.34}$$

Replacing τ in (9.25) by a constant gives random variables approximating (unstopped) maximal order statistics from a Poisson distribution, as in Theorem 5.22. In that setting the first Gamma factor in the sum (9.31) disappears, making the fluctuations somewhat larger. The intensity measure of our approximating process $\Xi^{\uparrow}$ has masses decreasing geometrically, so naturally the maxima in the occupancy scheme behave similarly to the maxima in samples from a geometric distribution (see for the latter [34] and [26, Example 4.3]). Making this comparison, it should be noted that in the occupancy regime of interest here the parameter of Poisson distribution (the mean number of balls) changes together with n (the number of boxes), but the parameter of the asymptotic geometric distribution is a fixed value $1-e^{-1}$ that does not depend on n (see (4.15)).

Similar Fourier series with small coefficients (typically involving a Gamma function) are also known from many different problems, see e.g. the examples in [41] and [26, Sections 2 and 4] and the references there. In the present situation, the terms in the sum in (9.32) contain a product of two Gamma functions, which makes the coefficients even smaller than in many other similar examples.

The technique from [26] may be further applied to obtain formulas for the variance and higher moments of $Z_{n,j}$.

10. Multiplicity of the maximum

Finally we consider the multiplicity of the stopped maximum occupancy count,

$$Q_n := \min\{i : M_{n,i} > M_{n,i+1}\} = \mu_{n,r^*}(\tau_n), \tag{10.1}$$

where $r^* = \max\{r : \mu_{n,r}(\tau_n) > 0\}$. The distribution of Q_n does not converge, because of oscillations, but we can obtain a good approximation by turning to its counterpart for a randomly shifted exponential process.

For a shift parameter $u\in\mathbb{R}$ let $\chi_j(u)$ be the probability that the rightmost atom of $\Xi^{\uparrow}_{(e-1)\tau+u}$ has multiplicity j.

Theorem 10.1. *As $n\to\infty$, for every fixed $j\in\mathbb{Z}_{>0}$,*

$$\mathbb{P}[Q_n=j] = \chi_j(a_n)+o(1) = \chi_j(c_n)+o(1), \tag{10.2}$$

where a_n and $c_n=\{a_n\}$ are given by (4.11)–(4.12), and χ_j introduced above is a continuous, 1-periodic function, representable by the Fourier series

$$\chi_j(u) = \frac{(1-e^{-1})^j}{j}\left(1+\sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{\Gamma(j-2\pi \mathrm{i}k)\Gamma(\ell+1-2\pi(e-1)k\mathrm{i})}{(j-1)!\ell!}e^{2\pi \mathrm{i}ku}\right). \tag{10.3}$$

Proof. The approximability (10.2) follows straight by Theorem 5.26 (or Corollary 8.3). The continuity and 1-periodicity of χ_j both follow from (3.8) and the exponential intensity (3.10), since for $c\in\mathbb{Z}$ the shift

$$\Xi^{\uparrow}_{b+c} \overset{\mathrm{d}}{=} \Xi^{\uparrow}_{b} + c \tag{10.4}$$

preserves the multiplicity of the maximum; in particular, $\chi_j(a_n) = \chi_j(c_n)$.

To prove (10.3) we start with evaluating a simpler probability, denoted $q_j(u)$, of the event that the rightmost atom of $\Xi_u^{\uparrow}$ has multiplicity $j \geqslant 1$. Arguing as in Corollary 8.3, and manipulating the Poisson probabilities,

$$\begin{aligned}
q_j(u) &= \sum_{k\in\mathbb{Z}} p_j((e-1)e^{u-k})p_0(e^{u-k}) \\
&= (1-e^{-1})^j \sum_{k\in\mathbb{Z}} p_0((e-1)e^{u-k})p_0(e^{u-k})\frac{e^{(u-k+1)j}}{j!} \\
&= (1-e^{-1})^j \sum_{k\in\mathbb{Z}} p_0(e^{u-k+1})\frac{e^{(u-k+1)j}}{j!} \\
&= (1-e^{-1})^j \sum_{k\in\mathbb{Z}} p_j(e^{u-k+1}) \\
&= (1-e^{-1})^j \sum_{k\in\mathbb{Z}} p_j(e^{u-k}). \qquad (10.5)
\end{aligned}$$

We cannot evaluate this sum explicitly, but it is easy to find its Fourier transform:

$$\begin{aligned}
\widehat{q}_j(k) &:= \int_0^1 e^{-2\pi\mathrm{i}ku} q_j(u)\,\mathrm{d}u \\
&= \frac{\left(1-e^{-1}\right)^j}{j!} \sum_{r\in\mathbb{Z}} \int_0^1 e^{j(u-r)}e^{-e^{u-r}}e^{-2\pi\mathrm{i}ku}\,\mathrm{d}u \\
&= \frac{\left(1-e^{-1}\right)^j}{j!} \int_{-\infty}^{\infty} e^{ju}e^{-e^{u}}e^{-2\pi\mathrm{i}ku}\,\mathrm{d}u \\
&= \frac{\left(1-e^{-1}\right)^j}{\ell!} \int_0^{\infty} e^{-v}v^{j-2\pi\mathrm{i}k-1}\,\mathrm{d}v \\
&= \frac{\left(1-e^{-1}\right)^j}{j!}\Gamma(j-2\pi\mathrm{i}k). \qquad (10.6)
\end{aligned}$$

The Fourier coefficient (10.6) appeared in [9] without proof (in Lemma 4.3 of that paper the value $\lambda = 1$ corresponds to the case of sampling from Geometric$(1-e^{-1})$), and was also identified in [34] by the calculus of residues.

We return to χ_j and note that, for $u \in \mathbb{R}$,

$$\chi_j(u) = \mathbb{E}\left[q_j((e-1)\tau + u)\right]. \qquad (10.7)$$

We calculate the Fourier coefficients again: for $k \in \mathbb{Z}$,

$$\begin{aligned}
\widehat{\chi}_j(k) &= \int_0^1 e^{-2\pi\mathrm{i}ku}\chi_j(u)\,\mathrm{d}u = \mathbb{E}\int_0^1 e^{-2\pi\mathrm{i}ku}q_j((e-1)\tau+u)\,\mathrm{d}u \\
&= \mathbb{E}\int_0^1 e^{-2\pi\mathrm{i}k(v-(e-1)\tau)}q_j(v)\,\mathrm{d}v = \widehat{q}_j(k)\cdot\mathbb{E}\left[e^{2\pi\mathrm{i}k(e-1)\tau}\right] \\
&= \frac{\left(1-e^{-1}\right)^j}{j!}\Gamma(j-2\pi k\mathrm{i})\frac{\Gamma(\ell+1-2\pi(e-1)k\mathrm{i})}{\ell!}, \qquad (10.8)
\end{aligned}$$

where by the change of variable we used 1-periodicity, and for the last step we used (10.6) and the characteristic function (9.26) of $\tau \overset{\mathrm{d}}{=} \mathrm{Gumbel}(\ell+1)$. This completes the proof of (10.3).

The last step of the proof can also be interpreted as follows: The formula (10.7) implies that χ_j is the convolution of q_j and the density function of $-(e-1)\tau$, and (10.8) then is the standard fact that the Fourier transform of the convolution of two functions is the product of their Fourier transforms. $\square$

The Fourier series in (10.3) is similar to (9.31). As there, the Fourier coefficients in (10.3) decrease rapidly as $|k|$ increases, and the sum is a very small oscillating term. For example, for $j=1$ and $\ell=0$, we have already for $k=\pm1$, similarly to (9.33),

$$\begin{aligned}\bigl|\widehat{\chi}_1(\pm1)\bigr| &= (1-e^{-1})\bigl|\Gamma(1\mp 2\pi\mathrm{i})\Gamma(1\mp 2\pi(e-1)\mathrm{i})\bigr| \\ &= (1-e^{-1})\Bigl(\frac{\pi\cdot 2\pi}{\sinh(2\pi^2)}\cdot\frac{\pi\cdot 2\pi(e-1)}{\sinh(2(e-1)\pi^2)}\Bigr)^{1/2} \\ &\approx 4\pi^2(e-1)^{3/2}e^{-e\pi^2-1} \doteq 0.730\cdot 10^{-10}. \end{aligned} \tag{10.9}$$

Hence, $\chi_1(u)$ varies about its 'mean' $1-e^{-1}$ with small amplitude of the order 10^{-10}. The oscillations are somewhat larger for larger j, but still small, and thus $\chi_j(u)$ is well approximated by its mean $(1-e^{-1})^j/j$. Hence, Theorem 10.1 implies that for large n, the distribution of Q_n is for practical purposes well approximated by the logarithmic distribution

$$\mathbb{P}[Q_n=j]\approx\frac{(1-e^{-1})^j}{j},\qquad j=1,2,\dots \tag{10.10}$$

In particular, the maximum is unique with probability close to $1-e^{-1}$.

As mentioned in the proof, q_j has appeared in connection with the multiplicity of the maximum in a sample from Geometric$(1-e^{-1})$. For this case Brands, Steutel and Wilms [9, Remark 2] observe that the fluctuations of q_1 are on the scale 10^{-4}. For the stopped maximum occupancy count these are smaller, again due to the smoothing resulting from the randomisation.

11. Numerics

In Table 1, we compare, for the CCP case $m=\ell=0$, the results of simulations of M_n with the expectation E_n of the approximation in (4.29). By (4.12) and (9.25), this approximation can be written as $b_n+Z_{n,1}$; thus, $E_n=b_n+\mathbb{E}[Z_{n,1}]$. Furthermore, by (9.31) and (9.33), E_n equals

$$a_n+\gamma e+\frac{1}{2} \tag{11.1}$$

within the precision used here; see also (9.32). Due to computational complexity, we used 100,000 Monte Carlo simulations for $n=10$ and $n=100$; 10,000 simulations for $n=10^3,10^4,10^5,150{,}000$; 1,000 simulations for $n=200{,}000; 250{,}000; 500{,}000$; and 300 simulations for $n=10^6$. For each considered value of n, we report the mean of the simulated value $\widetilde{M_n}$ of M_n, and the standard deviation of this mean, in columns three and four, respectively.

In Table 2, we compare, for various pairs (ℓ,m) and for $n=10^2,10^3,10^4,10^5$, the results of simulations of M_n with the expected value E_n from the approximation in (4.29). In a similar manner to Table 1, and using (9.31) and (9.33), the quantity E_n equals

$$a_n+\gamma e-(e-1)\sum_{j=1}^{\ell}\frac{1}{j}+\frac{1}{2} \tag{11.2}$$

within the level of precision used here; see also (9.32). The number of Monte Carlo simulations is abbreviated by MC and reported in the table. For each considered triple (n,ℓ,m), we report

Table 1

n	E_n	mean($\widetilde{M_n}$)	std(mean($\widetilde{M_n}$))
10	5.95083	5.71614	0.00677
100	11.86333	11.66791	0.00750
1,000	17.91967	17.76950	0.02410
10,000	24.03490	23.90320	0.02471
100,000	30.18241	30.10330	0.02494
150,000	31.26727	31.17440	0.02457
200,000	32.03735	32.09200	0.08153
250,000	32.63485	32.45600	0.07721
500,000	34.49189	34.44200	0.08071
1,000,000	36.35032	36.12333	0.15089

the mean of the simulated value $\widetilde{M_n}$ of M_n in the sixth column, and the standard deviation of this mean in the seventh column.

Table 2

n	MC	ℓ	m	E_n	mean($\widetilde{M_n}$)	std(mean($\widetilde{M_n}$))
100	10^5	0	1	14.48746	14.83162	0.00818
		0	2	15.92055	17.48246	0.00876
		0	3	16.65695	19.86110	0.00922
		5	0	7.93992	7.94814	0.00412
		10	0	6.83053	6.77567	0.00358
		25	0	5.30644	5.03896	0.00289
1,000	10^4	0	1	21.24050	21.48190	0.02576
		0	2	23.37030	24.58420	0.02748
		0	3	24.80341	27.27850	0.02844
		10	0	12.88688	12.88840	0.01240
		50	0	10.18877	9.94790	0.01108
		100	0	9.00629	8.54890	0.00926
10,000	10^3	0	1	27.85005	28.16300	0.08306
		0	3	32.40160	34.61200	0.08886
		0	5	34.88438	39.89800	0.09507
		10	0	19.00211	19.00500	0.04169
		100	0	15.12153	14.89000	0.03272
		1,000	0	11.17276	10.08900	0.02649
100,000	10^2	0	1	34.38098	34.72000	0.23956
		0	2	37.38853	38.31000	0.32495
		0	3	39.69937	41.60000	0.26967
		0	5	42.94900	47.39000	0.25776
		100	0	21.26903	21.23000	0.13015
		1,000	0	17.32026	16.73000	0.09832
		10,000	0	13.36454	11.47000	0.08343

Acknowledgements. The first author is indebted to Alexander Marynych for useful discussions and pointers to the literature. Special thanks go to Petr Akhmet'ev for locating and sending us a copy of [21].

Funding. The second author is funded by the Knut and Alice Wallenberg Foundation and the Swedish Research Council (VR). The third author is supported in part by BSF grant 2020063.

References

[1] Anderson, C. W. (1970). Extreme value theory for a class of discrete distributions, with applications to some stochastic processes, *J. Appl. Probab.* **7**, 99–113.

[2] Anderson, C.W., Coles, S.G. and Hüsler, J. (1997). Maxima of Poisson-like variables and related triangular arrays, *Ann. Appl. Probab.* **7**, 953–971.

[3] Baccelli, F., Borovkov, A. and Mairesse, J. (2000). Asymptotic results on infinite tandem queueing networks, *Probab. Theory Relat. Fields* **118**, 365–405.

[4] Barbour, A.D., Holst, L. and Janson, S. (1992). *Poisson Approximation*, Oxford University Press, Oxford.

[5] Baryshnikov, Y., Eisenberg, B. and Stengle, G. (1995) A necessary and sufficient condition for the existence of the limiting probability of a tie for first place, *Statist. Probab. Lett.* **23**, 203–209.

[6] Basrak, B. and Planinić, H. (2019). A note on vague convergence of measures, *Statistics and Probability Letters*, **153**, 180–186.

[7] Berestycki, N. (2009). *Recent Progress in Coalescent Theory*, Ensaios Matemáticos **16**, Sociedade Brasileira de Matemática, Rio de Janeiro.

[8] Billingsley, P. (1968). *Convergence of Probability Measures*, 1st ed., Wiley, New York.

[9] Brands, J.J.A.M., Steutel, F.W. and Wilms, R.J.G. (1994). On the number of maxima in a discrete sample, *Statistics and Probability Letters* **20**, 209–217.

[10] Davydov, Y.A. and Novikov. S.M. (2021). Remarks on asymptotic independence, *Theory Probab. Appl.* **66**, 44–57.

[11] De Moivre, A. (1718). *The Doctrine of Chances*, London.

[12] Derbazi, Z., Gnedin, A. and Marynych, A. (2024). Records in the infinite occupancy scheme, *ALEA, Lat. Am. J. Probab. Math. Stat.* **21**, 1475–1493.

[13] Dorogovtsev, A. (2024). *Measure-valued Processes and Stochastic Flows*, de Gruyter.

[14] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag.

[15] Gnedin, A., Hansen, B. and Pitman, J. (2007). Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws, *Probability Surveys* **4**, 146–171.

[16] Gnedin, A. and Stark, D. (2023). Random permutations and queues, *Advances in Applied Mathematics* **149**, no. 102549.

[17] Gumbel, E.J. (1935). Les valeurs extrêmes des distributions statistiques, *Annales Inst. H. Poincaré 5* **5** (2). 115–158.

[18] Ilienko, A. (2019). Convergence of point processes associated with coupon collector's and Dixie cup problems, *Electron. Commun. Probab.* **24** no. 51, 1–9.

[19] Ivchenko, G.I. (1971). On limit distributions of order statistics of the polynomial scheme, *Theory Probab. Appl.* **16** (1), 102–115.

[20] Ivchenko, G.I. (1971). Limit theorems in an occupancy problem, *Theory Probab. Appl.*, **16** (2), 293–307.

[21] Ivchenko, G.I. (1973). Waiting time and ordered series of frequencies in the polynomial scheme (Russian), *Tr. Mosk. Inst. Elektron. Mashinostr.* **32**, 39–64.

[22] Ivchenko, G.I. (1976). On some distributions connected with waiting time in a polynomial scheme, *Theory Probab. Appl.* **20** (3), 545–559.

[23] Ivchenko, G.I. (1993). The waiting time and related statistics in the multinomial scheme: a survey, *Discrete Math. Appl.* **3** (5), 451–482.

[24] Janson S., Łuczak T. and Ruciński A. (2000). *Random Graphs.* Wiley, New York.

[25] Janson, S. (2003). Cycles and unicyclic components in random graphs, *Combinatorics, Probability and Computing* **12**, 27–52.

[26] Janson, S. (2006). Rounding of continuous random variables and oscillatory asymptotics, *Annals of Probability* **34**, 1807–1826.

[27] Janson, S. (2010). Asymptotic equivalence and contiguity of some random graphs, *Random Structures Algorithms* **36** (1), 26–45.

[28] Johnson, N.L. and Kotz, S. (1977). *Urn Models and their Application*, Wiley.

[29] Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd ed., Springer-Verlag, New York.

[30] Kallenberg, O. (2007). *Random Measures, Theory and Applications.* Springer, Cham, Switzerland.

[31] Karlin, S. (1967). Central limit theorems for certain infinite urn schemes, *Journal of Mathematics and Mechanics* **17**, 373–401.

[32] Kelly, F.P. (2011). *Reversibility and Stochastic Networks*, Cambridge University Press.

[33] Kingman, J.F.K. (1993). *Poisson Processes*, Oxford Univ. Press, Oxford.

[34] Kirschenhofer, P. and Prodinger, H. (1993). A result in order statistics related to probabilistic counting, *Computing* **51**, 15–27.

[35] Kirschenhofer, P. and Prodinger, H. (1996). The number of winners in a discrete geometrically distributed sample, *Ann. Appl. Probab.* **6**, 687–694.

[36] Kolchin, V.F., Sevastyanov, B.A. and Chistyakov, V.P. (1978). *Random Allocations*, Halsted Press.

[37] Last, G. and Penrose, M. (2018). *Lectures on the Poisson Process*, Cambridge Univ. Press, Cambridge.

[38] Massey, W.A. and Whitt, W. (1993). Networks of infinite-server queues with nonstationary Poisson input, *Queueing Systems* **13**, 183–250.

[39] Mladenović, P. (2024). *Extreme Values in Random Sequences*, Springer.

[40] *NIST Handbook of Mathematical Functions* (2010). F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, editors. U.S. Department of Commerce, National Institute of Standards and Technology, Washington, DC; Cambridge University Press, Cambridge. Also available as *NIST Digital Library of Mathematical Functions*, `http://dlmf.nist.gov/`

[41] Prodinger, H. (2004). Periodic oscillations in the analysis of algorithms and their cancellations, *J. Iranian Statist. Soc.* **3** (2), 251–270.

[42] Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics*, Springer.

[43] Resnick, S. (2008). *Extreme Values, Regular Variation, and Point Processes*, Springer.

[44] Sevastyanov, B.A. (1967). Convergence of the distribution of the number of empty boxes to Gaussian and Poisson processes in the classical ball problem, *Probab. Th. Appl.* **12** (1), 144–154.

[45] Victorova, I.I. and Sevastyanov, B.A. (1967). On the limit behaviour of the maximum in the polynomial scheme, *Matem. Zametki* **1** (3), 331–338.

[46] Whitt, W. (2002). *Stochastic-Process Limits*, Springer.