



ON BERNOULLI TRIALS WITH UNEQUAL HARMONIC SUCCESS PROBABILITIES

Thierry Huillet, Martin Moehle

► To cite this version:

Thierry Huillet, Martin Moehle. ON BERNOULLI TRIALS WITH UNEQUAL HARMONIC SUCCESS PROBABILITIES. 2022. hal-03830538

HAL Id: hal-03830538

<https://hal.science/hal-03830538>

Preprint submitted on 26 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ON BERNOULLI TRIALS WITH UNEQUAL HARMONIC SUCCESS PROBABILITIES

THIERRY HUILLET AND MARTIN MÖHLE

ABSTRACT. A Bernoulli scheme with unequal harmonic success probabilities is investigated, together with some of its natural extensions. The study includes the number of successes over some time window, the times to (between) successive successes and the time to the first success. Large sample asymptotics, statistical parameter estimation, and relations to Sibuya distributions and Yule–Simon distributions are discussed. This toy model is relevant in several applications including reliability, species sampling problems, record values breaking and random walks with disasters.

1. INTRODUCTION

Introduce two weights $w_1 > 0$ and $w_2 \geq 0$, put $w := w_1 + w_2 > 0$, and let I_1, I_2, \dots be independent Bernoulli random variables with ‘harmonic’ success probabilities

$$(1) \quad \mathbf{P}(I_m = 1) := \frac{w_1}{w + m - 1}, \quad m \in \mathbb{N} := \{1, 2, \dots\},$$

decreasing inversely proportional to the number m of the trial. We note the following property of Bernoulli trials with such success probabilities: the first success time $K_1^+ := \inf\{m \in \mathbb{N} : I_m = 1\}$ is either a small units number or a very large one due to power-law tails of this random variable, see (19) below. In words indeed, if the I_m ’s fail to take the value 1 in the first steps, this tendency will be enhanced in the forthcoming steps resulting, for such models, in large (heavy-tailed) values of K_1^+ . So K_1^+ either will take small values close to 1 (the mode of K_1^+ is at 1 with probability mass w_1/w decreasing with w_2/w_1 if $w_2 > 0$) or very large values (responsible of its heavy-tailedness with tail index w_1): small values of w_2/w_1 favors early first success time while small values of w_1 favors late first success. So, the larger the number of steps for which no success was observed, the smaller the probability to see a success in the next step even though this probability is relatively large (harmonic decay in our case). This may be seen from the following argument:

Let $J_m := 1 - I_m$, $m \in \mathbb{N}$, and let $M_n = \prod_{m=1}^n J_m$. The event $M_n = 1$ is realized when no success was observed till time n . M_n is a multiplicative random walk

$$M_{n+1} = M_n J_{n+1}, \quad M_0 = 1,$$

for which the probability of a success at step $n + 1$ given no success till n is $\mathbf{P}(M_{n+1} = 0 \mid M_n = 1) = \mathbf{P}(I_{n+1} = 1)$. If $w_2 = 0$, then the probability

Date: April 14, 2022.

2020 Mathematics Subject Classification. Primary: 60J10; Secondary: 60C05.

Key words and phrases. Bernoulli variables, Ewens–Pitman sampling formula, Markov chains, Rényi’s records, Sibuya distribution, Stirling numbers, Yule–Simon distribution.

$\mathbf{P}(M_n = 1) = \prod_{m=1}^n \mathbf{P}(I_m = 0)$ of no success by time $n \in \mathbb{N}$ is obviously equal to 0, since $I_1 = 1$ in this case. If $w_2 > 0$ then this probability is equal to

$$\prod_{m=1}^n \mathbf{P}(I_m = 0) = \prod_{m=1}^n \frac{w_2 + m - 1}{w + m - 1} = \frac{\Gamma(w)\Gamma(w_2 + n)}{\Gamma(w_2)\Gamma(w + n)} \sim \frac{\Gamma(w)}{\Gamma(w_2)} \frac{1}{n^{w_1}}, \quad n \rightarrow \infty,$$

since $\Gamma(c + n) \sim n^c \Gamma(n)$ as $n \rightarrow \infty$ for any $c > 0$. Thus, the probability of no success by time n is small for large n , since $w_1 > 0$.

Examples of such enhancement mechanisms are

- $I_m = 1$ if some paper is cited the day m after its publication. Oversight.
- $I_m = 1$ if some new species is discovered the day m after a systematic daily sampling campaign. Rareness.
- $I_m = 1$ if some new word is used (or created) as the m -th word of some ongoing book. Scarcity.
- $I_m = 1$ if some individual renews its support to some political party the day (month) m after its creation. Weariness.
- Time unit increases by 1 when some athlete attempts to improve some record previously established. $I_m = 1$ if he/she succeeds at m -th trial: higher records become more and more difficult to break.

In several situations a success is actually a failure. Examples are

- $I_m = 1$ if some device breaks down the day m after it was put into service. Resilience.
- $I_m = 1$ if some population collapses the day m after it came to birth. Resilience.
- $I_m = 1$ if some patient contracts some illness the day m after birth date. Immunity.
- $I_m = 1$ if some driver has an accident the day m after obtaining its driving licence. Experience.

The number n of observations can be finite (possibly large though, depending on the time scale) or infinite. For instance, a typical driver only has finitely many driving days in his life (possibly randomly finite), but the attempts to break a record are potentially infinitely many.

For ‘harmonic’ Bernoulli sequences of the form (1) we study the number $S_n := \sum_{m=1}^n I_m$ of successes among the first $n \in \mathbb{N}_0$ trials, the time $K_l^+ := \inf\{m \in \mathbb{N} : S_m = l\}$ of the l -th success, $l \in \mathbb{N}_0$, and the times $L_l^+ := K_l^+ - K_{l-1}^+$ elapsed between successive successes, $l \in \mathbb{N}$, and analyse the associated Markov chains. It turns out that Sibuya distributions play an important role in this context. The two-parameter (w_1, w_2) -Sibuya distribution arises as the distribution of the waiting time till the first success. The shifted (w_1, w_2) -Sibuya distribution has many appealing properties, among them discrete self-decomposability and heavy-tailedness, [11]. It includes the ‘bare’ Sibuya distribution ($w_1 + w_2 = 1$, see [21]) and the Yule–Simon distribution ($w_2 = 1$, see [27]). The case $w_2 = 0$ is degenerate as far as the waiting time for the first success is concerned, but it appears to make sense from the point of view of the number of successes in the Ewens species sampling problem [4]. The case $(w_1, w_2) = (1, 0)$ also appears in the study of the number of record values

stemming from an arbitrary independent and identically distributed (iid) sequence of observations, see [13, 18, 19].

2. NUMBER OF SUCCESSES

In this section we are mainly interested in the number $S_n := \sum_{m=1}^n I_m$ of successes among the first $n \in \mathbb{N}_0$ trials. Note that $0 \leq S_n \leq n$ for $n \in \mathbb{N}_0$. In particular, $S_0 = 0$.

In the following, $s(n, k)$, $n, k \in \mathbb{N}_0 := \{0, 1, \dots\}$, denote the Stirling numbers of the first kind. Recall that the unsigned Stirling numbers of the first kind $|s(n, k)|$ are characterized via $[z]_n = \sum_{k \geq 0} |s(n, k)| z^k$, $z \in \mathbb{R}$, $n \in \mathbb{N}_0$, where $[z]_0 := 1$ and $[z]_n := z(z+1) \cdots (z+n-1)$, $n \in \mathbb{N}$. These numbers satisfy the recursion $|s_{n+1, k}| = n|s_{n, k}| + |s_{n, k-1}|$ with $|s_{n, k}| = 0$ for $k > n$, $|s_{n, n}| = 1$ and $|s_{n, 0}| = \delta_{n, 0}$ (Kronecker symbol).

2.1. The Markov chain $(S_n, n \in \mathbb{N}_0)$.

Clearly, $(S_n, n \in \mathbb{N}_0)$ is a time-inhomogeneous Markov chain with state-space \mathbb{N}_0 and transition probabilities

$$(2) \quad \mathbf{P}(S_{n+1} = k+1 \mid S_n = k) = 1 - \mathbf{P}(S_{n+1} = k \mid S_n = k) = \frac{w_1}{w+n}, \quad n, k \in \mathbb{N}_0.$$

Note that the probability (2) that the chain moves from state k at time n to state $k+1$ at time $n+1$ does not depend on the current state k . The increments $S_n - S_{n-1} = I_n$, $n \in \mathbb{N}$, are independent but not identically distributed. The chain $(S_n, n \in \mathbb{N}_0)$ also coincides with the chain studied in the restaurant process with a cocktail bar [12, Section 6.1] with parameters $(\alpha, \theta_1, \theta_2) := (0, w_1, w)$, where S_n counts the number of occupied tables after n customers have entered the restaurant. The probability-generating function (pgf) $z \mapsto f_n(z) := \mathbf{E}(z^{S_n})$ of S_n is given by

$$f_n(z) = \prod_{m=0}^{n-1} \frac{w_1 z + w_2 + m}{w + m} = \frac{[w_1 z + w_2]_n}{[w]_n} = \frac{[w_1(z-1) + w]_n}{[w]_n}, \quad z \in \mathbb{R},$$

Clearly, f_n is a polynomial of degree n of the form

$$f_n(z) = \frac{1}{[w]_n} \sum_{l=0}^n |s_{n, l}| (w_1 z + w_2)^l = \frac{1}{[w]_n} \sum_{k=0}^n z^k w_1^k \sum_{l=k}^n \binom{l}{k} |s_{n, l}| w_2^{l-k}.$$

Denoting by $[z^k] f_n(z)$ the coefficient in front of z^k of f_n yields

$$(3) \quad \mathbf{P}(S_n = k) = [z^k] f_n(z) = \frac{w_1^k}{[w]_n} \sum_{l=k}^n \binom{l}{k} |s_{n, l}| w_2^{l-k}, \quad k \in \{0, \dots, n\}.$$

With $(n)_0 := 1$ and $(n)_l := n(n-1) \cdots (n-l+1)$ for $l \in \mathbb{N}$, S_n has the l -th descending factorial moment

$$(4) \quad \mathbf{E}((S_n)_l) = l! [(z-1)^l] f_n(z) = \frac{w_1^l}{[w]_n} \sum_{k=l}^n (k)_l |s_{n, k}| w^{k-l}, \quad l \in \mathbb{N}_0.$$

Note that $\mathbf{E}((S_n)_l) = 0$ for $l > n$. The distribution $\pi_n(k) := \mathbf{P}(S_n = k)$ of S_n can be recursively computed via $\pi_0(k) = \delta_{k, 0}$ and

$$(5) \quad \pi_{n+1}(k) = \frac{w_1}{w+n} \pi_n(k-1) + \frac{w_2+n}{w+n} \pi_n(k), \quad n, k \in \mathbb{N}_0.$$

Note that $\pi_n(k) = 0$ for $k \notin \{0, \dots, n\}$. Comparing (5) with the recursion [7, Theorem 1] $s_r(n+1, k) = s_r(n, k-1) + (n+r)s_r(n, k)$ for the generalized Stirling numbers $s_r(n, k) := S(n, k; -1, 0, r)$, $n, k \in \mathbb{N}_0$, $r \in \mathbb{R}$, in the notation of [7] having vertical generating functions [7, Theorem 2] $k! \sum_{n \geq 0} s_r(n, k) t^n / n! = (1-t)^{-r} (-\log(1-t))^k$, $r \in \mathbb{R}$, $k \in \mathbb{N}_0$, $|t| < 1$, it follows that (3) can be alternatively expressed in terms of these generalized Stirling numbers as

$$(6) \quad \pi_n(k) = \frac{w_1^k}{[w]_n} s_{w_2}(n, k), \quad k \in \{0, \dots, n\},$$

in agreement with [12, Eq. (14)] for $(\alpha, \theta_1, \theta_2) := (0, w_1, w)$. Similarly, (4) can be written as

$$(7) \quad \mathbf{E}((S_n)_l) = \frac{w_1^l}{[w]_n} l! s_w(n, l), \quad l \in \mathbb{N}_0.$$

Introducing the superdiagonal stochastic transition matrices

$$\Pi_n := \begin{pmatrix} \frac{w_2+n}{w+n} & \frac{w_1}{w+n} & 0 & \cdots \\ 0 & \frac{w_2+n}{w+n} & \frac{w_1}{w+n} & 0 \\ 0 & 0 & \frac{w_2+n}{w+n} & \cdots \\ \vdots & \vdots & 0 & \ddots \end{pmatrix}, \quad n \in \mathbb{N}_0,$$

the distributions $\pi_n := (\pi_n(k), k \in \mathbb{N}_0)$ of S_n , $n \in \mathbb{N}_0$, satisfy the recursion $\pi_{n+1} = \pi_n \Pi_n$, $n \in \mathbb{N}_0$. Thus, $\pi_n = \pi_0 \prod_{m=0}^{n-1} \Pi_m$, $n \in \mathbb{N}_0$, with $\pi_0 = (1, 0, 0, \dots)$.

Remark 1. The pgf f_n of S_n has only real zeros $-(w_2+m)/w_1$, $m \in \{0, \dots, n-1\}$. By [15, Proposition 1], $(\pi_n(0), \dots, \pi_n(n))$ is a Pólya frequency sequence. Thus, the infinite matrix $M := (\pi_n(k-l))_{k,l \in \mathbb{N}_0}$ (where $\pi_n(k) = 0$ for $k \notin \{0, \dots, n\}$) is totally positive of any arbitrary order, i.e., all principal minors of any arbitrary order of M have nonnegative determinant.

2.2. Special cases.

- $w = 1$: $f_n(z) = \mathbf{E}(z^{S_n}) = \frac{[w_1(z-1)+1]_n}{n!} = \frac{1}{n!} \sum_{k=0}^n |s_{n+1,k+1}| w_1^k (z-1)^k$ showing that S_n has k -th descending factorial moment

$$\mathbf{E}((S_n)_k) = k! [(z-1)^k] f_n(z) = \frac{k!}{n!} |s_{n+1,k+1}| w_1^k, \quad k \in \mathbb{N}_0.$$

Note that in that case, necessarily $w_1 \in (0, 1)$.

- $w_2 = 1$: $f_n(z) = \mathbf{E}(z^{S_n}) = \frac{[w_1 z + 1]_n}{[w]_n} = \frac{[w_1 z]_{n+1}}{z[w_1]_{n+1}} = \frac{1}{[w_1]_{n+1}} \sum_{k=0}^{n+1} |s_{n+1,k}| w_1^k z^{k-1}$ showing that $(|s_{n,0}| = \delta_{n,0})$

$$\pi_n(k) = \mathbf{P}(S_n = k) = \frac{|s_{n+1,k+1}| w_1^{k+1}}{[w_1]_{n+1}}, \quad k \in \{0, \dots, n\}.$$

- $w_2 = 0$: $f_n(z) = \mathbf{E}(z^{S_n}) = \frac{[w_1 z]_n}{[w]_n} = \frac{1}{[w_1]_{n+1}} \sum_{k=0}^n |s_{n,k}| w_1^k z^k$ showing that

$$\pi_n(k) = \mathbf{P}(S_n = k) = \frac{|s_{n,k}| w_1^k}{[w_1]_n}, \quad k \in \{0, \dots, n\}.$$

If in addition $w_1 = 1$, then S_n is the number of record values of an arbitrary iid sequence of observations appearing before n ; [13, 18, 19]. In this case the law $\pi_n(k) = |s_{n,k}|/n!$, $k \in \{0, \dots, n\}$, of S_n coincides with the distribution of the number of cycles of a permutation of size n chosen uniformly at random.

2.3. Poisson approximation.

Clearly, $\mu_n := \mathbf{E}(S_n) = \sum_{m=1}^n \mathbf{P}(I_m = 1) = w_1 \sum_{m=0}^{n-1} 1/(w+m) = w_1 \log n + O(1)$ and

$$\sigma_n^2 := \text{Var}(S_n) = \sum_{m=1}^n \mathbf{P}(I_m = 0)\mathbf{P}(I_m = 1) = w_1 \sum_{m=0}^{n-1} \frac{w_2 + m}{(w+m)^2} \sim w_1 \log n$$

as $n \rightarrow \infty$. The law of S_n is in total variation distance close to the law of $N_n \stackrel{d}{\sim} \text{Poi}(\mu_n)$, (see [15] and [24]), because $\mu_n - \sigma_n^2 = w_1^2 \sum_{m=0}^{n-1} 1/(w+m)^2 \ll \mu_n$ (see [2, Theorems 1 and 2]) with LeCam Poisson approximation of the total variation distance $d_{TV}(S_n, N_n) := \frac{1}{2} \sum_{k \geq 0} |\pi_n(k) - \mu_n^k e^{-\mu_n} / k!|$ given by (see [20])

$$(8) \quad \frac{1}{32} \min(1, \mu_n^{-1})(\mu_n - \sigma_n^2) \leq d_{TV}(S_n, N_n) \leq (1 - e^{-\mu_n}) \frac{\mu_n - \sigma_n^2}{\mu_n}.$$

Therefore (and also by the Lindeberg–Feller central limit theorem), $(S_n - \mu_n)/\sigma_n \rightarrow \mathcal{N}(0, 1)$ in distribution as $n \rightarrow \infty$, consistently with the fact that $(N_n - \mu_n)/\sigma_n \rightarrow \mathcal{N}(0, 1)$ in distribution as $n \rightarrow \infty$. Since $\sum_{n \geq 2} \mathbf{P}(I_n = 0)\mathbf{P}(I_n = 1)/(\log n)^2 < \infty$, it follows from well-known law of large numbers results for sums of independent, but not identically distributed random variables, that $(\log n)^{-1} \sum_{m=1}^n (I_m - \mathbf{E}(I_m)) \rightarrow 0$ almost surely or, equivalently, that $S_n/\mu_n \rightarrow 1$ almost surely as $n \rightarrow \infty$.

2.4. Maximum likelihood estimation.

With $i := (i_1, \dots, i_n) \in \{0, 1\}^n$ an observed sequence of $I := (I_1, \dots, I_n)$ and $k := \sum_{m=1}^n i_m$,

$$(9) \quad \mathbf{P}(I = i) = \prod_{m=1}^n \frac{w_1^{i_m} (w_2 + m - 1)^{1-i_m}}{w + m - 1} = \frac{w_1^k \prod_{m=1}^n (w_2 + m - 1)^{1-i_m}}{[w]_n}.$$

This probability is not symmetric in i_1, \dots, i_n since the random variables I_1, \dots, I_n are not exchangeable. Using $\partial_w \log[w]_n = \sum_{m=0}^{n-1} 1/(w+m) = \Psi(w+n) - \Psi(w)$, where Ψ denotes the digamma function obeying $\Psi(z) = \log z - 1/(2z) + O(1/z^2)$ as $z \rightarrow \infty$, the two equations $\partial_{w_j} \log \mathbf{P}(I = i) = 0$, $j \in \{1, 2\}$, yield

$$(10) \quad \frac{k}{\hat{w}_1} = \sum_{m=0}^{n-1} \frac{1}{\hat{w} + m} = \Psi(\hat{w} + n) - \Psi(\hat{w})$$

and

$$(11) \quad \sum_{m=0}^{n-1} \frac{1 - i_{m+1}}{\hat{w}_2 + m} = \Psi(\hat{w} + n) - \Psi(\hat{w}),$$

where (\hat{w}_1, \hat{w}_2) is the maximum likelihood estimator (MLE) of (w_1, w_2) based on the observed sequence $i = (i_1, \dots, i_n)$ and $\hat{w} := \hat{w}_1 + \hat{w}_2$. It is easily checked that the 2×2 Hesse matrix J of the map $(w_1, w_2) \mapsto \mathbf{P}(I = i)$ is negative semi-definite at (\hat{w}_1, \hat{w}_2) . As $n \rightarrow \infty$, asymptotic normality of (\hat{w}_1, \hat{w}_2) is expected at rate $n^{-1/2}$, the limiting normal law having mean (w_1, w_2) and covariance matrix either the inverse of the expected Fisher information matrix or the inverse J^{-1} of the observed information matrix J evaluated at (\hat{w}_1, \hat{w}_2) .

If the model has two independent parameters (w_1, w_2) that have to be estimated, the first equation (10) gives $\hat{w}_1 = k/(\Psi(\hat{w} + n) - \Psi(\hat{w}))$ as a function of \hat{w} (and k) and so $\hat{w}_2 = \hat{w} - \hat{w}_1$ as a function of \hat{w} . Plugging this expression of \hat{w}_2 into the

second equation (11) yields an equation in the single variable \hat{w} that can be solved from the data $i = (i_1, \dots, i_n)$. An expression of both \hat{w}_1 and \hat{w}_2 then follows.

When $w = w_1 + w_2 = 1$, there is only one parameter to estimate, say w_1 , and (10) and (11) yield

$$\frac{k}{\hat{w}_1} = \sum_{m=1}^n \frac{1 - i_m}{m - \hat{w}_1} \quad (\text{entailing } \hat{w}_1 \in (0, 1)).$$

When $w_2 = 1$ or 0, only this first equation (10) is needed and the searched \hat{w}_1 solves

$$\frac{k}{\hat{w}_1} = \sum_{m=0}^{n-1} \frac{1}{\hat{w}_1 + m + 1} \quad \text{or} \quad \frac{k}{\hat{w}_1} = \sum_{m=0}^{n-1} \frac{1}{\hat{w}_1 + m}.$$

Note that $\Psi(\hat{w} + n) - \Psi(\hat{w}) \sim \log n$ as $n \rightarrow \infty$. Thus, by (10), $\hat{w}_1 \sim k/\log n$ as $n \rightarrow \infty$ and, by (11), a large n approximation for \hat{w}_2 is the solution w of the equation $\sum_{m=0}^{n-1} (1 - i_{m+1})/(w + m) = \log n$.

Remark 2. (Random number of observations) It can be natural to assume that the number of observations is finite but random (and independent of I_1, I_2, \dots). In this case one has to replace n by a random variable N taking values in \mathbb{N} , and $\mathbf{E}(z^{S_N}) = \sum_{n \geq 1} \mathbf{E}(z^{S_n}) \mathbf{P}(N = n)$ yields the law of the number of successes over the time window N with supposedly (or not) known mean $\mathbf{E}(N)$. For example, N could be geometrically distributed $\mathbf{P}(N = n) = p(1-p)^{n-1}$, $n \in \mathbb{N}$, with parameter $p \in (0, 1)$. For instance, it can be a good modeling issue to infer that there are only finitely many days in a species sampling campaign, geometrically distributed (without any further information but its mean number). The random variable S_N then counts the total number of sampled species over the observation window N .

3. TIMES TO SUCCESSIVE SUCCESSES

For $l \in \mathbb{N}$ let $K_l^+ := \inf\{n \in \mathbb{N} : S_n = l\}$ be the time elapsed till the l -th success. Furthermore, put $K_0^+ := 0$. The process $(K_l^+, l \in \mathbb{N}_0)$ is called the first-passage time process of the random walk $(S_n, n \in \mathbb{N}_0)$. Such processes have been studied extensively in the literature. We refer the reader to [3] and the references therein. We have $\mathbf{P}(K_l^+ > n) = \mathbf{P}(S_n < l)$ as the laws of (K_l^+, S_n) are mutual inverse in the sense of inverse sampling ([9, p. 192–194]). It follows from this, (8), and the works [18, 19] (see also [12, Proposition 1]), that

$$(12) \quad \frac{w_1 \log K_l^+}{l} \xrightarrow{\text{a.s.}} 1 \text{ as } l \rightarrow \infty \text{ and } \frac{w_1 \log K_l^+ - l}{\sqrt{l}} \xrightarrow{d} \mathcal{N}(0, 1) \text{ as } l \rightarrow \infty,$$

and the law of iterated logarithm for the $\log K_l^+$'s. And similarly for the time elapsed between contiguous successes, while replacing K_l^+ by $L_l^+ := K_l^+ - K_{l-1}^+$ in (12) with the notable exception that the first almost sure convergence is now a convergence in probability [13].

3.1. The laws of the times to successive successes and times elapsed between contiguous successes.

The law of K_l^+ is easily obtained as follows. Clearly, $\{K_l^+ = n\} = \{S_{n-1} = l-1, I_n = 1\}$. The independence of S_{n-1} and I_n thus yields

$$(13) \quad \mathbf{P}(K_l^+ = n) = \mathbf{P}(I_n = 1) \mathbf{P}(S_{n-1} = l-1) = \frac{w_1}{w + n - 1} \pi_{n-1}(l-1).$$

Using (3) the law of K_l^+ is therefore given by

$$(14) \quad \mathbf{P}(K_l^+ = n) = \frac{w_1^l}{[w]_n} \sum_{k=l-1}^{n-1} \binom{k}{l-1} |s_{n-1,k}| w_2^{k-l+1}, \quad n \geq l.$$

We also conclude that $L_{l+1}^+ := K_{l+1}^+ - K_l^+ = i$ is realized if and only, for some $n \geq l$: $S_{n-1} = l-1$ and I_n is a success and $S_{n+i-1} = l$ and I_{n+i} is a success. Hence, with $i \geq 1$,

$$(15) \quad \mathbf{P}(L_l^+ = i) = \sum_{n \geq l} \frac{w_1}{w+n-1} \frac{w_1}{w+n+i-1} \pi_{n-1}(l-1) \pi_{n+i-1}(l),$$

where $\pi_n(l)$ is given by (3). When $(w_1, w_2) = (1, 0)$, it follows from (3) in [13], developing problem 32 on p. 268 in [10], that

$$\mathbf{P}(L_l^+ > i) = \sum_{k=0}^i (-1)^k \binom{i}{k} (1+k)^{-l}.$$

The law of K_l^+ can be obtained on a computer by launching a three-term recursion. Indeed, from (13), the recursion (5) on $\pi_n(l)$ yields a recursion for $\mathbf{P}(K_l^+ = n)$ with $\mathbf{P}(K_l^+ = n) = 0$ if $n < l$. With $n \geq l$, this is

$$(16) \quad \mathbf{P}(K_{l+1}^+ = n+1) = \frac{w_1}{w+n} \mathbf{P}(K_l^+ = n) + \frac{w_2 + n - 1}{w+n} \mathbf{P}(K_{l+1}^+ = n).$$

Introducing the lower-triangular matrix $P = (P_{n,l})$, where $P_{n,l} := \mathbf{P}(K_l^+ = n)$, $l \leq n$, we see that $P_{n+1,l+1}$ can be obtained from its north-west and north neighbors. With the knowledge of the first column of P and its diagonal, this recursion becomes effective, starting from $P_{3,2}$ obtained from $P_{2,2}$ and $P_{2,1}$. For $n = l$, Eq. (16) reduces to $\mathbf{P}(K_{l+1}^+ = l+1) = (w_1/(w+l)) \mathbf{P}(K_l^+ = l)$, which yields the diagonal terms $\mathbf{P}(K_l^+ = l) = \prod_{m=0}^{l-1} w_1/(w+m)$. The entries $\mathbf{P}(K_1^+ = n)$ of the first column of P are given in (20) below.

3.2. Markov structure of $(K_l^+, l \in \mathbb{N})$.

The homogeneous Markov structure of the sequence $(K_l^+, l \in \mathbb{N})$ follows from

$$\mathbf{P}(K_{l+1}^+ - m > n \mid K_l^+ = m) = \prod_{k=0}^{n-1} \frac{w_2 + m + k}{w + m + k} = \frac{[w_2 + m]_n}{[w + m]_n} = \prod_{k=m}^{m+n-1} \frac{w_2 + k}{w + k},$$

where $m \geq l$ and $n > 0$. The random variable $L_{l+1}^+ := K_{l+1}^+ - K_l^+ \geq 1$ is the ‘time-lag’ elapsed between the l -th and the $(l+1)$ -th success. Its law depends on K_l^+ . It is thus expected that, for each $l \geq m$, the larger m is, the larger is $\mathbf{P}(K_{l+1}^+ - m > n \mid K_l^+ = m)$, because

$$\frac{\mathbf{P}(K_{l+1}^+ - (m+1) > n \mid K_l^+ = m+1)}{\mathbf{P}(K_{l+1}^+ - m > n \mid K_l^+ = m)} = \frac{w_2 + m + n}{w + m + n} \frac{w + m}{w_2 + m} > 1.$$

The chain $(K_l^+, l \in \mathbb{N})$ therefore obeys a sort of reinforcement property. For general information on random processes with reinforcement we refer the reader to [14].

From Stirling's formula, $\Gamma(z+b)/\Gamma(z+a) \sim z^{b-a}$ as $z \rightarrow \infty$. For fixed $m \ll n$, for each $m \geq l$, we indeed get

$$\mathbf{P}(K_{l+1}^+ - m > n \mid K_l^+ = m) = \frac{[w_2 + m]_n}{[w + m]_n} = \frac{\Gamma(w + m)}{\Gamma(w_2 + m)}(n^{-w_1} + O(n^{-(w_1+1)})),$$

translating that, given $K_l^+ = m$, the tails of L_{l+1}^+ have a tail index w_1 . Given the l -th record occurred at $m \ll n$, the waiting time till the $(l+1)$ -th has power-law tails with exponent w_1 . Note however that the probability that $K_{l+1}^+ - m = 1$ is $w_1/(w+m)$ which is small only if $m \gg 1$. Introducing $c_m := \Gamma(w+m)/\Gamma(w_2+m)$, for each $l \leq m$, $c_{m+1}/c_m = (w+m)/(w_2+m) > 1$ translating that the tails of L_{l+1}^+ get heavier as m increases, but without affecting the tail index itself, only the prefactor.

With $m' > m \geq l \geq 1$, we similarly get

$$\begin{aligned} \mathbf{P}(K_{l+1}^+ = m' \mid K_l^+ = m) &= \frac{w_1}{w + m' - 1} \prod_{n=m}^{m'-2} \frac{w_2 + n}{w + n}, \\ \mathbf{P}(K_{l+1}^+ = m') &= \sum_{m \geq l} \mathbf{P}(K_{l+1}^+ = m' \mid K_l^+ = m) \mathbf{P}(K_l^+ = m), \end{aligned}$$

with initial condition $\mathbf{P}(K_1^+ = m')$ given below in (20) if $w_2 \neq 0$. The homogeneous Markov structure of $(K_l^+, l \in \mathbb{N})$ appears more clearly, recalling $\mathbf{P}(K_l^+ = m)$ is given by (14). For $w_2 = 0$ the initial condition should start with $\mathbf{P}(K_2^+ = m')$ given in (23).

Introducing the excess time to l -th failure $K_l := K_l^+ - l \geq 0$, now a shifted random variable taking values in \mathbb{N}_0 ,

$$(17) \quad \mathbf{P}(K_{l+1} = n) = \frac{w_1}{w + n} \mathbf{P}(K_l = n) + \frac{w_2 + n}{w + n} \mathbf{P}(K_{l+1} = n - 1).$$

Replacing n by $n + l$ in (14) shows that the excess time K_l has distribution

$$\mathbf{P}(K_l = n) = \frac{w_1^l}{[w]_{n+l}} \sum_{k=l-1}^{n+l-1} \binom{k}{l-1} |s_{n+l-1,k}| w_2^{k-l+1}, \quad l \in \mathbb{N}, n \in \mathbb{N}_0.$$

Alternatively, the three-terms recursion (17) can be solved numerically using initially $\mathbf{P}(K_1 = n)$ and observing $\mathbf{P}(K_l = 0) = w_1^l/[w]_l$. Consequently, for all $n, n' \geq 0$,

$$(18) \quad \mathbf{P}(K_{l+1} > n' \mid K_l = n) = \prod_{m=n+1}^{2n+n'+1} \frac{w_2 + m}{w + m}$$

and, with

$$\mathbf{P}(K_{l+1} = n' \mid K_l = n) = \frac{w_1}{w + 2n + n' + 1} \prod_{m=n+1}^{2n+n'} \frac{w_2 + m}{w + m},$$

$$\mathbf{P}(K_{l+1} = k') = \sum_{k \geq 0} \mathbf{P}(K_{l+1} = k' \mid K_l = k) \mathbf{P}(K_l = k),$$

emphasizing the inhomogeneous Markov structure of $(K_l, l \in \mathbb{N})$ as well. Setting $n, n' = 0$ in (18), we get in particular $\mathbf{P}(K_{l+1} = 0 \mid K_l = 0) = w_1/(w+1)$. Note

that K_l also represents the number of failures till the observation of the l -th success, a generalized version of the negative binomial distribution.

4. TIME TO FIRST SUCCESS

If $l = 1$, with $K_0^+ := 0$, the distribution of the time to the first success reads ($L_1^+ = K_1^+ - K_0^+ = K_1^+$)

$$(19) \quad \mathbf{P}(K_1^+ > n) = [z^0] \mathbf{E}(z^{S_n}) = \frac{[w_2]_n}{[w]_n} \sim \frac{\Gamma(w)}{\Gamma(w_2)} n^{-w_1}, \quad n \rightarrow \infty.$$

and

$$(20) \quad \mathbf{P}(K_1^+ = n) = \frac{w_1}{w + n - 1} \frac{[w_2]_{n-1}}{[w]_{n-1}} = \frac{w_1}{w} \frac{[w_2]_{n-1}}{[w + 1]_{n-1}}, \quad n \in \mathbb{N}.$$

It is easily seen that the law of K_1^+ is unimodal with mode at $n = 1$ having mass w_1/w .

Upon shifting, $K_1 = K_1^+ - 1 \geq 0$ has a generalized (heavy tailed with index w_1) Sibuya distribution [11] with probability generating function (pgf)

$$(21) \quad \mathbf{E}(z^{K_1}) = \frac{w_1}{w} F(1, w_2; w + 1; z)$$

observing $(w + n)[w]_n = w[w + 1]_n$, where $F := {}_2F_1$ is the Gauss hypergeometric function $F(a, b; c; z) := \sum_{n \geq 0} ([a]_n [b]_n / [c]_n) (z^n / n!)$. The initial condition to the recursion (17) giving $\mathbf{P}(K_l = k)$ is

$$\mathbf{P}(K_1 = k) = \frac{w_1}{w + k} \frac{[w_2]_k}{[w]_k} = \frac{w_1}{w} \frac{[w_2]_k}{[w + 1]_k}, \quad k \in \mathbb{N}_0.$$

Remark 3. (Time to first success in a N -Bernoulli trial with N finite and Geometric (p)).

In that case, $\mathcal{K}_l^+ = \inf\{n \in \{1, \dots, N\} : S_n = l\}$ and $\mathcal{K}_1^+ = \inf\{m \in \{1, \dots, N\} : I_m = 1\}$. Therefore, $\mathcal{K}_1^+ = \infty$ with probability $\mathbf{P}(S_N = 0) = \mathbf{E}(\prod_{m=1}^N \mathbf{P}(I_m = 0))$ and $\mathbf{P}(\mathcal{K}_1^+ > n) = \mathbf{P}(S_n = 0 \mid N \geq n)$ with probability $\mathbf{P}(S_N > 0)$, where

$$\mathbf{P}(S_n = 0 \mid N \geq n) = \mathbf{P}(S_n = 0) = \frac{[w_2]_n}{[w]_n}.$$

So, if $w_2 > 0$, the new \mathcal{K}_1^+ has an atom at ∞ with mass

$$\mathbf{P}(S_N = 0) = q \sum_{n \geq 1} \frac{[w_2]_n}{[w]_n} p^{n-1} = \frac{q}{p} [F(1, w_2; w; p) - 1]$$

translating that no success was registered before N .

4.1. Special cases.

- Sibuya: $w = 1 \Rightarrow w_1, w_2 = 1 - w_1 \in (0, 1)$ with $\mathbf{E}(z^{K_1}) = w_1 F(1, 1 - w_1; 2; z) = z^{-1}(1 - (1 - z)^{w_1})$, equivalently, $\mathbf{P}(K_1 = k) = w_1 [1 - w_1]_k / (k + 1)!$, $k \geq 0$.

- Yule-Simon: $w_2 = 1, w_1 > 0$ with $\mathbf{E}(z^{K_1}) = \frac{w_1}{w_1 + 1} F(1, 1; w_1 + 2; z)$, equivalently, $\mathbf{P}(K_1 = k) = w_1 \frac{k!}{[w_1 + 1]_{k+1}}$.

- Ewens: $w_2 = 0, w_1 > 0$: this is a singular case for which $\mathbf{P}(K_1^+ = n) = \delta_{n,1}$.

In view of $F(a, b; c; z) = F(b, a; c; z)$, the Yule–Simon distribution with $a = b = 1$ and $c = w_1 + 2$ is the only one in the class (21) to be identifiable (different parameters yield different distributions).

4.2. Falling factorial moments of K_1 .

$K_1^+ = K_1 + 1$ is an important random variable if one considers that the first occurrence of a success may lead to a stop of some ongoing process.

With $a = 1$, $b = w_2$, $c = w + 1$, i integer, using the special values and differential identities

$$\begin{aligned} F(a, b; c; 1) &= \frac{[c - a]_a}{[c - a - b]_a} \\ \frac{d^i}{dz^i} F(a, b; c; z) &= \frac{[a]_i [b]_i}{[c]_i} F(a + i, b + i; c + i; z), \end{aligned}$$

evaluated at $z = 1$, with $(K_1)_i = K_1(K_1 - 1) \cdots (K_1 - i + 1)$, when $i < w_1$, we get the descending i -th factorial moments of K_1 as

$$\mathbf{E}[(K_1)_i] = \varphi^{(i)}(1) = \frac{i! [w_2]_i}{[w_1 - i]_i}, \quad i < w_1,$$

where $\varphi(z) := \mathbf{E}(z^{K_1}) = \frac{w_1}{w} F(1, w_2; w + 1; z)$. In particular, if $w_1 > 1$, $\mathbf{E}(K_1) = w_2/(w_1 - 1) < \infty$ and, if $w_1 > 2$,

$$\text{Var}(K_1) = \varphi''(1) + \varphi'(1) - (\varphi'(1))^2 = \frac{w_1(w_1 - 1)\mathbf{E}(K_1)}{(w_1 - 1)(w_1 - 2)} < \infty.$$

Overdispersion holds. The mean $\mathbf{E}(K_1^+) = \frac{w_1 - 1}{w_1 - 1} > 1$ and the variance $\text{Var}(K_1^+) = \text{Var}(K_1)$ of K_1^+ (if they exist) may be used to estimate (w_1, w_2) by the method of moments provided empirical values of these quantities are available.

4.3. MLE estimator of (w_1, w_2) from K_1^+ .

If we have an L -sample (n_1, \dots, n_L) for the time K_1^+ to first success,

$$\mathbf{P}(K_1^+(1) = n_1, \dots, K_1^+(L) = n_L) = w_1^L \prod_{l=1}^L \frac{1}{w + n_l - 1} \frac{[w_2]_{n_l - 1}}{[w]_{n_l - 1}}.$$

Considering $\partial_{w_k} \log \mathbf{P}(K_1^+(1) = n_1, \dots, K_1^+(L) = n_L) = 0$ for $k \in \{1, 2\}$ yields a MLE (\hat{w}_1, \hat{w}_2) for (w_1, w_2) based on the histogram of the observed time to first failure sample (n_1, \dots, n_L) . With $\hat{w} = \hat{w}_1 + \hat{w}_2$, we get

$$\frac{L}{\hat{w}_1} - \sum_{l=1}^L \frac{1}{\hat{w} + n_l - 1} - \sum_{l=1}^L (\Psi(\hat{w} + n_l - 1) - \Psi(\hat{w})) = 0$$

and

$$- \sum_{l=1}^L \frac{1}{\hat{w} + n_l - 1} - \sum_{l=1}^L (\Psi(\hat{w} + n_l - 1) - \Psi(\hat{w}) - \Psi(\hat{w}_2 + n_l - 1) + \Psi(\hat{w}_2)) = 0.$$

The first equation gives \hat{w}_1 as a function of \hat{w} (and the data) and so $\hat{w}_2 = \hat{w} - \hat{w}_1$ as a function of \hat{w} . Plugging this expression of \hat{w}_2 into the second equation yields an equation in the single variable \hat{w} that can be solved from the data. A separate

expression of both \hat{w}_1 and \hat{w}_2 then follows. Asymptotic normality of this estimator is proved in [11], together with an expression of the Fisher information matrix.

4.4. The Ewens case $w_2 = 0$.

In a sampling problem from a Poisson–Dirichlet partition $\text{PD}(\theta)$ of the unit interval modeling species abundances, the law of the number $S_n = \sum_{m=1}^n I_m$ of distinct sampled species for a size n uniform sample obeys (5), [4], [1] and [24], with $w_1 = \theta$, $w_2 = 0$ and $S_1 = 1$, corresponding to $K_1^+ = 1$. Because sampling is modeled as uniform throws on a partition of the unit interval, necessarily on day $n = 1$, a new species is sampled but new species with smaller abundance become increasingly unlikely to be subsequently sampled. The $\text{PD}(\theta)$ partition of the unit interval has countably many pieces, so the sampling process potentially never stops. Here K_l^+ ($l \geq 2$) is the sample size till l new species have been sampled with, from (14)

$$(22) \quad \mathbf{P}(K_l^+ = n) = w_1 [z^{l-1}] \frac{[w_1 z]_{n-1}}{[w_1]_n} = w_1^{l-1} \frac{|s_{n-1, l-1}|}{[w_1 + 1]_{n-1}}, \quad n \geq l.$$

This distribution seems to be new. Note the resulting ‘vertical’ identity for the $|s_{n, l}|$ ’s: $\sum_{n \geq l} \frac{|s_{n, l}|}{[w_1 + 1]_n} = w_1^{-l}$ for all $w_1 > 0$.

The random variable K_2^+ is the time to second non-trivial discovery of a new species (after $K_1^+ = 1$), with, recalling $|s_{n-1, 1}| = (n-2)!$,

$$(23) \quad \mathbf{P}(K_2^+ = n) = w_1 \frac{(n-2)!}{[w_1 + 1]_{n-1}}, \quad n \geq 2,$$

reducing to $\mathbf{P}(K_2^+ = n) = 1/(n(n-1))$ when $w_1 = 1$. With $K_2^+ - 1 \geq 0$ the time elapsed since $K_1^+ = 1$, we thus have

$$\mathbf{E}(z^{K_2^+ - 1}) = w_1 \int_0^z \frac{F(1, 1; w_1 + 1; t) - 1}{t} dt.$$

The above theory applies to this fundamental Ewens model. Given $S_n = k$, the probability to discover a new species at time $n + 1$ is $w_1/(w_1 + n)$, decreasing inversely proportional to n and independently of k . Recall from (10) that the MLE \hat{w}_1 for w_1 is characterized by $k/\hat{w}_1 = \Psi(\hat{w}_1 + n) - \Psi(\hat{w}_1)$ and hence only depends on $k = i_1 + \dots + i_n$. See [22, p. 41, Eq. (3.7.7)].

5. AN EXTENSION OF THE HARMONIC BERNOULLI TRIAL

With $\alpha > 0$, consider the inhomogeneous Bernoulli trial with $\mathbf{P}(I_m = 1) = w_1/(w + m^\alpha - 1)$, $m \in \mathbb{N}$.

For $\alpha \in (0, 1)$ the successful events are more frequent than for $\alpha = 1$. Then, $\mu_n := \mathbf{E}(S_n) = \sum_{m=1}^n \mathbf{P}(I_m = 1) = w_1 \sum_{m=0}^{n-1} 1/(w + m^\alpha) \sim \frac{w_1}{1-\alpha} n^{1-\alpha}$ as $n \rightarrow \infty$ and

$$\sigma_n^2 := \text{Var}(S_n) = \sum_{m=1}^n \mathbf{P}(I_m = 1)\mathbf{P}(I_m = 0) = w_1 \sum_{m=0}^{n-1} \frac{w_2 + m^\alpha}{(w + m^\alpha)^2} \sim \frac{w_1}{1-\alpha} n^{1-\alpha}$$

and the law of S_n is close in the sense of total variation distance to $P_n \stackrel{d}{\sim} \text{Poi}(\mu_n)$ for this new μ_n now growing algebraically with n .

Clearly also,

$$(24) \quad \frac{w_1(K_l^+)^{1-\alpha}}{l(1-\alpha)} \xrightarrow{\text{a.s.}} 1 \text{ as } l \rightarrow \infty \text{ and } \frac{w_1(K_l^+)^{1-\alpha}/(1-\alpha) - l}{\sqrt{l}} \xrightarrow{d} \mathcal{N}(0, 1) \text{ as } l \rightarrow \infty.$$

The time to the l -th success occurs much sooner than when $\alpha = 1$.

If $\alpha > 1$, then S_n converges in distribution to a Poisson random variable with finite mean $\mu_\infty := \lim_{n \rightarrow \infty} \mu_n = w_1 \sum_{m=0}^{\infty} 1/(w + m^\alpha)$.

6. A RELATED RANDOM WALK WITH DISASTERS

Bernoulli trials with unequal harmonic success probabilities are also relevant in the context of growth-collapse random walks with disasters. Discrete-time integral-valued growth-collapse processes where long periods of linear growth alternate with rare catastrophic events occur in a large variety of systems. A collapse or catastrophic event is when the size of some population shrinks by a random number of units, not exceeding the current system's size. A total disaster is when the size of the system shrinks instantaneously to zero (a massive extinction event). Disastrous growth-collapse models occur as models for population growth subject to rare catastrophic extinction events.

A one-parameter version of such discrete-time models was investigated in [8]. Here, holding probabilities were allowed (with some probability the system's size can be left unchanged) and pure reflection at the origin was assumed (once in state zero, the system's size grows by one unit with probability 1). Whenever zero is a reflection/absorption barrier, pomp periods will alternate with periods of scarcity. We herewith focus on disastrous growth-collapse models with no holding probability and with zero either standing for a reflection or an absorption barrier. The probabilities of either growth or disastrous events will be chosen to be dependent on the current state as in the Bernoulli model with harmonic success probabilities, and this will favor large populations in the long run.

With $\alpha > 0$, define $q_n := w_1/(w + n^\alpha)$ and $p_n := 1 - q_n$, $n \in \mathbb{N}_0$. With $(U_m, m \in \mathbb{N})$ an iid sequence of uniforms,

$$(25) \quad N_{m+1} := (N_m + 1)\mathbf{1}(U_{m+1} \leq p_{N_m}), \quad N_0 \geq 0,$$

defines a time-homogeneous Markov chain $(N_m, m \in \mathbb{N}_0)$ that moves from state n to state $n + 1$ with probability p_n or is sent from state n to state 0 with probability q_n (a disaster event).

The transition matrix of this Markov chain with state-space \mathbb{N}_0 is

$$P = \begin{pmatrix} q_0 & p_0 & & & \cdots \\ q_1 & 0 & p_1 & & \cdots \\ \vdots & \vdots & \ddots & \ddots & \cdots \\ q_n & 0 & \cdots & 0 & p_n & \cdots \\ \vdots & \vdots & & & \ddots & \ddots \end{pmatrix}.$$

Let us distinguish two cases.

Case 1. Assume that $w_2 = 0$. In this case state 0 is absorbing. Let $n \in \mathbb{N}$. The probability $\mathbf{P}(N_m \rightarrow \infty | N_0 = n) = \prod_{m \geq n} p_m$ is equal to 0 if and only if

$\sum_{m \geq n} q_m = \infty$, which in turn holds if and only if $\alpha \leq 1$. Thus, for $\alpha \leq 1$ the chain $(N_m, m \in \mathbb{N}_0)$, started from state $N_0 \equiv n$, will eventually go extinct. For $\alpha > 1$ the chain, started from state n , will tend to infinity with probability $\prod_{m \geq n} p_m > 0$ and go extinct with complementary probability $1 - \prod_{m \geq n} p_m$. The extinction time $\tau_{n,0} := \inf\{m \in \mathbb{N}_0 : N_m = 0, N_0 = n\}$ has pgf $\mathbf{E}(z^{\tau_{n,0}}) = \sum_{m \geq n} q_m z^m \prod_{k=n}^{m-1} p_k$, $|z| < 1$, and $\tau_{n,0}$ takes the value ∞ with probability $\prod_{m \geq n} p_m$ being strictly positive if and only if $\alpha > 1$.

Case 2. Assume that $w_2 > 0$. Then state 0 is reflecting and all states are communicating since $w_1 > 0$ by assumption. The chain $(N_m, m \in \mathbb{N}_0)$ is hence irreducible and obviously aperiodic. This is a small variation of a Markov chain whose salient statistical features were studied in [5]. From the study in [5] we conclude that:

- For $\alpha > 1$ the chain is transient. After a finite number of returns to 0 (excursions) the chain drifts to infinity.
- For $\alpha < 1$ the chain is positive recurrent with invariant probability measure $\pi_n = \pi_0 \prod_{k=0}^{n-1} p_k$, $n \in \mathbb{N}_0$, where the normalizing constant π_0 is determined by $\sum_{n=0}^{\infty} \pi_n = 1$.
- For $\alpha = 1$ (critical case) the chain is null-recurrent if $0 < w_1 \leq 1$ and positive recurrent if $w_1 > 1$. For the latter case $w_1 > 1$ the invariant probability measure is given by $\pi_n = \pi_0 [w_2]_n / [w]_n$, $n \in \mathbb{N}_0$, with normalizing constant $\pi_0 := (w_1 - 1) / (w - 1)$, having heavy tails with index $w_1 > 1$.

In the recurrent case ($\alpha \leq 1$) the sample paths of $(N_m, m \in \mathbb{N}_0)$, started at $N_0 = 0$, are made of iid excursions through state 0. The first excursion has length L_1^+ and height $L_1^+ - 1$, where $L_1^+ := \inf\{m \in \mathbb{N} : N_m = 0, N_0 = 0\}$ is the time elapsed till the first disaster. Clearly, in the positive recurrent case ($\alpha < 1$ or $\alpha = 1$ and $w_1 \leq 1$) the invariant probability measure has the general form $\pi_n = \mathbf{P}(L_1^+ > n) / \mathbf{E}(L_1^+)$, $n \in \mathbb{N}_0$.

With $(L_i^+ - 1, i \in \mathbb{N})$ iid copies of the first excursion height $L_1^+ - 1$, of interest for the control of overcrowding are the random variables

$$T_1(n) := \inf\{m \in \mathbb{N} : N_m > n \mid N_0 = n_0\} \text{ and } \inf\{i \in \mathbb{N} : \max_{j \in \{1, \dots, i\}} (L_j^+ - 1) > n\},$$

corresponding to the first (overcrossing) time the chain N_m exceeds n given $N_0 = n_0 < n$ and the number of the corresponding excursion.

Let $P_{(n)}$ be the truncated upper-left corner with size $(n+1, n+1)$ of the full irreducible transition matrix P of N_m (its north-west part). With $\mathbf{1}' = (1, \dots, 1)$ and $\mathbf{e}'_{n_0} = (0, \dots, 0, 1, 0, \dots, 0)$ transpose row vectors with size $n+1$ (with 1 in position $n_0 + 1$ for \mathbf{e}'_{n_0}), it follows from Propositions 11 and 12 of [6] that

$$(26) \quad \mathbf{P}_{n_0}(T_1(n) > l) = \mathbf{e}'_{n_0} P_{(n)}^l \mathbf{1},$$

where $P_{(n)}^l$ is the l -th power of $P_{(n)}$. $\mathbf{P}(T_1(n) > l) = 1$ for $l \in \{1, \dots, n - n_0\}$. At this time $T_1(n)$, the state of the chain N_m is $n+1$ because the overshoot can only be 1. So $T_1(n)$ has geometric tails with decay-rate parameter the spectral radius of $P_{(n)}$ and

$$\mathbf{E}_{n_0}(T_1(n)) = \mathbf{e}'_{n_0} (I - P_{(n)})^{-1} \mathbf{1}.$$

Clearly, given $N_0 = n_0 < n$, with $N_l^* = \max_{m \leq l} N_m$ the extremal process of N_m , the events $N_l^* \leq n$ and $T_1(n) > l$ coincide, so (26) also gives the marginal law $\mathbf{P}_{n_0}(N_l^* \leq n)$ of N_l^* .

The extremal chain N_l^* only grows (by one unit) at the record times $R_k := \inf\{r \in \mathbb{N} : r > R_{k-1}, N_r > N_{R_{k-1}}\}$ of N_m .

7. A MORE GENERAL MARKOV MODEL FOR THE NUMBER OF SUCCESSES

As before, let $w_1 > 0$ and $w_2 \geq 0$ and define $w := w_1 + w_2$. A more general model can be introduced by taking an additional third parameter $\alpha \in [0, 1]$ and assuming that the number S_n of successes forms a Markov chain $(S_n, n \in \mathbb{N}_0)$ satisfying $S_0 = 0$ and

$$\mathbf{P}(S_{n+1} = k + 1 \mid S_n = k) = 1 - \mathbf{P}(S_{n+1} = k \mid S_n = k) := \frac{w_1 + k\alpha}{w + n}, \quad n \in \mathbb{N}_0.$$

In this case S_n coincides with the number of occupied tables in the restaurant process with a cocktail bar [12] after n customers have entered the restaurant. For $\alpha = 0$ we are back to the model studied before. For $\alpha > 0$ the transition probabilities of the random walk $(S_n, n \in \mathbb{N})$ now depend not only on the time n but also on the current state $S_n = k$. The distribution of S_n can be expressed as (see [12, Eq. (14)])

$$\mathbf{P}(S_n = k) = \frac{[w_1|\alpha]_k}{[w]_n} S(n, k; -1, -\alpha, w_2), \quad k \in \{0, \dots, n\},$$

where $[w_1|\alpha]_0 := 1$, $[w_1|\alpha]_k := \prod_{i=0}^{k-1} (w_1 + i\alpha)$ for $k \in \mathbb{N}$ and $S(n, k; -1, -\alpha, w_2)$ denote the generalized Stirling numbers in the notation of Hsu and Shiue [7], which can be calculated as follows. For $\alpha = 0$ it follows from (3) that $S(n, k; -1, 0, w_2) = \sum_{l=k}^n \binom{l}{k} w_2^{l-k} |s_{n,l}|$, $k \in \{0, \dots, n\}$. For $\alpha \neq 0$, the Dobiński-type formula [7, Theorem 4] yields

$$S(n, k; -1, -\alpha, w_2) = \frac{\alpha^{-k}}{k!} \sum_{l=0}^k (-1)^l \binom{k}{l} [w_2 - l\alpha]_n, \quad k \in \{0, \dots, n\}.$$

Note that $\mathbf{P}(S_n = 0) = [w_2]_n/[w]_n$ does not depend on $\alpha \in [0, 1]$. In particular, for any $n \in \mathbb{N}$, $\mathbf{P}(S_n = 0) = 0$ if and only if $w_2 = 0$. Formulas for the moments of S_n are provided in [12, Section 6.1] for $\alpha = 0$ and in [12, Corollary 1] for $\alpha > 0$. The behavior of S_n for $\alpha > 0$ differs substantially from the case $\alpha = 0$. For $\alpha > 0$, as $n \rightarrow \infty$, S_n/n^α converges almost surely and in L^p for any $p > 0$ to a limiting random variable being three-parameter (α, β, γ) -Mittag-Leffler distributed, where $\beta := w$ and $\gamma := w_1/\alpha$, see [12, Theorem 3]. We refer the reader to [12, Section 7] for further details on the three-parameter Mittag-Leffler distribution $\text{ML}(\alpha, \beta, \gamma)$. For $\alpha = 1$ the limiting distribution $\text{ML}(1, w, w_1) = \beta(w_1, w_2)$ is the beta distribution with parameters w_1 and w_2 , in agreement with well-known results for standard Pólya urns.

If $w_2 = 0$ then S_n counts the number of distinct species in a sample of size n taken from Pitman and Yor's [17] two-parameter stick-breaking $\text{PD}(\alpha, w_1)$ -partition of the unit interval, extending the Ewens case. We refer the reader to Chapter 3 of Pitman's lecture notes [16] for further information on the two-parameter model and

to Yamato and Sibuya [25] and Yamato, Sibuya and Nomachi [26] for some further related works.

Assume now that $w_2 > 0$. In this case S_n may no longer be seen, *stricto sensu*, as the number of new species in a sample of size n taken from a partition of the unit interval. However (see [12, Theorem 2]), S_n is the number of new species (excluding a ‘fictitious species’ 0 with beta distributed ‘abundance’ $B_0 \stackrel{d}{=} \beta(w_2, w_1)$) in a sample of size n drawn from a kind of three-parameter Poisson–Dirichlet partition $\text{PD}(\alpha, w_1, w_2) := (B_0, (1 - B_0)\text{PD}(\alpha, w_1))$, where B_0 is independent of $\text{PD}(\alpha, w_1)$.

Note that $K_1^+ := \inf\{n \in \mathbb{N} : S_n = 1\}$ has distribution

$$\mathbf{P}(K_1^+ = n) = \mathbf{P}(S_{n-1} = 0)\mathbf{P}(S_n = 1 \mid S_{n-1} = 0) = w_1 \frac{[w_2]_{n-1}}{[w]_n}, \quad n \in \mathbb{N},$$

so that $S_n^+ := S_{n+K_1^+-1}$ (with $S_1^+ = 1$) coincides (in law) with the number of new species from a $\text{PD}(\alpha, w_1)$ -partition of the unit interval. Whenever a sample hits the ‘fictitious species’ 0, sampling simply fails to draw any new species: this event thus represents the possibility of a failure of the sampling process from scratch. The probability that in a sample of size n there are n_0 failure events clearly is the beta binomial probability mass function $\binom{n}{n_0} [w_2]_{n_0} [w_1]_{n-n_0} / [w]_n$, $n_0 \in \{0, \dots, n\}$. If $\alpha = 0$ then S_n is the number of new species (excluding the ‘fictitious species’ 0 with ‘abundance’ B_0) in a sample of size n drawn from the partition $\text{PD}(0, w_1, w_2) = (B_0, (1 - B_0)\text{PD}(0, w_1))$, extending the Ewens case.

Let $n_0 \in \mathbb{N}_0$ and $n_1, \dots, n_k \in \mathbb{N}$ and put $n := n_0 + \dots + n_k$. Note that

$$\begin{aligned} \mathbf{P}(S_n = k, N_n(0) = n_0, N_n(1) = n_1, \dots, N_n(k) = n_k) \\ (27) \quad = n! \frac{[w_1|\alpha]_k}{[w]_n} \frac{[w_2]_{n_0}}{n_0!} \prod_{l=1}^k \frac{[1-\alpha]_{n_l-1}}{(n_l-1)! \sum_{j=l}^k n_j} \end{aligned}$$

is the joint distribution that there are n_0 visits to the reservoir set with size B_0 (accounting for early failure events of the sampling process, or missed samples) and $S_n = k$ distinct visited species in order of appearance with positive sample sizes n_1, \dots, n_k not in the reservoir. For $w_2 = 0$, (27) reduces to the two-parameter Donnelly–Tavaré–Griffiths distribution $\text{DTG}(w_1, \alpha)$ (see [26, Theorem 1])

$$(28) \quad \mathbf{P}(S_n = k, N_n(1) = n_1, \dots, N_n(k) = n_k) = n! \frac{[w_1|\alpha]_k}{[w_1]_n} \prod_{l=1}^k \frac{[1-\alpha]_{n_l-1}}{(n_l-1)! \sum_{j=l}^k n_j}.$$

For $\alpha = 0$, (28) reduces to

$$\mathbf{P}(S_n = k, N_n(1) = n_1, \dots, N_n(k) = n_k) = n! \frac{w_1^k}{[w_1]_n} \prod_{l=1}^k \frac{1}{\sum_{j=l}^k n_j},$$

which is [23, Eq. (1)] with α there replaced by w_1 .

Summing (27) over all $n_1, \dots, n_k \in \mathbb{N}$ with $n_1 + \dots + n_k = n - n_0$, the joint probability that, in a sample of size n , there are $S_n = k$ new sampled species and $n_0 \leq n$ visits to the ‘fictitious species’ is thus

$$\mathbf{P}(N_n(0) = n_0, S_n = k) = \binom{n}{n_0} \frac{[w_2]_{n_0} [w_1|\alpha]_k}{[w]_n} S(n - n_0, k; -1, -\alpha, 0).$$

Observing that $\sum_{k=0}^{n-n_0} S(n-n_0, k; -1, -\alpha, 0)[w_1|\alpha]_k = [w_1]_{n-n_0}$, the probability that, in a sample of size n , there are $n_0 \leq n$ visits to the ‘fictitious species’ is thus the beta-binomial probability $\mathbf{P}(N_n(0) = n_0) = \binom{n}{n_0} [w_2]_{n_0} [w_1]_{n-n_0} / [w]_n$, in agreement with the explanations above.

ACKNOWLEDGMENT

T. Huillet acknowledges partial support from the Chair ‘*Modélisation mathématique et biodiversité*’ of Veolia-Ecole Polytechnique-MNHN-Fondation X and from the labex MME-DII Center of Excellence (*Modèles mathématiques et économiques de la dynamique, de l’incertitude et des interactions*, ANR-11-LABX-0023-01 project). This work was also funded by CY Initiative of Excellence (grant ‘*Investissements d’Avenir*’ ANR-16-IDEX-0008), Project ‘EcoDep’ PSI-AAP2020-0000000013.

REFERENCES

- [1] ARRATIA, R., BARBOUR, A. D. AND TAVARÉ, S. (1992). Poisson process approximations for the Ewens sampling formula. *Ann. Appl. Probab.* **2**(3), 519–535. MR1177897
- [2] BARBOUR, A. D. AND HALL, P. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.* **95**(3), 473–480. MR0755837
- [3] DENISOV, D., SAKHANENKO, A. AND WACHTEL, V. (2018). First-passage times for random walks with nonidentically distributed increments. *Ann. Probab.* **46**(6), 3313–3350. MR3857857
- [4] EWENS, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoret. Population Biol.* **3**, 87–112; erratum, *ibid.* **3**, 240; erratum, *ibid.* **3**, 376. MR0325177
- [5] GONCALVES, B. AND HUILLET, T. (2020). Scaling features of two special Markov chains involving total disasters. *J. Stat. Phys.* **178**(2), 499–531. MR4055249
- [6] GONCALVES, B. AND HUILLET, T. (2021). Keeping random walks safe from extinction and overpopulation in the presence of life-taking disasters. *Math. Popul. Stud.* Online first publication: <https://doi.org/10.1080/08898480.2021.1976476>.
- [7] HSU, L. C. AND SHIUE, P. J.-S. (1998). A unified approach to generalized Stirling numbers. *Adv. in Appl. Math.* **20**(3), 366–384. MR1618435
- [8] HUILLET, T. E. (2011). On a Markov chain model for population growth subject to rare catastrophic events. *Physica A* **390**, no 23–24, 4073–4086.
- [9] JOHNSON, N. L. AND KOTZ, S. (1977). *Urn Models and Their Application*. Wiley, New York. MR0488211
- [10] KARLIN, S. (1966). *A First Course in Stochastic Processes*. Academic Press, New York-London. MR0208657
- [11] KOZUBOWSKI, T. J. AND PODGÓRSKI, K. (2018). A generalized Sibuya distribution. *Ann. Inst. Statist. Math.* **70**(4), 855–887. MR3830290
- [12] MÖHLE, M. (2021). A restaurant process with cocktail bar and relations to the three-parameter Mittag–Leffler distribution. *J. Appl. Probab.* **58**(4), 978–1006. MR4342591
- [13] NEUTS, M. F. (1967). Waitingtimes between record observations. *J. Appl. Probab.* **4**(1), 206–208. MR0208652
- [14] PEMANTLE, R. (2007). A survey of random processes with reinforcement. *Probab. Surv.* **4**, 1–79. MR2282181
- [15] PITMAN, J. (1997). Probabilistic bounds on the coefficients of polynomials with only real zeros. *J. Combin. Theory Ser. A* **77**(2), 279–303. MR1429082
- [16] PITMAN, J. (2006). *Combinatorial Stochastic Processes, Lecture Notes in Mathematics* **1875**, Springer, Berlin. MR2245368
- [17] PITMAN, J. AND YOR, M. (1997). The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25**(2), 855–900. MR1434129
- [18] RÉNYI, A. (1962). On outstanding values of a sequence of observations. In: *Selected papers of A. Rényi*, Vol. 3, pp. 50–65, Akadémiai Kiadó, Budapest.
- [19] RÉNYI, A. (1962). Théorie des éléments saillants d’une suite d’observations. *Ann. Fac. Sci. Univ. Clermont-Ferrand*, **8**(2), 7–13. MR0286162

- [20] SEVAST'YANOV, B. A. (1972). Poisson limit law for a scheme of sums of dependent random variables. *Theory Probab. Appl.* **17**(4), 695–699. (Russian original reviewed in MR0310943).
- [21] SIBUYA, M. (1979). Generalized hypergeometric, digamma and trigamma distributions. *Ann. Inst. Statist. Math.* **31**(3), 373–390. MR0574816
- [22] TAVARÉ, S. AND ZEITOUNI, O. (2004). *Lectures on Probability Theory and Statistics*. Lecture Notes in Mathematics **1837**, Springer, Berlin. MR2071629
- [23] YAMATO, H. (1997). On the Donnelly–Tavaré–Griffiths formula associated with the coalescent. *Commun. Statist. Theory Methods* **26**(3), 589–599. MR1436290
- [24] YAMATO, H. (2017). Poisson approximations for sum of Bernoulli random variables and its application to Ewens sampling formula. *J. Japan Statist. Soc.* **47**(2), 187–195. MR3791201
- [25] YAMATO, H. AND SIBUYA, M. (2000). Moments of some statistics of Pitman sampling formula. *Bull. Inform. Cybernet.* **32**(1), 1–10. MR1792352
- [26] YAMATO, H., SIBUYA, M. AND NOMACHI, T. (2001). Ordered sample from two-parameter GEM distribution. *Statist. Probab. Lett.* **55**(1), 19–27. MR1860188
- [27] YULE, G. U. (1925). A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis, F.R.S., *Philos. Trans. Roy. Soc. London Ser. B* **213**, 21–87.

THIERRY HUILLET
LABORATOIRE DE PHYSIQUE THÉORIQUE ET MODÉLISATION
CY CERGY PARIS UNIVERSITY, CNRS UMR-8089
SITE DE SAINT MARTIN, 2 AVENUE ADOLPHE-CHAUVIN
95302 CERGY-PONTOISE, FRANCE

E-mail address: `thierry.huillet@cyu.fr`

MARTIN MÖHLE
FACHBEREICH MATHEMATIK
EBERHARD KARLS UNIVERSITÄT TÜBINGEN
AUF DER MORGENSTELLE 10
72076 TÜBINGEN, GERMANY

E-mail address: `martin.moehle@uni-tuebingen.de`