The Annals of Applied Probability
2024, Vol. 34, No. 6, 5809–5841
https://doi.org/10.1214/24-AAP2108

A TWO-TABLE THEOREM FOR A DISORDERED CHINESE RESTAURANT PROCESS

BY JAKOB E. BJÖRNBERG[1,2,a], CÉCILE MAILLER[3,b], PETER MÖRTERS[4,c] AND DANIEL UELTSCHI[5,d]

[1]*Department of Mathematics, Chalmers University of Technology*
[2]*Department of Mathematics, The University of Gothenburg,* [a]*jakob.bjornberg@gu.se*
[3]*Department of Mathematical Sciences, University of Bath,* [b]*c.mailler@bath.ac.uk*
[4]*Department Mathematik/Informatik, Universität zu Köln,* [c]*moerters@math.uni-koeln.de*
[5]*Department of Mathematics, University of Warwick,* [d]*daniel@ueltschi.org*

We investigate a disordered variant of Pitman's Chinese restaurant process where tables carry i.i.d. weights. Incoming customers choose to sit at an occupied table with a probability proportional to the product of its occupancy and its weight, or they sit at an unoccupied table with a probability proportional to a parameter $\theta > 0$. This is a system out of equilibrium where the proportion of customers at any given table converges to zero almost surely. We show that for weight distributions in any of the three extreme value classes, Weibull, Gumbel or Fréchet, the proportion of customers sitting at the largest table converges to one in probability, but not almost surely, and the proportion of customers sitting at either of the largest two tables converges to one almost surely.

1. Introduction. Markets out of equilibrium often follow a winner-takes-all dynamics by which competition allows the best performers to rise to the top at the expense of the losers [7]. In expanding markets, as time passes, more competitive performers emerge and take the place of the current winner. In this paper we study a simple model of this phenomenon, exploring the way in which new competitors take over from the current winners. In our model, a quality is attached to any product put on the market. When a new customer enters the market, a product is selected on the basis of its quality and on the number of customers that have chosen the product so far. This model is a disordered variant of the Chinese restaurant process of Dubins and Pitman [16], see also [1]. In this analogy customers enter a fictitious Chinese restaurant and choose a table to sit on; there is competition between tables in order to attract customers. We use this terminology throughout the paper.

More precisely, at first occupancy, a positive random "fitness", or "weight", is attached to each table, independently of everything else, according to a fixed distribution μ. A new customer either joins an already occupied table, with probability proportional to both its fitness and the number of customers already sitting there, or sits at a new table, with probability proportional to a fixed parameter $\theta > 0$. The proportion of customers at each table in the disordered Chinese restaurant process generates a dynamic random partition, representing the market share of each product in our earlier interpretation, parametrised by a positive real number θ, and a probability distribution μ on the interval $(0, \infty)$. The aim of this paper is to understand the evolution of the largest tables in the disordered Chinese restaurant process, representing the market share of the leading products.

In the classical model of [1] and [16], the random partition (with elements in decreasing order) converges in distribution to a Poisson–Dirichlet distribution of parameter θ. For more

Received March 2023; revised May 2024.
MSC2020 subject classifications. Primary 60K37; secondary 60F15, 60G70, 60J10, 60K35.
Key words and phrases. Winner-takes-all, random environment, extreme value theory, disorder, nonequilibrium, Chinese restaurant process, almost sure limit theorem, Poisson limit theorem.

information on the classical Chinese restaurant process, we refer the reader to, for example, [16], and the references therein. The introduction of the disorder radically changes the behaviour of the process since, contrary to what happens in the classical case, the proportion of customers sitting at any fixed table converges almost surely to zero as time goes to infinity. This is because fitter and fitter tables keep entering the system. This paper aims at answering the following questions:

What proportion of customers sit at the largest table at time n, that is, when there are n customers in the restaurant? What is the weight of this table? When was this table first occupied?

Our two main results are that:

- The proportion of customers sitting at the largest table converges to one *in probability* as the number of customers grows to infinity, see Theorem 1.2. This result does not hold almost surely.
- The proportion of customers sitting at the largest table *or at the second largest table* converges to one *almost surely* as the number of customers grows to infinity, see Theorem 1.3.

We call Theorem 1.3 the "two-table" theorem, as a reference to the parabolic Anderson "two-city" theorem, see [10]. Although the parabolic Anderson model is not at all related to the Chinese restaurant process, our results are reminiscent of those of [10], which they describe intuitively as follows: "*at a typical large time, the mass, which is thought of as a population, inhabits one site, interpreted as a city. At some rare times, however, word spreads that a better site has been found, and the entire population moves to the new site, so that at the transition times part of the population still lives in the old city, while another part has already moved to the new one*". A similar interpretation holds in our setting, with tables replacing cities, and customers replacing the elements of the population.

The proofs of our results rely on embedding the disordered Chinese restaurant process into continuous time. In this embedding, new tables are created at the jump times of a Poisson process of parameter θ, and the number of customers at each table is a Yule process whose parameter equals the weight of the table. This is reminiscent of the continuous-time embedding of the preferential attachment graph with fitnesses of Bianconi and Barabási [3, 5]. Our proof of Theorem 1.2 relies on methods developed in [6, 13] for the study of the Bianconi–Barabási model. It holds under a quite general assumption on the fitness distribution μ, we just ask that it belongs to an extreme value class, see Assumption 2.1. In particular, we allow the fitness distribution to have unbounded support. We are also able to give estimates of when the largest table at time n was first occupied, and of its weight. For the proof of Theorem 1.3, the "two-table theorem", a much refined analysis is needed. Theorem 1.3 holds under stronger assumptions on μ, see Assumptions 2.3, 2.4, and 2.5, depending on which extreme value class μ belongs to; in Appendix B we show that these assumptions are satisfied by a number of special cases of fitness distributions. We next give a formal definition of our model (Section 1.1) and state our main results (Section 1.2).

1.1. *Mathematical definition of the model.* The weighted Chinese restaurant process is a Markov process $(S_i(n)\colon i \geq 1)_{n\geq 0}$ taking values in the set of all sequences $(s_i)_{i\geq 1}$ of nonnegative integers such that there exists $k \in \mathbb{N}$ with $s_i = 0$ if and only if $i > k$. For all n, we call $S_i(n)$ the size of the ith table at time n, and $K_n = \max\{i \geq 1\colon S_i(n) \neq 0\}$ the number of occupied tables in the restaurant at time n. We sample a sequence $(W_i)_{i\geq 1}$ of i.i.d. random variables of distribution μ, the weights or fitnesses. Given this sequence, the process is recursively defined. At time zero, $S_1(0) = 1$ and $S_i(n) = 0$ for all $i \geq 2$. Given the configuration at time n, that is, $(S_i(n))_{i\geq 1}$ either:

- the $(n+1)$th customer enters and sits at the ith table, meaning that $S_i(n+1) = S_i(n) + 1$ and $S_j(n+1) = S_j(n)$ for $j \neq i$, with probability proportional to $W_i S_i(n)$;
- or the $(n+1)$th customer sits at a new table (table number $K_n + 1$), meaning that $S_{K_n+1}(n+1) = 1$ and $S_i(n+1) = S_i(n)$ for $i \leq K_n$, with probability proportional to θ.

The classical case of Pitman's process arises when the fitnesses are deterministic, that is, all tables have the same fitness. The case of interest for us is when μ has no mass at its essential supremum (which may be finite or infinite) so that fitter tables keep emerging. Under this assumption, the following basic properties hold; see Appendix A for the proof.

PROPOSITION 1.1 (Basic properties of the weighted Chinese restaurant process).

(i) *The number of occupied tables K_n when the nth customer enters the restaurant satisfies*

$$\lim_{n\to\infty} \frac{K_n}{\log n} = \frac{\theta}{\operatorname{essup}\mu} \quad \textit{almost surely},$$

where the right hand side is interpreted as zero if the fitnesses are unbounded.

(ii) *For every $k \geq 1$*

$$S_k(n) \to \infty \quad \textit{and} \quad \frac{S_k(n)}{n} \to 0 \quad \textit{almost surely as } n \to \infty.$$

Hence every fixed table has microscopic occupancy.

(iii) *There is no persistence of the table with maximal occupancy. In other words, the time B_n at which the most occupied table at time n gets its first occupany goes to infinity almost surely.*

(iv) *The proportion of customers sitting at the largest table*

$$\max_{i\geq 1} \frac{S_i(n)}{n}$$

does not converge to one, almost surely.

1.2. *Main results.* Here we briefly summarise our main results, postponing precise formulations of our assumptions to the next section. Our first result is a "one-table-theorem" and states that, in probability, the largest table "takes it all". It holds under Assumption 2.1, stated below, which essentially says that the weights W_i belong to the maximum domain of attraction of an extreme value distribution (Weibull, Gumbel or Fréchet):

THEOREM 1.2. *Assume that the distribution μ of the weights W_i satisfies Assumption 2.1, stated in Section 2.1 below. Then*

$$\max_{i\geq 1} \frac{S_i(n)}{n} \to 1, \quad \textit{in probability as } n \to \infty.$$

Recall that the convergence of Theorem 1.2 does not hold almost surely. Our second main result states that there are never more than two tables of macroscopic size. For this result we need a strengthened version of our basic Assumption 2.1.

THEOREM 1.3. *Assume that the distribution μ of the weights W_i satisfies Assumption 2.3, 2.4 or 2.5, stated in Section 2.1 below. Let $S^{(1)}(n)$ and $S^{(2)}(n)$ denote the occupancy of the largest two tables when there are n customers in the restaurant. Then*

$$\frac{S^{(1)}(n) + S^{(2)}(n)}{n} \to 1, \quad \textit{almost surely as } n \to \infty.$$

Technically, it is more convenient to prove our main results for a continuous-time version of our process and then transfer them to the discrete-time process. We thus give the proofs of Theorems 1.2 and 1.3 at the end of Section 2, in which we introduce the embedding of our process into continuous time and state their continuous-time analogues.

2. The process in continuous time. The disordered Chinese restaurant process is defined in the Introduction as a discrete time process. It can also be embedded into continuous time and this embedding is a major technical tool for us.

We first sample and fix a sequence $(W_i)_{i\geq 1}$ of i.i.d. random variables of distribution μ, where W_i constitutes the weight of table number i. At time $t = 0$, there is one customer in the restaurant, sitting at table number 1. Intuitively, given the weights $(W_i)_{i\geq 1}$, each customer sitting at table i carries an exponential clock of parameter W_i, and when one of these clocks rings, a new customer enters the restaurant and sits at the ith table. In addition, customers enter the restaurant and open new tables at rate θ. All exponential clocks are independent of each other.

More formally, we define $Z_i(t)$, the size of the ith table at time t in terms of an independent Yule process $(Y_i(t))_{t\geq 0}$, where we recall that a *Yule process* of parameter $\beta > 0$ is a continuous-time branching process where each individual is immortal and gives birth to one more individual at rate β, independently of each other. Writing $(Y_i)_{i\geq 1}$ for a sequence of i.i.d. Yule processes of parameter one, independent also of $(W_i)_{i\geq 1}$, we define

$$Z_i(t) = Y_i\big(W_i(t - \tau_i)\big)\mathbf{1}_{t\geq\tau_i}, \tag{2.1}$$

where $\tau_0 = 0$ and the τ_i's for $i \geq 1$ are the jump-times of an independent Poisson counting process of rate θ. To see that $(Z_i(t) : i \geq 1)_{t\geq 0}$ is a continuous time embedding of the discrete time process $(S_i(n) : i \geq 1)_{n\geq 0}$, we denote $(\mathscr{F}_t : t \geq 0)$ the filtration generated by $(Z_i(t) : i \geq 1)_{t\geq 0}$. Given $\mathscr{F}_t$ the next change of the random vector $(Z_i(t) : i \geq 1)$ is either the establishment of a new table if an exponential clock of parameter θ rings before the exponential clocks attached to the customers already present ring, and this happens with probability proportional to θ, or the next customer joins an existing customer at their table if her clock rings first, which happens with a probability proportional to their table's fitness.

The major advantage of the embedding comes from the fact that, by elementary properties of the Yule process (see, e.g., [2], Chapter III), there exists a sequence $(\zeta_i)_{i\geq 1}$ of i.i.d. random variables of exponential distribution of parameter 1 such that, for any fixed $i \geq 1$, $\mathrm{e}^{-t}Y_i(t) \to \zeta_i$ almost surely as $t \uparrow \infty$. Therefore,

$$Z_i(t) \sim \zeta_i \exp\big(W_i(t - \tau_i)\big) \quad \text{almost surely as } t \uparrow \infty. \tag{2.2}$$

Thus the relative table sizes are primarily determined by the relative sizes of the "exponents" $W_i(t - \tau_i)$. This intuition is central to much of our analysis and will be made rigorous later.

2.1. *Notation and setting.* Recall that μ denotes the distribution of table weights. We assume that μ belongs to the maximum domain of attraction of a distribution ν on $\mathbb{R}$, meaning that there are functions $(A(t))_{t\geq 0}$ and $(B(t))_{t\geq 0}$ such that

$$\frac{\max_{i=1..n} W_i - A(n)}{B(n)} \Rightarrow \nu, \quad \text{in distribution as } n \to \infty. \tag{2.3}$$

In fact, we assume the following. If μ has bounded support, we assume without loss of generality that its essential supremum is 1 and we define $M = 1$. If the support of μ is unbounded, we set $M = \infty$. Throughout this paper we will assume that μ is absolutely continuous. Then our standing assumption is as follows.

TABLE 1
The functions Φ, u_t, v_t and w_t for the three possible distributions ν. Here $\alpha > 0$ and $L_0(t), \dots, L_3(t)$ denote slowly varying functions

Weibull	$\Phi(x) = \lvert x\rvert^{\alpha} 1_{x<0}$	$u_t = t^{\frac{\alpha}{\alpha+1}} L_0(t)$ $v_t = 1$ $w_t = t^{-\frac{1}{\alpha+1}} L_0(t)$	$L_0(t) \to 0$
Gumbel	$\Phi(x) = \mathrm{e}^{-x}$	$u_t = t L_1(t)$ $v_t = L_2(t)$ $w_t = L_1(t) L_2(t)$	$L_1(t) \to 0$ $L_2(t) \to M$
Fréchet	$\Phi(x) = \infty \mathbf{1}_{x \le 0} + x^{-\alpha} \mathbf{1}_{x>0}$	$u_t = t$ $v_t = 0$ $w_t = t^{\frac{1}{\alpha}} L_3(t)$	

ASSUMPTION 2.1 (first part). There are two continuous functions $(A(t))_{t\ge 0}$, $(B(t))_{t\ge 0}$ and a probability distribution ν on $\mathbb{R}$ such that, for all $x \in \mathbb{R}$,

$$t\mu\big((A(t) + xB(t), M)\big) \to -\log \nu(-\infty, x) =: \Phi(x), \qquad \text{as } t \uparrow \infty. \tag{2.4}$$

Also, Φ is nonincreasing, A is either constant or increasing, and either $A(t) = 0$ for all $t \ge 0$, or $A(t)/B(t)$ is nondecreasing and tends to infinity as $t \uparrow \infty$.

By classical extreme value theory (see, e.g., [4], Section 8.13), the stated properties of A and B hold without loss of generality. Also ν is either a Weibull, a Gumbel, or a Fréchet distribution, and we can choose B nonnegative and Φ as in Table 1. Also, we can choose $A = 1$ in the Weibull case, A bounded from zero, increasing, and converging to M in the Gumbel case, and $A = 0$ in the Fréchet case. The Weibull case occurs only if $M = 1$ and the Fréchet case only if $M = \infty$, while in the Gumbel case we can have either $M = 1$ or $M = \infty$.

In the Weibull and Gumbel cases, to control the size of high-weighted tables that are created late in the process, we need the convergence of (2.4) to also hold in L^1. This holds if the sequence of functions $(u \mapsto n\mu((A(n) + uB(n), M))_{n\ge 1}$ is uniformly integrable, which is the case in all explicit examples we have considered; see also Appendix B.

ASSUMPTION 2.1 (Continued). If Φ is either the Weibull or the Gumbel distribution, then, for all $x > 0$,

$$\int_x^{\infty} t\mu\big((A(t) + uB(t), M)\big)\,\mathrm{d}u \to \int_x^{\infty} \Phi(u)\,\mathrm{d}u, \qquad \text{as } t \uparrow \infty. \tag{2.5}$$

Further, in the Weibull and Gumbel cases, we define u_t as the solution of

$$tB(u_t) = u_t A(u_t), \tag{2.6}$$

and we set $v_t = A(u_t)$, $w_t = B(u_t)$. The existence of such u_t is proved in Lemma 2.2 below. In particular, we have

$$u_t v_t = t w_t. \tag{2.7}$$

In the Fréchet case, we set $u_t = t$, $v_t = 0$, and $w_t = B(t)$. The motivation for these definitions is as follows:

- The largest tables at time t were created at times of order u_t.
- The weights of the largest tables at time t are of order $v_t + \Theta(w_t)$.

(See Theorem 2.6 for a rigorous statement.)

Recall that a function $L(t)$ is called *slowly varying* as $t \to \infty$ if $L(ct)/L(t) \to 1$ as $t \to \infty$ for any fixed $c > 0$. A function $f(t)$ is called *regularly varying of index* β if $f(t) = t^\beta L(t)$ for some slowly varying $L(t)$.

LEMMA 2.2. *Under Assumption 2.1, in the Weibull and Gumbel cases, for all t large enough, equation (2.6) has a unique solution u_t. Furthermore, u_t is nondecreasing in a neighbourhood of infinity, $u_t \to \infty$, and $u_t = o(t)$ as $t \to \infty$. Further*:

(i) *in the Weibull case, $v_t = 1$, and (u_t) is regularly varying with index $\frac{\alpha}{\alpha+1}$ and (w_t) is regularly varying with index $\frac{-1}{\alpha+1}$*;

(ii) *in the Gumbel case (u_t) is regularly varying of index 1, while (v_t) and (w_t) are slowly varying. Moreover, (v_t) is bounded from zero for large enough t*;

(iii) *in the Fréchet case, (w_t) is regularly varying with index $\frac{1}{\alpha}$.*

PROOF. By Assumption 2.1, $A(u)/B(u)$ is nondecreasing in u, which implies that the function $f(u) = uA(u)/B(u)$ is increasing to infinity. This and continuity imply that, for all large enough t, (2.6) has a unique solution $u_t = f^{-1}(t)$. Also, $u_t = f^{-1}(t) \uparrow \infty$ as $t \to \infty$. Finally, $tB(u_t) = u_t A(u_t)$ implies that $u_t/t = B(u_t)/A(u_t) \to 0$, as $t \uparrow \infty$.

(i) By [4], Theorem 8.13.3, B is regularly varying with index $-1/\alpha$. Hence, by [4], Theorem 1.5.12, the function (u_t) is regularly varying with index $\frac{\alpha}{\alpha+1}$. And as $w_t = B(u_t)$ we get that (w_t) is regularly varying with index $\frac{-1}{\alpha+1}$.

(ii) Note that, in the Gumbel case, the functions $A(t)$ and $B(t)$ are both slowly varying. This can be deduced from [4], Theorem 8.13.4, and its proof as follows. Using their notation, with $H(x) = -\log \mathbb{P}(X > x)$, we have that $A(t) = H^{\leftarrow}((\log t) + 1) - H^{\leftarrow}(\log t)$. This is slowly varying by condition (iii) in [4], Theorem 8.13.4. In the proof of the same theorem it is verified that $B(t) = H^{\leftarrow}(\log t)$ is in the de Haan class and therefore also slowly varying.

(iii) See [4], Theorem 8.13.2. □

We introduce the function

$$\Phi_t(x) := u_t \mu(v_t + x w_t, M) \quad (\text{where } \Phi_t(x) = 0 \text{ if } v_t + x w_t \ge M). \tag{2.8}$$

By Assumption 2.1, we have that $\Phi_t(x) \to \Phi(x)$ as $t \to \infty$ for any $x \in \mathbb{R}$. Also note that $\Phi_t(x)$ is decreasing in x for any fixed t.

Theorem 1.3 requires different assumptions on μ than Assumption 2.1. In the Weibull case, Assumption 2.1 implies that $\mu(1 - x, 1) = x^\alpha \ell(x)$ for some function ℓ that is slowly varying at zero and some $\alpha > 0$; see, for example, [17]. We introduce the following stronger assumption on α in this case.

ASSUMPTION 2.3 (Weibull). μ is supported on $(0, 1)$ and $\mu(1 - x, 1) = x^\alpha \ell(x)$ where ℓ is slowly varying at zero and $\alpha > 1$.

Analogously to the Weibull case, in the Fréchet case Assumption 2.1 implies that $\mu(x, \infty)$ is a regularly varying function, this time at infinity. In this case, we actually do not need a stronger assumption on the index of variation.

ASSUMPTION 2.4 (Fréchet). () μ is supported on $(0, \infty)$ and $\mu(x, \infty) = x^{-\alpha} L(x)$ where $L(x)$ is slowly varying at infinity and $\alpha > 0$.

In the Gumbel case, the assumption needed for the two-table theorem is more complicated. Recall the function $\Phi_t(x)$ in (2.8) and that $\Phi(x) = \mathrm{e}^{-x}$ in this case.

ASSUMPTION 2.5 (Gumbel). In addition to (2.4):

(i) There exist $c_1, c_2 > 0$ such that for all t large enough,

$$\begin{cases} \Phi_t(x) \geq \mathrm{e}^{-x-c_1x^2/\log t} & \text{for all } x \in (-c_2 \log t, c_2 \log t), \\ \Phi_t(x) \leq \mathrm{e}^{-x+c_1x^2/\log t} & \text{for all } x \in \left(-c_2 \log t, \dfrac{M - v_t}{w_t}\right). \end{cases}$$

(ii) The slowly varying function $L_1(t) = w_t/v_t = u_t/t$ satisfies

$$L_1(t) \log\log t \to 0 \quad \text{as } t \to \infty.$$

We give examples of distributions μ satisfying the assumptions in Appendix B.

2.2. *Results in continuous time.* We let $M(t)$ denote the number of nonempty tables at time t. Recall that τ_n are the times of creation of tables, and W_n their weights. The key step to get a one-table theorem in probability is to show the following point process convergence. Recall the concept of vague convergence of measures: if $\gamma, \gamma_1, \gamma_2, \ldots$ are measures on a complete separable metric space $\mathcal{S}$, then γ_n converge vaguely to γ if $\int f\, \mathrm{d}\gamma_n \to \int f\, \mathrm{d}\gamma$ for all nonnegative, continuous, compactly supported functions $f: \mathcal{S} \to \mathbb{R}$. The topology of vague convergence makes the set of Radon measures γ on $\mathcal{S}$ a Polish space. Thus the standard theory of convergence in distribution applies to random variables with values in this space. Let PPP(λ) denote a Poisson point process with σ-finite intensity measure λ, which is represented as a random variable taking values in the set of Radon measures.

THEOREM 2.6. *Let*

$$\mathcal{S} := \begin{cases} [0, \infty] \times [-\infty, \infty] \times (-\infty, \infty] & \textit{in the Weibull and Gumbel cases}, \\ [0, 1] \times [0, \infty] \times (-\infty, \infty] & \textit{in the Fréchet case}. \end{cases}$$

Under Assumption 2.1, *the random variables*

$$\Gamma_t := \sum_{n=1}^{M(t)} \delta\left(\frac{\tau_n}{u_t}, \frac{W_n - v_t}{w_t}, \frac{\log Z_n(t) - t v_t}{t w_t}\right) \tag{2.9}$$

taking values in the space of Radon measures on $\mathcal{S}$ equipped with the vague topology, converge in distribution as $t \to \infty$, to $\Gamma_\infty := \mathrm{PPP}(\mathrm{d}\zeta(s, y, z))$, where

$$\mathrm{d}\zeta(s, y, z) := \begin{cases} \theta\, \mathrm{d}s \otimes -\Phi'(y)\, \mathrm{d}y \otimes \delta_{y-s}(\mathrm{d}z) & \textit{in the Weibull and Gumbel cases}, \\ \theta\, \mathrm{d}s \otimes -\Phi'(y)\, \mathrm{d}y \otimes \delta_{y(1-s)}(\mathrm{d}z) & \textit{in the Fréchet case}. \end{cases}$$

The proof of this theorem appears at the end of Section 3. It shows that the largest tables at time t were created around time $\Theta(u_t)$, have fitness of order $v_t + \Theta(w_t)$, and thus, their size at time t is of order $\exp(t v_t + \Theta(t w_t))$. Indeed, the mass of all points with $\tau_n/u_t \to 0$ (corresponding to "older" tables) concentrates asymptotically on the subset of $\mathcal{S}$ where the first coordinate is zero. As this set has no mass under the intensity measure of the limiting Poisson process, these points must leave every compact subset of $\mathcal{S}$ and, because of the compactification of the intervals in the definition of $\mathcal{S}$, this can only happen by their third coordinate going to $-\infty$. Hence none of these points corresponds to the largest table. This argument, which is crucial in the proof, also applies when $\tau_n/u_t \to \infty$, or $\frac{W_n - v_t}{w_t}$ goes to infinity, or to zero in the Fréchet, or $-\infty$ in the Weibull or Gumbel case.

As promised, the point process convergence of Theorem 2.6 implies a one-table theorem in probability.

COROLLARY 2.7 (One-table theorem). *Let $N(t)$ denote the number of customers in the restaurant at time t. If Assumption 2.1 holds, then*

$$\max_{1\le i\le M(t)} \frac{Z_i(t)}{N(t)} \to 1, \quad \textit{in probability when } t\to\infty. \tag{2.10}$$

PROOF. Let $Z^{(1)}(t)$ and $Z^{(2)}(t)$ denote the sizes of the largest and second largest tables at time t. Also let

$$W^{(1)}(t) = \frac{\log Z^{(1)}(t) - tv_t}{tw_t} \quad \text{and} \quad W^{(2)}(t) = \frac{\log Z^{(2)}(t) - tv_t}{tw_t}.$$

By Theorem 2.6, we have, for all $z_1, z_2 > 0$,

$$\begin{aligned}\mathbb{P}(W^{(1)}(t)\ge z_1, W^{(2)}(t)\ge z_2) &= \mathbb{P}(\Gamma_t(\widehat{\mathcal{S}}\times[z_1,\infty])\ge 1, \Gamma_t(\widehat{\mathcal{S}}\times[z_2,\infty])\ge 2)\\ &\to \mathbb{P}(\Gamma_\infty(\widehat{\mathcal{S}}\times[z_1,\infty])\ge 1, \Gamma_\infty(\widehat{\mathcal{S}}\times[z_2,\infty])\ge 2)\end{aligned}$$

where we have set

$$\widehat{\mathcal{S}} = \begin{cases}[0,\infty]\times[-\infty,\infty] & \text{in the Weibull and Gumbel cases,}\\ [0,1]\times[0,\infty] & \text{in the Fréchet case.}\end{cases}$$

This implies that, as $t\to\infty$, we have $(W^{(1)}(t), W^{(2)}(t)) \Rightarrow (W^{(1)}, W^{(2)})$, where $W^{(1)}$ and $W^{(2)}$ are two almost-surely finite random variables such that $W^{(1)} > W^{(2)}$ almost surely. Clearly

$$N(t) = \sum_{i=1}^{M(t)} Z_i(t) = Z^{(1)}(t) + \left(\sum_{i=1}^{M(t)} Z_i(t) - Z^{(1)}(t)\right). \tag{2.11}$$

Our aim is to show that the second term is negligible in front of $Z^{(1)}(t)$. Almost surely for all $t\ge 0$,

$$0 \le \sum_{i=1}^{M(t)} Z_i(t) - Z^{(1)}(t) \le M(t)Z^{(2)}(t) = M(t)Z^{(1)}(t)\exp[(W^{(2)}(t) - W^{(1)}(t))tw_t].$$

Since $M(t)$ is Poisson-distributed with parameter θt, we have $W^{(2)}(t) - W^{(1)}(t) \Rightarrow W^{(2)} - W^{(1)} < 0$, and $\log t = o(tw_t)$, we indeed get that, in probability as $t\uparrow\infty$,

$$\sum_{i=1}^{M(t)} Z_i(t) - Z^{(1)}(t) = o(Z^{(1)}(t)),$$

which, by (2.11), implies $N(t) = (1+o(1))Z^{(1)}(t)$ and thus concludes the proof. □

From Corollary 2.7 it is a small step to Theorem 1.2.

PROOF OF THEOREM 1.2. Writing T_n for the time of arrival of the nth customer, we have $S_i(n) = Z_i(T_n)$. Then $T_n\to\infty$ almost surely as $n\to\infty$ (indeed, $\{\sup_{n\ge1} T_n < \infty\}$ is equivalent to $\{\exists t_\infty\colon N(t_\infty)=\infty\}$, which has probability zero), so that $\max_{i\ge1} S_i(n)/n = \max_{1\le i\le M(T_n)} Z_i(T_n)/N(T_n) \to 1$ in probability. □

The following result states that, almost surely as $t\uparrow\infty$, no more than two tables can have macroscopic sizes at time t.

THEOREM 2.8. *Assume that Assumption 2.3 (Weibull), Assumption 2.4 (Fréchet), or Assumption 2.5 (Gumbel) hold. Denote by $Z^{(1)}(t)$ the size of the largest table at time t, and by $Z^{(2)}(t)$ the size of the second largest table at time t. Then,*

$$\frac{Z^{(1)}(t)+Z^{(2)}(t)}{N(t)} \to 1, \qquad \text{almost surely as } t\to\infty,$$

where $N(t)$ is the total number of customers in the restaurant at time t.

The proof of this theorem is given in Section 4. We now show how to deduce Theorem 1.3.

PROOF OF THEOREM 1.3. As in the proof of Theorem 1.2, we let T_n be the time of arrival of the nth customer; we have $T_n \uparrow \infty$ almost surely as $n\uparrow\infty$, and $S_i(n) = Z_i(T_n)$, for all $n\geq 1$, $i\geq 1$. Thus, by Theorem 2.8,

$$\frac{S^{(1)}(n)+S^{(2)}(n)}{n} = \frac{Z^{(1)}(T_n)+Z^{(2)}(T_n)}{N(T_n)} \to 1,$$

almost surely as $n\uparrow\infty$. □

It remains to prove Theorems 2.6 and 2.8; this is done in Sections 3 and 4, respectively.

3. One-table result: Proof of Theorem 2.6. The proof of Theorem 2.6 is done in two steps. First, in Section 3.1 we prove convergence of Γ_t (see (2.9)) on the space of measures on

$$\mathcal{W} := \begin{cases} [0,\infty)\times(-\infty,\infty]\times[-\infty,\infty] & \text{in the Weibull and Gumbel cases,}\\ [0,1)\times(0,\infty]\times[-\infty,\infty] & \text{in the Fréchet case.}\end{cases} \tag{3.1}$$

Note that this differs from the claim of Theorem 2.6, where convergence is on the space of measures on the space $\mathcal{S}$ which differs from $\mathcal{W}$ at the endpoints of several of the intervals. Second, and this is the most difficult part of the proof, in Section 3.2, we prove that young tables ($\tau_i \gg u_t$) as well as unfit tables ($W_i - v_t \ll w_t$) are both too small to contribute to the limit. This allows us to "close the brackets" in the first two coordinates of (3.1); in doing so however, the mass corresponding to tables that do not contribute to the limit instead "escapes" to $-\infty$ in the third coordinate. We thereby transfer the convergence on $\mathcal{W}$ to convergence on $\mathcal{S}$.

3.1. *Local convergence.* We prove the following convergence for the space $\mathcal{W}$.

LEMMA 3.1. *In distribution as $t\to\infty$,*

$$\Gamma_t \to \mathrm{PPP}\big(\mathrm{d}\zeta(s,y,z)\big),$$

where

$$\mathrm{d}\zeta(s,y,z) = \begin{cases} \theta\mathrm{d}s\otimes -\Phi'(y)\mathrm{d}y\otimes\delta_{y-s}(\mathrm{d}z) & \textit{in the Weibull and Gumbel cases,}\\ \theta\mathrm{d}s\otimes -\Phi'(y)\mathrm{d}y\otimes\delta_{y(1-s)}(\mathrm{d}z) & \textit{in the Fréchet case,}\end{cases}$$

on the space of measures on $\mathcal{W}$ equipped with the vague topology.

To prove Lemma 3.1, we first prove that

$$\Psi_t := \sum_{n=1}^{M(t)} \delta\left(\frac{\tau_n}{u_t}, \frac{W_n - v_t}{w_t}, \frac{W_n(t-\tau_n)-tv_t}{tw_t}\right) \to \mathrm{PPP}\big(\mathrm{d}\zeta(s,y,z)\big), \tag{3.2}$$

on $\mathcal{W}$, and then prove that this implies convergence of Γ_t on the same space. The difference between Γ_t and Ψ_t is that in the third coordinate we have replaced $\log Z_n(t)$ with its conditional mean $W_n(t-\tau_n)$. We will show that equation (3.2) is a direct consequence of the following lemma.

LEMMA 3.2. *For all $t \geq 0$, in distribution, as $t \to \infty$,*

$$\widehat{\Psi}_t := \sum_{n=1}^{M(t)} \delta\left(\frac{\tau_n}{u_t}, \frac{W_n - v_t}{w_t}\right) \to \mathrm{PPP}\big(\theta \mathrm{d}s \otimes -\Phi'(y)\mathrm{d}y\big),$$

on the space of measures on $\widehat{\mathcal{W}}$ equipped with the vague topology, where

$$\widehat{\mathcal{W}} := \begin{cases} [0,\infty) \times (-\infty,\infty] & \text{in the Weibull and Gumbel cases,} \\ [0,1) \times (0,\infty] & \text{in the Fréchet case.} \end{cases}$$

Before proving Lemma 3.2, we show how to deduce (3.2) from it: If we set $s_{n,t} = \tau_n/u_t$ and $y_{n,t} = (W_n - v_t)/w_t$, then

$$\begin{aligned} W_n(t-\tau_n) &= (v_t + y_{n,t} w_t)(t - s_{n,t} u_t) \\ &= t v_t + y_{n,t} t w_t - s_{n,t} u_t v_t - s_{n,t} y_{n,t} u_t w_t. \end{aligned} \tag{3.3}$$

In the Weibull and Gumbel cases, we have $u_t v_t = t w_t$, and thus

$$W_n(t-\tau_n) = t v_t + (y_{n,t} - s_{n,t}) t w_t - s_{n,t} y_{n,t} u_t w_t,$$

which implies

$$\frac{W_n(t-\tau_n) - t v_t}{t w_t} = y_{n,t} - s_{n,t} - s_{n,t} y_{n,t} \frac{u_t}{t}.$$

Because $u_t/t \to 0$ as $t \uparrow \infty$, this concludes the proof of (3.2) in the Weibull and Gumbel cases. In the Fréchet case, because $u_t = t$ and $v_t = 0$, (3.3) gives

$$W_n(t-\tau_n) = y_{n,t} t w_t - s_{n,t} y_{n,t} t w_t = y_{n,t}(1 - s_{n,t}) t w_t,$$

which concludes the proof of (3.2).

PROOF OF LEMMA 3.2. Invoking Kallenberg's theorem [17], Prop. 3.22, it is enough to prove that for all compact boxes $B = [0,a] \times [b,\infty]$, where $b \in \mathbb{R}$ in the Weibull and Gumbel cases, and $b > 0$ in the Fréchet case, we have:

- $\mathbb{P}(\widehat{\Psi}_t(B) = 0) \to \exp(\int_B \theta \Phi'(y) \mathrm{d}s \mathrm{d}y) = \exp(-\theta a \Phi(b))$,
- $\mathbb{E}[\widehat{\Psi}_t(B)] \to -\int_B \theta \Phi'(y) \mathrm{d}s \mathrm{d}y = \theta a \Phi(b)$.

We let $I(t)$ be the set of all n such that $\tau_n \leq a u_t$; so that $|I(t)|$ is Poisson-distributed with parameter $a\theta u_t$. We have

$$\mathbb{P}\big(\widehat{\Psi}_t(B) = 0\big) = \mathbb{P}\big(\forall 1 \leq n \leq |I(t)|, W_n < v_t + b w_t\big) = \mathbb{E}\big[\big(1 - \mu(v_t + b w_t, M)\big)^{|I(t)|}\big],$$

where we recall that $M \in \{1, \infty\}$ is the essential supremum of μ.

Since $|I(t)|$ is Poisson-distributed with parameter $a\theta u_t$, we get

$$\mathbb{P}\big(\widehat{\Psi}_t(B) = 0\big) = \exp\big(-a\theta u_t \mu(v_t + b w_t, M)\big) = \exp\big(-a\theta u_t \mu\big(A(u_t) + bB(u_t), M\big)\big),$$

since, by definition, $v_t = A(u_t)$ and $w_t = B(u_t)$. By Assumption 2.1,

$$\mathbb{P}\big(\widehat{\Psi}_t(B) = 0\big) \to \exp\big(-a\theta \Phi(b)\big),$$

as $t \uparrow \infty$, which concludes the proof of the first assumption of Kallenberg's theorem. For the second assumption, note that

$$\begin{aligned}\mathbb{E}[\widehat{\Psi}_t(B)] &= \mathbb{E}\Big[\sum_{n\in I(t)} \mathbf{1}_{W_n \geq v_t + bw_t}\Big] = \mathbb{E}\big[|I(t)|\big]\mu(v_t + bw_t, M) \\ &= a\theta u_t \mu\big(A(u_t) + bB(u_t), M\big) \to a\theta\Phi(b), \quad \text{as } t \uparrow \infty,\end{aligned}$$

by Assumption 2.1. This concludes the proof. □

Lemma 3.1 is an immediate consequence of the following result and Lemma 3.2, which established convergence of Ψ_t.

LEMMA 3.3. *For all continuous, compactly supported functions $f : \mathcal{W} \to \mathbb{R}$, we have*

$$\left|\int f \,\mathrm{d}\Gamma_t - \int f \,\mathrm{d}\Psi_t\right| \to 0,$$

in distribution when $t \to \infty$.

PROOF. First note that, by density of the set of Lipschitz-continuous, compactly supported functions in the set of continuous, compactly supported functions with respect to L^∞-norm, we may assume that f is Lipschitz-continuous. Let $a > 0$ and $b \in \mathbb{R}$ in the Weibull and Gumbel cases, respectively $a \in [0, 1)$ and $b > 0$ in the Fréchet case, and let $f : [0, a] \times [b, \infty] \times [-\infty, \infty]$ be a Lipschitz-continuous function of Lipschitz constant κ. We have

$$\left|\int f \,\mathrm{d}\Gamma_t - \int f \,\mathrm{d}\Psi_t\right| \leq \kappa \sum_{n\in I(t)} \frac{|\log Z_n(t) - W_n(t-\tau_n)|}{tw_t},$$

where $I(t)$ is the set of all integers n such that

$$\tau_n \in [0, au_t] \quad \text{and} \quad W_n \geq v_t + bw_t.$$

For all $n \geq 1$ and $s \geq 0$, we set $R_n(s) = \sup_{t\geq s} |\log Y_n(t) - t|$, where we recall from (2.1) that Y_n is the Yule process such that $Z_n(t) = Y_n(W_n(t - \tau_n))\mathbf{1}_{t\geq\tau_n}$. By definition, $v_t + bw_t \to M \geq 1$ and $u_t \leq t$ (see equation (2.6)). This means that there is $\delta > 0$ such that $W_n(t - \tau_n) \geq (v_t + bw_t)(t - au_t) \geq \delta t$ for all t large enough (we can take $\delta = (1 - a)/2$ in the Fréchet case and $\delta = 1/2$ in the Weibull and Gumbel cases). We thus get that, almost surely for all t large enough,

$$\left|\int f \,\mathrm{d}\Gamma_t - \int f \,\mathrm{d}\Psi_t\right| \leq \frac{\kappa}{tw_t} \sum_{n\in I(t)} R_n(\delta t).$$

For all integers n, note that $R_n(\delta t) \to |\log \zeta_n|$ almost surely as $t \uparrow \infty$. Moreover, we have $\liminf_{t\to\infty} tw_t = \infty$. Since, in addition, by Lemma 3.2 and its proof, $|I(t)| = \hat{\Psi}_t([0, a] \times [b, \infty])$ converges in distribution to an almost-surely finite random variable independent of (ζ_n) this concludes the proof. □

3.2. *New and unfit tables do not contribute.* To get convergence of Γ_t on $\mathcal{S}$ rather than $\mathcal{W}$ we prove that "new" tables, as well as tables with small weight, are too small to contribute to the limit. We start with the new tables.

LEMMA 3.4. *For all $\varepsilon, \kappa > 0$, there exists $x_0 < M$ such that, for all sufficiently large t, for all $x \geq x_0$,*

$$\mathbb{P}\Big(\max_{n\leq M(t)} (\log Z_n(t))\mathbf{1}_{\{\tau_n \geq x u_t\}} \geq \ell_\kappa(t)\Big) \leq \varepsilon,$$

where

$$\ell_\kappa(t) := \begin{cases} t v_t - \kappa t w_t & \text{in the Weibull and Gumbel cases,} \\ \kappa t w_t & \text{in the Fréchet case.} \end{cases} \tag{3.4}$$

PROOF. Recall the Yule processes Y_n from (2.1). For all $n \geq 1$, set

$$A_n = \sup_{s\geq 0} Y_n(s)\mathrm{e}^{-s} = \sup_{s\geq \tau_n} Z_n(s)\mathrm{e}^{-W_n(s-\tau_n)}. \tag{3.5}$$

Note that the A_n are i.i.d. and that A_n is in fact independent of W_n as it only depends on Y_n. Let $A = \sup_{s\geq 0} Y(s)\mathrm{e}^{-s}$ be a random variable with the distribution of the A_n. Then we have the following tail-bound: for some $C > 0$

$$\mathbb{P}(A > u) \leq C\mathrm{e}^{-u/2}, \quad \text{for all } u > 0. \tag{3.6}$$

This is proved using the maximal inequality for the submartingale $\exp(\theta Y(s)\mathrm{e}^{-s})$, where $\theta \in (0, 1)$, and that $\mathbb{E}[\exp(\theta Y(s)\mathrm{e}^{-s})]$ is uniformly bounded, which may be verified using the explicit distribution, $\mathbb{P}(Y(s) = k) = \mathrm{e}^{-s}(1 - \mathrm{e}^{-s})^{k-1}$ for $k \geq 1$.

Let $I_x(t)$ be the set of all integers n such that $\tau_n \geq x u_t$; using a union bound in the second inequality, we get

$$\begin{aligned}
&\mathbb{P}\Big(\max_{n\leq M(t)} \log Z_n(t)\mathbf{1}\{\tau_n \geq x u_t\} \geq \ell_\kappa(t)\Big) \\
&\quad\leq \mathbb{P}\big(\exists n \in I_x(t) \colon A_n \geq \exp(\ell_\kappa(t) - W_n(t-\tau_n))\big) \\
&\quad\leq \mathbb{E}\Big[\sum_{n\in I_x(t)} \mathbb{P}\big(A_n \geq \exp(\ell_\kappa(t) - W_n(t-\tau_n))\,|\,(\tau_n)\big)\Big].
\end{aligned}$$

As $(\tau_n)_{n\geq 1}$ is a Poisson process of parameter θ, independent of (A_n) and (W_n),

$$\mathbb{P}\Big(\max_{n\leq M(t)} \log Z_n(t)\mathbf{1}\{\tau_n \geq x u_t\} \geq \ell_\kappa(t)\Big) \leq \theta \int_{x u_t}^{t} \mathrm{d}s\,\mathbb{P}\big(A \geq \exp(\ell_\kappa(t) - W(t-s))\big),$$

where A is a copy of A_1 and W a copy of W_1, independent of each other. Thus,

$$\begin{aligned}
&\mathbb{P}\Big(\max_{n\leq M(t)} \log Z_n(t)\mathbf{1}\{\tau_n \geq x u_t\} \geq \ell_\kappa(t)\Big) \\
&\quad\leq \theta \int_{x u_t}^{t} \mathrm{d}s \int_0^\infty \mathrm{d}\mu(w)\mathbb{P}\big(A \geq \exp(\ell_\kappa(t) - w(t-s))\big) \qquad (3.7)\\
&\quad= \theta \int_x^{t/u_t} \mathrm{d}a \int_{-v_t/w_t}^\infty \mathrm{d}\tilde{\mu}_t(u)\mathbb{P}\big(A \geq \exp(\ell_\kappa(t) - (v_t + u w_t)(t - a u_t))\big),
\end{aligned}$$

where $\mathrm{d}\tilde{\mu}_t(u) := u_t\,\mathrm{d}\mu(v_t + u w_t)$ and we have used the changes of variable $s = a u_t$ and $w = v_t + u w_t$. We treat the rest of the proof separately for the Weibull and Gumbel cases on the one hand, and the Fréchet case on the other hand.

The Weibull and Gumbel cases. In these cases, $\ell_\kappa(t) = tv_t - \kappa t w_t$ and $u_t v_t = t w_t$, which implies that

$$
\begin{aligned}
&\mathbb{P}\Big(\max_{n\le M(t)} \log Z_n(t)\mathbf{1}\{\tau_n \ge x u_t\} \ge \ell_\kappa(t)\Big)\\
&\quad\le \theta \int_x^{t/u_t} \mathrm{d}a \int_{-v_t/w_t}^{\infty} \mathrm{d}\tilde{\mu}_t(u)\mathbb{P}\big(A \ge \exp\big(-(\kappa+u)tw_t + a u_t(v_t + u w_t)\big)\big)\\
&\quad\le \theta \int_x^{t/u_t} \mathrm{d}a \int_{-v_t/w_t}^{\infty} \mathrm{d}\tilde{\mu}_t(u)\mathbf{1}\big\{a(v_t+uw_t) \le (2\kappa+u)v_t\big\}\\
&\qquad+ \theta \int_x^{t/u_t} \mathrm{d}a \int_{-v_t/w_t}^{\infty} \mathrm{d}\tilde{\mu}_t(u)\mathbf{1}\big\{a(v_t+uw_t) > (2\kappa+u)v_t\big\}\mathbb{P}\big(A \ge \mathrm{e}^{\kappa t w_t}\big)\\
&\quad\le \theta \int_x^{t/u_t} \mathrm{d}a \int_{-v_t/w_t}^{\infty} \mathrm{d}\tilde{\mu}_t(u)\mathbf{1}\big\{a(v_t+uw_t) \le (2\kappa+u)v_t\big\} + C\mathrm{e}^{-\frac{1}{2}\exp(\kappa t w_t)} t w_t.
\end{aligned}
\tag{3.8}
$$

In the last step, we used that there exists a constant $C > 0$ such that $\mathbb{P}(A \ge u) \le C\mathrm{e}^{-u/2}$ for all $u \ge 0$, and also that $\int_{-\infty}^{\infty} \mathrm{d}\tilde{\mu}_t(u) = u_t w_t$. Since $tw_t \to \infty$, we get that the second term above tends to zero as $t \uparrow \infty$. For the first term, note that, for all $a < t/u_t = v_t/w_t$,

$$
a(v_t + uw_t) \le (2\kappa + u)v_t \quad\Leftrightarrow\quad u \ge \frac{a - 2\kappa}{1 - a w_t/v_t} \quad\Rightarrow\quad u \ge a - 2\kappa,
$$

and thus, for all $x > 2\kappa$,

$$
\begin{aligned}
&\theta \int_x^{t/u_t} \mathrm{d}a \int_{-v_t/w_t}^{\infty} \mathrm{d}\tilde{\mu}_t(u)\mathbf{1}\big\{a(v_t+uw_t) \le (2\kappa+u)v_t\big\}\\
&\quad\le \theta \int_x^{\infty} \mathrm{d}a \int_{a-2\kappa}^{\infty} \mathrm{d}\tilde{\mu}_t(u) = \theta \int_x^{\infty} \mathrm{d}a\Phi_t(a-2\kappa) \to \theta \int_x^{\infty} \mathrm{d}a\Phi(a-2\kappa),
\end{aligned}
\tag{3.9}
$$

as $t \uparrow \infty$, by Assumption 2.1, see (2.5).

We look at the two different possibilities for Φ: in the Weibull case, Φ is zero on $(0,\infty)$, and thus $\int_x^\infty \theta\, \mathrm{d}a\Phi(a-2\kappa) = 0$ as soon as $x > 2\kappa$. In the Gumbel case, we have $\Phi(u) = \mathrm{e}^{-\alpha u}$ for some $\alpha > 0$, and thus

$$
\int_x^{\infty} \theta\, \mathrm{d}a\Phi(a-2\kappa) = \int_x^{\infty} \theta\, \mathrm{d}a\mathrm{e}^{-\alpha(a-2\kappa)} = \frac{1}{\alpha}\mathrm{e}^{-\alpha(x-2\kappa)},
$$

which tends to zero as $x \to \infty$. In both the Weibull and Gumbel cases, we thus get that for all $\delta > 0$, for all x large enough, $\int_x^\infty \theta\, \mathrm{d}a\Phi(a-2\kappa) \le \delta/2$. Therefore, by (3.8) and (3.9), for all x large enough, for all t large enough,

$$
\mathbb{P}\Big(\max_{n\le M(t)} \log Z_n(t)\mathbf{1}\{\tau_n \ge x u_t\}\Big) \le \delta,
\tag{3.10}
$$

which concludes the proof.

The Fréchet case. In the Fréchet case, $v_t = 0$, $u_t = t$, and $\ell_\kappa(t) = \kappa t w_t$. Thus, (3.7) becomes

$$\begin{aligned}
&\mathbb{P}\Big(\max_{n\le M(t)} \log Z_n(t)\mathbf{1}\{\tau_n \ge xt\}\Big) \\
&\quad\le \theta \int_x^1 \mathrm{d}a \int_0^\infty \mathrm{d}\tilde{\mu}_t(u)\mathbb{P}\big(A \ge \exp\big((\kappa-(1-a)u)tw_t\big)\\
&\quad\le \theta \int_x^1 \mathrm{d}a \int_0^\infty \mathrm{d}\tilde{\mu}_t(u)\mathbf{1}\big\{(1-a)u \ge \kappa/2\big\} \qquad (3.11)\\
&\qquad + C\mathrm{e}^{-\frac{1}{2}\exp(\kappa t w_t/2)}\theta \int_x^1 \mathrm{d}a \int_0^\infty \mathrm{d}\tilde{\mu}_t(u)\mathbf{1}\big\{(1-a)u < \kappa/2\big\}\\
&\quad\le \theta \int_0^\infty \Big(1-\frac{\kappa}{2u}-x\Big)_+ \mathrm{d}\tilde{\mu}_t(u) + C\theta t w_t \mathrm{e}^{-\frac{1}{2}\exp(\kappa t w_t/2)},
\end{aligned}$$

because $\tilde{\mu}(0,\infty) = tw_t$. The second term goes to zero as $t\uparrow\infty$ for all $\kappa > 0$. For the first term, we get

$$\begin{aligned}
\int_0^\infty \theta\Big(1-\frac{\kappa}{2u}-x\Big)_+ \mathrm{d}\tilde{\mu}_t(u) &\le \theta \int_{\frac{\kappa}{2(1-x)}}^\infty \mathrm{d}\tilde{\mu}_t(u) = \theta\Phi_t\Big(\frac{\kappa}{2(1-x)}\Big)\\
&= (\theta + o(1))\Phi\Big(\frac{\kappa}{2(1-x)}\Big),
\end{aligned}$$

as $t\uparrow\infty$, by Assumption 2.1. Thus, making x close to 1, one can make the first term of (3.11) as small as desired, which concludes the proof in the Fréchet case. □

In the following lemma, we control the contributions of the small-weight tables.

LEMMA 3.5. *For all $\varepsilon, \kappa > 0$, there exists y_0 such that, for sufficiently large t, for all $y \ge y_0$,*

$$\mathbb{P}\Big(\max_{n\le M(t)} \log Z_n(t)\mathbf{1}\{W_n \le v_t - yw_t\} \ge \ell_\kappa(t)\Big) \le \varepsilon,$$

where $\ell_\kappa(t)$ is defined in (3.4).

PROOF. This proof is very similar to the proof of the previous lemma. Note that, for all $n \ge 1$, $t \ge 0$, if $\log Z_n(t) \ge \ell_\kappa(t)$ and $W_n \le v_t - yw_t$, then

$$\begin{aligned}
\log Z_n(t) - W_n(t-\tau_n) &\ge \ell_\kappa(t) - (v_t - yw_t)t\\
&= \begin{cases} (y-\kappa)tw_t & \text{in the Weibull and Gumbel cases,}\\ (y+\kappa)tw_t & \text{in the Fréchet case.}\end{cases}
\end{aligned}$$

Therefore, using the independence of $M(t)$ and (A_n),

$$\mathbb{P}\Big(\max_{n\le M(t)} \log Z_n(t)\mathbf{1}\{W_n \le v_t - yw_t\} \ge \ell_\kappa(t)\Big) \le \mathbb{E}\big[M(t)\big]\mathbb{P}\big(A_1 \ge \exp\big((y-\kappa)tw_t\big)\big).$$

Recall that $M(t)$ is Poisson distributed of parameter θt, and thus

$$\mathbb{P}\Big(\max_{n\le M(t)} \log Z_n(t)\mathbf{1}\{W_n \le v_t - yw_t\} \ge \ell_\kappa(t)\Big) \le C_0\theta t \exp\Big(-\frac{1}{2}\exp\big((y-\kappa)tw_t\big)\Big),$$

where we used that $\mathbb{P}(A \ge x) \le C_2\mathrm{e}^{-x/2}$. Since $w_t \to \infty$, in the Fréchet case, t can be made large enough so that $C_0\theta t\exp(-\frac{1}{2}\exp((y+\kappa)tw_t)) \le \varepsilon$. In the Weibull and Gumbel cases, for all $y > \kappa$, t can be made large enough so that $C_0\theta t \exp(-\frac{1}{2}\exp((y-\kappa)tw_t)) \leq \varepsilon$. This completes the proof in all three cases. □

We now show how to deduce Theorem 2.6 from Lemmas 3.1, 3.4 and 3.5:

PROOF OF THEOREM 2.6. We give details of the proof in the Weibull and Gumbel cases, as the Fréchet case is identical, except that the first coordinate takes values in $[0, 1]$ instead of $[0, \infty]$, and the third in $(0, \infty]$ instead of $(-\infty, \infty]$.

Let $f : [0, \infty] \times [-\infty, \infty] \times (-\infty, \infty] \to \mathbb{R}$ be a nonnegative, continuous and compactly supported function. Let $\kappa > 0$ such that $\{f \neq 0\} \subseteq [0, \infty] \times [-\infty, \infty] \times [-\kappa, \infty] =: \mathcal{A}(\kappa)$. We aim to prove that, in distribution as $t \uparrow \infty$,

$$\int f \, \mathrm{d}\Gamma_t \to \int f \, \mathrm{d}\Gamma_\infty \tag{3.12}$$

Fix $\eta > 0$. By Lemma 3.4, there exists $x_0 = x_0(\kappa, \eta)$ such that, for all $x \geq x_0$,

$$\liminf_{t\uparrow\infty} \mathbb{P}\big(\Gamma_t\big(\mathcal{B}(x, \kappa)\big) = 0\big) \geq 1 - \eta, \tag{3.13}$$

where we have set $\mathcal{B}(x, \kappa) = (x, \infty] \times [-\infty, \infty] \times [-\kappa, \infty]$. Furthermore, by Lemma 3.5, there exists $y_0 = y_0(\kappa, \eta)$ such that, for all $y \geq y_0$,

$$\liminf_{t\uparrow\infty} \mathbb{P}\big(\Gamma_t\big(\mathcal{C}(y, \kappa)\big) = 0\big) \geq 1 - \eta, \tag{3.14}$$

where we have set $\mathcal{C}(y, \kappa) = [0, \infty] \times [-\infty, -y) \times [-\kappa, \infty]$. For all $t \geq 0$,

$$\int f \, \mathrm{d}\Gamma_t = \int_{\mathcal{A}(\kappa)} f \, \mathrm{d}\Gamma_t = \int_{\mathcal{A}(\kappa)\cap\mathcal{B}(x,\kappa)^c\cap\mathcal{C}(y,\kappa)^c} f \, \mathrm{d}\Gamma_t + R(t),$$

where

$$0 \leq R(t) \leq \int_{\mathcal{B}(x,\kappa)} f \, \mathrm{d}\Gamma_t + \int_{\mathcal{C}(y,\kappa)} f \, \mathrm{d}\Gamma_t.$$

By (3.13) and (3.14), for all t large enough, with probability at least $1 - 2\eta$, $\Gamma_t(\mathcal{B}(x, \kappa)) = \Gamma_t(\mathcal{C}(y, \kappa)) = 0$, implying that $R(t) = 0$. To conclude, note that

$$\mathcal{A}(\kappa) \cap \mathcal{B}(x, \kappa)^c \cap \mathcal{C}(y, \kappa)^c = \big(x, \infty] \times [-\infty, -y\big) \times [-\kappa, \infty],$$

and thus, by Lemma 3.1, in distribution as $t \to \infty$,

$$\int_{\mathcal{A}(\kappa)\cap\mathcal{B}(x,\kappa)^c\cap\mathcal{C}(y,\kappa)^c} f \, \mathrm{d}\Gamma_t \to \int_{\mathcal{A}(\kappa)\cap\mathcal{B}(x,\kappa)^c\cap\mathcal{C}(y,\kappa)^c} f \, \mathrm{d}\Gamma_\infty.$$

Making x and y large enough, because Γ_∞ has no atom, we can make the right-hand side arbitrarily close to $\int_{\mathcal{A}(\kappa)} f \, \mathrm{d}\Gamma_\infty = \int f \, \mathrm{d}\Gamma_\infty$, which concludes the proof. □

4. Two-table theorem: Proof of Theorem 2.8. For the proof of Theorem 2.8 we treat the three cases (Weibull, Gumbel, and Fréchet) in parallel. Although technical details differ, the general strategy is the same for all cases. We first work on the "exponents" instead of the table sizes. That is, we set, for all $t \geq 0$ and all $1 \leq i \leq M(t)$,

$$\Theta_n(t) := W_n(t - \tau_n). \tag{4.1}$$

Recall from (2.2) that $Z_n(t) \sim \zeta_n \exp(\Theta_n(t))$ almost surely as $t \uparrow \infty$, where $(\zeta_n)_{n\geq1}$ is a sequence of i.i.d. random variables of exponential distribution of parameter 1. This is why we call the $\Theta_n(t)$ the "exponents". We also introduce the order statistics of this sequence, $\Theta^{(1)}(t) \geq \Theta^{(2)}(t) \geq \Theta^{(3)}(t) \geq \dots$ and we let $m_i = m_i(t)$ be the index such that

$$\Theta^{(i)}(t) = \Theta_{m_i(t)}(t)$$

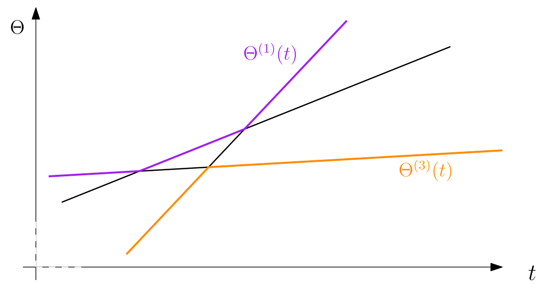

FIG. 1. *Schematic picture of the largest exponents at a time of transition. In Proposition 4.1, we bound the gap between the largest exponent $\Theta^{(1)}(t)$ (in purple) and third largest exponent $\Theta^{(3)}(t)$ (in orange).*

(see Fig. 1). Then $\tau_{m_i(t)}$ denotes the time of creation of the table which at time t has the ith largest exponent. In what follows we often suppress the t-dependence of $m_i(t)$ from the notation.

Recall the function $(w_t)_{t\geq 0}$ given in Lemma 2.2. In this section, we establish the following result, which, by the Borel–Cantelli lemma, gives the existence of a diverging sequence of times $(t_k)_{k\geq 0}$ at which, almost surely, the largest and third-largest exponents are well separated.

PROPOSITION 4.1. *Let $t_k = k^\eta$ and $\lambda_t = t^{-\kappa}$, where $\eta, \kappa > 0$ satisfy $2\kappa\eta > 1$. Then under Assumption 2.1 for the Weibull and Fréchet cases, respectively Assumption 2.5 for the Gumbel case, we have*

$$\sum_{k=1}^{\infty} \mathbb{P}\big(\Theta^{(1)}(t_k) - \Theta^{(3)}(t_k) \leq \lambda_{t_k} t_k w_{t_k}\big) < \infty. \tag{4.2}$$

We prove Proposition 4.1 in Section 4.2. Here is a brief summary of how, in Sections 4.3 to 4.5, we conclude the proof of Theorem 2.8 once Proposition 4.1 has been established. We argue in two steps that the fraction $\sum_{j=3}^{M(t)} Z_{m_j}(t)/Z_{m_1}(t)$ converges to zero almost surely. First, in Section 4.3 we show that it suffices to consider the process at times $(t_k)_{k\geq 1}$, which are sufficiently dense and, by (4.2), at these times, the largest and third-largest exponents are well-separated. Second, in Section 4.4, we show that $Z_{m_j}(t)$ indeed grows like $\exp(\Theta^{(j)}(t))$, using the large deviations estimate for Yule processes given in Lemma C.1. Therefore the fraction $\sum_{j=3}^{M(t)} Z_{m_j}(t_k)/Z_{m_1}(t_k)$ is bounded by $M(t)\exp(\Theta^{(3)}(t_k) - \Theta^{(1)}(t_k))$ and by (4.2) the exponent is smaller than $-\lambda_{t_k} t_k w_{t_k}$ almost surely for all k large enough. In Section 4.5, we deduce the same result for $\sum_{j=3}^{M(t)} Z_{n_3}(t)/Z_{n_1}(t)$, where $n_i = n_i(t)$ is the index of the ith largest table at time t (which may be different from the index of the ith largest exponent).

4.1. *Potter bounds.* In the proofs, the following Potter bounds for slowly varying functions will be useful, see Theorem 1.5.6(i) in [4].

- If $L(x)$ is positive and slowly varying as $x \to \infty$, then for any $\delta, C_1, C_2 > 0$, there exists $x_0 = x_0(\delta, C_1, C_2) > 0$ such that, for all $x \geq x_0$,

$$C_1 x^{-\delta} \leq L(x) \leq C_2 x^{\delta}. \tag{4.3}$$

- If $\ell(x)$ is positive and slowly varying as $x \to 0$, then, for any $\delta, c_1, c_2 > 0$, there exists $x_0 = x_0(\delta, c_1, c_2) > 0$ such that, for all $|x| \leq x_0$,

$$c_1 x^{\delta} \leq \ell(x) \leq c_2 x^{-\delta}. \tag{4.4}$$

Below, we will typically write $L, L_1, L_2, \ldots$ for functions slowly varying at infinity and $\ell, \ell_1, \ell_2, \ldots$ for functions slowly varying at zero.

4.2. *Proof of Proposition* 4.1. To prove (4.2) we consider the following normalised version of the exponents. For all $t \geq 0$, $1 \leq n \leq M(t)$, let

$$\xi_n(t) = \frac{W_n(t-\tau_n) - tv_t}{tw_t} = \frac{\Theta_n(t) - tv_t}{tw_t} \tag{4.5}$$

and introduce also their order statistics, $\xi^{(1)}(t) \geq \xi^{(2)}(t) \geq \xi^{(3)}(t) \geq \dots$ We aim to find an upper bound for $\mathbb{P}(\xi^{(1)}(t) - \xi^{(3)}(t) \leq \lambda_t)$. Note that the $\xi_n(t)$ are all negative in the Weibull case (since $v_t = 1$), all positive in the Fréchet case (since $v_t = 0$), and can be either positive or negative in the Gumbel case.

For all $t \geq 0$ and $x \in \mathbb{R}$, we let

$$A_t(x) = \{(s, w) \in [0, t] \times [0, M) : w(t-s) > tv_t + xtw_t\}. \tag{4.6}$$

Then the event that $\xi_n(t) > x$ is the same as the event that $(\tau_n, W_n) \in A_t(x)$. We let $\Pi := ((\tau_n, W_n))_{n \geq 1}$, which is a Poisson point process on $[0, \infty) \times [0, M)$. We write $\pi := \theta \mathrm{d}s \otimes \mathrm{d}\mu$ for the intensity measure of Π.

Recall from (2.8) that $\Phi_t(x) = u_t \mu(v_t + xw_t, M)$, which is nonincreasing in x.

LEMMA 4.2. *Let* $x > -v_t/w_t$. *Then*

$$\pi(A_t(x)) = \theta(v_t + xw_t) \frac{tw_t}{u_t} \int_x^\infty \frac{\Phi_t(z)}{(v_t + zw_t)^2} \,\mathrm{d}z, \tag{4.7}$$

and for $\varepsilon > 0$,

$$0 \leq \pi(A_t(x)) - \pi(A_t(x+\varepsilon)) \leq \varepsilon\theta \frac{tw_t}{u_t} \frac{\Phi_t(x)}{v_t + xw_t}. \tag{4.8}$$

PROOF. For (4.7) we use that

$$\pi(A_t(x)) = \int_0^t \int_0^\infty \theta \,\mathrm{d}s \,\mathrm{d}\mu(w) \mathbf{1}_{(s,w)\in A_t(x)} = \int_0^t \theta \,\mathrm{d}s \mu\left(\frac{tv_t + xtw_t}{t-s}, M\right)$$

$$= \theta(v_t + xw_t) \int_x^\infty \frac{tw_t \mu(v_t + zw_t, M)}{(v_t + zw_t)^2} \,\mathrm{d}z \tag{4.9}$$

$$= \theta(v_t + xw_t) \frac{tw_t}{u_t} \int_x^\infty \frac{\Phi_t(z)}{(v_t + zw_t)^2} \,\mathrm{d}z, \tag{4.10}$$

where we used the change of variable $\frac{tv_t + xtw_t}{t-s} = v_t + zw_t$, to go from s to z. For (4.8), using $A_t(x+\varepsilon) \subseteq A_t(x)$ and (4.7), discarding a term which is ≤ 0, we get

$$0 \leq \pi(A_t(x)) - \pi(A_t(x+\varepsilon)) \leq \theta(v_t + xw_t) \frac{tw_t}{u_t} \int_x^{x+\varepsilon} \frac{\Phi_t(z)}{(v_t + zw_t)^2} \,\mathrm{d}z. \tag{4.11}$$

We then use the fact that the integrand is nonincreasing in z (because $z \geq x > -v_t/w_t$) to get the result. □

LEMMA 4.3. *Under Assumption* 2.1 *we have the following bounds*:

- *In the Weibull case, let* $x_t > 0$ *such that* $x_t w_t \to 0$, *for any* $\delta, C > 0$, *for all* t *large enough*,

$$\pi(A_t(-x_t)) \geq C x_t^{1+\alpha+\delta} t^{-\delta(1+\alpha+\delta+\frac{1}{1+\alpha})}, \tag{4.12}$$

and, whenever $\xi - \lambda \geq -x_t$, *we have*

$$\pi(A_t(\xi - \lambda)) - \pi(A_t(\xi)) \leq C\lambda x_t^{\alpha-\delta} t^{\delta(1+\alpha+\delta+\frac{1}{1+\alpha})}. \tag{4.13}$$

- *In the Fréchet case, let $x_t > 0$ such that $x_t w_t \to \infty$, for any $\delta, C > 0$, for all t large enough,*

$$\pi\big(A_t(x_t)\big) \ge C x_t^{-(\alpha+\delta)} t^{-\delta(\alpha+\frac{1}{\alpha}+\delta)}, \tag{4.14}$$

and, whenever $\xi - \lambda \ge x_t$, we have

$$\pi\big(A_t(\xi-\lambda)\big) - \pi\big(A_t(\xi)\big) \le C\lambda x_t^{-(1+\alpha-\delta)} t^{\delta(1/\alpha+\alpha-\delta)}. \tag{4.15}$$

- *In the Gumbel case, let $x_t > 0$ such that $x_t w_t / v_t \to 0$, if $M = 1$ then $\frac{x_t w_t}{1-v_t} \to 0$,*

$$\pi\big(A_t(-x_t)\big) \ge \big(\theta + o(1)\big) \int_{-x_t}^{x_t} \Phi_t(z)\mathrm{d}z, \tag{4.16}$$

as $t \uparrow \infty$, and, whenever $\xi - \lambda \ge -x_t$, we have

$$\pi\big(A_t(\xi-\lambda)\big) - \pi\big(A_t(\xi)\big) \le C\lambda \Phi_t(-x_t). \tag{4.17}$$

PROOF. We argue separately for the three cases. It is helpful to refer to Table 1.

- In the Weibull case, $M = 1$ and $v_t = 1$. Then $\Phi_t(z) = 0$ as soon as $z \ge 0$. Also $u_t = t w_t$. For any $-1/w_t < x < 0$, by (4.7),

$$\pi\big(A_t(x)\big) = \theta(1 + x w_t) \int_x^0 \frac{\Phi_t(z)}{(1 + z w_t)^2}\mathrm{d}z \ge \theta(1 + x w_t) \int_x^0 \Phi_t(z)\mathrm{d}z. \tag{4.18}$$

Replacing x with $-x_t$ and using that $1 - x_t w_t = 1 + o(1)$ we get

$$\pi\big(A_t(-x_t)\big) \ge \big(\theta + o(1)\big) \int_0^{x_t} \Phi_t(-z)\mathrm{d}z. \tag{4.19}$$

Now recall that $\mu(1-\varepsilon, 1) = \varepsilon^\alpha \ell(\varepsilon)$, $u_t = t^{\frac{\alpha}{1+\alpha}} L(t)$, and $w_t = t^{-\frac{1}{1+\alpha}} L(t)$ (see Assumption 2.3 and Lemma 2.2). This, combined with the Potter bounds (4.3) and (4.4), gives for $0 \le z \le x_t$

$$\begin{aligned} \Phi_t(-z) &= z^\alpha L(t)^{1+\alpha} \ell(z w_t) \ge C_1 z^\alpha t^{-\delta(1+\alpha)} (z w_t)^\delta \\ &\ge C_2 z^{\alpha+\delta} t^{-\delta(1+\alpha+\delta+\frac{1}{1+\alpha})}. \end{aligned} \tag{4.20}$$

Then

$$\pi\big(A_t(-x_t)\big) \ge C_2\big(\theta + o(1)\big) t^{-\delta(1+\alpha+\delta-\frac{1}{1+\alpha})} \int_0^{x_t} z^{\alpha+\delta}\mathrm{d}z, \tag{4.21}$$

so (4.12) follows. For (4.13), we have from (4.8) that

$$\begin{aligned} \pi\big(A_t(\xi-\lambda)\big) - \pi\big(A_t(\xi)\big) &\le \theta\lambda \frac{\Phi_t(\xi-\lambda)}{1 + (\xi-\lambda)w_t} \le \theta\lambda \frac{\Phi_t(-x_t)}{1 - x_t w_t} \\ &= \big(\lambda\theta + o(1)\big)\Phi_t(-x_t). \end{aligned} \tag{4.22}$$

Similar to (4.25), the Potter bounds (4.3) give

$$\Phi_t(-x_t) \le C_3 x_t^{\alpha-\delta} t^{\delta(1+\alpha-\delta+\frac{1}{1+\alpha})}, \tag{4.23}$$

which gives (4.13).

- In the Fréchet case, $M = \infty$, $u_t = t$, $v_t = 0$, and $w_t = t^{\frac{1}{\alpha}} L(t)$, so for $x > 0$, (4.7) simplifies to

$$\pi\big(A_t(x)\big) = \theta x \int_x^\infty \frac{\Phi_t(z)}{z^2}\,\mathrm{d}x. \tag{4.24}$$

Moreover, $\mu(x,\infty) = x^{-\alpha} L_1(x)$. Using the Potter bounds (4.3) we get, for any $\delta > 0$, $z \geq x_t$, and t large enough,

$$\begin{aligned}\Phi_t(z) &= t(zw_t)^{-\alpha} L_1(zw_t) \geq C_1 t(zw_t)^{-\alpha-\delta} = C_1 z^{-\alpha-\delta} t^{-\delta/\alpha} L(t)^{-\alpha-\delta} \\ &\geq C_2 z^{-\alpha-\delta} t^{-\delta/\alpha-\delta(\alpha+\delta)}.\end{aligned} \tag{4.25}$$

Then

$$\begin{aligned}\pi\big(A_t(x_t)\big) &\geq C_2\big(\theta + o(1)\big) x_t t^{-\delta/\alpha-\delta(\alpha+\delta)} \int_{x_t}^{\infty} z^{-(2+\alpha+\delta)} \mathrm{d}u \\ &\geq C_3 t^{-\delta/\alpha-\delta(\alpha+\delta)} x_t^{-(\alpha+\delta)},\end{aligned} \tag{4.26}$$

as claimed in (4.14). Next, from (4.8) we have

$$\pi\big(A_t(\xi-\lambda)\big) - \pi\big(A_t(\xi)\big) \leq \lambda\theta \frac{\Phi_t(\xi-\lambda)}{\xi-\lambda} \leq \lambda\theta \frac{\Phi_t(x_t)}{x_t}. \tag{4.27}$$

Similar to (4.25), using the Potter bounds (4.3)

$$\Phi_t(x_t) \leq C_4 x_t^{-\alpha+\delta} t^{\delta(1/\alpha+\alpha-\delta)}, \tag{4.28}$$

which gives (4.15).

- In the Gumbel case, we use that $u_t v_t = t w_t$ and that $\Phi_t(z) = 0$ if $z \geq (M - v_t)/w_t$ to see that, by (4.7), for any $0 < x < \frac{v_t}{w_t}$,

$$\pi\big(A_t(-x)\big) = \theta\left(1 - x\frac{w_t}{v_t}\right) \int_{-x}^{(M-v_t)/w_t} \frac{\Phi_t(z)}{(1+z\frac{w_t}{v_t})^2} \mathrm{d}z. \tag{4.29}$$

We now note that $x_t \leq (M - v_t)/w_t$ for all t large enough. This is immediate if $M = \infty$, while if $M = 1$ then it follows from the assumption $x_t w_t/(1 - v_t) \to 0$. Since the integrand in (4.29) is nonnegative, we get that, as $t \uparrow \infty$,

$$\pi\big(A_t(-x_t)\big) \geq \frac{\theta(1 - x_t w_t/v_t)}{(1 + x_t w_t/v_t)^2} \int_{-x_t}^{x_t} \Phi_t(z)\,\mathrm{d}z \geq \big(\theta + o(1)\big) \int_{-x_t}^{x_t} \Phi_t(z)\,\mathrm{d}z, \tag{4.30}$$

because $(1 + z\frac{w_t}{v_t})^{-2} \geq (1 + x_t\frac{w_t}{v_t})^{-2}$ for all $z \leq x_t$, and because $x_t = o(v_t/w_t)$ as $t \uparrow \infty$. Next, from (4.8), we get that, for all ξ and λ such that $\xi - \lambda \geq -x_t$,

$$\begin{aligned}\pi\big(A_t(\xi-\lambda)\big) - \pi\big(A_t(\xi)\big) &\leq \lambda\theta \frac{\Phi_t(\xi-\lambda)}{1 + (\xi-\lambda)\frac{w_t}{v_t}} \leq \lambda\theta \frac{\Phi_t(-x_t)}{1 - x_t\frac{w_t}{v_t}} \\ &\leq \lambda\big(\theta + o(1)\big)\Phi_t(-x_t),\end{aligned} \tag{4.31}$$

as $t \uparrow \infty$, as required for (4.17). □

Using Lemma 4.3, we deduce the following key estimates on $\xi^{(1)}(t) - \xi^{(3)}(t)$.

LEMMA 4.4. *Under Assumption 2.1 for the Weibull and Fréchet cases, and Assumption 2.5 for the Gumbel case, let $\lambda_t = t^{-\kappa}$ where $\kappa > 0$. Then for any $\gamma, C > 0$, for all t large enough,*

$$\mathbb{P}\big(\xi^{(1)}(t) - \xi^{(3)}(t) \leq \lambda_t\big) \leq C t^{\gamma} \lambda_t^2. \tag{4.32}$$

PROOF. For any $y_t \in \mathbb{R}$ we have the decomposition

$$\begin{aligned}&\mathbb{P}\big(\xi^{(1)}(t) - \xi^{(3)}(t) \leq \lambda_t\big) \\ &\quad\leq \mathbb{P}\big(\xi^{(1)}(t) \leq y_t\big) + \mathbb{P}\big(\xi^{(1)}(t) - \xi^{(3)}(t) \leq \lambda_t \text{ and } \xi^{(1)}(t) > y_t\big).\end{aligned} \tag{4.33}$$

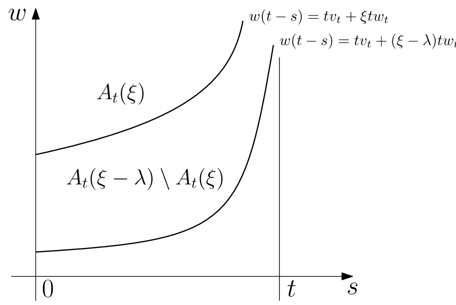

FIG. 2. *Intuition behind* (4.35).

We will use this for $y_t > -v_t/w_t + \lambda_t$. Note that

$$\mathbb{P}(\xi^{(1)}(t) \le y_t) = \mathbb{P}(\Pi(A_t(y_t)) = \varnothing) = \exp(-\pi(A_t(y_t))). \tag{4.34}$$

For the other term in (4.33), note that $(\xi_n(t))_{n\ge 0}$ is a Poisson point process and let $\rho_t(\cdot)$ denote its intensity measure. Then using Mecke's formula (see, e.g., [11], Theorem 4.1) and simple properties of Poisson random variables, we have for all $t > 0$,

$$\begin{aligned} &\mathbb{P}(\xi^{(1)}(t) - \xi^{(3)}(t) \le \lambda_t \text{ and } \xi^{(1)}(t) > y_t) \\ &\quad = \int_{y_t}^{\infty} \mathrm{d}\rho_t(\xi)\mathbb{P}(\Pi(A_t(\xi)) = \varnothing)\mathbb{P}(|\Pi(A_t(\xi - \lambda_t)) \setminus \Pi(A_t(\xi))| \ge 2) \\ &\quad \le \int_{y_t}^{\infty} \mathrm{d}\rho_t(\xi)\mathbb{P}(\Pi(A_t(\xi)) = \varnothing)(\pi(A_t(\xi - \lambda_t)) - \pi(A_t(\xi)))^2. \end{aligned} \tag{4.35}$$

The intuition behind the first equality is that we integrate over all possible values ξ for $\xi^{(1)}(t)$. For ξ to be maximal, there needs to be one point of the point process at ξ, and none larger (hence the term $\mathbb{P}(\Pi(A_t(\xi)) = \varnothing)$); for $\xi^{(1)}(t) - \xi^{(3)}(t) \le \lambda_t$, there needs to be at least two points in $A_t(\xi - \lambda_t)) \setminus A_t(\xi)$. (See Figure 2.) Note that, using Mecke's formula again,

$$\int_{y_t}^{\infty} \mathrm{d}\rho_t(\xi)\mathbb{P}(\Pi(A_t(\xi)) = \varnothing) = \mathbb{P}(\xi^{(1)}(t) > y_t) \le 1. \tag{4.36}$$

We proceed using Lemma 4.3.

- In the Weibull case, we set $x_t = t^{\varepsilon}$ for $0 < \varepsilon < \frac{1}{1+\alpha}$ and in the decomposition (4.33) we set $y_t = -\frac{1}{2}x_t$. As $w_t = t^{-\frac{1}{1+\alpha}}L(t)$ we have $x_t w_t \to 0$. Then, by (4.12), for all t large enough

$$\mathbb{P}(\xi^{(1)}(t) \le y_t) \le \exp(-Ct^{(\varepsilon-\delta)(1+\alpha+\delta)-\frac{\delta}{1+\alpha}}). \tag{4.37}$$

  Using (4.13) in (4.35) and applying (4.36) we get that, for all t large enough,

$$\mathbb{P}(\xi^{(1)}(t) - \xi^{(3)}(t) \le \lambda_t \text{ and } \xi^{(1)}(t) > y_t) \le C\lambda_t^2 t^{2\varepsilon(\alpha-\delta)+2\delta(1+\alpha+\delta+\frac{1}{1+\alpha})}. \tag{4.38}$$

  Clearly we may select $\varepsilon, \delta > 0$ small enough that (4.37) and (4.38) are both at most $Ct^{\gamma}\lambda_t^2$, for any $\gamma > 0$.
- In the Fréchet case, we set $x_t = t^{-\varepsilon}$ for $0 < \varepsilon < \frac{1}{\alpha}$ and $\varepsilon < \kappa$ and we set $y_t = x_t$. Since $w_t = t^{\frac{1}{\alpha}}L(t)$, we have $x_t w_t \to \infty$. By (4.14), for all t large enough,

$$\mathbb{P}(\xi^{(1)}(t) \le y_t) \le \exp(-Ct^{\varepsilon(\alpha+\delta)-\delta(\alpha+1/\alpha+\delta)}). \tag{4.39}$$

  Using (4.15) in (4.35) and applying (4.36) we get that, for all t large enough,

$$\mathbb{P}(\xi^{(1)}(t) - \xi^{(3)}(t) \le \lambda_t \text{ and } \xi^{(1)}(t) > y_t) \le C\lambda_t^2 t^{2\varepsilon(1+\alpha-\delta)+2\delta(1/\alpha+\alpha-\delta)}. \tag{4.40}$$

  Clearly we may select $\varepsilon, \delta > 0$ small enough that (4.39) and (4.40) are both at most $Ct^{\gamma}\lambda_t^2$, for any $\gamma > 0$.

- In the Gumbel case, let us for ease of reference recall Assumption 2.5(i):

$$\begin{cases} \Phi_t(x) \ge \mathrm{e}^{-x-c_1x^2/\log t} & \text{for all } x \in (-c_2\log t, c_2\log t), \\ \Phi_t(x) \le \mathrm{e}^{-x+c_1x^2/\log t} & \text{for all } x \in \left(-c_2\log t, \dfrac{M-v_t}{w_t}\right). \end{cases} \tag{4.41}$$

Set $x_t = 2\log\log t$. By Assumption 2.5(ii), $x_t w_t / v_t \to 0$ and thus Lemma 4.3 applies; together with the lower bound in (4.41), this gives, for all t large enough,

$$\begin{aligned} \mathbb{P}(\xi^{(1)}(t) \le -x_t) &\le \exp\left[-(\theta+o(1))\int_{-x_t}^{x_t}\Phi_t(z)\,\mathrm{d}z\right] \\ &\le \exp\left[-(\theta+o(1))\exp(-c_1x_t^2/\log t)\int_{-x_t}^{x_t}\mathrm{e}^{-z}\,\mathrm{d}z\right]. \end{aligned} \tag{4.42}$$

Since $x_t^2/\log t \to 0$ we get that, as $t\uparrow\infty$,

$$\mathbb{P}(\xi^{(1)}(t) \le -x_t) \le \exp(-(\theta+o(1))\mathrm{e}^{x_t}) = \mathrm{e}^{-(\theta+o(1))(\log t)^2} \le \lambda_t^2, \tag{4.43}$$

for all t large enough, because $\lambda_t = t^{-\kappa}$. Next, (4.17) gives that, for $\xi - \lambda_t \ge -x_t$,

$$\begin{aligned} \pi\big(A_t(\xi-\lambda_t)\big) - \pi\big(A_t(\xi)\big) &\le \lambda_t\big(\theta+o(1)\big)\Phi_t(-x_t) \le \lambda_t\big(\theta+o(1)\big)\mathrm{e}^{x_t+c_1x_t^2/\log t} \\ &= \lambda_t\exp\big(x_t(1+o(1))\big), \end{aligned} \tag{4.44}$$

because $x_t/\log t = o(1)$ as $t\uparrow\infty$. Thus, in total,

$$\mathbb{P}\big(\xi^{(1)}(t) - \xi^{(3)}(t) \le \lambda_t \text{ and } \xi^{(1)}(t) > -x_t\big) \le \lambda_t^2\exp\big(2x_t(1+o(1))\big).$$

As $\mathrm{e}^{2x_t(1+o(1))} \le t^\gamma$, for t large enough, this concludes the proof. □

PROOF OF PROPOSITION 4.1. By Lemma 4.4, for any $\gamma, C > 0$, for all k large enough,

$$\mathbb{P}(\xi^{(1)}(t_k) - \xi^{(3)}(t_k) \le \lambda_{t_k}) \le C t_k^{2\kappa-\gamma} = Ck^{-(2\kappa\eta-\eta\gamma)}. \tag{4.45}$$

Since $2\kappa\eta > 1$ we can choose $\gamma > 0$ so that $2\kappa\eta - \eta\gamma > 1$. It follows that $\mathbb{P}(\xi^{(1)}(t_k) - \xi^{(3)}(t_k) \le \lambda_{t_k})$ are summable, as required. □

4.3. *Interpolation.* By Proposition 4.1 and the Borel–Cantelli lemma, almost surely, there exists k_0 such that, for all $k \ge k_0$,

$$\Theta^{(1)}(t_k) - \Theta^{(3)}(t_k) > \lambda_{t_k} t_k w_{t_k}. \tag{4.46}$$

We now show that we can "interpolate" between the times t_k:

PROPOSITION 4.5. *As in Proposition 4.1, set $\lambda_t = t^{-\kappa}$ and $t_k = k^\eta$ with $\kappa, \eta > 0$ satisfying $2\kappa\eta > 1$. Assume further that:*

- *In the Weibull case, that $\frac{1}{\eta} > \kappa + \frac{1}{1+\alpha}$,*
- *In the Fréchet case, $\frac{1}{\eta} > \kappa + \frac{1}{\alpha}$,*
- *In the Gumbel case, Assumption 2.5.*

Then, almost surely, there exists k_1 such that, for all $k \ge k_1$,

$$\inf_{t\in[t_{k-1},t_k]}\big(\Theta^{(1)}(t) - \Theta^{(3)}(t)\big) > \frac{1}{2}\lambda_{t_k} t_k w_{t_k}. \tag{4.47}$$

To prove Proposition 4.5, we use the following:

LEMMA 4.6. *Let $(t_k)_{k\geq 0}$ be an increasing sequence such that $t_0 = 0$. For all $k \geq 1$, for all $t \in [t_{k-1}, t_k)$,*

$$\Theta^{(1)}(t) - \Theta^{(3)}(t) \geq \Theta^{(1)}(t_k) - \Theta^{(3)}(t_k) - W_{m_1(t_k)}(t_k - t_{k-1}). \tag{4.48}$$

PROOF. Each $\Theta_n(t)$ is an increasing (affine) function of t. Hence, for all $i \geq 1$, $\Theta^{(i)}(t)$ is increasing in t. In particular, for $t \in [t_{k-1}, t_k)$,

$$\begin{aligned}\Theta^{(1)}(t) - \Theta^{(3)}(t) &\geq \Theta^{(1)}(t_{k-1}) - \Theta^{(3)}(t_k)\\ &= \Theta^{(1)}(t_k) - \Theta^{(3)}(t_k) - \big[\Theta^{(1)}(t_k) - \Theta^{(1)}(t_{k-1})\big].\end{aligned} \tag{4.49}$$

Because the largest exponent at time t_k can only be larger than the largest exponent at time t_{k-1}, we have $0 \leq \Theta^{(1)}(t_k) - \Theta^{(1)}(t_{k-1})$. Furthermore,

$$0 \leq \Theta^{(1)}(t_k) - \Theta^{(1)}(t_{k-1}) \leq W_{m_1(t_k)}(t_k - t_{k-1}). \tag{4.50}$$

Indeed, the second inequality comes from the fact that

$$\Theta^{(1)}(t_k) = W_{m_1(t_k)}(t_k - \tau_{m_1(t_k)}) = W_{m_1(t_k)}(t_k - t_{k-1}) + W_{m_1(t_k)}(t_{k-1} - \tau_{m_1(t_k)}).$$

If $\tau_{m_1(t_k)} > t_{k-1}$, then $\Theta^{(1)}(t_k) \leq W_{m_1(t_k)}(t_k - t_{k-1})$ and we indeed have (4.50). Otherwise, $W_{m_1(t_k)}(t_{k-1} - \tau_{m_1(t_k)})$ is at most equal to the largest exponent at time t_{k-1}, which is, by definition, $\Theta^{(1)}(t_{k-1})$. This indeed implies (4.50). □

In the Gumbel case, we need the following facts about slowly varying functions.

LEMMA 4.7. *Let $L : (1, \infty) \to (0, \infty)$ be a nondecreasing function, slowly varying at infinity, such that $L(x) \uparrow \infty$ as $x \uparrow \infty$, and let L^{-1} be its generalised inverse.*

1. *For any $\beta > 0$,*

$$\sum_{n\geq 1} \frac{n}{L^{-1}(n^\beta)} < \infty.$$

2. *For any $\varepsilon > 0$, as $n \uparrow \infty$,*

$$\frac{n}{L^{-1}(L(n^{1+\varepsilon}))} \to 0,$$

PROOF. *(1)* By the Potter bounds, for any $\delta > 0$, there exists $x_0 = x_0(\delta)$ such that, for all $x \geq x_0$, $L(x) \leq x^\delta$. Because L is nondecreasing, so is L^{-1}, and we get that

$$L^{-1}\big(L(x)\big) \leq L^{-1}(x^\delta),$$

which implies, because $L^{-1}(L(x)) \geq x$, that $x \leq L^{-1}(x^\delta)$. Taking $y = x^\delta$, we get that $L^{-1}(y) \geq y^{1/\delta}$. Taking δ large enough such that $\beta/\delta > 2$ concludes the proof.

(2) For any $K > 0$, for all n large enough, $L(n)^\varepsilon \geq K$. Thus, because L^{-1} is nondecreasing, $L^{-1}(L(n)^{1+\varepsilon}) \geq L^{-1}(K L(n))$. By [4], Theorem 2.7(i), L^{-1} is rapidly varying, which, by definition, implies that

$$\frac{L^{-1}(L(n))}{L^{-1}(K L(n))} \to 0, \qquad \text{as } n \uparrow \infty.$$

Thus,

$$\frac{n}{L^{-1}(L(n)^{1+\varepsilon})} \leq \frac{L^{-1}(L(n))}{L^{-1}(L(n)^{1+\varepsilon})} \to 0, \qquad \text{as } n \uparrow \infty,$$

which concludes the proof. □

In the Fréchet case we also need the following almost sure estimate for the maximum weight of the n first tables.

LEMMA 4.8. *Under Assumption 2.1 in the Fréchet case, for any* $\varepsilon > 0$, *almost surely for* n *large enough*, $\max_{1\le i\le n} W_i \le n^{2/\alpha+\varepsilon}$.

PROOF. Using that $\mu(x,\infty) = x^{-\alpha}L(x)$, where $L(x)$ is slowly varying at ∞, we get

$$\mathbb{P}\Big(\max_{1\le i\le n} W_i > n^{2/\alpha+\varepsilon}\Big) \le n\mathbb{P}\big(W_1 \ge n^{2/\alpha+\varepsilon}\big) = n^{-1-\varepsilon\alpha}L\big(n^{2/\alpha+\varepsilon}\big). \tag{4.51}$$

Then, by the Potter bounds, $\sum_{n\ge1}\mathbb{P}(\max_{1\le i\le n} W_i > n^{2/\alpha+\varepsilon}) < \infty$ so the result follows from the Borel–Cantelli lemma. □

PROOF OF PROPOSITION 4.5. Note that (4.46) combined with Lemma 4.6 gives that, for all $k \ge k_0$ and $t \in [t_{k-1}, t_k)$,

$$\Theta^{(1)}(t) - \Theta^{(3)}(t) \ge \lambda_{t_k} t_k w_{t_k} - \overline{W}(t_k)(t_k - t_{k-1}), \tag{4.52}$$

where $\overline{W}(t)$ is the largest table weight up to time t. We argue that

$$\frac{\lambda_{t_k} t_k w_{t_k}}{\overline{W}(t_k)(t_k - t_{k-1})} \to \infty \quad \text{almost surely}, \tag{4.53}$$

which gives the claim. Note that $t_k - t_{k-1} = k^\eta - (k-1)^\eta = (\eta + o(1))t_k^{1-1/\eta}$ as $k \to \infty$. For what follows, it is useful to refer to Table 1 for expressions for w_t.

- In the Weibull case, $\overline{W}(t) \le 1$ almost surely, and $w_t = t^{-\frac{1}{1+\alpha}}L_0(t)$ where $L_0(t)$ is slowly varying at infinity. Then, almost surely as $k \to \infty$,
$$\frac{\lambda_{t_k} t_k w_{t_k}}{\overline{W}(t_k)(t_k - t_{k-1})} \ge \big(1/\eta + o(1)\big)t_k^{\frac{1}{\eta}-\kappa-\frac{1}{1+\alpha}}L_0(t_k) \to \infty, \quad \text{since } \frac{1}{\eta} > \kappa + \frac{1}{1+\alpha}.$$
- In the Fréchet case, first note that $M(t_k) \to \infty$ almost surely, and by large-deviations estimates for Poisson random variables, almost surely for all k large enough, $M(t_k) \le 2\theta t_k$. It follows from Lemma 4.8 that
$$\overline{W}(t_k) \le (2\theta t_k)^{2/\alpha+\varepsilon} \quad \text{almost surely for all } k \text{ large enough.} \tag{4.54}$$
Also, $w_t = t^{1/\alpha}L_3(t)$ where $L_3(t)$ is slowly varying at infinity. Then, as $k \uparrow \infty$,
$$\frac{\lambda_{t_k} t_k w_{t_k}}{\overline{W}(t_k)(t_k - t_{k-1})} \ge \big(1/\eta + o(1)\big)t_k^{\frac{1}{\eta}-\kappa-\frac{1}{\alpha}-\varepsilon}L_3(t_k). \tag{4.55}$$
Since $\frac{1}{\eta} > \kappa + \frac{1}{\alpha}$ we can find $\varepsilon > 0$ such that $\frac{1}{\eta} > \kappa + \frac{1}{\alpha} + \varepsilon$. Then (4.53) follows.
- In the Gumbel case we get
$$\frac{\lambda_{t_k} t_k w_{t_k}}{\overline{W}(t_k)(t_k - t_{k-1})} \ge \big(1/\eta + o(1)\big)\frac{t_k^{\frac{1}{\eta}-\kappa}L_1(t_k)L_2(t_k)}{\overline{W}(t_k)} \tag{4.56}$$
where $L_1(t) \to 0$ and $L_2(t) \to M$ are both slowly varying at infinity. In the bounded case $M = 1$, (4.53) follows for any $\kappa > 0$ picking any $\frac{1}{\eta} \in (\kappa, 2\kappa)$. In the unbounded case $M = \infty$, for any $\rho > 0$, and $n \ge 1$,
$$\mathbb{P}\Big(\max_{1\le i\le n} W_i > n^\rho\Big) \le n\mu(n^\rho, \infty). \tag{4.57}$$
By Assumption 2.1, we have $A^{-1}(n^\rho)\mu(A(A^{-1}(n^\rho)), \infty) \to 1$. Because, in the Gumbel case, A is increasing, we get
$$\mu(n^\rho, \infty) = \frac{1 + o(1)}{A^{-1}(n^\rho)}.$$

Thus, by (4.57),

$$\mathbb{P}\Big(\max_{1\le i\le n} W_i > n^\rho\Big) \le \frac{n(1+o(1))}{A^{-1}(n^\rho)},$$

which, by Lemma 4.7 is summable, because, in the Gumbel case, A is slowly varying (as proved in the proof of Lemma 2.2). Arguing similar to the Fréchet case, we get that $\overline{W}(t_k) \le M(t_k)^\rho \le (2\theta t_k)^\rho$ almost surely for k large enough. Choosing $\rho > 0$ such that $\kappa < \frac{1}{\eta} + \rho < 2\kappa$ (4.53) follows. □

With the results obtained so far we get the following proposition.

PROPOSITION 4.9. *Under the same assumptions as for Proposition 4.5,*

$$\frac{\exp(\Theta^{(1)}(t)) + \exp(\Theta^{(2)}(t))}{\sum_{n=1}^{M(t)} \exp(\Theta_n(t))} \to 1 \quad \textit{almost surely}. \tag{4.58}$$

PROOF. Fix $t > 0$. We have

$$0 \le 1 - \frac{\exp(\Theta^{(1)}(t)) + \exp(\Theta^{(2)}(t))}{\sum_{n=1}^{M(t)} \exp(\Theta_n(t))} \le \frac{M(t)\mathrm{e}^{\Theta^{(3)}(t)}}{\mathrm{e}^{\Theta^{(1)}(t)}}. \tag{4.59}$$

Let $k = k(t)$ be such that $t \in [t_{k-1}, t_k]$. Then, almost surely for all t large enough, by Proposition 4.5 and a large-deviations bound for $M(t_k)$,

$$\frac{M(t)\mathrm{e}^{\Theta^{(3)}(t)}}{\mathrm{e}^{\Theta^{(1)}(t)}} \le 2\theta t_k \exp\Big(-\frac{1}{2}\lambda_{t_k} t_k w_{t_k}\Big). \tag{4.60}$$

We now check that the right-hand side goes to zero as t (and thus $k = k(t)$) goes to infinity:

- In the Weibull case, $\lambda_{t_k} t_k w_{t_k} = t_k^{1-\kappa-\frac{1}{1+\alpha}} L_0(t_k)$ so (4.60) goes to zero provided we select $\kappa < 1 - \frac{1}{1+\alpha}$ and then η as in Proposition 4.5.
- In the Fréchet case, $\lambda_{t_k} t_k w_{t_k} = t_k^{1-\kappa+\frac{1}{\alpha}} L_3(t_k)$ so (4.60) goes to zero provided we select $\kappa < 1 + \frac{1}{\alpha}$ and then η as in Proposition 4.5.
- In the Gumbel case, $\lambda_{t_k} t_k w_{t_k} = t_k^{1-\kappa} L_1(t_k) L_2(t_k)$ so (4.60) goes to zero provided we select $\kappa < 1$ and then η as in Proposition 4.5. □

Proposition 4.9 can be seen as an analog of Theorem 2.8 where we have replaced the tables with their growth rates. To establish Theorem 2.8 we need to argue that $\exp(\Theta^{(j)}(t))$ is a good approximation of the size of the jth largest table.

4.4. *Approximating the table sizes.*

PROPOSITION 4.10. *Let $t_k = k^\eta$ with $\eta > 0$ and let $\varphi \in (0,1)$. Then, almost surely, for all k large enough,*

$$\textit{for all } m \textit{ with } \tau_m \le t_k, \sup_{t \ge \tau_m} \big|\log Z_m(t) - \Theta_m(t)\big| \le t_k^{1-\varphi}.$$

PROOF. We aim at using the Borel–Cantelli lemma, and thus prove that $\mathbb{P}(A_k)$ is summable where

$$A_k = \Big\{\exists m \colon \tau_m \le t_k \text{ and } \sup_{t \ge \tau_m} \big|\log Z_m(t) - \Theta_m(t)\big| > t_k^{1-\varphi}\Big\}.$$

We have

$$
\begin{aligned}
\mathbb{P}(A_k) &\le \mathbb{E}\Big[\sum_{\tau_m\le t_k}\mathbb{P}\Big(\sup_{t\ge\tau_m}\big|\log Z_m(t)-\Theta_m(t)\big|>t_k^{1-\varphi}\mid W_m,\tau_m\Big)\Big] \\
&= \mathbb{E}\Big[\sum_{\tau_m\le t_k}\mathbb{P}\Big(\sup_{t\ge\tau_m}\big|\log Z_m(t)-W_m(t-\tau_m)\big|>t_k^{1-\varphi}\mid W_m,\tau_m\Big)\Big].
\end{aligned}
\tag{4.61}
$$

We now use Lemma C.1 with $\lambda = W_m$ and $R = t_k^{1-\varphi}$ to get that, for all integers m such that $\tau_m \le t_k$,

$$
\mathbb{P}\Big(\sup_{t\ge\tau_m}\big|\log Z_m(t)-W_m(t-\tau_m)\big|>{t_k}^{1-\varphi}\mid W_m,\tau_m\Big)\le 2\Gamma(1/2)\mathrm{e}^{-t_k^{1-\varphi}/2}. \tag{4.62}
$$

Thus,

$$
\mathbb{P}(A_k)\le 2\Gamma(1/2)\mathrm{e}^{-t_k^{1-\varphi}/2}\mathbb{E}\big[M(t_k)\big]=2\theta\Gamma(1/2)t_k\mathrm{e}^{-t_k^{1-\varphi}/2}.
$$

Because $t_k = k^\eta$, this is summable as soon as $\varphi < 1$. □

4.5. *Proof of Theorem* 2.8. Recall that $m_j = m_j(t)$ denotes the index of the jth largest exponent Θ. Our first aim is to prove that

$$
\sup_{t\in[t_{k-1},t_k]}\frac{\sum_{j=3}^{M(t)}Z_{m_j}(t)}{Z_{m_1}(t)}\to 0,\qquad \text{almost surely as } k\to\infty. \tag{4.63}
$$

We do this before showing how to deduce the same claim about the largest tables, that is, Theorem 2.8. In the proof of (4.63), a delicate issue is to verify that there exists a choice of the parameters κ, η, and φ that satisfies all necessary assumptions.

PROOF OF (4.63). By Proposition 4.10, almost surely for all k large enough,

$$
\begin{aligned}
\sup_{t\in[t_{k-1},t_k]}\frac{\sum_{j=3}^{M(t)}Z_{m_j}(t)}{Z_{m_1}(t)} &\le \sup_{t\in[t_{k-1},t_k]}\frac{\sum_{j=3}^{M(t)}\exp(\Theta_{m_j}(t)+t_k^{1-\varphi})}{\exp(\Theta_{m_1}(t)-t_k^{1-\varphi})} \\
&\le \sup_{t\in[t_{k-1},t_k]} M(t)\exp\big(-(\Theta_{m_1}(t)-\Theta_{m_3}(t))+2t_k^{1-\varphi}\big) \\
&\le M(t_k)\exp\Big(-\frac{1}{2}t_k\lambda_{t_k}w_{t_k}+2t_k^{1-\varphi}\Big),
\end{aligned}
\tag{4.64}
$$

by Proposition 4.5. Using the fact that $M(t_k)\le 2\theta t_k$ almost surely for all k large enough (by a large deviation estimate for Poisson random variables, because $M(t_k)$ is a Poisson random variable of parameter θt_k), we get that, almost surely for all k large enough,

$$
\sup_{t\in[t_{k-1},t_k]}\frac{\sum_{j=3}^{M(t)}Z_{m_j}(t)}{Z_{m_1}(t)}\le 2\theta t_k\exp\Big(-\frac{1}{2}t_k\lambda_{t_k}w_{t_k}+2t_k^{1-\varphi}\Big) \tag{4.65}
$$

We need to check that the right-hand side of (4.65) converges to zero, that is, that $t_k\lambda_{t_k}w_{t_k}\gg t_k^{1-\varphi}$ as $k\uparrow\infty$. Recall that $\lambda_t = t^{-\kappa}$.

- Weibull case: $w_t = t^{-\frac{1}{1+\alpha}}L_0(t)$ so (4.63) follows as soon as $-\kappa+\varphi-\frac{1}{1+\alpha}>0$ that is,

$$
\varphi>\kappa+\frac{1}{1+\alpha}. \tag{4.66}
$$

- Fréchet case: $w_t = t^{\frac{1}{\alpha}} L_3(t)$ so (4.63) follows as soon as $-\kappa + \varphi + \frac{1}{\alpha} > 0$ that is,

$$\varphi > \kappa - \frac{1}{\alpha}. \tag{4.67}$$

- Gumbel case: w_t is slowly varying so (4.63) follows as soon as

$$\varphi > \kappa. \tag{4.68}$$

To conclude the proof, we need to check that there exists a choice of the parameters κ, η, ρ, and φ that satisfies all our assumptions. In all cases (Weibull, Gumbel, and Fréchet), we have $\lambda_t = t^{-\kappa}$ and $t_k = k^\eta$. Our first assumption is that $2\kappa\eta > 1$ coming from Proposition 4.1. In addition, for the three extreme-value distributions we have the following assumptions:

- Weibull: For Proposition 4.5, we need $\kappa + \frac{1}{1+\alpha} < \frac{1}{\eta} < 2\kappa$. For Proposition 4.10, we need $\varphi \in (0, 1)$, and for (4.66), we need $\varphi > \kappa + \frac{1}{1+\alpha}$. These inequalities can only be consistent if $\alpha > 1$, which is indeed the contents of Assumption 2.3. Assuming that $\alpha > 1$, we can satisfy all the inequalities as follows. Since $\frac{2}{1+\alpha} < 1$ we can pick some $\kappa > \frac{1}{1+\alpha}$ satisfying $\frac{2}{1+\alpha} < \kappa + \frac{1}{1+\alpha} < 1$. We can then pick any φ satisfying $\kappa + \frac{1}{1+\alpha} < \varphi < 1$, and any η satisfying $\frac{1}{\eta} < 2\kappa$.
- Fréchet: For Proposition 4.5, we need $\kappa + \frac{1}{\alpha} < \frac{1}{\eta} < 2\kappa$; for Proposition 4.10, we need $\varphi \in (0, 1)$; and for (4.67), we need $\varphi > \kappa - \frac{1}{\alpha}$. These inequalities are consistent for any $\alpha > 0$, which is why we do not need a stronger assumption in the Fréchet case. To show that they can all be satisfied, we start by picking κ such that $\frac{1}{\alpha} < \kappa < \frac{1+\alpha}{\alpha}$. Note that we then have that $\kappa + \frac{1}{\alpha} < 2\kappa$ and $\kappa - \frac{1}{\alpha} < 1$. We pick η such that $\kappa + \frac{1}{\alpha} < \frac{1}{\eta} < 2\kappa$ and φ such that $\kappa - \frac{1}{\alpha} < \varphi < 1$.
- Gumbel: Proposition 4.5 places no restrictions on the parameters. For Proposition 4.10, we need $\varphi \in (0, 1)$ and for (4.68) we need $\varphi > \kappa$. In this case we simply pick any κ, η such that $0 < \kappa < 1$ and $\kappa < \frac{1}{\eta} < 2\kappa$, and any $\varphi \in (\kappa, 1)$.

Having shown that the various inequalities can all be simultaneously satisfied, we conclude that the right-hand side of (4.64) goes to 0, which means that

$$\mathbb{P}\left(\frac{\sum_{j=3}^{M(t)} Z_{m_j}(t)}{Z_{m_1(t)}} \to 0 \quad \text{as } t \to \infty\right) = 1. \tag{4.69}$$

□

Now we show how to deduce Theorem 2.8.

PROOF OF THEOREM 2.8. Let

$$\mathcal{G} = \left\{\frac{\sum_{j=3}^{M(t)} Z_{m_j}(t)}{Z_{m_1}(t)} \to 0 \text{ as } t \to \infty\right\}. \tag{4.70}$$

Then (4.69) implies that $\mathbb{P}(\mathcal{G}) = 1$. Let $n_i = n_i(t)$ denote the index of the ith largest table. We claim that, on $\mathcal{G}$,

$$\frac{\sum_{j=3}^{M(t)} Z_{n_j}(t)}{Z_{n_1}(t)} \to 0 \quad \text{as } t \to \infty. \tag{4.71}$$

First note that $n_1(t) \in \{m_1(t), m_2(t)\}$ for all large enough t, since if $n_1(t) \notin \{m_1(t), m_2(t)\}$ then

$$\frac{\sum_{j=3}^{M(t)} Z_{m_j}(t)}{Z_{m_1}(t)} \geq \frac{Z_{n_1}(t)}{Z_{n_1}(t)} = 1,$$

which is not true on $\mathcal{G}$, for t large enough. Assume from now on that $n_1(t) \in \{m_1(t), m_2(t)\}$. Consider the sets

$$\mathcal{N}(t) = \{n_j(t) : 3 \le j \le M(t)\} \quad \text{and} \quad \mathcal{M}(t) = \{m_j(t) : 3 \le j \le M(t)\}.$$

These two sets have the same size, and neither contains $n_1(t)$. We have two cases: either $n_2(t) \in \mathcal{M}(t)$ or $n_2(t) \notin \mathcal{M}(t)$. If $n_2(t) \in \mathcal{M}(t)$ then there is some $j_0 \ge 3$ such that $n_{j_0}(t) \notin \mathcal{M}(t)$. If $n_2(t) \notin \mathcal{M}(t)$ then $\mathcal{M}(t) = \mathcal{N}(t)$, in which case we set $j_0 = 2$. In either case, since $j_0 \ge 2$ we have

$$\sum_{j=3}^{M(t)} Z_{m_j}(t) = \sum_{j=3}^{M(t)} Z_{n_j}(t) - Z_{n_{j_0}}(t) + Z_{n_2}(t) \ge \sum_{j=3}^{M(t)} Z_{n_j}(t).$$

Thus, using also that, by definition of $n_1(t)$, $Z_{n_1}(t) \ge Z_{m_1}(t)$, we get that, on $\mathcal{G}$,

$$\frac{\sum_{j=3}^{M(t)} Z_{n_j}(t)}{Z_{n_1}(t)} \le \frac{\sum_{j=3}^{M(t)} Z_{m_j}(t)}{Z_{m_1}(t)} \to 0,$$

as claimed. □

5. Further discussion.

Related models. Other variants of the Chinese restaurant process perturbed by a disorder have been considered by various authors.

• In [13], the authors discuss a model where customer $n+1$ chooses to sit at table i with random weight $0 < W_i < 1$ with probability $\frac{1}{n} S_i(n) W_i$ and occupies a new table with the remaining probability. As in our case the random weights are i.i.d. If the weight distribution has no atom at 1, the authors prove that, irrespective of the extreme value type of the weight distribution, the tables have microscopic occupancy and the ratio R_n of the largest and second largest table satisfies $\lim_{n\to\infty} \mathbb{P}(R_n \ge x) = 1/x$ for all $x \ge 1$.

• In [18], the authors introduce a "weighted" Chinese restaurant process, in which the customers are weighted instead of tables. In this model, the nth customer has weight W_n, and a new customer joins a table with probability proportional to the sum of the weights of the customers already sitting at that table, and they create a new table with probability proportional to a parameter θ. The focus in [18] is on cases where the weight distribution has an atom at the essential supremum. Even if this is not the case, at least for light-tailed weight distributions, we expect the tables to have macroscopic occupancy in this model, just as in the classical case. If $\theta = W_0$ is also a random weight, then the tables in this model can be seen as the subtrees of the root in the weighted random recursive tree (see, e.g., [19]) where this random tree is introduced and studied. The fact that tables in the original Chinese restaurant process can be seen as the subtrees of the root in the (nonweighted) random recursive tree is shown in [9].

• In the statistics literature (see, e.g., [8] and the references therein), a weighted Chinese restaurant process has been studied. In this model "*customers each have a fixed affiliation and are biased to sit at tables with other customers having similar affiliations*", see [12]. Affiliations can be seen as weights, and they are carried by the customers; however, their effect on the probability to join a given table is different from the model described in the second bullet point just above.

Further results. • In [15] an algorithm that gives access to queries about the Chinese restaurant process in sublinear time is presented. This algorithm is suitable for our model.

Open problems. An interesting challenge is to describe the length of the periods, in which the largest table remains the same as a function of time. We conjecture that, for all fitness distributions μ, these periods are stochastically increasing in time, a phenomenon known as *ageing*. As done in [14] for the parabolic Anderson model, one can describe this phenomenon in the weak sense, by looking at the asymptotic probability of a change of the largest table in a given time window, and in the strong sense, by identifying an almost sure upper envelope for the process of the time remaining until the next change of profile. For the winner takes all market this corresponds to an analysis of the slowing down in the rate of innovation as the market expands.

APPENDIX A: PROOF OF PROPOSITION 1.1

In this section we prove Proposition 1.1 of the Introduction. We use the continuous time embedding, in which our statement becomes

$$\lim_{t\to\infty} \frac{M(t)}{\log N(t)} = \frac{\theta}{\text{essup}\mu}.$$

Recall that from (2.1) that, in continuous time, $Z_i(t) = Y_i(W_i(t - \tau_i))$ where $(Y_i)_{i\geq 1}$ is a sequence of i.i.d. Yule processes of parameter 1 and, for all $i \geq 1$, τ_i is the time at which table i is first occupied. Also, by (2.2), almost surely as $t \uparrow \infty$ $Z_i(t) \sim \zeta_i \exp(W_i(t - \tau_i))$, where $(\zeta_i)_{i\geq 1}$ is a sequence of i.i.d. standard exponential random variables. We also recall that, by definition of the model,

$$M(t) \sim \theta t \quad \text{almost surely as } t \uparrow \infty. \tag{A.1}$$

First note that, for all $a < \text{essup}\mu$, there exists a random index $j \geq 1$ such that $W_j > a$. Thus, by (2.2), for all t large enough, $Z_j(t) \geq \exp(at)$. Hence, by (A.1), for all $\varepsilon > 0$, for all t is large enough, $M(t)/\log N(t) \leq (1 + \varepsilon)\theta/a$. If $\text{essup}\mu = \infty$, this concludes the proof, since one can make a arbitrarily large and conclude that $M(t)/\log N(t) \to 0$ almost surely as $t \uparrow \infty$, as claimed. In the case when $a := \text{essup}\mu < \infty$, note that, by (2.2), for all t large enough, $N(t) \leq 2\Xi_t \exp(at)$, where Ξ_t is the sum of $M(t)$ independent standard exponentials. Hence, for all $\varepsilon > 0$, for all sufficiently large t, $\log N(t) \leq (1 + \varepsilon)at$ and $M(t) \geq (1 - \varepsilon)t$, by (A.1), which implies $M(t)/\log N(t) \geq \frac{(1-\varepsilon)\theta}{(1+\varepsilon)a}$. Since $\varepsilon > 0$ is arbitrary, this implies (i).

Now fix a table number $i \in \mathbb{N}$. Recall that, by (2.2), $Z_i(t) \sim \zeta_i \exp(W_i(t - \tau_i))$, which clearly implies that $Z_i(t) \to \infty$ as $t \uparrow \infty$, because $\tau_i < \infty$ almost surely. Because μ has no atom at its essential supremum, there exists almost surely a random index $j \neq i$ such that $W_j > W_i$. Using (2.2) again, we get that $Z_i(t)/Z_j(t) \to 0$ as $t \uparrow \infty$ almost surely. If $N(t)$ denotes the number of customers in the restaurant at time t, then $Z_i(t)/N(t) \leq Z_i(t)/Z_j(t) \to 0$ as $t \uparrow \infty$ almost surely, so that table i cannot have macroscopic occupancy, as claimed in (ii) and (iii).

To see (iv), assume that the proportion of customers at the largest table converges almost surely to one. On this event, there exists $N > 4$ such that

$$\max_{i\geq 1} \frac{S_i(n)}{n} > \frac{3}{4} \quad \text{for all } n \geq N.$$

Let i_N denote the index of the unique largest table at time N: the function $f(n) := S_{i_N}(n)/n$ takes a value larger than $3/4$ at $n = N$ and, by (iii), it goes to zero as $n \to \infty$. Note that, for all $m \geq N$, $|f(m + 1) - f(m)| \leq \frac{1}{N}$ and hence there exists some $M \geq N$ such that

$$\left| f(M) - \frac{1}{2} \right| \leq \frac{1}{N}.$$

Hence, i_N is not the index of the largest table at time M, and for the index i_M of the largest table at time M we have $S_{i_M}(M)/M \leq (M - S_{i_N}(M))/M \leq 1/2 + \frac{1}{N}$, contradicting our assumption.

APPENDIX B: EXAMPLES OF WEIGHT DISTRIBUTIONS

B.1. Examples satisfying Assumption 2.1. We give four examples of probability distributions μ that satisfy Assumption 2.1; for each of these, we give formulas for $A(t)$, $B(t)$, u_t, v_t and w_t.

EXAMPLE B.1 (Weibull). For $\alpha > 0$ let $\mu(1-x, 1) = x^\alpha$ for all $x \in [0, 1]$. Then, for $x \geq 0$,

$$t\mu\big(1 - xt^{-1/\alpha}, 1\big) = x^\alpha,$$

and thus Assumption 2.1 is satisfied with $A(t) = 1$, $B(t) = t^{-1/\alpha}$ and $\Phi(x) = |x|^\alpha$ for all $x \leq 0$ and $\Phi(x) = 0$ otherwise. We get from (2.6) that

$$u_t = t^{\frac{\alpha}{1+\alpha}}, \qquad v_t = 1, \qquad w_t = t^{-\frac{1}{1+\alpha}}. \tag{B.1}$$

Since there is equality in (2.4), the convergence in L^1 of (2.5) holds straightforwardly.

EXAMPLE B.2 (Gumbel bounded). For $\alpha > 0$ let $\mu(1-x, 1) = \exp(1 - x^{-\alpha})$ for all $x \in [0, 1]$. Then, for all $x \in \mathbb{R}$,

$$t\mu\big(1 - (1 + \log t)^{-\frac{1}{\alpha}} + x(1 + \log t)^{-\frac{1}{\alpha}-1}/\alpha, 1\big) \to \mathrm{e}^{1-\alpha x}.$$

Thus, Assumption 2.1 is satisfied with $A(t) = 1 - (1+\log t)^{-\frac{1}{\alpha}}$, $B(t) = \frac{1}{\alpha}(1+\log t)^{-\frac{1}{\alpha}-1}$ and $\Phi(x) = \mathrm{e}^{-x}$ for all $x \in \mathbb{R}$. We identify u_t as in the proof of Lemma 2.2, namely $u_t = f^{-1}(t)$ where

$$f(u) = uA(u)/B(u) = u(\log u)\big((\log u)^{1/\alpha} - 1\big).$$

This implies that $u_t = t(\log t)^{-\frac{\alpha+1}{\alpha}}(1/\alpha + o(1))$, and thus $v_t = 1 - (\log t - (1 + 1/\alpha) \times \log\log t)^{-\frac{1}{\alpha}}(1 + o(1))$, and $w_t = (\log t)^{-\frac{\alpha+1}{\alpha}}(1/\alpha + o(1))$. We now check that (2.5) holds: for all $x > 0$,

$$\begin{aligned}
&t\mu\big(A(t) + uB(t), 1\big)\,\mathrm{d}u \\
&\quad = \int_x^{1+\log t} t\exp\bigg(1 - \bigg((1 + \log t)^{-1/\alpha} - \frac{u}{\alpha}(1 + \log t)^{-1-1/\alpha}\bigg)^{-\alpha}\bigg)\,\mathrm{d}u \\
&\quad = \int_x^{1+\log t} t\exp\bigg(1 - (1 + \log t)\bigg(1 - \frac{u}{\alpha(1 + \log t)}\bigg)^{-\alpha}\bigg)\,\mathrm{d}u.
\end{aligned}$$

To use the dominated convergence theorem note that, for all $x \leq u \leq 1 + \log t$,

$$\begin{aligned}
0 &\leq t\exp\bigg(1 - (1 + \log t)\bigg(1 - \frac{u}{\alpha(1 + \log t)}\bigg)^{-\alpha}\bigg) \\
&\leq t\exp\bigg(1 - (1 + \log t)\bigg(1 + \frac{u}{1 + \log t}\bigg)\bigg) = \mathrm{e}^{-u},
\end{aligned}$$

because, for all $w \in (0, 1)$, $(1 - w)^{-\alpha} \geq 1 + \alpha w$. As $u \mapsto \mathrm{e}^{-u}$ is integrable on $[x, \infty)$, the dominated convergence theorem applies and we can conclude that (2.5) holds.

EXAMPLE B.3 (Gumbel unbounded). For $\alpha > 0$ let $\mu(x, \infty) = \exp(-x^\alpha)$ for all $x \geq 0$. Then

$$t\mu\big((\log t)^{\frac{1}{\alpha}} + x(\log t)^{\frac{1}{\alpha}-1}/\alpha, \infty\big) \to \mathrm{e}^{-x}.$$

Thus, Assumption 2.1 is satisfied with $A(t) = (\log t)^{\frac{1}{\alpha}}$, $B(t) = \frac{1}{\alpha}(\log t)^{\frac{1}{\alpha}-1}$ and $\Phi(x) = \mathrm{e}^{-x}$ for all $x \in \mathbb{R}$. Similarly to before we have $u_t = f^{-1}(t)$ where this time $f(u) = u(\log u)$. This implies that $u_t = (1 + o(1))t/\log t$, and thus $v_t = (\log t)^{1/\alpha} - (\log\log t)(\log t)^{1/\alpha-1}(1/\alpha + o(1))$, and $w_t \sim \frac{1}{\alpha}(\log t)^{\frac{1}{\alpha}-1}$. Checking (2.5) is similar to Example B.2.

EXAMPLE B.4 (Fréchet). For $\alpha > 0$ let $\mu(x,\infty) = x^{-\alpha}$ for all $x \geq 1$. Then

$$t\mu\big(xt^{1/\alpha}, \infty\big) = x^{-\alpha},$$

and thus Assumption 2.1 is satisfied with $A(t) = 0$, $B(t) = t^{1/\alpha}$, and $\Phi(x) = x^{-\alpha}$ for all $x > 0$, and $\Phi(x) = \infty$ for all $x \leq 0$. As discussed, in this case we take $v_t = 0$ and we take $u_t = t$ instead of taking it as a solution of (2.6). We get that $w_t = B(t) = t^{1/\alpha}$.

B.2. Examples satisfying Assumption 2.5. We list a few examples satisfying Assumption 2.5. When $M = 1$ we write $\mu(x, 1) = \exp(-m(x))$ for all $x \in [0, 1)$. Then the following weight distributions, given by a suitable function m, all satisfy Assumption 2.5.

(a) $m(x) = (1-x)^{-\alpha} - 1$ for $\alpha > 0$;
(b) $m(x) = \mathrm{e}^{\frac{1}{1-x}} - \mathrm{e}$;
(c) $m(x) = \frac{x}{1-x}$;
(d) $m(x) = \mathrm{e}^{\frac{1}{\sqrt{1-x}}} - \mathrm{e}$;
(e) $m(x) = \tan(\pi x/2)$.

Here (a–e) also satisfy von Mises' condition [17], Proposition 1.1(b), which is a sufficient condition for μ to belong to the domain of attraction of the Gumbel distribution. Note that we are unable to prove that Assumption 2.5 is satisfied by all distributions that satisfy the von Mises condition. We are also unable to provide an example of weight distribution that belongs to the domain of attraction of the Gumbel distribution, satisfies Assumption 2.5, and does not satisfies the von Mises condition. However, the function $m(x) = \log(\frac{\mathrm{e}}{1-x})\log\log(\frac{\mathrm{e}}{1-x})$, for all $x \in [0, 1)$, corresponds to a weight distribution μ that is in the domain of attraction of the Gumbel distribution and does not satisfy Assumption 2.5 (this distribution does not satisfy the von Mises condition). Examples (a–e) are all bounded weight distributions. The following is an unbounded example:

(f) $\mu(x,\infty) = \exp(-x^{\alpha})$ for any $\alpha > 1$.

We prove next that (a) satisfies Assumption 2.5. The others are similar. Recall that, in this example, $\mu(1-x, 1) = \exp(1 - x^{-\alpha})$ for some $\alpha > 0$ and all $x \in (0, 1]$. Assumption 2.1 is satisfied with

$$A(t) = 1 - (1 + \log t)^{-\frac{1}{\alpha}} \quad \text{and} \quad B(t) = \frac{1}{\alpha}(1 + \log t)^{-\frac{\alpha+1}{\alpha}}.$$

We also set $\hat{A}(t) = 1 - A(t) = (1 + \log t)^{-\frac{1}{\alpha}}$. For all $t \geq 0$ and all $x \in \mathbb{R}$, we have

$$t\mu\big(A(t) + xB(t), 1\big) = t\exp\bigg[1 - \hat{A}(t)^{-\alpha}\bigg(1 - \frac{xB(t)}{\hat{A}(t)}\bigg)^{-\alpha}\bigg].$$

Now note that, for all $y < 1$, $(1-y)^{-\alpha} \geq 1 + \alpha y$. Thus, for all $x < \hat{A}(t)/B(t) = \alpha(1 + \log t)$, we have

$$t\mu\big(A(t) + xB(t), 1\big) \leq t\exp\bigg[1 - \hat{A}(t)^{-\alpha}\bigg(1 + \alpha\frac{xB(t)}{\hat{A}(t)}\bigg)\bigg] = t\exp\big(1 - (1 + \log t) - x\big) = \mathrm{e}^{-x}.$$

Making the change of variables $t \mapsto u_t$, this says:

$$\text{if } x \in \big(-\infty, \alpha(1+\log t)\big) \text{ then } \Phi_t(x) \le \mathrm{e}^{-x},$$

which establishes the upper bound in Assumption 2.5(i). For the lower bound, note that there exists a constant $C > 0$ such that, for all $y \in [-1, 1/2]$, $(1-y)^{-\alpha} \le 1 + \alpha y + Cy^2$. Therefore, for all $x \in (-\hat{A}(t)/B(t), \hat{A}(t)/2B(t))$ we have

$$\begin{aligned} t\mu\big(A(t)+xB(t),1\big) &\ge t\exp\left[1-\hat{A}(t)^{-\alpha}\left(1+\frac{\alpha x B(t)}{\hat{A}(t)}+C\frac{x^2B(t)^2}{\hat{A}(t)^2}\right)\right] \\ &= \exp\left(-x - C\frac{x^2B(t)^2}{\hat{A}(t)^{2+\alpha}}\right). \end{aligned}$$

Note that

$$\frac{B(t)^2}{\hat{A}(t)^{2+\alpha}} = \frac{1}{\alpha^2(1+\log t)},$$

thus after the change of variables $t \mapsto u_t$ we have

$$\text{if } x \in \left(-\alpha(1+\log t), \frac{1}{2}\alpha(1+\log t)\right) \text{ then } \Phi_t(x) \ge \mathrm{e}^{-x}\exp\left(-x^2\frac{C}{\alpha^2(1+\log t)}\right)$$

which concludes the proof of the lower bound in Assumption 2.5(i).

For Assumption 2.5(ii), recall that u_t is defined as the unique solution of

$$\alpha u_t\big(1-(1+\log u_t)^{-\frac{1}{\alpha}}\big) = t(1+\log u_t)^{-\frac{\alpha}{\alpha+1}}.$$

Hence $\log u_t \sim \log t$ as $t \uparrow \infty$ and $u_t = t\hat{u}_t$ with $\log \hat{u}_t = o(\log t)$. Thus, $\alpha\hat{u}_t \sim (\log t)^{-\frac{\alpha}{\alpha+1}}$ and so $\hat{u}_t \sim \frac{1}{\alpha}(\log t)^{-\frac{\alpha}{\alpha+1}}$. This implies

$$u_t = \big(1/\alpha + o(1)\big)t(\log t)^{-\frac{\alpha}{\alpha+1}}.$$

Therefore

$$L_1(t) = u_t/t = \frac{1/\alpha + o(1)}{(\log t)^{\frac{\alpha}{1+\alpha}}}$$

so clearly $L_1(t)\log\log t \to 0$.

APPENDIX C: A LARGE DEVIATIONS BOUND FOR THE YULE PROCESS

LEMMA C.1. *Let $(Y_t : t \ge 0)$ be a Yule process with parameter $\lambda > 0$ and let $R > 0$. Then*,

$$\mathbb{P}\Big(\sup_{t\ge 0}|\log Y_t - \lambda t| \ge R\Big) \le 2\Gamma(1/2)\mathrm{e}^{-R/2}.$$

PROOF. First note that, for any $T \ge 0$,

$$\begin{aligned} &\mathbb{P}\Big(\sup_{t\in[0,T]}|\log Y_t - \lambda t| \ge R\Big) \\ &\quad\le \mathbb{P}\Big(\sup_{t\in[0,T]}\log Y_t - \lambda t \ge R\Big) + \mathbb{P}\Big(\inf_{t\in[0,T]}\log Y_t - \lambda t \le -R\Big) \\ &\quad= \mathbb{P}\Big(\sup_{t\in[0,T]}\frac{Y_t}{\mathrm{e}^{\lambda t}} \ge \mathrm{e}^R\Big) + \mathbb{P}\Big(\inf_{t\in[0,T]}\frac{Y_t}{\mathrm{e}^{\lambda t}} \le \mathrm{e}^{-R}\Big). \end{aligned} \tag{C.1}$$

Now, $(Y_t/\mathrm{e}^{\lambda t})_{t\geq 0}$ is a martingale started at 1. Thus by Doob's maximal inequality, and using $\mathbb{E}[Y_T/\mathrm{e}^{\lambda T}]=1$, we have

$$\mathbb{P}\Big(\sup_{t\in[0,T]}\frac{Y_t}{\mathrm{e}^{\lambda t}}\geq \mathrm{e}^R\Big)\leq \frac{\mathbb{E}[\frac{Y_T}{\mathrm{e}^{\lambda T}}]}{\mathrm{e}^R}=\mathrm{e}^{-R}. \tag{C.2}$$

On the other hand, for any $0<\varepsilon<1$,

$$\mathbb{P}\Big(\inf_{t\in[0,T]}\frac{Y_t}{\mathrm{e}^{\lambda t}}\leq \mathrm{e}^{-R}\Big)=\mathbb{P}\Big(\sup_{t\in[0,T]}\Big(\frac{Y_t}{\mathrm{e}^{\lambda t}}\Big)^{-\varepsilon}\geq \mathrm{e}^{\varepsilon R}\Big).$$

Since $x\mapsto x^{-\varepsilon}$ is convex, $(\frac{Y_t}{\mathrm{e}^{\lambda t}})^{-\varepsilon}$ is a submartingale. Thus, by Doob's maximal inequality again,

$$\mathbb{P}\Big(\sup_{t\in[0,T]}\Big(\frac{Y_t}{\mathrm{e}^{\lambda t}}\Big)^{-\varepsilon}\geq \mathrm{e}^{\varepsilon R}\Big)\leq\frac{\mathbb{E}[(\frac{Y_T}{\mathrm{e}^{\lambda T}})^{-\varepsilon}]}{\mathrm{e}^{\varepsilon R}}=\mathrm{e}^{-\varepsilon R+\varepsilon\lambda T}\mathbb{E}\big[Y_T^{-\varepsilon}\big].$$

To finish the proof we recall [2], Section III.5, that Y_T has the geometric distribution with parameter $p=\mathrm{e}^{-\lambda T}$. By Proposition C.2 below,

$$\mathbb{E}\big[Y_T^{-\varepsilon}\big]\leq\frac{\mathrm{e}^{-\varepsilon\lambda T}}{1-\mathrm{e}^{-\lambda T}}\Gamma(1-\varepsilon).$$

Thus,

$$\mathbb{P}\Big(\inf_{t\in[0,T]}\frac{Y_t}{\mathrm{e}^{\lambda t}}\leq \mathrm{e}^{-R}\Big)\leq\frac{\mathrm{e}^{-\varepsilon R}}{1-\mathrm{e}^{-\lambda T}}\Gamma(1-\varepsilon).$$

Taking $\varepsilon=1/2$ and combining with (C.1) and (C.2), we get that

$$\mathbb{P}\Big(\sup_{t\in[0,T]}|\log Y_t-\lambda t|\geq R\Big)\leq\frac{2\Gamma(1/2)\mathrm{e}^{-R/2}}{1-\mathrm{e}^{-\lambda T}}.$$

As the event on the left is increasing in T, letting $T\uparrow\infty$ concludes the proof. □

PROPOSITION C.2. *Let Y be geometrically distributed with parameter $p\in[0,1]$ and let $\varepsilon\in(0,1)$. Then*

$$\mathbb{E}\big[Y^{-\varepsilon}\big]\leq\frac{p^\varepsilon}{1-p}\Gamma(1-\varepsilon).$$

PROOF. Using the change of variables $u=x\log(1/(1-p))$:

$$\begin{aligned}\mathbb{E}\big[Y^{-\varepsilon}\big]&=\sum_{k=1}^\infty\frac{p(1-p)^{k-1}}{k^\varepsilon}\leq\frac{p}{1-p}\sum_{k=1}^\infty\int_{k-1}^k \mathrm{d}x\,\frac{(1-p)^x}{x^\varepsilon}\\&=\frac{p}{1-p}\int_0^\infty \mathrm{d}x\,x^{-\varepsilon}\exp\big(-x\log\big(1/(1-p)\big)\big)\\&=\frac{p}{1-p}\Big(\log\Big(\frac{1}{1-p}\Big)\Big)^{\varepsilon-1}\int_0^\infty \mathrm{d}u\,u^{-\varepsilon}\mathrm{e}^{-u}=\frac{p}{1-p}\Big(\log\Big(\frac{1}{1-p}\Big)\Big)^{\varepsilon-1}\Gamma(1-\varepsilon)\\&\leq\frac{p^\varepsilon}{1-p}\Gamma(1-\varepsilon),\end{aligned}$$

where the last step used $1-p\leq \mathrm{e}^{-p}$. □

Acknowledgments. The authors would like to thank three anonymous reviewers for valuable comments which helped improve the paper.

Funding. The research of JEB was supported by *Vetenskapsrådet, grant 2019-04185*, by *Ruth och Nils Erik Stenbäcks stiftelse*, and by the Sabbatical Program at the Faculty of Science, University of Gothenburg.

REFERENCES

[1] ALDOUS, D. J. (1985). Exchangeability and related topics. In *École D'été de Probabilités de Saint-Flour, XIII*—1983 (P.L. Hennequin, ed.). *Lecture Notes in Math*. **1117** 1–198. Springer, Berlin. MR0883646 https://doi.org/10.1007/BFb0099421

[2] ATHREYA, K. B. and NEY, P. E. (1972). *Branching Processes*. *Die Grundlehren der Mathematischen Wissenschaften*, *Band* 196. Springer, New York. MR0373040

[3] BIANCONI, G. and BARABÁSI, A.-L. (2001). Bose-Einstein condensation in complex networks. *Phys*. *Rev*. *Lett*. **86** 5632.

[4] BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1989). *Regular Variation*. *Encyclopedia of Mathematics and Its Applications* **27**. Cambridge Univ. Press, Cambridge. MR1015093

[5] BORGS, C., CHAYES, J., DASKALAKIS, C. and ROCH, S. (2007). First to market is not everything: An analysis of preferential attachment with fitness. In *STOC*'07—*Proceedings of the* 39*th Annual ACM Symposium on Theory of Computing* 135–144. ACM, New York. MR2402437 https://doi.org/10.1145/1250790.1250812

[6] DEREICH, S., MAILLER, C. and MÖRTERS, P. (2017). Nonextensive condensation in reinforced branching processes. *Ann*. *Appl*. *Probab*. **27** 2539–2568. MR3693533 https://doi.org/10.1214/16-AAP1268

[7] FRANK, R. and COOK, P. J. (1995). *The Winner-Take-All Society*: *Why the Few at the Top Get so Much More than the Rest of Us*. Penguin Books, New York.

[8] ISHWARAN, H. and JAMES, L. F. (2003). Generalized weighted Chinese restaurant processes for species sampling mixture models. *Statist*. *Sinica* **13** 1211–1235. MR2026070

[9] JANSON, S. (2019). Random recursive trees and preferential attachment trees are random split trees. *Combin*. *Probab*. *Comput*. **28** 81–99. MR3917907 https://doi.org/10.1017/S0963548318000226

[10] KÖNIG, W., LACOIN, H., MÖRTERS, P. and SIDOROVA, N. (2009). A two cities theorem for the parabolic Anderson model. *Ann*. *Probab*. **37** 347–392. MR2489168 https://doi.org/10.1214/08-AOP405

[11] LAST, G. and PENROSE, M. (2018). *Lectures on the Poisson Process*. *Institute of Mathematical Statistics Textbooks* **7**. Cambridge Univ. Press, Cambridge. MR3791470

[12] LINDSEY, R. V., KHAJAH, M. and MOZER, M. C. (2014). Automatic discovery of cognitive skills to improve the prediction of student learning. *Adv*. *Neural Inf*. *Process*. *Syst*. **27**.

[13] MAILLER, C., MÖRTERS, P. and SENKEVICH, A. (2021). Competing growth processes with random growth rates and random birth times. *Stochastic Process*. *Appl*. **135** 183–226. MR4226439 https://doi.org/10.1016/j.spa.2021.02.003

[14] MÖRTERS, P., ORTGIESE, M. and SIDOROVA, N. (2011). Ageing in the parabolic Anderson model. *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **47** 969–1000. MR2884220 https://doi.org/10.1214/10-AIHP394

[15] MÖRTERS, P., SOHLER, C. and WALZER, S. (2022). A sublinear local access implementation for the Chinese restaurant process. In *Approximation*, *Randomization*, *and Combinatorial Optimization*. *Algorithms and Techniques* (A. Chakrabarti and C. Swamy, eds.). *LIPIcs*. *Leibniz Int*. *Proc*. *Inform*. **245** Art. No. 28, 18. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR4494348 https://doi.org/10.4230/lipics.approx/random.2022.28

[16] PITMAN, J. (2006). *Combinatorial Stochastic Processes*. *Lecture Notes in Math*. **1875**. Springer, Berlin. MR2245368

[17] RESNICK, S. (2013). *Extreme Values*, *Regular Variation and Point Processes*. Springer.

[18] SARIEV, H., FORTINI, S. and PETRONE, S. (2023). Infinite-color randomly reinforced urns with dominant colors. *Bernoulli* **29** 132–152. MR4497242 https://doi.org/10.3150/21-bej1452

[19] SÉNIZERGUES, D. (2021). Geometry of weighted recursive and affine preferential attachment trees. *Electron*. *J*. *Probab*. **26** Paper No. 80, 56. MR4269210 https://doi.org/10.1214/21-ejp640